Screening in digital monopolies*

Pietro Dall'Ara† Elia Sartori‡

January 28, 2026

Abstract

A defining feature of digital goods is that replication and degradation are costless: once a high-quality good is produced, low-quality versions can be created and distributed at no additional cost. This paper studies quality-based screening in markets for digital goods. Production costs depend only on the highest quality supplied, unlike in standard screening models à la Mussa and Rosen (1978). The monopolist allocation exhibits two interdependent inefficiencies. First, a productive inefficiency arises: the monopolist underinvests in the highest quality relative to the efficiency benchmark. Second, due to a distributional inefficiency, certain buyers receive degraded versions of the produced good. Competition exacerbates productive inefficiency, but improves distributional efficiency relative to monopoly.

Keywords: Screening, Monopoly, Digital Goods.
JEL codes: D42, D82, L12.

Contents

*This paper is based on the first chapter of Elia Sartori's PhD dissertation at Princeton University. We thank Alberto Bennardo, Francis Bloch, Roberto Corrao, Vincenzo Denicolò, Tangren Feng, Joel Flynn, Daniel Gottlieb, Faruk Gul, Paul Heidhues, Andreas Kleiner, Sophie Kreutzkamp, Stephen Morris, Franz Ostrizek, Marco Pagnozzi, Eduardo Perez-Richet, Wolfgang Pesendorfer, Giorgio Saponaro, Alex Smolin, and Tymon Tatur for comments and discussions. Financial support from the Unicredit and Universities Foundation is gratefully acknowledged. An earlier version of this paper bore the title: "Competitive provision of digital goods."

†CSEF and University of Naples Federico II; contact: pietro.dallara@unina.it.

‡CSEF and University of Naples Federico II; contact: elia.sartori@unina.it.

1 Introduction

Quality-based screening is pervasive across markets: sellers routinely offer multiple versions of the same product at different prices, effectively discriminating among consumers with different willingness to pay. For example, car dealers distinguish between models with and without alloy wheels, and software companies sell premium and basic product versions. The seminal contribution of Mussa and Rosen (1978) provides a framework that captures this logic. In the model, the optimum for the seller involves providing, for every type of consumer, a quality equating a virtual value of the type with the marginal cost of producing this quality. The wedge between the type and its virtual value accounts for the rents that the quality generates for other types.

This "marginal cost equal adjusted marginal utility" condition reveals the core force in quality screening: the quality of a type is under-provided in order to reduce the information rents of any higher types. This equality is due to an important feature of the model: all the interdependence across types can be handled by transforming the revenues generated by each type. Effectively, this procedure reduces the problem of the seller to a sequence of separate tradeoffs, one for each type. In particular, there are no cost interdependencies. Importantly, costs are assumed to accumulate by summing a cost for every good sold. We refer to this standard assumption as "cost separability."

This cost separability, however, is not well suited for every industry. The model provides a compelling description of the car industry, in which the production cost of each car is sensitive to the particular version that is sold. In the software example, instead, separability provides a less accurate description: it is not the case that the seller pays a cost for each unit of high-quality software sold and a smaller cost for each unit of low-quality software. A better approximation of production is that a high-quality version is built once, and then reproduced and degraded at zero cost. In this context, there is not any primitive cost capturing the effect of increasing the quality going to a single type. Only the maximum quality determines total cost. What are the inefficiencies of the monopolist allocation with such costs? How does the ability to degrade the produced good impact the production decision? Does the monopolist sell an efficient quality level to the top type? This paper studies these questions.

In this paper, we develop a model to analyze screening when production features the "invest, then damage" costs. We study screening when production costs depend only on the maximum quality produced, so, costs are not separable across delivered qualities. The seller first chooses a *cap quality* and then replicates and damages the chosen-quality good at zero cost. Producing below the cap is costless. The demand side is standard: there is a continuum of buyers with one-dimensional private information (type) and increasing-differences utility.

The digital economy offers several examples, because low replication costs are a key feature of the production of digital goods (Goldfarb and Tucker, 2019). Consider a word processing application offered via subscription, such as Microsoft

Word. The developer builds a feature-complete application at a certain cost, and deploys the application across cloud servers. A basic-tier subscriber receives the identical codebase as a premium subscriber, but certain functions—collaboration tools, design templates, or revision history—are deactivated through licensing flags, without additional development or infrastructure cost. For many products, the same technology that enables costless replication makes damaging nearly free. For example, the same computer that can copy a dataset can also delete observations, the compiler that generates executable code can disable functions, the servers that store content available for multiple viewers can stream content at low resolution.

Two benchmarks organize the analysis (Proposition 1). First, a planner that maximizes total surplus allocates the same undamaged quality to all buyer types. The efficient allocation is a constant function of the type, because giving the highest quality of any allocation to all types increases total utility leaving costs unchanged. Second, the efficient quality equates the marginal cost of production to the average marginal utility in the population of buyers.

The monopolist allocation features two kinds of inefficiencies (Proposition 2). The seller chooses a cap q^M, and allocates to each type either the cap or a lower, damaged, quality. We show that a *productive inefficiency* arises: the monopolist underinvests in the highest quality. Specifically, q^M is lower than the efficient quality. The productive inefficiency interacts with the *distributional inefficiency* that arises from information rents and occurs via damaging.

The characterization of the monopolist allocation uses a decomposition of the monopolist problem into a family of "separable" problems (Lemma 1). For a fixed cap q, we define the cap-constrained problem as a canonical zero-cost virtual-surplus problem subject to the upper bound q on the feasible quality allocations. The optimal allocation for this auxiliary problem maximizes the virtual surplus truncated at q. Thus, a region of high types are *bunched* because they all receive the cap quality, and every low type receives the damaged version of the cap that maximizes virtual surplus conditional on her type.

The seller jointly solves the cap-constrained problem and chooses the cap by equating the marginal production costs with the marginal revenue from relaxing the cap. The marginal revenue equals the marginal utility of the lowest type that is bunched, the *marginally bunched* type, weighted by the mass of the bunching region. Intuitively, the extra revenue given by a higher cap comes from charging a higher price for the good purchased by all bunched types, because the allocation of low types is unchanged.

The marginally bunched type has a special role: she determines the price of the top quality increment and splits the market into two regions. Above the marginally bunched type, there is distributional efficiency because everyone gets the undamaged quality, but productive inefficiency is active because the top quality is lower than the efficient quality. Below this type, there is downgrading, but no additional productive loss relative to a monopolist who replicates and damages an efficient-quality good.

In certain settings, regulatory or technological constraints make versioning infeasible. A ban on screening increases the share of buyers receiving an undamaged good. The marginal revenues in this case account for both the price increment and the inframarginal types, similarly to the screening case. The key difference lies in how the relevant cutoff type is determined. Without screening, the cutoff type makes the seller indifferent between serving her or not; with screening, the cutoff type—marginally bunched—makes the seller indifferent between damaging her quality or not. As a result, the no-screening seller underinvests in quality with respect to the screening case, so productive inefficiency is strengthened (Proposition 5).

We introduce competition in a stylized framework by considering multiple firms that commit to a cap before a pricing stage. The familiar Bertrand-pricing force drives the price of the second-highest quality produced to 0, so only one firm emerges as a monopolist. Specifically, the equilibrium adds a "competitive" constraint in the problem of the endogenous monopolist, stating that: any quality feasible for at least one competitor must be offered for free. The emergence of a monopoly induces a war-of-attrition feature in the production stage, so the production strategies are mixed in equilibrium. Every equilibrium induces a worse productive inefficiency and a more prevalent distributional efficiency than the monopolist allocation (Proposition 6).

Related literature The idea that a seller may want to damage its products for screening purposes was introduced by Deneckere and McAfee (1996), who consider a seller paying damaging costs, possibly zero, to create inferior versions of an available product. We study a multi-product monopolist engaged in quality-based screening (Mussa and Rosen, 1978; Maskin and Riley, 1984; Wilson, 1993), who uses damaging following the logic of Deneckere and McAfee. The screening models in Oren, Smith, and Wilson (1985), Grubb (2009), and Corrao, Flynn, and Sastry (2023) feature consumers who damage goods through underutilization. The seller in Hahn (2006) and Inderst (2008) offers damaged goods over time to mitigate the Coasean commitment problem. Our model introduces a production cost that depends only on the highest quality supplied. The cost structure is not separable in the sense that the total cost cannot be expressed as the sum of costs incurred by interacting separately with every type, contrary to the cited work.[1] This model captures a distinctive property of digital goods.[2]

Information goods are freely replicable (Bergemann and Ottaviani, 2021) and can be versioned cheaply (Shapiro and Varian, 1998). In the typical approach to modeling information markets, a seller replicates and garbles Blackwell experiments at zero cost, but information goods exhibit other important properties, such as non-excludability (Admati and Pfleiderer, 1986), payoff externalities (Ichihashi, 2021; Bonatti, Dahleh, Horel, and Nouripour, 2024; Rodríguez Olivera, 2024), and multi-dimensionality (Bergemann et al., 2018, in contrast with pure vertical differentiation.) Our model complements the literature by considering the production stage and identifying a source of monopoly inefficiency—the productive inefficiency and its interaction with the distributive properties of the monopoly allocation—although abstracting from other properties of information. The literature on excludable public goods studies goods that are freely replicable and excludable through a price system (e.g., Moulin, 1994.)

Our costs can be viewed as generating a production externality, because the cost of allocating a quality to a type depends on the quality allocated to other types. The literature on screening with externalities considers demand externalities that arise if the allocation of a buyer affects the utility that can be extracted from other buyers (Segal, 1999; Jehiel, Moldovanu, and Stacchetti, 1999; Segal and Whinston, 2003). Recently, Halac, Lipnowski, and Rappaport (2024) study the profit guarantee with network goods.

In the single-agent interpretation of the model, replication is irrelevant, and

[1] A kind of nonseparability arises with inventory costs, in which any distribution covered by the inventory has zero costs and other distributions are infeasible (Loertscher and Muir, 2022; Bergemann, Heumann, and Morris, 2025b). Our costs are sensitive to the highest quality supplied, and so are not of the inventory type; however, the distribution stage of the monopolist problem can be seen as a monopolist problem with inventory cost (Lemma 1). Capacity costs can be interpreted as not separable across consumers. For instance, the capacity cost in Boiteux (1960) is not separable across time, in a model without information asymmetries, and the monopoly airline in Gale and Holmes (1993) pays capacity costs, over-investing compared to the efficient level.

[2] Research on digital economics spans from the study of platforms (Rochet and Tirole, 2003) to recent work on the pricing of cloud computing and language models (Bergemann, Bonatti, and Smolin, 2025a; Bergemann and Deb, 2025).

our cost structure represents a seller who produces before eliciting the type of the buyer and damaging the produced good. Hence, the comparison with the workhorse screening model identifies the role of the timing of production, because the seller produces after eliciting the type in Mussa and Rosen (1978).

The game with which we introduce competition has a natural timing given the two-stage interpretation of our monopolist problem. The two-stage game resembles the price-and-quantity competition of Kreps and Scheinkman (1983), in which every firm commits to a capacity before competing in prices, given a rationing rule and absent buyer private information. In our model, buyers have single-unit demand and private information, so the nature of competition is starkly different. We discuss alternative models of competition in Section 4.2.

Outline The main model and results are in Section 2. Section 3 offers interpretations and applications; in Section 4 we introduce a no-screening constraint and competition. The Appendix contains a more general setup and all proofs.

2 Model and results

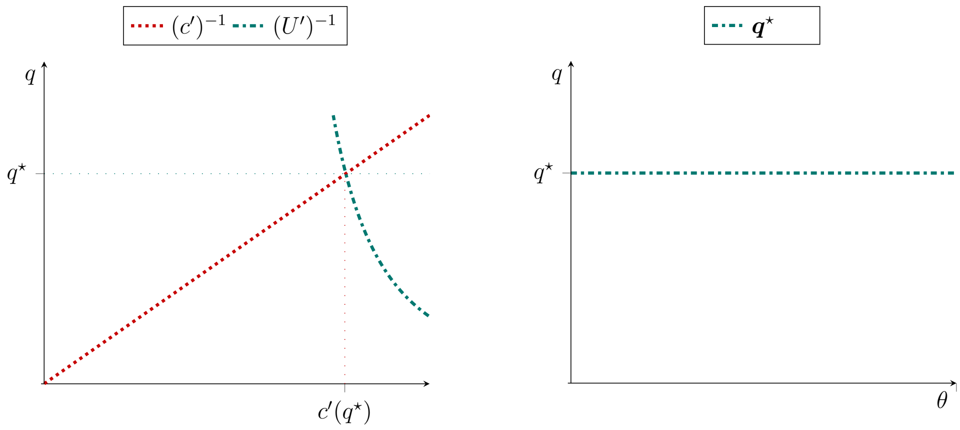
2.1 Model

A seller (she) faces a unit mass of buyers. Each buyer (he) has a type θ that is distributed according to the distribution function F; F is $\mathcal{C}^2$, has support $\Theta := [0,1]$, and has increasing virtual value $\varphi\colon \theta \mapsto \theta - (1-F(\theta))/F'(\theta)$. The payoff of type θ from quality $q \in Q := \mathbb{R}_+$ and transfer $t \in \mathbb{R}$ is $u(q,\theta) - t$, in which the utility from quality is $u(q,\theta) = g(q) + \theta q$, for a $\mathcal{C}^2$, increasing, and strictly concave $g\colon \mathbb{R} \to \mathbb{R}$ with $g(0) = 0$. An *allocation* is a measurable $\boldsymbol{q}\colon \Theta \to Q$, and the cost of producing and distributing according to $\boldsymbol{q}$ is $c(\sup_{\theta\in\Theta} \boldsymbol{q}(\theta))$, for an increasing, strictly convex, and differentiable $c\colon \mathbb{R} \to \mathbb{R}_+$; we denote $\sup_{\theta\in\Theta} \boldsymbol{q}(\theta)$ by $\sup \boldsymbol{q}$. We assume that g and c satisfy the Inada conditions $\lim_{q\to\infty} g'(q) = c'(0) = 0$ and $\lim_{q\to\infty} c'(q) = g'(0) = \infty$.

2.2 Efficiency

An allocation $\boldsymbol{q}$ is *efficient* if it maximizes welfare, which is given by $\int_\Theta u(\boldsymbol{q}(\theta),\theta)\,\mathrm{d}F(\theta) - c(\sup \boldsymbol{q})$.

Proposition 1. *Let $q^\star$ be the quality q solving $g'(q) + \int_\Theta \theta\,\mathrm{d}F(\theta) = c'(q)$. The efficient allocation is given by $\boldsymbol{q}^\star(\theta) = q^\star$ for all θ.*



(a) The quality $q^\star$ equates marginal production cost to average marginal utility. The marginal social value of quality is $U' \colon q \mapsto g'(q) + \mathbb{E}\{\theta\}$.

(b) The efficient allocation $\boldsymbol{q}^\star$ gives quality $q^\star$ to every type.

Figure 1: Panel (a) illustrates the efficient quality $q^\star$; Panel (b) illustrates the efficient allocation. For these graphs and the following ones, θ is uniformly distributed, $c(q) = \frac{1}{2}q^2$, and $g(q) = \sqrt{q}$, unless specified otherwise.

The efficiency benchmark in markets for digital goods is characterized by two features. First, the allocation is constant, which follows from the fact that any non-constant allocation can be improved upon by giving the maximum quality of the allocation to every type. Total costs are unchanged, and total utility increases. This feature is a departure from traditional markets in which the efficient allocation is increasing. Second, the quality $q^\star$ is characterized by equating the marginal cost of production with its marginal social value: the average marginal utility in the population. Figure 1 illustrates the efficient allocation. We benchmark the monopolist allocation both against the efficient quality $q^\star$ and the property that no type receives a damaged good under the efficient allocation.

Remark 1. If the seller observes the type of the buyers, then she can charge price $u(q^\star, \theta)$ for quality $q^\star$ to every type θ. In this way, the seller implements the efficient allocation and extracts all the surplus. In the rest of the paper, we assume that the type of a buyer is her private information.

2.3 The monopolist allocation

The seller maximizes profits by choice of a direct mechanism: a pair of an allocation $\boldsymbol{q}$ and a transfer function $t\colon \Theta \to \mathbb{R}$. The seller's problem $\mathcal{P}^M$ is

$$\sup_{\boldsymbol{q},t} \int_\Theta t(\theta)\,\mathrm{d}F(\theta) - c(\sup \boldsymbol{q}) \text{ subject to:}$$
$$u(\boldsymbol{q}(\theta),\theta) - t(\theta) \geq u(\boldsymbol{q}(\theta'),\theta) - t(\theta') \text{ and } u(\boldsymbol{q}(\theta),\theta) \geq 0, \text{ for all } (\theta,\theta') \in \Theta^2.$$

The allocation in a solution to $\mathcal{P}^M$ determines the associated transfers by standard arguments, so we identify the solution to the monopolist problem with its allocation, denoted by $\boldsymbol{q}^M$. The characterization of $\boldsymbol{q}^M$ does not follow directly from known arguments because we cannot rely on "pointwise" analysis. We address this non-separability by transforming $\mathcal{P}^M$ into a family of separable problems, $(\mathcal{P}(q))_{q\geq 0}$, whose values can be weighted by their associated costs and compared across problems.

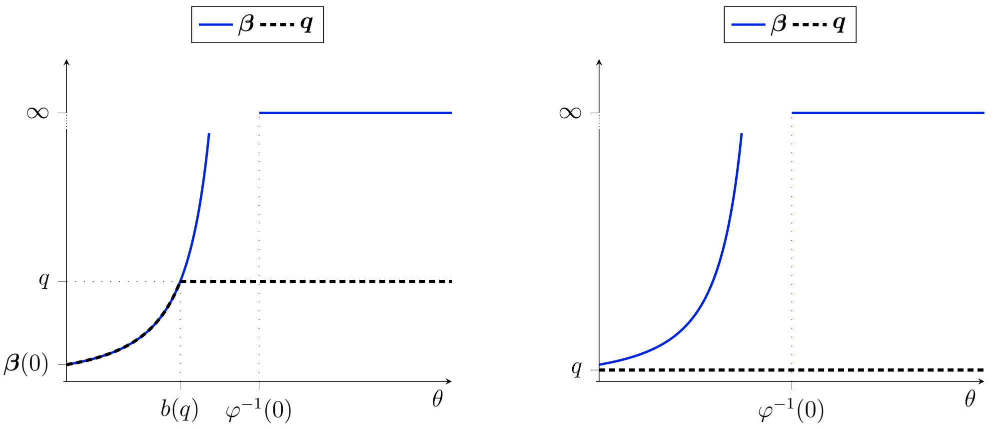
Lemma 1. *For quality q, let the q-constrained problem $\mathcal{P}(q)$ be*

$$V(q) = \sup_{\boldsymbol{q}} \int_\Theta g(q) + \varphi(\theta) q\,\mathrm{d}F(\theta) \textit{ subject to:}$$
$$\boldsymbol{q} \textit{ is nondecreasing, } \boldsymbol{q}(\theta) \leq q, \textit{ for all } \theta \in \Theta.$$

The allocation $\boldsymbol{q}$ solves $\mathcal{P}^M$ if and only if: $\boldsymbol{q}$ solves $\mathcal{P}(q^M)$, where q^M solves $\max_{\hat{q}} V(\hat{q}) - c(\hat{q})$.

The value of the q-constrained problem is the maximal revenue that a monopolist obtains by damaging and replicating a quality-q good at no cost. The characterization of $V(q)$ uses the standard approach of replacing the incentive constraints with a monotonicity constraint and obtaining the virtual surplus in the objective. Hence, the "pointwise" problem $\mathcal{P}(q)$ can be solved using traditional techniques. The cost of doing so is that we need to solve a collection of constrained problems, and compare their values with the quality acquisition costs, which requires to characterize the returns from relaxing the upper-bound constraint.

Lemma 1 is important for two reasons. First, the decomposition uncovers a timing in the monopoly provision of digital goods: the seller produces a single good of some quality q, and, then, she damages and replicates the quality-q good at no extra costs. Second, since every $\mathcal{P}(q)$ is a separable problem and the acquisition step reduces to a unidimensional maximization, Lemma 1 provides



(a) If $q > \boldsymbol{\beta}(0)$, then the marginally bunched type is interior; i.e., $b(q) \in (0, 1)$. Every type $\theta > b(q)$ receives q, and type $\theta < b(q)$ receives a damaged quality.

(b) If $q \leq \boldsymbol{\beta}(0)$, then full bunching occurs: all types receive q. The marginally bunched type is $b(q) = 0$.

Figure 2: The solution $\boldsymbol{q}$ to $\mathcal{P}(q)$ caps the virtual-surplus maximizer $\boldsymbol{\beta}$ at q. Panel (a) illustrates the solution for $q > \boldsymbol{\beta}(0)$; Panel (b) illustrates the solution for $q \leq \boldsymbol{\beta}(0)$.

a fundamental simplification of the monopolist problem. To characterize the monopolist allocation, we define the maximizer of the virtual surplus of a zero-cost monopolist, $\boldsymbol{\beta}$, and its right-continuous inverse b (Figure 2).

Definition 1. *The virtual-surplus maximizer is* $\boldsymbol{\beta}\colon \theta \mapsto \arg\max_q g(q) + \varphi(\theta)q$, *with* $\boldsymbol{\beta}(\theta) = \infty$ *if* $\varphi(\theta) \geq 0$*; the generalized inverse of the virtual-surplus maximizer is:* $b\colon q \mapsto \inf\{\theta \mid \boldsymbol{\beta}(\theta) \geq q\}$.

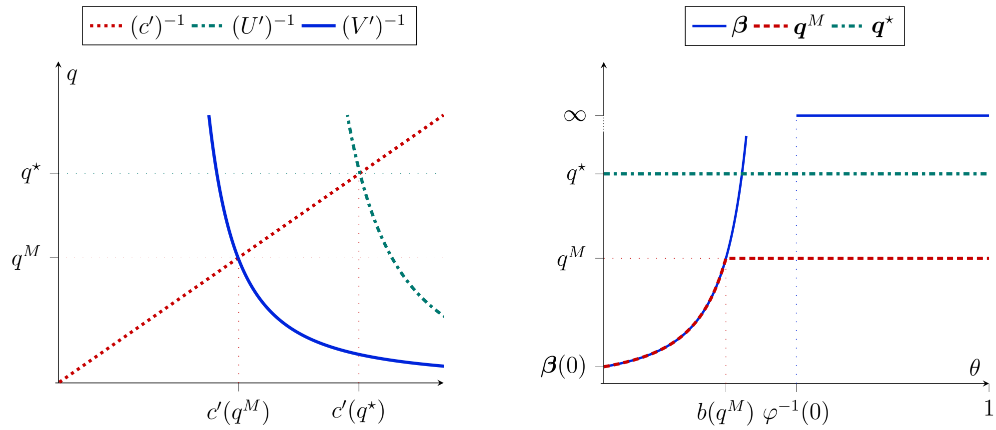
The relationship between the allocation $\boldsymbol{\beta}$ and b is governed by

$$g'(q) + \varphi(\theta) = 0. \tag{2.1}$$

Specifically, $b(q)$ is the type θ solving the above equation, if $g'(q) \leq -\varphi(0)$; and $\boldsymbol{\beta}(\theta)$ is the quality q solving the above equation, if $\varphi(\theta) < 0$. These objects are derived solely from the preference primitives and fully characterize the monopolist allocation via the decomposition in Lemma 1.

Proposition 2. *Let* q^M *be the quality* q *solving* $(1 - F(b(q)))(b(q) + g'(q)) = c'(q)$. *The monopolist allocation is given by* $\boldsymbol{q}^M(\theta) = \min\{\boldsymbol{\beta}(\theta), q^M\}$ *for all* θ. *Moreover, it holds that* $q^M < q^\star$.

Here is the intuition for the result. The solution to the q-constrained problem is determined by capping the virtual-surplus maximizer at the quality q. This



(a) The monopolist cap q^M equates marginal revenues and marginal costs. For all q, the marginal revenue at q is lower than the marginal social value of q, so it holds that $q^M < q^\star$.

(b) The monopolist allocation $\boldsymbol{q}^M$ caps the virtual-surplus maximizer at the cap q^M. Type $\theta \geq b(q^M)$ receives an undamaged good of quality lower than $q^\star$; type $\theta < b(q^M)$ receives a damaged good of the same quality as under the solution to $\mathcal{P}(q^\star)$.

Figure 3: Panel (a) illustrates the monopolist quality q^M; Panel (b) illustrates the monopolist allocation.

operation implies that $\boldsymbol{q}^M$ takes the form of $\theta \mapsto \min\{\boldsymbol{\beta}(\theta), q^M\}$ for a cap q^M that trades off the revenues $V(q^M)$ with the acquisition cost $c(q^M)$, i.e., the value of $\mathcal{P}(q^M)$ with the cost of a relaxation of the associated upper-bound constraint (Figure 3). The rest of Proposition 2 follows from the characterization of the marginal value V' from a quality investment that we describe in what follows.

The function $\boldsymbol{\beta}$ has three properties: first, $\boldsymbol{\beta}(0) > 0$; second, $\boldsymbol{\beta}$ is increasing on $[0, \varphi^{-1}(0))$; third, $\boldsymbol{\beta}(\theta) = \infty$ if $\theta \geq \varphi^{-1}(0)$ (Figure 2.) These properties lead to the following features of the monopolist allocation. First, the monopolist engages in full bunching if the highest quality q^M is below $\boldsymbol{\beta}(0)$, due to the capping structure of the solution to $\mathcal{P}(q)$. Instead, low types get a damaged version of the quality-q^M good whenever $q^M > \boldsymbol{\beta}(0)$. In general, the monopolist allocates a damaged quality to types lower than $b(q^M)$, and "bunches" types above $b(q^M)$, so we refer to type $b(q^M)$ as the *marginally bunched type*.

The key to understand the expression for $V'(q)$ is the observation that, following an increment in the cap q, the seller makes the same revenues from selling all qualities below q. The reason is that qualities below q are allocated to the same types and at the same price. The fact that low qualities go to the same types follows from the solution $\theta \mapsto \min\{\boldsymbol{\beta}(\theta), q\}$ to $\mathcal{P}(q)$: if the cap is not binding, an increase in the cap is immaterial. Low qualities sell at the same price because rents accumulate from below by incentive compatibility: if the allocation of types below θ does not change, then the transfer $t(\theta)$ does not change. Hence, the extra revenues come from serving types in the bunching region: the increment in q is distributed undamaged to all types above $b(q)$, and with an extra price equal to the marginal utility of type $b(q)$. Therefore, marginal revenues are

$$V'(q) = (1 - F(b(q)))(g'(q) + b(q)), \tag{2.2}$$

in which the marginal utility of $b(q)$ is weighted by the mass of the bunching region. The function b is crucial. First, b determines production because it shapes the marginal revenues in Equation 2.2. Second, b impacts the distribution because, effectively, the monopolist allocation $\boldsymbol{q}^M$ is defined by the cap q^M and two distributional conditions: every quality q below the cap q^M goes to its "natural" type $b(q^M)$, and the cap goes to all types above $b(q^M)$.

For the "moreover" part, we argue that the marginal revenues at q are lower than the average marginal utility from quality q. By Markov's inequality, we have $b(q)(1 - F(b(q))) < \mathbb{E}\{\theta\}$, so we conclude that $V'(q) < g'(q) + \mathbb{E}\{\theta\}$. Hence marginal revenues are uniformly below the marginal social value of a quality investment, i.e., V' crosses the marginal cost below $q^\star$.

Proposition 2 uses the fact that V is concave and differentiable (Proposition A.2). Under regularity, full bunching does not preclude differentiability of V, because V' is continuously pasted at $\boldsymbol{\beta}(0)$. The revenue function V is not differentiable if F is not regular. Intuitively, suppose that types in (θ', θ'') are bunched "at" $q > 0$ by a monopolist who solves $\mathcal{P}(\overline{q})$ for $\overline{q} > q$, and the monopolist allocation is increasing for all other types. In this case, the extra revenues of adding $\varepsilon > 0$ to q come from bunching at the top only types that strictly exceed θ'', whereas the revenues lost from degrading q by ε are due to including types lower than θ' to the bunching region. We conclude that $\lim_{\varepsilon \to 0^+} V'(q-\varepsilon) - V'(q+\varepsilon) > 0$. In the Appendix, we show that Proposition A.2 holds without regularity.[3]

Remark 2. Let's argue that the type $b(q)$ maximizes $\theta \mapsto (1 - F(\theta))(\theta + g'(q))$. The q-constrained problem can be equivalently stated as a choice of a tariff $T\colon [0, q] \to \mathbb{R}$ determining the price of every quality (Guesnerie and Laffont, 1984). The tariff T is optimal only if the *price increment at* q, $T'(q)$, solves $\max_p p \Pr(\{\theta : u_q(q, \theta) \geq p\})$, by known arguments under regularity conditions (Wilson, 1993). Intuitively, every quality increment is priced as if it constitutes a separate market. From the optimal tariff corresponding to the solution $(\boldsymbol{q}, t)$

[3] Specifically, the shape of the solution to $\mathcal{P}(q)$ and the fact that $q^M < q^\star$ hold with more general F and increasing-differences u (Proposition A.2 and B.1).

to $\mathcal{P}(q)$—i.e., the mapping $T\colon q \mapsto t(b(q))$—we compute the price increment at q, that is $T'(q) = g'(q) + b(q)$. Heuristically, if type θ purchases an interior quality, then the first-order condition holds for the given tariff. Hence, the price increment $g'(q) + b(q)$ solves $\max_p p(1 - F(p - g'(q)))$. Equivalently, type $b(q)$ solves $\max_\theta (1 - F(\theta))(\theta + g'(q))$.

Inefficiencies For high types, distributional efficiency holds, because all types above $b(q^M)$ are allocated an undamaged quality. However, productive efficiency does not hold, because the undamaged quality q^M is lower than $q^\star$. Under separable costs, the top type does not impose information rents to other types, so his quality is efficient because his information rent is traded-off with the cost of producing only his quality. For digital goods, instead, the monopolist accounts for the productive externality that the top-type quality has for the quality of the other types. Hence, the highest quality determines the information rents of types in the bunching region. As a result, the standard "efficiency at the top" observation is weakened to distributional efficiency. Moreover, the distributional efficiency is more "prevalent" than with separable costs, in which case efficiency holds only for the top type, because all types above $b(q^M)$ receive the undamaged quality.

The types who are subject to the two inefficiencies can be read off the graph in Figure 3. The marginally bunched type partitions the type space into two regions. Each region is associated with an auxiliary economy in which one inefficiency is shut down: (i) the seller produces q^M and distributes q^M to everyone, and (ii) the seller produces $q^\star$ and distributes damaged qualities to maximize revenues, i.e., as in the solution to $\mathcal{P}(q^\star)$. Types below $b(q^M)$ receive a damaged good, so they are subject to distributional inefficiency. Nonetheless, they are unaffected by productive inefficiency, because they receive the same quality by a monopolist producing the efficient quality as under $\boldsymbol{q}^M$. No type above $b(q^M)$ is subject to distributional inefficiency, because he receives an undamaged good; however, he is subject to productive inefficiency, because he receives a quality exceeding q^M by a seller that produces efficiently.

Remark 3. With linear utility, i.e., $u(q, \theta) = \theta q$, the marginally bunched type is $b(q) = \varphi^{-1}(0)$; and we have: $\boldsymbol{\beta}(\theta) = 0$ for all $\theta < \varphi^{-1}(0)$ (Figure 4). Damaging takes the trivial form of excluding low types, because $\boldsymbol{q}^M(\theta) = q^M$ if $\theta \geq \varphi^{-1}(0)$, and $\boldsymbol{q}^M(\theta) = 0$ if $\theta \leq \varphi^{-1}(0)$.

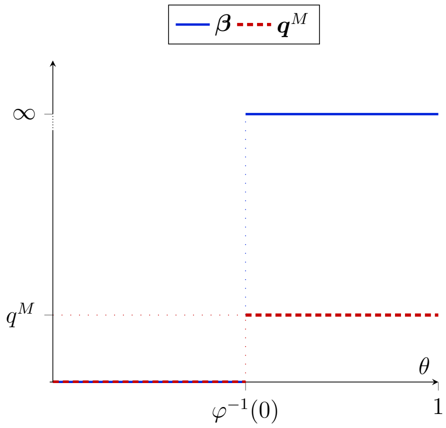


Figure 4: With linear utility, the problem $\mathcal{P}(q)$ is solved by excluding types lower than the zero of the virtual value φ, and allocating q to higher types.

2.4 Discussion of the model primitives

In markets with our cost structure, richness of versions is driven by preferences alone. In particular, the monopolist allocation with linear preferences (Remark 3) can be implemented by offering a single quality at a posted price. The reason is that the curvature of the cost function does not directly impact the distributive properties of the monopolist allocation (Lemma 1), and only determines the highest quality. Hence, the rich variety of versions observed in markets for digital goods (Bergemann and Bonatti, 2019) should be attributed to extra curvature in the utility from quality than that allowed by linear preferences.

The type-independent utility shifter given by g is a parsimonious addition to the multiplicative preferences in Mussa and Rosen (1978) that permits our screening model to be applicable to digital markets. The reason is that the seller finds it more profitable to discriminate across types as we climb the quality ladder, for reasons that are not related to costs: the ratio of the marginal utility of type θ to the marginal utility of a lower type θ' is increasing in quality, i.e., $\frac{g'(q)+\theta}{g'(q)+\theta'}$ increases in q. This pattern aligns with the way digital goods are employed in basic and advanced tasks, whenever buyers are more heterogeneous in advanced tasks than basic ones. In this interpretation, the heterogeneity in relative marginal utilities vanishes around 0 because quality increments in a neighborhood of 0 are "infinitely" valuable for the accomplishment of the basic task. Instead, the heterogeneity becomes more pronounced as q grows, because quality increments are used almost exclusively in the accomplishment of the professional task.

Remark 4. In a model of separable screening, the curvature of the cost function directly affects the shape of the monopolist allocation. Under linear preferences, the optimality of a single-quality menu arises only if costs are linear. Sandmann (2025), and related literature, provides sufficient conditions and intuition for the optimality of single-quality menues in general setups with separable costs.

Comparative statics We introduce the parameter $\kappa = (\kappa_c, \kappa_g)$ that enters the model via $c_\kappa(q) \coloneqq \kappa_c c(q)$ and $u_\kappa(q, \theta) \coloneqq \kappa_g g(q) + \theta q$. The value of κ_c shifts the marginal costs of production, whereas κ_g shifts the importance of the common curvature of the utility function. Hence, an increase in κ_g reduces the importance of preference heterogeneity. By Proposition 2, the monopolist allocation is determined as follows. First, we define $b_{\kappa_g} : Q \to \Theta$ such that $\kappa_g g'(q) + \varphi(b_{\kappa_g}(q)) = 0$ for all $q \in Q$ and $b_{\kappa_g}(q) = 0$ if $\kappa_g g'(q) \leq -\varphi(0)$, with right-continuous inverse $\boldsymbol{\beta}_{\kappa_g}$. The monopolist allocation $\boldsymbol{q}_\kappa^M$ is defined the following conditions,

$$\begin{cases} q_\kappa^M \in Q \text{ is such that } \kappa_c c'(q_\kappa^M) = (1 - F(b_{\kappa_g}(q_\kappa^M)))(\kappa_g g'(q_\kappa^M) + b_{\kappa_g}(q_\kappa^M)), \\ \text{for all } \theta,\ \boldsymbol{q}_\kappa^M(\theta) = \min\{\boldsymbol{\beta}_{\kappa_g}(\theta), q_\kappa^M\}. \end{cases}$$

As κ_c increases, the acquired quality decreases and the function b_{κ_g} does not change. Hence, the allocation of all but the low types is pushed downwards, and the bunching region expands: $b_{\kappa_g}(q_\kappa^M)$ is decreasing in κ_c.

An increase in κ_g affects the allocation both through the distribution, shifting the virtual-surplus maximizer upwards, and the acquisition condition, via marginal revenues. The acquisition effect has a direct channel—the marginal utility is parametrized by κ_g—and an indirect channel—the function b_{κ_g} depends on κ_g and determines the marginal revenue. Overall, the monopolist cap increases and the bunching region expands as κ_g grows.

Proposition 3. *The cap q_κ^M is decreasing in κ_c and nondecreasing in κ_g. Moreover, b_{κ_g} is nonincreasing in κ_g pointwise; $b_{\kappa_g}(q_\kappa^M)$ is nonincreasing in κ_g, decreasing if interior, and there exists $\overline{\kappa}_g$ such that $b_{\kappa_g}(q_\kappa^M) = 0$ for all $\kappa_g > \overline{\kappa}_g$.*

The economically relevant case of non-degenerate screening is due to intermediate curvature in the utility form quality: preferences reduce to the linear benchmark if $\kappa_g = 0$, and screening collapses to a pooling contract as κ_g grows. Therefore, rich contracts in digital monopolies require some, but not overwhelming, curvature in utility.

The comparative statics implies that full bunching is optimal in two cases: a large κ_g and a large κ_c. Intuitively, the optimality of full bunching is related to the relative importance of the common curvature in the utility and the strength of the type heterogeneity. First, as κ_g increases, the common curvature becomes mechanically more important because $\kappa_g g'(q)$ increases, and full bunching is more profitable. Due to this channel, full bunching arises with a high monopolist cap because q^M_κ is increasing in κ_g. Second, as κ_c increases, the highest quality q^M_κ decreases. Hence, the term $\kappa_g g'(q)$ increases endogenously via the range of qualities that are allocated by the monopolist as κ_c grows. Due to this channel, full bunching arises with a low cap.

The model differs sharply from separable screening in its sensitivity to the type distribution. Because all bunched types receive the cap, a change in the density of high types does not affect the distributive properties of the monopolist allocation. Hence, any distributions F and G that coincide below the threshold given by the marginally bunched type corresponding to F induce the same monopolist allocation. If, in addition, the distributions have the same mean, then the efficient allocation is the same.[4]

The comparative statics clarifies the roles of primitives. Costs determine the reach of acquisition via the cap, whereas the curvature of the utility term g shapes both the acquisition and the distribution stages. Distributional perturbations matter via the assignment of quality—below the bunching region and through the mass at the cap—rendering a class of shape changes above some threshold irrelevant for allocation and welfare. These observations stand in contrast to the case of separable-cost screening, in which density shifts produce changes in allocation and welfare throughout the type space.

Remark 5. The main results (Proposition 1, 2, 5, and 6) hold with more general utility functions, as shown in the Appendix. In our comparative statics (Proposition 3), we leverage the simplicity of intuitions arising by expressing type-θ rents as $\int_{[0,\theta]} \boldsymbol{q}(\tilde{\theta})\, \mathrm{d}\tilde{\theta}$. The multi-principal model of Chade and Swinkels (2021) also leverages this tractability. The current u captures more general preferences, given by the gross utility function $q \mapsto g_1(q) + \theta g_2(q)$ for some invertible $g_2 \in \mathbb{R}^Q$ and a concave transformation $g_1 \in \mathbb{R}^Q$ of g_2, because quality does not admit a natural metric.

[4] Let's consider the effect of scaling the support of the type. Let $\alpha > 0$ and G be the distribution function of the random variable $\alpha\hat{\theta}$, using $\hat{\theta}$ for the r.v. induced by F. The implications of replacing F with G follow from Proposition 3; effectively, we multiply g and c by $\frac{1}{\alpha}$.

3 Interpretation

In this section, we discuss the implications of our results within the leading interpretation of the digital goods industry, and offer an alternative interpretation under which we compare our model with Mussa and Rosen (1978, MR).

3.1 Digital goods

A *cost mismatch* arises under the free-damaging and free-replication scenario of the digital-good motivation. Recall the maximand in the monopolist problem, $\int_{\Theta} t(\theta)\,\mathrm{d}F(\theta) - c(\sup \boldsymbol{q})$. The production cost has the same order of magnitude as the revenue term, and is infinitely larger than the surplus generated by a single buyer. We view this feature as capturing the correct economics for digital goods: the seller chooses how far up the quality ladder to climb, pays the development cost, and can then serve any number of buyers with any lower quality by throttling, feature-gating, or binning. For example, consider a game released in two editions: an all-inclusive "Deluxe" edition, and a basic edition with fewer features. The game studio only needs to build the Deluxe edition, whereas the basic one is created by removing maps, characters, or game modes. Producing the Deluxe edition is costly, whereas producing the basic edition simply involves withholding "pieces" of the existing game, rather than building a separate game, and costs almost nothing. The game studio pays the cost of development once, and can then freely replicate and degrade the Deluxe game for any number of players.

This interpretation leads to Lemma 1: the distribution stage amounts to a pure price discrimination problem on a fixed ladder, and occurs after production has taken place. A key feature of the interpretation is that the development cost is parametrized relative to the population size. Specifically, suppose that we perturb the mass of consumers α, away from its unit-mass benchmark in Section 2. The perturbation impacts profits only through the revenue term, contrary to a model à la MR in which α is a profit shifter: both revenues and costs are scaled by α in MR. Hence, an increase in α changes the weights of the two addenda in the objective of the seller. However, the comparative statics in Proposition 3 applies, because an increase in α is equivalent to a decrease in development costs; i.e., a decrease in κ_c with the notation of Proposition 3. Under this interpretation, a seller that faces a single buyer does not find it optimal to produce: as $\alpha \to 0$ the revenue following any investment vanishes.

The role of the production cost is to select the highest quality. On a mechanical level, this selection induces a quality under-provision that interacts with the familiar screening distortions in the distribution stage. The acquisition margin delivers an economic insight that is absent under the "interim" perspective that assumes an exogenous top version and studies the distribution stage under free replication and damaging.

Because we work with a "global" development cost rather than a per-unit cost, the direct comparison with standard screening in this multi-agent interpretation is artificial. We compare our results with the conclusion of a model à la MR under an alternative interpretation in Section 3.2.

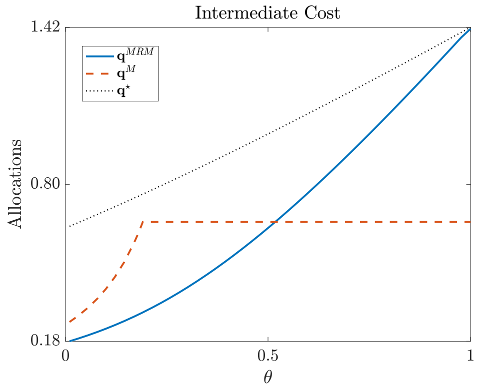
3.2 Single agent

In this section, we interpret costs as expressed in the same units as the utility of a single buyer—recall that costs are expressed in per-agent units in the leading interpretation of the model in Section 3.1. This interpretation can be understood as follows: there is a *single buyer*, and quality cannot be produced after the type is elicited—contrary to the single-agent interpretation of MR. When the buyer enters the store, the seller has already produced a version of the good, and screening only occurs through offering damaged versions of the available good.[5] With this interpretation, a comparison of the monopolist allocation $\boldsymbol{q}^M$ with the monopolist allocation under separable costs (given by c) is meaningful.

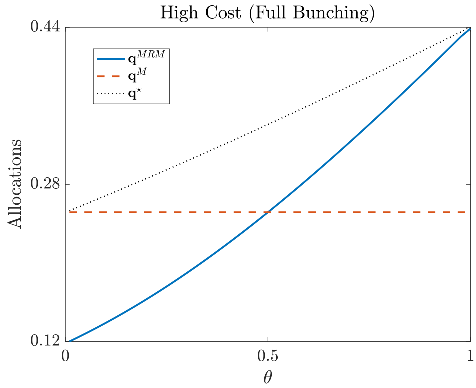
Remark 6. If the costs c are separable, then the seller maximizes $\int_\Theta t(\theta) - c(\boldsymbol{q}(\theta))\,\mathrm{d}F(\theta)$. The pointwise maximization of virtual surplus yields the optimal allocation, which is $\boldsymbol{q}^{MRM}\colon \theta \mapsto \max\{(g' - c')^{-1}(-\varphi(\theta)), 0\}$, as in MR.

The comparison speaks to the effects of the timing of production. In the car-dealer example (Section 1), cost separability provides an accurate description of the production process. However, if the interaction with the buyer occurs mainly through a dealer, who only offers multiple versions that are assembled through add-ons (e.g., sound system, paint finishes, leather seats,) then the production costs are such that a baseline "top version" is made available by the manufacturer, and screening occurs via damaging. Hence, the comparison

[5] Replication is irrelevant with a single buyer, and there is no use of damaging if production occurs after the type elicitation. With a single buyer, it is the interaction of free damaging and pre-production that creates a non-trivial model: adding free damaging but keeping the MR timing does not change the MR solution because there is no need to over produce; if seller has to pre-produce and cannot damage, then she offers a single quality, solving the problem analyzed in Section 4.1.



(a) For this figure: $c(q) = \frac{1}{2}q^2$, so that $b(q^M) > 0$.



(b) For this figure: $c(q) = 2q^2$, so that $b(q^M) = 0$.

Figure 5: Panel (a) compares the monopolist allocation with $\boldsymbol{q}^{MRM}$ if the monopolist does not engage in full bunching, so $b(q^M) > 0$; Panel (b) compares the monopolist allocation with $\boldsymbol{q}^{MRM}$ when the monopolist optimally offers a pooling contract, so $b(q^M) = 0$. The two figures have different allocations only because of different cost functions, see Proposition 3.

between $\boldsymbol{q}^M$ and $\boldsymbol{q}^{MRM}$ is attributed to the difference between screening after or before production.[6]

In our model, the "produce-then-damage" timing is a derived property of the cost structure; see Lemma 1. The current interpretation, instead, takes this timing as a primitive assumption, and leads to the same profit-maximization problem as in Section 2. With this interpretation, the efficient allocation is given by the allocation $\boldsymbol{q}^{MRE}$ satisfying $u_q(\boldsymbol{q}^{MRE}(\theta), \theta) = c'(\boldsymbol{q}^{MRE}(\theta))$ for all θ, as in MR.[7]

Let's compare the monopolist allocation $\boldsymbol{q}^M$ with the MR allocation $\boldsymbol{q}^{MRM}$ (see Figure 5). By incentive compatibility, both allocations are nondecreasing and pointwise below the efficient allocation $\boldsymbol{q}^{MRE}$. For types below the marginally bunched $b(q^M)$, the monopolist allocation $\boldsymbol{q}^M$ provides a higher quality than $\boldsymbol{q}^{MRM}$; the reason is that our monopolist sets marginal virtual surplus to zero, whereas the MR monopolist sets marginal virtual surplus net of costs to zero. The MR allocation exhibits efficiency at the top, differently from $\boldsymbol{q}^M$, so the

[6]Our results apply to "cost-separable" industries in which screening only uses "simple" damaging of a baseline product. On a level of interpretation, which of the two models provides a better approximation for a given application depends on how distant the buyer-seller interaction is from production decisions, and how good the large number of buyers approximates the continuum of buyers.

[7]Proposition 1 characterizes a "constrained" efficiency benchmark, in which the planner faces two constraints: producing before eliciting the type of the buyer, and distributing the same quality to every type.

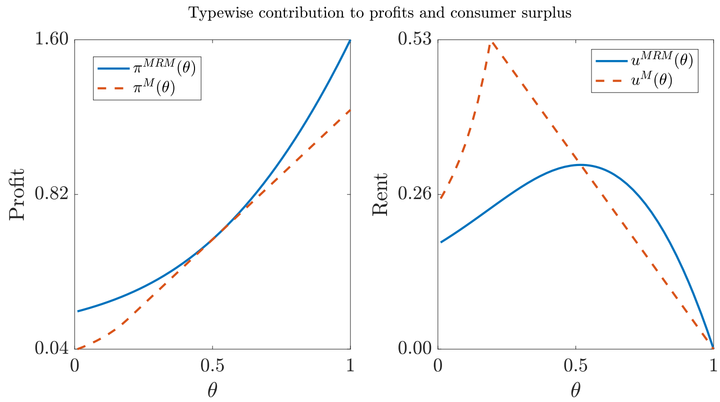


Figure 6: The left panel shows the profit of the seller given the monopolist allocation $\boldsymbol{q}^M$ and given the MR allocation with separable costs. The seller is better off under separability. The right panel shows the consumer surplus given $\boldsymbol{q}^M$ and $\boldsymbol{q}^{MRM}$, using $u^M(\theta) = \int_{[0,\theta]} \boldsymbol{q}^M(s)\,\mathrm{d}s$ and $u^{MRM}(\theta) = \int_{[0,\theta]} \boldsymbol{q}^{MRM}(s)\,\mathrm{d}s$ for the information rent of type θ. The ranking of consumer surplus is ambiguous.

quality ranking is reversed for sufficiently high types. Interestingly, our full-bunching monopolist exhibits efficiency at the bottom, whereas the MR monopolist underprovides quality to the bottom type.[8] These comparisons are established in the following result.

Proposition 4. *The allocations $\boldsymbol{q}^M$ and $\boldsymbol{q}^{MRE}$ are such that: (i) if $b(q^M) > 0$, then $\boldsymbol{q}^M < \boldsymbol{q}^{MRE}$ pointwise; (ii) if $b(q^M) = 0$, then $q^M = \boldsymbol{q}^{MRE}(0)$. Moreover, the allocations $\boldsymbol{q}^M$ and $\boldsymbol{q}^{MRM}$ are such that: (i) $q^M < \boldsymbol{q}^{MRM}(1)$; (ii) if $\theta < b(q^M)$, then: $\boldsymbol{q}^M(\theta) > \boldsymbol{q}^{MRM}(\theta)$.*

The seller is worse off with nonseparable costs than with separable costs: she ends up overpaying relative to the quality of the good sold with some probability. The implications of nonseparability for consumer surplus are more delicate, and the fact that the allocations cross suggests that the ranking of consumer surplus is ambiguous (see Figure 6). Low types receive higher rents under $\boldsymbol{q}^M$ than with separable costs, because of the higher quality under nonseparability than in MR (Proposition 4). Instead, higher types can receive lower rents under nonseparability. Because the consumer surplus of allocation $\boldsymbol{q}$ is $S(\boldsymbol{q}) := \int_\Theta \boldsymbol{q}(\theta)(1 - F(\theta))\,\mathrm{d}\theta$, low types weigh "more" than high types. With a uniform type distribution, quadratic costs, and $g(q) = \kappa_g\sqrt{q}$, as in Section 2.4, the ranking of consumer surplus is determined by κ_g. Specifically, $S(\boldsymbol{q}^M) - S(\boldsymbol{q}^{MRM})$ is negative at $\kappa_g = 0$,

[8]The MR monopolist offers a rich menu because of the curvature of the cost function, even in the case in which our monopolist engages in full bunching.

increases with κ_g, and is positive for high values of κ_g. The intuition for this ranking builds upon Section 2.4. If $u(q,\theta)$ is approximately linear in q, then our monopolist excludes low types, as in Figure 4, whereas $\boldsymbol{q}^{MRM}$ is increasing, due to the curvature of c. This difference explains why the consumer surplus in MR is higher than under nonseparable costs in the case of linear preferences. As κ_g grows, the bunching region expands by Proposition 3, and the ranking reverses.

4 Extensions

4.1 Non-discriminating monopolist

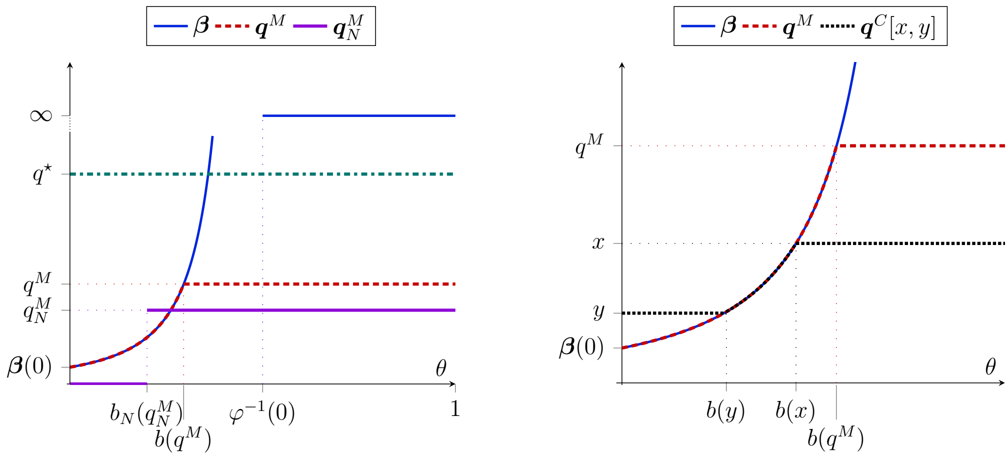
In certain settings, technological or regulatory constraints make it infeasible for the seller to degrade the good. If the monopolist allocation induces full bunching, then a no-damaging constraint—i.e., $\boldsymbol{q}(\Theta) \subseteq \{0, q\}$ for some $q \in Q$—is irrelevant for the monopolist problem.

To characterize the allocation $\boldsymbol{q}_N^M$ of a non-discriminating monopolist, fix a produced quality q. The monopolist effectively chooses a type $b_N(q)$ such that: lower types are excluded and higher types buy quality q. The implied price of the quality-q good makes type $b_N(q)$ indifferent between buying and the outside option. The resulting revenues are $V_N(q) = \max_\theta (1 - F(\theta))(g(q) + \theta q)$. The resulting marginal revenues account both for the price increment and the inframarginal types, because, by the envelope theorem, we have $V_N'(q) = (1 - F(b_N(q)))(g'(q) + b_N(q))$.

The structure of the marginal value of relaxing the q-constraint for the non-discriminating seller is the same as for the screening seller. The key difference lies in how the relevant cutoff type is determined. Without screening, the cutoff type makes the monopolist indifferent between serving and excluding him; i.e., $b_N(q)$ is the type θ solving $g(q) + \varphi(\theta)q = 0$. With screening, the cutoff type makes the monopolist indifferent between damaging his quality or not; i.e., $b(q)$ is the type θ solving $g'(q) + \varphi(\theta) = 0$. The screening ban mechanically shuts down damaging, so, intuitively, the monopolist sells the undamaged good to a larger set of types than under screening, for fixed cap q.[9]

Figure 7a illustrates the optimal allocation. Here is the intuition why the monopolist supplies a lower top quality than under screening. The marginally bunched type $b(q)$ is the type θ that maximizes the marginal utility $u_q(q,\theta)$

[9] In particular, we have $\frac{g(q)}{q} > g'(q)$, for all $q > 0$, by strict concavity of g; hence, we have $b_N(q) \leq b(q)$, with strict inequality if $b(q)$ is interior, for all $q > 0$.



(a) Without damaging, the monopolist sells am undamaged quality (q_N^M) to more types than with damaging, excludes low types, and strengthens productive inefficiency.

(b) In a subgame with highest and second-highest qualities x and y, resp., every type θ purchases quality $\boldsymbol{q}^C[x,y](\theta)$.

Figure 7: Panel (a) illustrates the monopolist allocation $\boldsymbol{q}_N^M$ under a no-damaging constraint; Panel (b) illustrates the allocation $\boldsymbol{q}^C[x,y]$ that prevails in a subgame in which the highest and second-highest quality are, resp., x and y.

weighted by the mass of types above θ, by Remark 2. However, the marginal revenues of the non-discriminating seller account for the fact that q is sold to non-excluded types at a price that increases with the marginal utility of *cutoff* type $b_N(q)$. Hence, productive inefficiency gets stronger with the addition of the no-damaging constraint: $V'(q) - V_N'(q) \geq 0$ for all q, with strict inequality whenever the screening seller does not engage in full bunching. Therefore, the highest quality without screening, q_N^M, is lower than q^M. By the conclusion of the preceding paragraph, distributional efficiency is more prevalent than with screening. In particular, the strengthening of productive inefficiency implies that a larger region of types gets the undamaged good, i.e., $b_N(q_N^M) \leq b(q^M)$.

Proposition 5. *If $b(q^M) > 0$, then it holds that $q_N^M < q^M$ and $b_N(q_N^M) < b(q^M)$.*

The ban may induce exclusion, which does not occur under screening, and makes low types worse off. In particular, the ban makes types in $[0, b_N(q_N^M)]$ worse off because these types are excluded under $\boldsymbol{q}_N^M$. Instead, the ban makes low types better off whenever the non-discriminating monopolist does not exclude the bottom type. In particular, the ban makes types in $[0, b(q_N^M)]$ better off if $b_N(q_N^M) = 0$, because $\boldsymbol{q}_N^M$ allocates q_N^M to all types lower than $b(q_N^M)$ and $q_N^M > \boldsymbol{q}^M(\theta)$ for $\theta < b(q_N^M)$.

4.2 Competition

In this section, we extend our analysis to a competitive setting. We show that a competitive market yields a stochastic allocation that "shrinks" the monopolist one: the lowest quality is higher, and is offered for free, more types receive an undamaged good, but at the same time the highest quality is lower. The reduction of the damaging inefficiency and the increase of the productive inefficiency have an ambiguous effects, which boils down to the shape of the cost function. We present the model and results, and then discuss our approach relative to the alternatives in the literature.

Description of the game We study a two-stage game of perfect information among countably many replicas, indexed by $i \in \{1, 2, \dots\}$, of the seller in Section 2. In the first stage of the game, every firm i simultaneously chooses an investment strategy—potentially mixed, so firm i chooses a distribution over Q—and pays the production cost of the realized quality q_i.[10] Then, the profile of realized investments becomes public information among firms. In the second stage, firms compete à la Bertrand: every firm i simultaneously chooses a pricing function p_i over the qualities that are feasible; i.e., in $[0, q_i]$. Finally, every buyer treats the firms as supplying homogeneous goods, and observes the pricing functions; effectively the buyers face the tariff given by the lower envelope of the pricing functions, $q \mapsto \min_i p_i(q)$. Every buyer maximizes his payoff purchasing a feasible quality q from a firm i or buying nothing for the outside option of 0 payoff. The revenues of firm i derive from the demand for the qualities it offers at the lowest price, and costs are sunk at the second stage. We study the subgame-perfect equilibria of this two-stage game of complete information, in which firms choose a distribution of their investment and a pricing function conditional on the realized investments.

Results We can study competition as a constrained monopolist problem, due to the structure of the game. Consider the subgame starting at a cap profile with first- and second-highest qualities given by x and y, respectively. Bertrand competition drives to zero the revenues from qualities in $[0, y]$. Hence, only one firm earns positive revenues: the firm i^* that produces x and sells qualities in the spectrum $[y, x]$ over which it holds market power. We establish that the ensuing allocation $\boldsymbol{q}^C[x, y]$ admits an intuitive characterization: it is obtained

[10] The definition of the game and the technical details are in the Appendix.

by "slicing" the virtual-surplus maximizer both from above at x and from below at y, yielding $\boldsymbol{q}^C[x, y]\colon \theta \mapsto \max\{\min\{\boldsymbol{\beta}(\theta), x\}, y\}$. The quality y is provided for free to all types below $b(y)$.[11] Qualities above y, instead, are available exclusively from the "interim" monopolist, the only firm earning positive revenues, given by $V(x) - V(y)$.

As an implication, the investment game reduces to an all-pay contest in which the second-highest bid takes away part of the gains from winning. From the point of view of firm i, the investment of competitors acts as a fixed cost in its production problem "at" the highest quality of the opponents—equal to $V(\max_{j \neq i} q_j)$. Only by "winning" does firm i earn positive revenues. This fixed cost affects only the "entry" decision, and not investment conditional on entry. Therefore, the best response is either q^M or 0.

The following result characterizes the equilibrium allocations. A firm is *active* in an equilibrium if it produces a positive quality with positive probability.

Proposition 6. *For every $n \in \{1, 2, \dots\}$, there exists an equilibrium with n active firms. In every equilibrium in pure strategies, only one firm is active. In every equilibrium with $n \geq 2$ active firms, the investment strategy of every active firm is the continuous distribution function with support $[0, q^M]$ given by $H_n(q) = (c'(q)/V'(q))^{1/(n-1)}$, for all $q \in [0, q^M]$.*

A monopoly arises in certain equilibria as a consequence of the implicit commitment of the timing and information structure of the game. Specifically, an active firm can be thought of as committing to fight "entry," because its pricing function conditions on the investment of competitors. This type of commitment is absent in models of competitive screening with simultaneous contract posting described in what follows.

A "competitive" equilibrium with $n \geq 2$ active firms induces the random allocation given by $\boldsymbol{q}^C[X, Y]$, in which X and Y are the first- and second-order statistics of n i.i.d. draws from H_n, respectively. The resulting allocation is a contraction of the monopolist allocation (Figure 7b). This observation leads to a qualitative effect of competition on the inefficiencies of the monopolist allocation. First, competition exacerbates the productive inefficiency; i.e., $X < q^M$ with probability 1. Second, competition increases the allocation of low types reducing the damaging inefficiency; i.e., $Y \geq \boldsymbol{\beta}(0)$ with probability 1, and with strict

[11] If a buyer is indifferent across firms, then he purchases from the lowest-index firm. In equilibrium, the firm serving type $\theta \leq b(y)$ is the firm with lowest index i such that $q_i \geq y$.

inequality with positive probability.[12]

Corollary 1. *In every equilibrium with $n \geq 2$ active firms, with probability one: the highest distributed quality is strictly lower than q^M, and the lowest distributed quality is greater than $\boldsymbol{\beta}(0)$. Moreover, the lowest distributed quality is strictly greater than $\boldsymbol{\beta}(0)$ with positive probability.*

In the Appendix, we shut down either effect of competition on the inefficiency via appropriate cost functions. Thus, we cannot go beyond the above qualitative assessment of the impact of competition. Specifically, all firms make 0 profits in a competitive equilibrium, so welfare amounts to consumer surplus. Moreover, the relevant welfare comparison is between duopoly and monopoly as welfare decreases in the number of active firms, (starting from $n \geq 2$ active firms.) If full bunching occurs in $\boldsymbol{q}^M$—e.g., for steep c by Proposition 3—then the productive inefficiency is the dominant difference between the monopoly and the duopoly allocation; in this case, welfare is higher under monopoly than under duopoly. If, instead, c approximate a fixed-cost function, then the distributive effect dominates, and duopoly dominates duopoly in terms of welfare.

Alternative models of competition In our model, firms compete for buyers after the investment stage, which determines feasible contracts, is complete and publicly observed. Given this timing assumption, an "incumbent" firms effectively commits to fight an "entrant" in stage 2. Models of competitive screening can lead to non-existence of equilibria if firms can offer any contract taking as given the contracts offered by the opponents, differently from our model (Rothschild and Stiglitz, 1976).

The timing of our model is also adopted in the quality-screening duopoly of Champsaur and Rochet (1989), in which firms first commit to a quality spectrum at no costs, and then price feasible qualities. Firms make positive profits in equilibrium by playing disjoint quality intervals, which are ruled out in our model in which feasible quality sets take the form $[0, q]$. The space of feasible quality intervals is a key difference between the setups: in Champsaur and Rochet (1989), a firm choosing $[a, b]$ in stage 1 effectively commits not to sell qualities worse than a in stage 2, whereas this commitment is absent in our model.

[12]The lowest monopolist quality $\boldsymbol{\beta}(0)$ can exceed the second-highest quality, y. In this case, the competitive allocation is the same as the monopolist allocation for low types; i.e., types in $[0, b(x)]$ receive the same quality under $\boldsymbol{q}^M$ and under $\boldsymbol{q}^C[x, y]$. If, instead, $y \geq \boldsymbol{\beta}(0)$, then every type in $[0, b(y))$ gets a higher quality under $\boldsymbol{q}^C[x, y]$ than under $\boldsymbol{q}^M$, and for a price of 0.

There are alternative ways to model multiple firms and privately informed buyers, see Stole (2007) for a survey. In Garrett, Gomes, and Maestri (2019), asymmetric information is two-sided: consumers are imperfectly informed about the posted contracts. Their model generates menu dispersion and competition can raise prices of low-quality goods. In Johnson and Myatt (2003), an entrant and an incumbent move simultaneously, but the entrant can only produce qualities lower than a given threshold. This exogenous constraint mimics the equilibrium in the pricing stage of our game. The notion of "relevance" introduced by Chade and Swinkels (2021) does not induce equilibrium existence with our nonseparable costs. In their "vertical" oligopoly, with a contract-posting game à la Rothschild and Stiglitz, a form of market power induces equilibrium existence: each firm has a cost advantage over a quality interval. However, if producing q makes all qualities lower than q available for free, then a cost advantage "for" low qualities does not play any role; in fact, costs are separable in the model of Chade and Swinkels.

5 Conclusion

This paper extends the workhorse screening model by studying an alternative cost structure in which costs depend on the maximum quality. The optimum for the seller involves both the usual virtual-surplus analysis, and the association of each quality to a marginally bunched type, in order to determine the marginal revenue of a quality investment. A natural extension is to consider costs such that providing a quality to a type depends both on her own quality—as in MR—and on the entire allocation—as in this paper, in which the dependence occurs via the maximum of the allocation. This analysis can be used to model, e.g., economies of scale, and damaging costs.

Innovation in digital markets proceeds dynamically, and firms adjust research and production decisions over time, for example as decisions of competitors are observed, or new information about demand is gathered. An extension that captures a dynamic model of competition can inform policies related to patent protection, innovation, and the regulation of natural monopolies in digital markets.

A Proofs

A.1 General model

The following model nests the model in the main body of the paper. The set of qualities is $Q := [0, \overline{q}]$, an allocation is a measurable $\boldsymbol{q}\colon \Theta \to Q$, the set of allocations is $\boldsymbol{Q}$, F is a distribution function whose support is $\Theta := [0,1]$, and is $\mathcal{C}^2$. The utility $u\colon Q \times \Theta \to \mathbb{R}$ and the virtual surplus $J\colon (q,\theta) \mapsto u(q,\theta) - \frac{1-F(\theta)}{F'(\theta)} u_2(q,\theta) - k(q)$ satisfy increasing differences, and are: $\mathcal{C}^2$, concave in q for all $\theta \in \Theta$, strictly quasiconcave in q for all $\theta \in \Theta \backslash S$, for a countable $S \subseteq \Theta$, and a convex, $\mathcal{C}^2$, and nondecreasing $k\colon Q \to \mathbb{R}$. The function $c\colon Q \to \mathbb{R}$ is increasing, strictly convex, differentiable, and satisfies: $\lim_{q \to 0} \left(\int_\Theta J_1(q,\theta)\,\mathrm{d}F(\theta) - c'(q) \right) > 0 > \lim_{q \to \overline{q}} \left(\int_\Theta J_1(q,\right.$ We maintain the convention that $\inf \emptyset = 1$.

The present model nests the model in the main body of the paper except for the finite quality upper bound $\overline{q}$, because Q is unbounded in the paper. The only role of $\overline{q}$ is to compactify the quality space, as shown in the rest of the Appendix (Proposition A.1, A.2). Hence, the results in the main text are implied by the results that are proved in what follows.

Preliminaries The allocation $\boldsymbol{q}$ is *efficient* if $\boldsymbol{q}$ solves $\max_{\boldsymbol{q} \in \boldsymbol{Q}} \int_\Theta u(\theta, \boldsymbol{q}(\theta)) - k(\boldsymbol{q}(\theta))\,\mathrm{d}F(\theta) - c(\sup \boldsymbol{q})$. We let $U(q)$ be the value of the problem $\mathcal{P}^\star(q)$, for $q \in Q$,

$$U(q) = \sup_{\boldsymbol{q} \in \boldsymbol{Q}} \int_\Theta u(\boldsymbol{q}(\theta), \theta) - k(\boldsymbol{q}(\theta))\,\mathrm{d}F(\theta) \text{ subject to: } \boldsymbol{q}(\theta) \leq q \text{ for all } \theta \in \Theta.$$

The *surplus maximizer* is the largest selection $\boldsymbol{\alpha}$ from $\theta \mapsto \operatorname{Argmax}_{q \in Q} u(\theta, q) - k(q)$. The generalized inverse $a\colon q \mapsto \inf\{\theta \in \Theta : \boldsymbol{\alpha}(\theta) \geq q\}$ of $\boldsymbol{\alpha}$ is nondecreasing because u has increasing differences (Topkis, 1978), and satisfies: $a(q) = 0$ if $q < \boldsymbol{\alpha}(0)$, and $a(q) = 1$ if $q > \boldsymbol{\alpha}(1)$. Moreover, a is continuously differentiable at $q \in (\boldsymbol{\alpha}(0), \boldsymbol{\alpha}(1))$, with $a'(q) = -\frac{u_{11}(q,a(q)) - k''(q)}{u_{12}(q,a(q))}$ by the implicit function theorem.

The set of direct mechanisms that are incentive-compatible and individually rational is $\boldsymbol{M} := \{(\boldsymbol{q}, t) \in \boldsymbol{Q} \times \mathbb{R}^\Theta : u(\theta, \boldsymbol{q}(\theta)) - t(\theta) \geq u(\theta, \boldsymbol{q}(\hat{\theta})) - t(\hat{\theta}) \text{and} u(\theta, \boldsymbol{q}(\theta)) - t(\theta) \geq 0 \text{for all} (\theta, \hat{\theta}) \in \Theta^2\}$. The allocation $\boldsymbol{q}$ is *monopolist* if there exists $t\colon \Theta \to \mathbb{R}$ such that $(\boldsymbol{q}, t)$ solves $\sup_{(\boldsymbol{q},t) \in \boldsymbol{M}} \int_\Theta t(\theta)\,\mathrm{d}F(\theta) - c(\sup \boldsymbol{q}(\Theta))$. We let $V(q)$ be the value of the problem $\mathcal{P}(q)$, for $q \in Q$,

$$V(q) = \sup_{(\boldsymbol{q},t)\in \boldsymbol{M}} \int_\Theta t(\theta)\,\mathrm{d}F(\theta) \text{ subject to: } \boldsymbol{q}(\theta) \le q \text{ for all } \theta \in \Theta.$$

The *virtual surplus maximizer* is the largest selection $\boldsymbol{\beta}$ from $\theta \mapsto \operatorname{Argmax}_{q\in Q} J(q,\theta)$. The generalized inverse $b\colon q \mapsto \inf\{\theta \in \Theta : \boldsymbol{\beta}(\theta) \ge q\}$ of $\boldsymbol{\beta}$ is nondecreasing because J satisfies increasing differences, and satisfies: $b(q) = 0$ if $q < \boldsymbol{\beta}(0)$, and $b(q) = 1$ if $q > \boldsymbol{\beta}(1)$. Moreover, b is continuously differentiable at $q \in (\boldsymbol{\beta}(0), \boldsymbol{\beta}(1))$, with $b'(q) = -\frac{J_{11}(q,b(q))}{J_{12}(q,b(q))}$.

A.2 Proofs for Section 2

Proposition 1, Lemma 1, and Proposition 2 are implied by Proposition A.1, Lemma A.2, and Proposition A.2, respectively; a technical argument in the proof of Proposition A.1 and A.2 is relegated to Appendix B.

Lemma A.1. *The allocation $\boldsymbol{q}$ is efficient if and only if: $\boldsymbol{q}$ solves $\mathcal{P}^\star(q^\star)$ for a quality $q^\star \in \operatorname{Argmax}_{q\in Q} U(q) - c(q)$.*

Proof. We first show that $\boldsymbol{q}$ is efficient only if $\boldsymbol{q}$ solves $\mathcal{P}^\star(q^\star)$ for a quality $q^\star \in \operatorname{Argmax}_{q\in Q} U(q) - c(q)$. Let $\boldsymbol{q}$ be efficient. We want to show that: there exists $q^\star \in Q$ such that: (1) $\sup \boldsymbol{q} \le q^\star$, (2) for all $\tilde{\boldsymbol{q}} \in \boldsymbol{Q}$ with $\sup \tilde{\boldsymbol{q}} \le q^\star$, $\int_\Theta u(\boldsymbol{q}(\theta),\theta) - k(\boldsymbol{q}(\theta))\,\mathrm{d}F(\theta) \ge \int_\Theta u(\tilde{\boldsymbol{q}}(\theta),\theta) - k(\tilde{\boldsymbol{q}}(\theta))\,\mathrm{d}F(\theta)$, and (3) for all $q \in Q$, $U(q^\star) \ge U(q)$. We claim that $\sup \boldsymbol{q}$ satisfies the three properties. First, consider property (2). By efficiency of $\boldsymbol{q}$, for every $\tilde{\boldsymbol{q}} \in \boldsymbol{Q}$,

$$\int_\Theta u(\boldsymbol{q}(\theta),\theta) - k(\boldsymbol{q}(\theta))\,\mathrm{d}F(\theta) - c(\sup \boldsymbol{q}) \ge \int_\Theta u(\tilde{\boldsymbol{q}}(\theta),\theta) - k(\tilde{\boldsymbol{q}}(\theta))\,\mathrm{d}F(\theta) - c(\sup \tilde{\boldsymbol{q}}).$$

As an implication of efficiency and the fact that c is increasing, we have that: for every $\tilde{\boldsymbol{q}} \in \boldsymbol{Q}$, if $c(\tilde{\boldsymbol{q}}) < c(\boldsymbol{q})$, then $\int_\Theta u(\boldsymbol{q}(\theta),\theta) - k(\boldsymbol{q}(\theta))\,\mathrm{d}F(\theta) > \int_\Theta u(\tilde{\boldsymbol{q}}(\theta),\theta) - k(\tilde{\boldsymbol{q}}(\theta))\,\mathrm{d}F(\theta)$. Hence, property (2) holds. Moreover, $U(\sup \boldsymbol{q}) = \int_\Theta u(\boldsymbol{q}(\theta),\theta) - k(\boldsymbol{q}(\theta))$, so (3) and (1) hold.

We proceed to show that $\boldsymbol{q}$ is efficient if $\boldsymbol{q}$ solves $\mathcal{P}^\star(q^\star)$ for a quality $q^\star \in \operatorname{Argmax}_{q\in Q} U(q) - c(q)$. Towards establishing an intermediate claim, let $X := \boldsymbol{Q}$, $Y := Q$, for all $(x,y) \in X \times Y$, $f(x,y) := \int_\Theta u(x(\theta),\theta) - k(x(\theta))\,\mathrm{d}F(\theta) - c(y)$, and $h(x) := \sup x(\Theta)$. Denote by v_1 the value of the efficient allocation, $v_1 := \sup\{f(x,h(x)) \mid x \in X\}$, and v_2 the value of the decomposed problem, $v_2 := \sup\{\sup\{f(x,y) \mid x \in X,\ h(x) = y\} \mid y \in Y\}$. By construction, we have that $v_1 = \sup\{f(x, y) \mid x \in X,\, y \in Y,\, h(x) = y\}$. We claim that $v_1 = v_2$. We show that $v_2 \geq v_1$. For all $y \in Y$, $x \in X$, if $h(x) = y$, then

$$f(x, y) \leq \sup\{f(\tilde{x}, y) \mid \tilde{x} \in X,\, h(\tilde{x}) = y\}.$$

Thus, for all $y \in Y$, $x \in X$, if $h(x) = y$, then $f(x, y) \leq v_2$. Therefore, we have that $v_2 \geq v_1$. We show that $v_1 \geq v_2$. For all $y \in Y$, $x \in X$, if $h(x) = y$, then

$$f(x, y) \leq v_1.$$

Therefore, it holds that $v_1 \geq v_2$. We conclude that $v_1 = v_2$.

Let $\boldsymbol{q}$ solve $\mathcal{P}^\star(q^\star)$ for a quality $q^\star \in \operatorname{Argmax}_{q \in Q} U(q) - c(q)$. By construction, we have that $f(\boldsymbol{q}, \sup \boldsymbol{q}) \geq f(\boldsymbol{q}, q^\star)$, $v_2 = f(\boldsymbol{q}, q^\star)$, and $v_1 \geq f(\boldsymbol{q}, \sup \boldsymbol{q})$. We conclude that:

$$v_2 \geq f(\boldsymbol{q}, \sup \boldsymbol{q}) \geq v_1.$$

Therefore, $\boldsymbol{q}$ is efficient. □

Proposition A.1. *Let $q^\star$ be the unique quality q such that $\int_{[a(q),1]} u_1(q, \theta) - k'(q)\, \mathrm{d}F(\theta) = c'(q)$. The allocation $\boldsymbol{q}$ is efficient if and only if: there exists an allocation $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma}(\theta) = \boldsymbol{\alpha}(\theta)$ almost everywhere and $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q^\star\}$ for all θ.*

Proof. The proof has three main steps. First, we solve $\mathcal{P}^\star(q)$ for all $q \in Q$; second, we show that U is continuously differentiable and concave; third, we show that an interior quality solves $\max_{q \in Q} U(q) - c(q)$. Then, the result follows from Lemma A.1.

Claim: for fixed $q \in Q$, $\boldsymbol{q}$ solves $\mathcal{P}^\star(q)$ iff $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q\}$ for all θ and some $\boldsymbol{\gamma} \in \boldsymbol{Q}$ such that $\boldsymbol{\gamma}(\theta) = \boldsymbol{\alpha}(\theta)$ almost everywhere. By Lemma B.1, $\boldsymbol{q}$ solves $\mathcal{P}^\star(q)$ iff $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q\}$ for all θ and some allocation $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma}(\theta) \in \operatorname{Argmax}_{\hat{q} \in Q} u(\hat{q}, \theta) - k(\hat{q})$ almost everywhere. We have $\boldsymbol{\gamma}(\theta) = \max \operatorname{Argmax}_{\hat{q} \in Q} u(\hat{q}, \theta) - k(\hat{q})$ a.e. by strict quasiconcavity of $u(\cdot, \theta) - k(\cdot)$, so the claim holds.

Claim: U is concave, continuously differentiable, and $U'(q) = \int_{[a(q),1]} u_1(q, \theta) - k'(q)\, \mathrm{d}F(\theta)$ for all $q \in (0, \overline{q})$. First, U is differentiable at $q < \boldsymbol{\alpha}(0)$ with $U'(q) = \int_{[0,1]} u_1(q, \theta) - k'(q)\, \mathrm{d}F(\theta)$ if $q > 0$; U is differentiable at $q \in (\boldsymbol{\alpha}(0), \boldsymbol{\alpha}(1))$ with $U'(q) = \int_{[a(q),1]} u_1(q, \theta) - k'(q)\, \mathrm{d}F(\theta)$ because a is continuously differentiable; finally, U is differentiable at $q > \boldsymbol{\alpha}(1)$ with $U'(q) = 0$ if $q < \overline{q}$. The previous derivatives are continuously pasted at $\boldsymbol{\alpha}(0)$ and $\boldsymbol{\alpha}(1)$, so U is continuously differentiable. For the claim, it suffices to establish that $U'(q_2) - U'(q_1) \leq 0$ for all $q_1, q_2 \in Q$ with $q_2 > q_1$. It holds that

$$U'(q_2) - U'(q_1) = \int_{[a(q_2),1]} u_1(q_2,\theta) - u_1(q_1,\theta) - k'(q_2) + k'(q_1)\,\mathrm{d}F(\theta) - \int_{[a(q_1),a(q_2))} u_1(q_1,\theta) - k'(q_1)\,\mathrm{d}F(\theta).$$

Moreover, $u_1(q_2,\theta) - u_1(q_1,\theta) - k'(q_2) + k'(q_1) \leq 0$ by concavity of $q \mapsto u(q,\theta) - k(q)$ for all $\theta \in \Theta$, and $u_1(q_1,\theta) - k'(q_1) \geq 0$ for all $\theta \geq a(q_1)$ by definition of a and $\boldsymbol{\alpha}$. Hence, U is concave.

Claim: U is right continuous at 0. For all $q > 0$, we have: $U(q) \geq \int_\Theta u(0,\theta) - k(0)\,\mathrm{d}F(\theta)$; moreover, if q is sufficiently small, we have $U(q) \leq \int_\Theta u(q,\theta) - k(0)\,\mathrm{d}F(\theta)$ by our assumption on c'. Hence, we have $U(q) \to \int_\Theta u(0,\theta) - k(0)\,\mathrm{d}F(\theta)$ as $q \to 0$ by continuity of u in quality.

By the properties of c', if $q \in \operatorname{Argmax}_{q\in Q} U(q) - c(q)$ then $q \in (0, \overline{q})$, so the proof is complete. □

Lemma A.2. *The allocation $\boldsymbol{q}$ is monopolist if and only if: $\boldsymbol{q}$ solves $\mathcal{P}(q^M)$ for a quality $q^M \in \operatorname{Argmax}_{q\in Q} V(q) - c(q)$.*

Proof. This proof follows the same steps as in the proof of Lemma A.1. Two preliminary observations follow from standard arguments (Carroll, 2023). First, we have

$$V(q) = \max_{\boldsymbol{q}\in\boldsymbol{Q}} \int_\Theta J(\boldsymbol{q}(\theta),\theta)\,\mathrm{d}F(\theta) \text{ subject to: } \boldsymbol{q}(\theta) \leq q \text{ for all } \theta \in \Theta,\ \boldsymbol{q} \text{ is nondecreasing;}$$

and $(\boldsymbol{q}, t)$ solves $\mathcal{P}(q)$ for some t iff $\boldsymbol{q}$ solves the above problem. Second, we have that: $\boldsymbol{q}$ is monopolist iff $\boldsymbol{q}$ solves

$$\max_{\boldsymbol{q}\in\boldsymbol{Q}} \int_\Theta J(\boldsymbol{q}(\theta),\theta)\,\mathrm{d}F(\theta) \text{ subject to: } \boldsymbol{q} \text{ is nondecreasing,}$$

We first show that $\boldsymbol{q}$ is monopolist only if $\boldsymbol{q}$ solves $\mathcal{P}(q^M)$ for a quality $q^M \in \operatorname{Argmax}_{q\in Q} V(q) - c(q)$. Let $\boldsymbol{q}$ be monopolist. We want to show that: there exists $q^M \in Q$ such that: (1) $\sup \boldsymbol{q} \leq q^M$, (2) for all nondecreasing $\tilde{\boldsymbol{q}} \in \boldsymbol{Q}$ with $\sup \tilde{\boldsymbol{q}} \leq q^M$, $\int_\Theta J(\boldsymbol{q}(\theta),\theta)\,\mathrm{d}F(\theta) \geq \int_\Theta J(\tilde{\boldsymbol{q}}(\theta),\theta)\,\mathrm{d}F(\theta)$, and (3) for all $q \in Q$, $V(q^M) \geq V(q)$. We claim that $\sup \boldsymbol{q}$ satisfies the three properties. First, consider property (2). For every $\tilde{\boldsymbol{q}} \in \boldsymbol{Q}$, we have

$$\int_{\Theta} J(\boldsymbol{q}(\theta), \theta) \, \mathrm{d}F(\theta) - c(\sup \boldsymbol{q}) \geq \int_{\Theta} J(\tilde{\boldsymbol{q}}(\theta), \theta) \, \mathrm{d}F(\theta) - c(\sup \tilde{\boldsymbol{q}}).$$

Because c is increasing, we have that: for every $\tilde{\boldsymbol{q}} \in \boldsymbol{Q}$, if $c(\tilde{\boldsymbol{q}}) < c(\boldsymbol{q})$, then $\int_{\Theta} J(\boldsymbol{q}(\theta), \theta) \, \mathrm{d}F(\theta) > \int_{\Theta} J(\tilde{\boldsymbol{q}}(\theta), \theta) \, \mathrm{d}F(\theta)$. Hence, property (2) holds. Moreover, $V(\sup \boldsymbol{q}) = \int_{\Theta} J(\boldsymbol{q}(\theta), \theta)$, so (3) and (1) hold.

We proceed to show that $\boldsymbol{q}$ is monopolist if $\boldsymbol{q}$ solves $\mathcal{P}(q^M)$ for a quality $q^M \in \operatorname{Argmax}_{q \in Q} V(q) - c(q)$. Towards establishing an intermediate claim, let $X := \{\boldsymbol{q} \in \boldsymbol{Q} \mid \boldsymbol{q} \text{ is nondecreasing}\}$, $Y := Q$, for all $(x, y) \in X \times Y$, $f(x, y) := \int_{\Theta} J(x(\theta), \theta) \, \mathrm{d}F(\theta) - c(y)$, and $h(x) := \sup x(\Theta)$. We denote by v_1 the profit of the monopolist allocation, $v_1 := \sup\{f(x, h(x)) \mid x \in X\}$, and v_2 the value of the decomposed problem, $v_2 := \sup\{\sup\{f(x, y) \mid x \in X,\ h(x) = y\} \mid y \in Y\}$. By the same steps as in the proof of Lemma A.1, we conclude that $v_1 = v_2$.

Let $\boldsymbol{q}$ solve $\mathcal{P}(q^M)$ for a quality $q^M \in \operatorname{Argmax}_{q \in Q} V(q) - c(q)$. By construction, we have that $f(\boldsymbol{q}, \sup \boldsymbol{q}) \geq f(\boldsymbol{q}, q^M)$, $v_2 = f(\boldsymbol{q}, q^M)$, and $v_1 \geq f(\boldsymbol{q}, \sup \boldsymbol{q})$. We conclude that

$$v_2 \geq f(\boldsymbol{q}, \sup \boldsymbol{q}) \geq v_1.$$

Therefore, $\boldsymbol{q}$ is monopolist. $\square$

Proposition A.2. *Let q^M be the unique quality q such that $\int_{[b(q),1]} J_1(q, \theta) \, \mathrm{d}F(\theta) = c'(q)$. The allocation $\boldsymbol{q}$ is monopolist if and only if: there exists a nondecreasing allocation $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma}(\theta) = \boldsymbol{\beta}(\theta)$ almost everywhere and $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q^M\}$ for all θ. Moreover, it holds that $0 < q^M < q^\star$.*

Proof. The proof has three main steps. First, we solve $\mathcal{P}(q)$ for all $q \in Q$; second, we show that V is continuously differentiable and concave; third, we show that an interior quality below $q^\star$ solves $\max_{q \in Q} V(q) - c(q)$. Then, the result follows from Lemma A.2.

Claim: for fixed $q \in Q$, $\boldsymbol{q}$ solves $\mathcal{P}(q)$ iff $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q\}$ for all θ and some nondecreasing $\boldsymbol{\gamma} \in \boldsymbol{Q}$ with $\boldsymbol{\gamma}(\theta) = \boldsymbol{\beta}(\theta)$ almost everywhere. For necessity, let $\boldsymbol{q} \in \boldsymbol{Q}$ be such that $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q\}$ for all $\theta \in \Theta$ and for some nondecreasing allocation $\boldsymbol{\gamma}$ with $\boldsymbol{\gamma}(\theta) = \boldsymbol{\beta}(\theta)$ almost everywhere. Then, $\boldsymbol{q}$ solves the problem without the monotonicity constraint (Lemma B.1) and satisfies the monotonicity constraint, so $\boldsymbol{q}$ solves $\mathcal{P}(q)$. For sufficiency, let $\boldsymbol{q}$ solve $\mathcal{P}(q)$ and be such that: there does not exist a nondecreasing allocation $\boldsymbol{\gamma}$ with $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q\}$

for all θ and $\boldsymbol{\gamma}(\theta) = \boldsymbol{\beta}(\theta)$ almost everywhere. By Lemma B.1 and the strict-quasiconcavity property of $J(\cdot, \theta)$, $\boldsymbol{q}$ does not solve the problem without the monotonicity constraint. We observe that the function $\theta \mapsto \min\{\boldsymbol{\beta}(\theta), q\}$ is an allocation that solves the problem without the monotonicity constraint (Lemma B.1) and satisfies the monotonicity constraint, so $\boldsymbol{q}$ does not solve $\mathcal{P}(q)$. The claim follows from the preliminary observation in the proof of Lemma A.2.

Claim: V is concave, continuously differentiable, and $V'(q) = \int_{[b(q),1]} J_1(q,\theta)\,\mathrm{d}F(\theta)$ for all $q \in (0, \overline{q})$. First, V is differentiable at $q < \boldsymbol{\beta}(0)$ with $V'(q) = \int_{[0,1]} J_1(q,\theta)\,\mathrm{d}F(\theta)$ if $q > 0$; U is differentiable at $q \in (\boldsymbol{\beta}(0), \boldsymbol{\beta}(1))$ with $V'(q) = \int_{[b(q),1]} J_1(q,\theta)\,\mathrm{d}F(\theta)$ because b is continuously differentiable; finally, V is differentiable at $q > \boldsymbol{\beta}(1)$ with $V'(q) = 0$ if $q < \overline{q}$. The previous derivatives are continuously pasted at $\boldsymbol{\beta}(0)$ and $\boldsymbol{\beta}(1)$. For the claim, it suffices to establish that $V'(q_2) - V'(q_1) \leq 0$ for all $q_1, q_2 \in Q$ with $q_2 > q_1$. It holds that

$$V'(q_2) - V'(q_1) = \int_{[b(q_2),1]} J_1(q_2,\theta) - J_1(q_1,\theta)\,\mathrm{d}F(\theta) - \int_{[b(q_1),b(q_2)]} J_1(q_1,\theta)\,\mathrm{d}F(\theta).$$

Moreover, $J_1(q_2,\theta) - J_1(q_1,\theta) \leq 0$ by concavity of $q \mapsto J(q,\theta)$ for all $\theta \in \Theta$, and $J_1(q_1,\theta) \geq 0$ for all $\theta \geq b(q_1)$ by definition of b, $\boldsymbol{\beta}$. Hence, V is concave.

Claim: V is right continuous at 0. For all $q > 0$, we have: $\int_\Theta u(q,\theta) - \frac{1-F(\theta)}{F'(\theta)} u_2(0,\theta) - k(0)\,\mathrm{d}F(\theta) \geq V(q) \geq \int_\Theta J(0,\theta)\,\mathrm{d}F(\theta)$. Hence, we have $V(q) \to \int_\Theta J(0,\theta)\,\mathrm{d}F(\theta)$ as $q \to 0$.

Claim: $\boldsymbol{\beta}(\theta) \leq \boldsymbol{\alpha}(\theta)$ for all $\theta \in \Theta$, and, additionally, $\boldsymbol{\beta}(\theta) < \boldsymbol{\alpha}(\theta)$ if: $\boldsymbol{\alpha}(\theta) > 0$, $\boldsymbol{\beta}(\theta) < \overline{q}$, and $\theta < 1$. It suffices to observe that $u_1(q,\theta) - k'(q) - J_1(q,\theta) = \frac{1-F(\theta)}{F'(\theta)} u_{12}(q,\theta)$ for all $(q,\theta) \in Q \times \Theta$.

Claim: $q^M < q^\star < \overline{q}$. First, we observe that, if $q \in (0, \overline{q})$, then: $a(q) < b(q)$, and, moreover,

$$U'(q) - V'(q) = \int_{[b(q),1]} \frac{1-F(\theta)}{F'(\theta)} u_{12}(q,\theta)\,\mathrm{d}F(\theta) + \int_{[a(q),b(q))} u_q(q,\theta) - k'(q)\,\mathrm{d}F(\theta).$$

It follows that $U'(q) - V'(q) \geq \int_{[a(q),b(q)]} u_q(q,\theta) - k'(q)\,\mathrm{d}F(\theta) > 0$ by the definition of a. Hence, we have that $q^M < q^\star < \overline{q}$.

To complete the proof, it remains to show that $q^M > 0$. We let the right-continuous inverse of $\boldsymbol{\beta}$ be $T^+ \colon q \mapsto \inf\{\theta \in \Theta : \boldsymbol{\beta}(\theta) > q\}$. Observe that: $\lim_{q \to 0} V'(q) = \int_{[T^+(0),1]} J_1(0,\theta)\,\mathrm{d}F(\theta)$. If $1 > T^+(0) > 0$, by increasing differences we have

$$\frac{1}{1 - F(T^+(0))} \lim_{q \to 0} V'(q) > \int_{[0,1]} J_1(0, \theta) \, \mathrm{d}F(\theta)$$

We note that $1 > T^+(0)$, because $J_1(0, 1) > 0$ and $\boldsymbol{\beta}$ is continuous. If, instead, $T^+(0) = 0$, then $\lim_{q \to 0} V'(q) = \int_{[0,1]} J_1(0, \theta) \, \mathrm{d}F(\theta)$. By the properties of c', it follows that $q^M > 0$. $\square$

A.3 Proofs for Section 3

Proof of Proposition 3. For this proof, we assume that u is as in the main text. Define, for all $q \in Q$, $G(q, \kappa_c, \kappa_g) := (1 - F(b_{\kappa_g}(q)))(b_{\kappa_g}(q) + \kappa_g g'(q)) - \kappa_c c'(q)$. By Proposition A.2, q^M_κ is interior, and is the unique quality q such that $G(q, \kappa_c, \kappa_g) = 0$.

Claim 1: The cap q^M_κ is decreasing in κ_c. $G_2(q, \kappa_c, \kappa_g) = -c'(q)$ and $G_1(q, \kappa_c, \kappa_g) < 0$ by strict concavity of $q \mapsto V(q) - \kappa_c c(q)$. By implicit function theorem: $\frac{\partial}{\partial \kappa_c} q^M_\kappa = -G_2(q^M_\kappa, \kappa_c, \kappa_g)/G_1(q^M_\kappa, \kappa_c, \kappa_g)$. Hence, q^M_κ is decreasing in κ_c.

Claim 2: The cap q^M_κ is nondecreasing in κ_g. Fix q with $b_{\kappa_g}(q) \in (0, 1)$. It holds that

$$\begin{aligned} G_3(q, \kappa_c, \kappa_g) &= (1 - F(b_{\kappa_g}(q)))(g'(q) + \frac{\partial}{\partial \kappa_g} b_{\kappa_g}(q)) - f(b_{\kappa_g}(q)) \frac{\partial}{\partial \kappa_g} b_{\kappa_g}(q)(b_{\kappa_g}(q) + \kappa_g g'(q)) \\ &= (1 - F(b_{\kappa_g}(q)))g'(q), \end{aligned}$$

using $\kappa_g g'(q) + \varphi(b_{\kappa_g}(q)) = 0$ for the second equality, and $G_1(q, \kappa_c, \kappa_g) < 0$ by strict concavity of $q \mapsto V(q) - \kappa_c c(q)$. By implicit function theorem: $\frac{\partial}{\partial \kappa_g} q^M_\kappa = -G_3(q^M_\kappa, \kappa_c, \kappa_g)/G_1(q^M_\kappa, \kappa_c, \kappa_g)$. Hence, q^M_κ is nondecreasing in κ_g.

Claim 3: b_{κ_g} is nonincreasing in κ_g pointwise. Fix q with $b_{\kappa_g}(q) \in (0, 1)$. From $\kappa_g g'(q) + \varphi(b_{\kappa_g}(q)) = 0$, by the implicit function theorem: $\frac{\partial}{\partial \kappa_g} b_{\kappa_g}(q) = -g'(q)/\varphi'(b_{\kappa_g}(q))$. Hence, for every $q \in Q$, $b_{\kappa_g}(q)$ is nonincreasing in κ_g.

Claim 4: $b_{\kappa_g}(q^M_\kappa)$ is nonincreasing in κ_g, decreasing if interior. This claim follows from claim 2 and 3.

Claim 5: there exists $\overline{\kappa}_g \geq 0$ such that $b_{\kappa_g}(q^M_\kappa) = 0$ for all $\kappa_g > \overline{\kappa}_g$. By the previous claims, $b_{\kappa_g}(q^M_\kappa) \to 0$ as $\kappa_g \to \infty$. If the claim does not hold, then $\kappa_g g'(q^M_\kappa) \to -\varphi(0)$, which is impossible by concavity of g and monotonicity of q^M_κ in κ_g. $\square$

Proof of Proposition 4. For this proof, we assume that u is as in the main text.

We show that, if $0 \leq \theta < b(q^M)$, then $\boldsymbol{q}^{MRE}(\theta) > \boldsymbol{q}^M(\theta)$. By hypothesis,

$$c'(\boldsymbol{\beta}(\theta)) < c'(q^M) = (1 - F(b(q^M)))u_1(q^M, b(q^M))$$

in which the rightmost inequality follows from Proposition A.2. By strict concavity of V, we have

$$(1 - F(b(q^M)))u_1(q^M, b(q^M)) < (1 - F(\theta))u_1(\boldsymbol{\beta}(\theta), \theta) \leq u_1(\boldsymbol{\beta}(\theta), \theta),$$

We conclude that $u_1(\boldsymbol{\beta}(\theta), \theta) - c'(\boldsymbol{\beta}(\theta)) > 0$. Therefore, $\boldsymbol{\beta}(\theta) < \boldsymbol{q}^{MRE}(\theta)$.

The rest of the proof follows from the argument sketched in the text and is omitted. $\square$

A.4 Proofs for Section 4

Proposition 5 and 6 are implied by Proposition A.3 and A.4, respectively.

A.4.1 Non-discriminating monopolist

The zero of the virtual surplus is $b_N : q \mapsto \inf\{\theta \in \Theta \mid J(q, \theta) \geq 0\}$. The allocation $\boldsymbol{q}$ is *no screening* if $\boldsymbol{q}$ solves

$$\max_{(\boldsymbol{q}, t) \in \boldsymbol{M}} \int_{\Theta} t(\theta) \, \mathrm{d}F(\theta) - c(\sup \boldsymbol{q}(\Theta)) \text{ subject to: } \boldsymbol{q}(\Theta) \subseteq \{0, \hat{q}\} \text{ for some } \hat{q} \in Q.$$

We let $V_N(q)$ be the value of the problem $\mathcal{P}_N(q)$, for quality $q \in Q$,

$$V_N(q) := \max_{(\boldsymbol{q}, t) \in \boldsymbol{M}} \int_{\Theta} t(\theta) \, \mathrm{d}F(\theta) \text{ subject to: } \boldsymbol{q}(\Theta) \subseteq \{0, \hat{q}\} \text{ for some } \hat{q} \in Q,$$
$$\boldsymbol{q}(\theta) \leq q \text{ for all } \theta \in \Theta.$$

We use Iverson brackets: $[P] = 1$ if the statement P is true, and $[P] = 0$ otherwise.

Proposition A.3. *Assume that $J(0, \theta) = 0$ for all θ, and $J(q, \cdot)$ is increasing for all $q > 0$; let q_N^M be the unique quality q such that $\int_{[b_N(q),1]} J_1(q, \theta) \, \mathrm{d}F(\theta) = c'(q)$. The allocation $\boldsymbol{q}_N^M$ is no screening if and only if: $\boldsymbol{q}_N^M(\theta) = [\theta \geq b_N(q_N^M)]q_N^M$ for all $\theta \in \Theta \setminus \{b_N(q_N^M)\}$, and $\boldsymbol{q}_N^M(b_N(q_N^M)) \in \{0, q_N^M\}$. Moreover, it holds that:*

1. *$0 < q_N^M \leq q^M$;*
2. *$q_N^M < q^M$ if $b(q^M) > b_N(q^M)$.*

Proof. The proof has three main steps. First, we solve $\mathcal{P}_N(q)$ for all $q \in Q$; second, we show that V_N is continuously differentiable and concave; third, we show that $q_N^M \leq q^M$, with strict inequality if $q^M > 0$. The following preliminary observations hold by known arguments:

$$V_N(q) = \max_{\boldsymbol{q} \in \boldsymbol{Q}} \int_\Theta J(\boldsymbol{q}(\theta), \theta) \, \mathrm{d}F(\theta) \text{ subject to: } \boldsymbol{q}(\theta) \leq q \text{ for all } \theta \in \Theta,$$

$$\boldsymbol{q} \text{ is nondecreasing, } \boldsymbol{q}(\Theta) \subseteq \{0, \hat{q}\} \text{ for some } \hat{q} \in Q;$$

$(\boldsymbol{q}, t)$ solves $\mathcal{P}_N(q)$ for some t iff $\boldsymbol{q}$ solves the above problem; the allocation $\boldsymbol{q}$ is no screening iff: $\boldsymbol{q}$ solves $\mathcal{P}_N(q_N^M)$ for a quality $q_N^M \in \operatorname{Argmax}_{q \in Q} V_N(q) - c(q)$.

Claim: $b(q) \geq b_N(q)$ for all $q \in Q$. It suffices to show that $\boldsymbol{\beta}(\theta) \geq q$ implies $J(q, \theta) \geq 0$ for fixed $(\theta, q) \in \Theta \times Q$. By concavity of $J(\cdot, \theta)$, we have $J(q, \theta) = \int_{[0,q]} J_1(\tilde{q}, \theta) \, \mathrm{d}\tilde{q}$, and $\boldsymbol{\beta}(\theta) \geq q$ holds iff $J_1(q, \theta) \geq 0$ holds. Hence, $\boldsymbol{\beta}(\theta) \geq q$ implies $J(q, \theta) \geq 0$.

Claim: for fixed $q \in Q$, $\boldsymbol{q}$ solves $\mathcal{P}_N(q)$ iff $\boldsymbol{q}(\theta) = [\theta \geq b_N(q)]q$ for all $\theta \in \Theta$. We first argue $\boldsymbol{q}$ solves $\mathcal{P}_N(q)$ only if $\boldsymbol{q}(\theta) = [\theta \geq b_N(q)]q$ almost everywhere. Fix $q \in (0, \overline{q}]$, a solution $\boldsymbol{q}_N^M$ to $\mathcal{P}_N(q)$, and suppose that there exists $(\theta', \theta'') \subseteq \Theta$ such that: $\boldsymbol{q}'(\theta) \neq \boldsymbol{q}_N^M(\theta)$ a.e. on (θ', θ''), for $\boldsymbol{q}' \colon \theta \mapsto [\theta \geq b_N(q)]q$. Define $S \coloneqq \{\theta \in (\theta', \theta'') : \boldsymbol{q}'(\theta) = q\}$ and $\Delta \coloneqq \int_{(\theta', \theta'')} J(\boldsymbol{q}'(\theta), \theta) - J(\boldsymbol{q}_N^M(\theta), \theta) \, \mathrm{d}F(\theta)$, it holds that

$$\Delta = \int_S \int_{[\boldsymbol{q}_N^M(\theta), q]} J_1(\tilde{q}, \theta) \, \mathrm{d}\tilde{q} \, \mathrm{d}F(\theta) - \int_{(\theta', \theta'') \setminus S} J(\boldsymbol{q}_N^M(\theta), \theta) \, \mathrm{d}F(\theta).$$

For $\hat{q} \in [\boldsymbol{q}_N^M(\theta), q]$, it holds that

$$J_1(\hat{q}, \theta) \geq J_1(q, \theta) \geq J_1(q, \inf S) \geq \frac{J(q, \inf S)}{q} \geq 0$$

a.e. on S, in which the inequalities follow from: concavity of $J(\cdot, \theta)$, increasing differences, concavity of $J(\cdot, \theta)$, and the definition of S, $b_N(q)$, from left to right. Moreover, it holds that $J(\boldsymbol{q}_N^M(\theta), \theta) \leq 0$ a.e. on $(\theta', \theta'') \setminus S$. Moreover, using increasing differences and the fact that $J(\hat{q}, \cdot)$ is increasing for $\hat{q} > 0$, we have that: either $\int_S \int_{[\boldsymbol{q}_N^M(\theta), q]} J_1(\tilde{q}, \theta) \, \mathrm{d}\tilde{q} \, \mathrm{d}F(\theta) > 0$, or $\int_{(\theta', \theta'') \setminus S} J(\boldsymbol{q}_N^M(\theta), \theta) \, \mathrm{d}F(\theta) < 0$, or both. It follows that $\Delta > 0$. By incentive compatibility, the solution to $\mathcal{P}_N(q)$ is nondecreasing, by the binary-image constraint, the solution does not differ from $\theta \mapsto [\theta \geq b_N^M(q)]$ if $\theta \neq b_N(q)$.

For the other direction, let $\boldsymbol{q}$ solve $\mathcal{P}_N(q)$ and not be equal to $\boldsymbol{q}'$ (identified in the previous paragraph) almost everywhere. By the previous argument, $\boldsymbol{q}'$ attains a strictly higher value of the maximand in $\mathcal{P}_N(q)$ than $\boldsymbol{q}$. Moreover, $\boldsymbol{q}'$ is nondecreasing and satisfies $\boldsymbol{q}'(\Theta) \subseteq \{0, q\}$. The claim follows.

Claim: b_N is nondecreasing. We proceed in three steps. First, we use the definition of $\boldsymbol{\beta}$ and the fact that b upper bounds b_N to show that: $J(q+\varepsilon, b_N(q)) \leq J(q, b_N(q))$, $(\varepsilon, q) \in (0, \overline{q} - q) \times (0, \overline{q})$. Suppose that $J(q+\varepsilon, b_N(q)) > J(q, b_N(q))$. Then, there exists $q' > q$ such that $J_1(q', b_N(q)) > 0$. Then, by concavity, either $\boldsymbol{\beta}(b_N(q)) = \overline{q}$, or there exists $q'' \geq q'$ such that $J_1(q'', b_N(q)) \leq 0$, or both. In all cases: $\boldsymbol{\beta}(b_N(q)) > q$, which contradicts the definition of b. Second, we observe that: if $J(q+\varepsilon, b_N(q)) < J(q, b_N(q))$, then $b_N(q+\varepsilon) > b_N(q)$ by increasing differences. Third, we show that $J(q+\varepsilon, b_N(q)) = J(q, b_N(q))$ only if $b_N(q+\varepsilon) = b_N(q)$ and distinguish among three exhaustive and mutually exclusive cases, for given $q \in Q$: (A) $J(q, 0) \geq 0$; (B) $J(q, 1) \leq 0$; (C) $b_N(q) \in (0, 1)$. In case (A), $b_N(q+\varepsilon) = b_N(q)$ because, otherwise: $b_N(q+\varepsilon) > b_N(q) = 0$, so $J(q+\varepsilon, b_N(q+\varepsilon)) > J(q+\varepsilon, b_N(q)) \geq 0$, which contradicts the definition of b_N. In case (B), $b_N(q+\varepsilon) = b_N(q)$ because, otherwise: $b_N(q+\varepsilon) < b_N(q) = 1$, so $J(q+\varepsilon, b_N(q+\varepsilon)) < J(q+\varepsilon, b_N(q)) \leq 0$, which contradicts the definition of b_N. In case (C), $b_N(q+\varepsilon) = b_N(q)$ by increasing differences. Hence, b_N is nondecreasing.

Claim: V_N is continuously differentiable, concave, and $V_N'(q) = \int_{[b_N(q),1]} J_1(q, \theta)\,\mathrm{d}F(\theta)$. First, it holds that b_N satisfies: $b_N(q) = 0$ if $q \leq c$, and $b_N(q) = 1$ if $q \geq d$, for some $c \leq d$, because b_N is nondecreasing. Moreover, b_N is continuously differentiable at $q \in (c, d)$, with $b_N'(q) = -\frac{J_1(q, b_N(q))}{J_2(q, b_N(q))}$, by the implicit function theorem. As an implication, by the same argument used in the proof of Proposition A.2, V_N is continuously differentiable at $q \in (0, \overline{q})$ with derivative $\int_{[b_N(q),1]} J_1(q, \theta)\,\mathrm{d}F(\theta)$. For the claim, it suffices to establish that $V_N'(q_2) - V_N'(q_1) \leq 0$ for all $q_1, q_2 \in Q$ with $q_2 > q_1$. It holds that

$$V_N'(q_2) - V_N'(q_1) = \int_{[b_N(q_2),1]} J_1(q_2, \theta) - J_1(q_1, \theta)\,\mathrm{d}F(\theta) - \int_{[b_N(q_1), b_N(q_2)]} J_1(q_1, \theta)\,\mathrm{d}F(\theta).$$

Moreover, $J_1(q_2, \theta) - J_1(q_1, \theta) \leq 0$ by concavity of $q \mapsto J(q, \theta)$ for all $\theta \in \Theta$, and $J_1(q_1, \theta) \geq 0$ for all $\theta \geq b_N(q_1)$ by concavity of $J_1(\cdot, \theta)$ and the definition of b_N. Hence, V_N' is concave; V_N is right continuous at 0 by the same argument used in the proof of Proposition A.2 to show that V is right continuous at 0.

Claim: $q_N^M \leq q^M$, with strict inequality if $b(q^M) > b_N(q^M)$. We show that $\Delta := V'(q) - V_N'(q) \leq 0$. It holds that $\Delta = -\int_{[b_N(q), b(q)]} J_1(q, \theta)\,\mathrm{d}F(\theta)$, and, by definition of $b(q)$, $J_1(q,\theta) \le 0$ for all $\theta \le b(q)$. Moreover, if $b(q) > b_N(q)$ and $q > 0$, then $\Delta > 0$. It remains to show that $q_N^M > 0$, which follows from the same argument used in the proof of Proposition A.2 to show that $q^M > 0$. □

A.4.2 Competition

We study a game played by N firms facing the same buyer population as in the main text, i.e., we assume that u is as in the main text. We use the following notation and results.

Setup We denote by $\mathcal{N}$ the set of players $\{1,\dots,N\}$, and we let $Q = \mathbb{R}_+$ be the set of qualities. Fix a profile of qualities $(\overline{q}_1,\dots,\overline{q}_N) \in Q^N$. The demand of type θ given the profile of tariffs $(p_1,\dots,p_N) \in \times_{i\in\mathcal{N}}\mathbb{R}^Q$ satisfying $p_i(q) = \infty$ if $q > \overline{q}_i$, $i \in \mathcal{N}$, is: $D_{(p_1,\dots,p_N)}(\theta) \coloneqq \operatorname{Argmax}_{q\in Q} u(\theta,q) - \min\{p_1(q),\dots,p_N(q)\}$. The player serving θ given the profile of tariffs $(p_1,\dots,p_N)$ is $\iota_{(p_1,\dots,p_N)}(\theta) \coloneqq \min \operatorname{Argmin}_{i\in\mathcal{N}} p_i(D_{(p_1,\dots,p_N)}(\theta))$; if multiple firms offer the same price for the demand, we break indifferences in favor of the low-index firms. Note that every type $\theta > 0$ has single-valued demand.

Timing and actions First, every player i simultaneously chooses a quality $\overline{q}_i \in Q$. In the second stage of the game, the *pricing* stage, every firm i observes the qualities $(\overline{q}_1,\dots,\overline{q}_N)$ and simultaneously chooses a pricing function $p_i\colon Q \to \mathbb{R}$ that only prices feasible qualities, i.e., with $p_i(q_i) = \infty$ if $q_i > \overline{q}_i$. Then, every buyer's type θ observes the pricing functions and, afterwards, he chooses a firm i, purchases a quality q_i at price $p_i(q_i)$, and nothing from the other firms, or does not purchase any good for a utility of 0.

Strategies, payoffs, and equilibria Firm i chooses a cap in stage 1 and a pricing function in stage 2. There is symmetric information among firms in stage 2, so every player i quotes a pricing function conditional on the known cap quality of every opponent. Hence, the set of pure strategies for i is $S_i \coloneqq Q \times \boldsymbol{P}_i$, letting $\boldsymbol{P}_i \subseteq (\mathbb{R}^Q)^{Q^N}$ be the set of "conditional" pricing functions of firm i; i.e., we define $\mathbf{P}_i = \{P\colon Q^N \to \mathbb{R}^Q \mid q > \overline{q}_i \text{ implies } P[\dots,\overline{q}_{i-1},\overline{q}_i,\overline{q}_{i+1},\dots](q) > g(\overline{\overline{q}}) + \overline{\theta}\overline{\overline{q}}\}$. The revenues of firm i given a profile of pricing functions $(p_1,\dots,p_N)$ are

$$R_i(p_1,\dots,p_N) \coloneqq \int_\Theta p_i(D_{(p_1,\dots,p_N)}(\theta))[\iota_{(p_1,\dots,p_N)}(\theta) = i]\,\mathrm{d}F(\theta).$$

The *payoff* of firm i from the profile of pure strategies $s = (\dots, (\overline{q}_i^s, P_i^s), \dots) \in \times_{i=1}^N S_i$ is $\Pi_i(s) \coloneqq R_i(P_1^s[\overline{q}^s], \dots, P_N^s[\overline{q}^s]) - c(\overline{q}_i^s)$. A strategy for player i is a pair of a cap distribution and conditional pricing function, (H_i, P_i), in which H_i is a distribution function with support contained in Q and $P_i \in \mathbf{P}_i$ for all i. A strategy profile $((H_1, P_1), \dots, (H_N, P_N))$ is an *equilibrium* if it is a subgame-perfect Nash equilibrium of the game, that is, the following two conditions hold.

1. The conditional pricing functions maximize conditional profits, that is, for all cap profiles $\overline{q} = (\overline{q}_1, \dots, \overline{q}_N)$,

$$R_i(P_i[\overline{q}], P_{-i}[\overline{q}]) \geq R_i(p_i', P_{-i}[\overline{q}]) \quad \text{for all } p_i' \in \mathbb{R}^Q,\ i \in \mathcal{N}.$$

2. The cap distributions maximize expected profits, that is, for all $\overline{q}_i$ in the support of H_i we have

$$\int_{Q^{N-1}} R_i(P_i[\overline{q}], P_{-i}[\overline{q}]) \,\mathrm{d}H_{-i}(\overline{q}_{-i}) - c(\overline{q}_i) \geq \int_{Q^{N-1}} R_i(P_i[\overline{q}_i', \overline{q}_{-i}], P_{-i}[\overline{q}_i', \overline{q}_{-i}]) \,\mathrm{d}H_{-i}(\overline{q}_{-i}) - c(\overline{q}_i')$$

for all $\overline{q}_i' \in Q$, letting H_{-i} denote the joint distribution of the caps of i's opponents under $((H_1, P_1), \dots, (H_N, P_N))$, and for all $i \in \mathcal{N}$.

Auxiliary monopoly problem The allocation $\boldsymbol{q} \in \boldsymbol{Q}$ is x-y *second best*, for qualities x, $y \in Q$ with $x \leq y$, if there exists a transfer function t such that: $(\boldsymbol{q}, t)$ solves

$$V^{x,y} = \sup_{(\boldsymbol{q},t)\in\mathcal{M}} \int t(\theta) \,\mathrm{d}F(\theta) \text{ subject to: } y \leq \boldsymbol{q}(\theta) \leq x \text{ for all } \theta \in \Theta.$$

Lemma A.3. *Let x, $y \in Q$ with $x \leq y$. The allocation $\boldsymbol{q}$ is x-y second best if and only if: there exists a nondecreasing allocation $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma}(\theta) = \boldsymbol{\beta}(\theta)$ almost everywhere and $\boldsymbol{q}(\theta) = \max\{\min\{\boldsymbol{\gamma}(\theta), x\}, y\}$ for all θ.*

Proof. The following result follows from the same argument as the proof of Proposition A.2. We establish that $\boldsymbol{q}$ solves the problem if and only if: $\boldsymbol{q}(\theta) = \max\{\min\{\boldsymbol{\gamma}(\theta), x\}, y\}$ for all θ and some nondecreasing $\boldsymbol{\gamma} \in \boldsymbol{Q}$ with $\boldsymbol{\gamma}(\theta) = \boldsymbol{\beta}(\theta)$ almost everywhere.

For necessity, let $\boldsymbol{q} \in \boldsymbol{Q}$ be such that $\boldsymbol{q}(\theta) = \max\{\min\{\boldsymbol{\gamma}(\theta), x\}, y\}$ for all $\theta \in \Theta$ and for some nondecreasing allocation $\boldsymbol{\gamma}$ with $\boldsymbol{\gamma}(\theta) = \boldsymbol{\beta}(\theta)$ almost everywhere. Then, $\boldsymbol{q}$ solves the problem of Lemma B.1 without the monotonicity constraint, and satisfies the monotonicity constraint, so $\boldsymbol{q}$ solves the problem of Lemma B.1. For sufficiency, let $\boldsymbol{q}$ solve the problem in the statement of the lemma and be such that: there does not exist a nondecreasing allocation $\boldsymbol{\gamma}$ with $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q\}$ for all θ and $\boldsymbol{\gamma}(\theta) = \boldsymbol{\beta}(\theta)$ almost everywhere. By Lemma B.1 and strict quasiconcavity of $J(\cdot, \theta)$, $\boldsymbol{q}$ does not solve the problem of Lemma B.1 without the monotonicity constraint. We observe that the function $\theta \mapsto \min\{\boldsymbol{\beta}(\theta), q\}$ is an allocation that solves the problem without the monotonicity constraint (Lemma B.1) and satisfies the monotonicity constraint, so $\boldsymbol{q}$ does not solve $\mathcal{P}(q)$. The claim follows from the preliminary observation in the proof of Lemma A.2. □

Auxiliary results about the pricing game The pricing game given the quality profile $(q_1, \ldots, q_N) \in Q^N$ is the strategic-form game $\Gamma(q_1, \ldots, q_N) = (\mathcal{N}, (\mathbb{R}^{[0,q_i]}, R_i(\cdot))_{i \in \mathcal{N}}$. The profile of pricing functions $(p_1, \ldots, p_N)$ is a $(q_1, \ldots, q_N)$ *equilibrium* if $(p_1, \ldots, p_N)$ is a Nash equilibrium of $\Gamma(q_1, \ldots, q_N)$.

We study the subgame starting at the given production profile $(q_1, \ldots, q_N) \in Q^N$ with $\max\{q_1, \ldots, q_N\} =: x$ and $y := \max\{q_1, \ldots, q_N\} \setminus \{x\}$.

Lemma A.4. *Let $q_i \in Q$ for all $i \in \mathcal{N}$. For every $(q_1, \ldots, q_N)$ equilibrium $(p_1, \ldots, p_N)$, it holds that:*

$$R_i(p_1, \ldots, p_N) = (V(q_i) - V(\max\{q_1, \ldots, q_n\} \setminus \{q_i\}))_+,$$

and

$$D_{(p_1, \ldots, p_N)}(\theta) = \max\{\min\{\boldsymbol{\gamma}(\theta), x\}, y\},$$

for all θ, in which $\boldsymbol{\gamma}$ is a nondecreasing allocation such that $\boldsymbol{\gamma}(\theta) = \boldsymbol{\beta}(\theta)$ almost everywhere.

Proof. (0) Quantities less than y come at zero price. Fix a $(q_1, \ldots, q_N)$ equilibrium $(p_1, \ldots, p_N)$. Suppose there exists a positive measure type set $A \subseteq \Theta$ such that $D_{(p_1, \ldots, p_N)}(\theta) \leq y$ and purchase at positive price: $p_{\iota_{(p_1, \ldots, p_N)}(\theta)} D_{(p_1, \ldots, p_N)}(\theta) > 0$ for all $\theta \in A$ and $\iota_{(p_1, \ldots, p_N)}(\theta) =: i$.

Firm j, that invested $q_j \geq y$ and is not i, has the following action available:

$$p'(q) = \begin{cases} p_j(q), & \text{if } q \notin D_{(p_1, \ldots, p_N)}(A), \\ p_i(q) - \varepsilon, & \text{if } q \in D_{(p_1, \ldots, p_N)}(A). \end{cases}$$

for sufficiently small $\varepsilon > 0$. Conditional on type $\theta \in A$, playing p' induces strictly higher profits than p_i. Conditional on type $\theta \notin A$, playing p' induces lower profits than p_i only if $\iota_{(p_1,\ldots,p_N)}(\theta) = i$, and by a negligible amount if ε is sufficiently small.

(1) Preliminaries: using the results from the x-y second best problem. We now compute the value of the x-y second best problem, defined above. For simplicity, we assume that $g'(0) > -\varphi(0)$. For quantities $0 \le y < x \le \overline{\overline{q}}$, it holds that

$$V^{x,y} = \int_{[\underline{\theta},a]} P(y)\,\mathrm{d}F(\theta) + \int_{(a,b)} g(\boldsymbol{\beta}(\theta)) + \boldsymbol{\beta}(\theta)\varphi(\theta)\,\mathrm{d}F(\theta) + \int_{[b,\overline{\theta}]} g(x) + x\varphi(\theta)\,\mathrm{d}F(\theta),$$

for a pricing function P, cutoff types $\underline{\theta} \le a < b \le \overline{\theta}$, and a quality allocation $\boldsymbol{q}$ that agrees with $\theta \mapsto \min\{\boldsymbol{\beta}(\theta), x\}$ on $[a, \overline{\theta}]$. (We note that $a = b(y)$ and $b = b(x)$).

Hence, we have

$$V^{x,y} = V(x) - V(y) + \int_{[\underline{\theta},b(y)]} P(y)\,\mathrm{d}F(\theta).$$

for a pricing function P.

(2) Bertrand-type argument. By step (1), the payoff to i if $q_i = x > y$ in equilibrium is

$$R_i(p_1,\ldots,p_N) = V(x) - V(y) + \int_{[\underline{\theta},b(y)]} [\iota_{(p_1,\ldots,p_N)} = i] P_i(D_{(p_1,\ldots,p_N)}(\theta))\,\mathrm{d}F(\theta).$$

By the argument in part (0), any quality less than y involves zero profits, i.e., the last term on the right-hand side of the above equality is 0. So, the formula holds. $\square$

Notation for the production game Every Nash equilibrium of $\Gamma(q_1,\ldots,q_N)$ induces the same revenues, i.e., for all $(q_1,\ldots,q_N)$ equilibria $(p_1^*,\ldots,p_N^*)$ and $(p_1^{**},\ldots,p_N^{**})$, we have $R_i(p_1^*,\ldots,p_N^*) = R_i(p_1^{**},\ldots,p_N^{**})$ (Lemma A.4). We define $\overline{R}_i(q_1,\ldots,q_N)$ the unique equilibrium revenues of $i \in \mathcal{N}$ in the pricing game $\Gamma(q_1,\ldots,q_N)$. By Lemma A.4, we have

$$\overline{R}_i(q_1,\ldots,q_N) = (V(q_i) - V(\max\{q_1,\ldots,q_n\} \setminus \{q_i\}))_+.$$

The *production game* is the strategic-form game $\Gamma = \big(\mathcal{N}, (Q, \overline{R}_i(\cdot) - c(\cdot))_{i\in\mathcal{N}}\big)$. By mixed strategies, we refer to elements of ΔQ, the set of probability distributions over Q identified by distribution functions, and we extend payoffs of Γ to mixed-strategy profiles as usual; i.e., defining

$$\Pi_i\colon (\Delta Q)^N \to \mathbb{R} : (H_1, \ldots, H_N) \mapsto \int_Q \cdots \int_Q \overline{R}_i(q_1, \ldots, q_N) - c(q_i)\, \mathrm{d}H_1(q_1) \cdots \mathrm{d}H_N(q_N).$$

Pure-strategy equilibria of the production game

Lemma A.5 (Pure-strategy equilibria)**.** *The quality profile* $(q_1, \ldots, q_N)$ *is a Nash equilibrium in pure strategies of the production game if and only if: there exists* $i \in \mathcal{N}$ *such that* $q_i = q^M$ *and* $q_j = 0$ *for all* $j \neq i$*.*

Proof. Let $q_i,\ q_j > 0$ for $i \neq j$. Then the player choosing $\min\{q_i, q_j\}$ has a strictly profitable deviation to 0. Thus, in every equilibrium, there exists at most one player choosing a positive quality q_i.

Let $(q_1, \ldots, q_N)$ be a Nash equilibrium. Then, $\max\{q_1, \ldots, q_n\} = q^M$, by the characterization of the monopolist quality q^M.

It is left to verify that $\overline{R}_2(q^M, 0, \ldots, 0) \geq \overline{R}_2(q^M, q, \ldots, 0)$ for all $q > 0$. First, we observe that $\overline{R}_2(q^M, 0, \ldots, 0) = 0 = \overline{R}_2(q^M, q, 0, \ldots, 0)$ for $q \leq q^M$, by the argument mentioned above. Second, we compute the payoff from qualities higher than q^M: $\overline{R}_2(q^M, q, 0, \ldots, 0) = V(q) - V(q^M) - c(q)$. By previous results, $V'(q) < c'(q)$ if $q > \overline{q}^M$. So, $q \mapsto \overline{R}_2(q^M, q, 0, \ldots, 0)$ is decreasing on $(q^M, \overline{q})$. Hence, our claim follows. □

Mixed-strategy equilibria of the production game A Nash equilibrium in mixed strategies $(H_1, \ldots, H_N)$ has *multiple active players* if $H_i(q) \geq H_j(q) > 0$ for $q > 0$ and $i,\ j \in \mathcal{N}$ with $i \neq j$. A Nash equilibrium in mixed strategies $(H_1, \ldots, H_N)$ is *symmetric among active players* if $H_i(q) = H_j(q)$ for all $q \in Q$ whenever $H_i(q) \geq H_j(q) > 0$ for some $q > 0$. Note that there are only two kinds of equilibria in mixed strategies that are symmetric among active players: those with at least 2 active players and the monopolistic equilibria.

Lemma A.6 (Necessary conditions for mixed-strategy equilibria)**.** *If* $(H_1, \ldots, H_N)$ *is a Nash equilibrium in mixed strategies with multiple active players, then*

$$\prod_{j \in \mathcal{N} \setminus \{i\}} H_j(q) = \begin{cases} \frac{c'(q)}{V'(q)}, & \text{if } q \in (0, q^M), \\ 1, & \text{if } q \geq q^M, \end{cases}$$

for all i *such that* $H_i(q) > 0$ *for some* $q > 0$*.*

Proof. For the rest of the proof, we fix an equilibrium with n active firms $(H_1, \ldots, H_N)$ and we fix a player i (we note that i can be active, i.e., $H_i(q) > 0$ for some $q > 0$, or idle.) We denote by H_{-i} the equilibrium distribution of the maximum of the caps of player i's opponents.

Part 1: Left- and right-limit of equilibrium expected payoff with respect to own quality. The expected payoff from quality q_i of firm i, $\Pi_i(q_i, H_{-i}) := \int_{[0,q_i]} V(q_i) - V(y)\,\mathrm{d}H_{-i}(y) - c(q_i)$, can be re-arranged as follows,

$$\Pi_i(q_i, H_{-i}) = V(q_i)H_{-i}(q_i) - \int_{[0,q_i]} V(y)\,\mathrm{d}H_{-i}(y) - c(q_i). \tag{A.1}$$

We also use the integration-by-parts formula

$$\begin{aligned}\Pi_i(q_i, H_{-i}) &= [(V(q_i) - V(y))H_{-i}(y)]|_0^{q_i} + \int V'(y)H_{-i}(y)\,\mathrm{d}y - c(q_i),\\ &= \int_{[0,q_i]} V'(y)H_{-i}(y)\,\mathrm{d}y - c(q_i)\end{aligned}$$

So, for $q_1,\ q_2 \in Q$, with $q_1 < q_2$, the following expression, which we leverage in the rest of the proof, holds:

$$\Pi_i(q_2, H_{-i}) - \Pi_i(q_1, H_{-i}) = \int_{[q_1,q_2]} V'(y)H_{-i}(y) - c'(y)\,\mathrm{d}y. \tag{A.2}$$

The following inequalities, implied by the above equality, are used throughout the proof: For $q_1,\ q_2 \in Q$, with $q_1 < q_2$, it holds that

$$\Pi_i(q_2, H_{-i}) - \Pi_i(q_1, H_{-i}) \in [(V'(q_2)H_{-i}(q_1) - c'(q_2))(q_2 - q_1), (V'(q_1)H_{-i}(q_2) - c'(q_1))(q_2 - q_1)]. \tag{A.3}$$

As an implication, i's payoff is continuous in q_i at $q > 0$ if H_{-i} is continuous, because it holds that

$$\lim_{\varepsilon\to 0} \frac{\Pi(q+\varepsilon, H_{-i}) - \Pi(q-\varepsilon, H_{-i})}{2\varepsilon} \in \Big[V'(q)\lim_{\varepsilon\to 0} H_{-i}(q-\varepsilon) - c'(q), V'(q)\lim_{\varepsilon\to 0} H_{-i}(q+\varepsilon) - c'(q)\Big].$$

Step 2: A differential equation holds almost everywhere in the interior of the support. We establish that: $[a, b] \subseteq \operatorname{supp} H_i$ *implies* $V'(y)H_{-i}(y) = c'(y)$ *for a.e.* $y \in [a, b]$.

Suppose $(c, d) \subseteq Q$ exists such that: $V'(y)H_{-i}(y) > c'(y)$ for all $y \in (c, d)$. Then, we have $\Pi(d, H_{-i}) - \Pi(c, H_{-i}) > 0$ by equality A.2. Thus, for all $[c, d] \subseteq Q$ with $c,\ d \in \operatorname{supp} H_i$, it holds that $V'(y)H_{-i}(y) \geq c'(y)$, otherwise player i strictly prefers to play $q_1 + \varepsilon$ than to play H_i, for sufficiently small $\varepsilon > 0$. Similarly, suppose $(c, d) \subseteq Q$ exists such that: $V'(y)H_{-i}(y) < c'(y)$ for all $y \in (c, d)$. Then, $\Pi(d, H_{-i}) - \Pi(c, H_{-i}) < 0$. Thus, for all $[c, d] \subseteq Q$ with $c,\ d \in \operatorname{supp} H_i$, it holds that $V'(y)H_{-i}(y) = c'(y)$.

Step 3: H_i does not have mass points in $\operatorname{supp} H_i \setminus \{0\}$. We establish that H_i does not have mass points in $\operatorname{supp} H_i \setminus \{0\}$, using the left and right derivatives from Step 1.

Suppose $H_i(q_1) > \limsup_{\varepsilon \to 0} H_i(q_1 - \varepsilon)$, for $q_1 \in \operatorname{supp} H_i$ with $q_1 > 0$ and all i. (The proof effectively consists in applying the same intuition as in Step 1 to the left and to the right of q_1.) Using the inequalities in A.3, playing $q_1 + \varepsilon$ accumulates profits at rate

$$V'(q_1) \lim_{\varepsilon \to 0} H_{-i}(q_1 + \varepsilon) - c'(q_1) = V'(q_1)H_{-i}(q_1) - c'(q_1).$$

Thus, we have $V'(q_1)H_{-i}(q_1) \leq c'(q_1)$, otherwise $q_1 + \varepsilon$ is a strictly profitable deviation to H_i, for sufficiently small $\varepsilon > 0$. Using the inequalities in A.3, playing $q_1 - \varepsilon$ accumulates profits at rate

$$V'(q_1) \lim_{\varepsilon \to 0} H_{-i}(q_1 - \varepsilon) - c'(q_1) \leq V'(q_1) \limsup_{\varepsilon \to 0} H_{-i}(q_1 - \varepsilon) - c'(q_1);$$

the above inequality is a heuristic argument, the formal result is that, for sufficiently small $\varepsilon > 0$,

$$\Pi_i(q_1 - \varepsilon, H_{-i}) - \Pi_i(q_1, H_{-i}) \geq \Pi_i(q_1, H_{-i}) - \left(V'(q_1) \limsup_{\varepsilon \to 0} H_{-i}(q_1 - \varepsilon) - c'(q_1) \right) \varepsilon.$$

By the hypothesis, $\limsup_{q \to q_1^-} H_{-i}(q) < H_{-i}(q_1)$, so we have

$$\Pi_i(q_1 - \varepsilon, H_{-i}) - \Pi_i(q_1, H_{-i}) > \Pi_i(q_1, H_{-i}) - (V'(q_1)H_{-i}(q_1) - c'(q_1))\varepsilon.$$

By the fact that $V'(q_1)H_{-i}(q_1) \leq c'(q_1)$, we have

$$\Pi_i(q_1 - \varepsilon, H_{-i}) - \Pi_i(q_1, H_{-i}) > \Pi_i(q_1, H_{-i}).$$

Thus, player i has a downward profitable deviation. As a result, there are no mass points above 0 in H_i.[13]

Step 4: There are no flat regions in H_i if i is active. We establish that there

[13] We establish that the expected payoffs are continuous in own quality, in order to claim that war-of-attrition discontinuity arguments cannot be applied. Note that, as $\varepsilon \to 0$, by expression

are no flat regions excluding 0 using no mass points and the differential equation.

Suppose $H_i(q_2) - H_i(q_1) = 0$ for $q_1 < q_2$ in the convex hull of the support of H_i. There exists one such maximal interval, in the set inclusion order, which we denote by (q_1, q_2), by no mass points. If $q_1 = 0$, pick another flat region, which does not include q_1, q_2. If there is no other flat region, then the claim holds.

Suppose, then, that $q_1 > 0$, for a maximal interval of some flat region $[q_1, q_2]$, with $q_2 > q_1$. (Note that, as an implication, player i is active.) Then, the differential equation — first part of Step 2 — holds on $(q_1 - \varepsilon) \cup (q_2 + \varepsilon)$ for sufficiently small $\varepsilon > 0$, by no mass points. So, by step 2, we have that $\liminf_{\varepsilon \to 0} H_i(q_2 + \varepsilon) - \limsup_{\varepsilon \to 0} H_i(q_1 - \varepsilon) > 0$. This implication contradicts Step 3, because it implies that every player has a mass point at q_2 or at q_1, or both.

Conclusion. In any symmetric equilibrium with multiple active players, the support of H_i is a convex interval and H_i is continuous on $\mathbb{R}_{++}$. As an implication, $\max \operatorname{supp} H_i = q^M$. We find the mass at 0 by solving the differential equation downward. The mass is null, because $c'(0) = 0$. Moreover, we have that $0 = \min \operatorname{supp} H_i$, because $c'(q) > 0$ for $q > 0$. $\square$

Lemma A.7 (Necessary and sufficient conditions for mixed-strategy equilibria)**.** *If $(H_1, \ldots, H_N)$ is a Nash equilibrium in mixed strategies of the production game that is symmetric among active players, then one, and only one, of the following statements holds.*

1. *There exists a number of active players $n \in \mathcal{N} \setminus \{1\}$ such that: for a set of*

A.1, we have

$$\begin{aligned}\Pi_i(q_1 + \varepsilon, H_{-i}) - \Pi_i(q_1, H_{-i}) &\to \\ &V(q_1)(H_{-i}(q_1 + \varepsilon) - H_{-i}(q_1)) - V(q_1)(H_{-i}(q_1 + \varepsilon) - H_{-i}(q_1)) - c(q_1 + \varepsilon) + c(q_1), \\ &\to -c'(q_1)\varepsilon,\end{aligned}$$

for all q_1, including possible mass points of H_{-i}. Similarly, we have

$$\begin{aligned}\Pi_i(q_2, H_{-i}) - \Pi_i(q_2 - \varepsilon, H_{-i}) &\to \\ &V(q_2)(H_{-i}(q_2) - H_{-i}(q_2 - \varepsilon)) - V(q_2)(H_{-i}(q_2) - H_{-i}(q_2 - \varepsilon)) - c(q_2) + c(q_2 - \varepsilon), \\ &\to -c'(q_2)\varepsilon,\end{aligned}$$

for all q_2, including possible mass points of H_{-i}. As an implication, $\Pi_i(\cdot, H_{-i})$ is continuous above 0.

active players $I \subseteq \mathcal{N}$ with $|I| = n$ we have

$$H_i \colon q \mapsto \begin{cases} \left(\frac{c'(q)}{V'(q)}\right)^{\frac{1}{n-1}}, & \text{if } q \in [0, q^M], \\ 1, & \text{if } q > q^M, \end{cases}$$

for all $i \in I$.

2. *There exists an active player i such that: $H_i \colon q \mapsto [q \geq q^M]$ and $H_j \colon q \to 0$ for all $j \in \mathcal{N} \setminus \{i\}$.*

Moreover, the strategy profiles described in both parts constitute Nash equilibria in mixed strategies that are symmetric among active players.

Proof. It is left to verify that the strategy profile described in part 1 constitutes a Nash equilibrium in mixed strategies. It follows from the definition of $\overline{q}^M$ and by a similar argument as the one used to characterize the pure-strategy equilibria. □

Proof of the main result

Proposition A.4. *For all $n \leq N$, there exists a symmetric n equilibrium. Moreover, every symmetric and competitive n equilibrium induces the random allocation given by $\boldsymbol{q}[\hat{x}, \hat{y}]$, in which the random variables $\hat{x}$ and $\hat{y}$ are, respectively, the first- and second-order statistics of a collection of n i.i.d. random variables, each with distribution function H_n given by*

$$H_n \colon q \mapsto \begin{cases} \left(\frac{c'(q)}{V'(q)}\right)^{\frac{1}{n-1}}, & \text{if } q \in [0, q^M], \\ 1, & \text{if } q \in (q^M, \infty). \end{cases} \tag{A.4}$$

Proof. This proof uses previously established results (Proposition A.7, Lemma A.4, and Lemma A.6) and the definitions in sections A.4.2.

Step 1: Existence. The first part of the result (i.e., existence) is established in Proposition A.7.

Step 2: Quality allocation. We establish that every allocation that is induced in equilibrium takes the stated form. Fix a value of $\hat{x} := \max\{\overline{q}_1, \ldots, \overline{q}_N\}$ and $\hat{y} := \max\{\overline{q}_1, \ldots, \overline{q}_N\} \setminus \{\hat{x}\}$ in Q, given the subgame starting after the choice of caps $(\overline{q}_1, \ldots, \overline{q}_N)$. The result follows from the last part of Lemma A.4.

Step 3: First- and second-order statistics. It is left to establish that the stated distribution of $\hat{x}$ and $\hat{y}$ is correctly induced by the equilibrium distribution of stage-1 caps $(\overline{q}_1, \ldots, \overline{q}_N)$. This result is established by Lemma A.6. In particular, as a result of the definitions in Section A.4.2, the profile of strategies $((H_1, P_1), \ldots, (H_N, P_N))$ is a symmetric and competitive equilibrium if, and only if, the profile of distribution functions $(H_1, \ldots, H_N)$ is a mixed-strategy Nash equilibrium of the production game with multiple active players that is symmetric among active players.[14] $\square$

Remark A.1. *The monopoly allocation arises as the unique outcome of the game under pure strategies. First, we argue that firm i producing q^M and all other firms staying idle constitutes an equilibrium. Firm i is a monopolist when her opponents are idle, so her best response is q^M due to Proposition A.2. If firm i produces q^M, then firm j can invest in qualities beyond q^M. Let's verify that becoming a monopolist for better qualities than q^M is not a best response when an opponent produces q^M. If firm j acquires quality $q > q^M$, her profits can be expressed as $\int_{[q^M,q]} V'(\tilde{q}) - c'(\tilde{q})\,\mathrm{d}\tilde{q} - c(q^M)$. So, i is better off staying idle than investing in quality $q > q^M$.*

Let's verify that there are no competitive equilibria in pure strategies. Specifically, we claim that, in all equilibria in pure strategies, one firm produces q^M and all other firms stay idle. First, in any equilibrium, the firm producing the lower quality earns no revenues and can save on acquisition costs by becoming idle. All firms staying idle does not occur in equilibrium as well, since any firm i prefers to be a monopolist (this observation follows from Proposition A.2.) Thus, the only pure-strategy equilibria are the monopoly ones.

[14] We note that, by Proposition A.7, for all $n > 2$ there exists a mixed-strategy Nash equilibrium of the production game with multiple active players that is symmetric among active players such that: for a set $I \subseteq \mathcal{N}$ with $|I| = n$ we have

$$H_i \colon q \mapsto \begin{cases} \left(\frac{c'(q)}{V'(q)}\right)^{\frac{1}{n-1}}, & \text{if } q \in [0, \overline{q}^M], \\ 1, & \text{if } q > \overline{q}^M, \end{cases}$$

for all $i \in I$. Moreover, if we have a mixed-strategy Nash equilibrium of the production game with multiple active players that is symmetric among active players, then there exists $n > 2$ such that: for a set $I \subseteq \mathcal{N}$ with $|I| = n$ we have

$$H_i \colon q \mapsto \begin{cases} \left(\frac{c'(q)}{V'(q)}\right)^{\frac{1}{n-1}}, & \text{if } q \in [0, \overline{q}^M], \\ 1, & \text{if } q > \overline{q}^M, \end{cases}$$

for all $i \in I$.

B Supplementary material

B.1 Auxiliary results

Lemma B.1. *Let $(y, x) \in Q^2$, $K\colon Q \times \Theta \to \mathbb{R}$ be continuously differentiable $K(\cdot, \theta)$ be concave for every $\theta \in \Theta$. Then, $\boldsymbol{q} \in \boldsymbol{Q}$ solves*

$$\mathcal{Q} : \max_{\boldsymbol{q} \in \boldsymbol{Q}} \int_{\Theta} K(\boldsymbol{q}(\theta), \theta) \,\mathrm{d}F(\theta) \text{ subject to: } y \leq \boldsymbol{q}(\theta) \leq x \text{ for all } \theta \in \Theta,$$

if and only if: there exists an allocation $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma}(\theta) \in \operatorname{Argmax}_{q \in Q} K(q, \theta)$ almost everywhere and $\boldsymbol{q}(\theta) = \max\{\min\{\boldsymbol{\gamma}(\theta), x\}, y\}$ for all θ.

Proof. Fix $(y, x) \in Q^2$. We establish the "if" direction first.

Fix $\boldsymbol{q}' \in \boldsymbol{Q}$ such that $y \leq \boldsymbol{q}'(\theta) \leq x$ for all θ, and $\boldsymbol{\gamma} \in \boldsymbol{Q}$ with $\boldsymbol{\gamma}(\theta) \in \operatorname{Argmax}_{q \in Q} K(q, \theta)$ for a.e. θ. Define $\Delta = \int_{\Theta} K(\boldsymbol{q}(\theta), \theta) - K(\boldsymbol{q}'(\theta), \theta) \,\mathrm{d}F(\theta)$ for $\boldsymbol{q}(\theta) = \max\{\min\{\boldsymbol{\gamma}(\theta), x\}, y\}$, and $S = \{\theta \in \Theta : \boldsymbol{q}(\theta) = x\}$, $T = \{\theta \in \Theta : \boldsymbol{q}(\theta) = y\}$. We note that

$$\begin{aligned}\Delta = \int_T K(y, \theta) - K(\boldsymbol{q}'(\theta), \theta) \,\mathrm{d}F(\theta) + \int_{\Theta \setminus (S \cup T)} K(\boldsymbol{\gamma}(\theta), \theta) - K(\boldsymbol{q}'(\theta), \theta) \,\mathrm{d}F(\theta) \\ + \int_S K(x, \theta) - K(\boldsymbol{q}'(\theta), \theta) \,\mathrm{d}F(\theta).\end{aligned}$$

By definition of $\boldsymbol{\gamma}$, we have $K(\boldsymbol{\gamma}(\theta), \theta) - K(\boldsymbol{q}'(\theta), \theta) \geq 0$ a.e. on $\Theta \setminus (S \cup T)$. Additionally, we have $K(x, \theta) - K(\boldsymbol{q}'(\theta), \theta) = \int_{[\boldsymbol{q}'(\theta), x]} K_1(\tilde{q}, \theta) \,\mathrm{d}\tilde{q}$, which is nonnegative a.e. on S by definition of $\boldsymbol{\gamma}$ and concavity of $K(\cdot, \theta)$; similarly, we have $K(y, \theta) - K(\boldsymbol{q}'(\theta), \theta) = -\int_{[y, \boldsymbol{q}'(\theta)]} K_1(\tilde{q}, \theta) \,\mathrm{d}\tilde{q}$, which is nonnegative a.e. on T by definition of $\boldsymbol{\gamma}$ and concavity of $K(\cdot, \theta)$. Hence, we have $\Delta \geq 0$, so $\boldsymbol{q}$ solves $\mathcal{Q}$.

It remains to establish the "only if" direction. Fix a solution $\boldsymbol{q}^*$ to $\mathcal{Q}$, and suppose that there exists $(\theta', \theta'') \subseteq \Theta$ such that: (for every $\boldsymbol{\gamma} \in \boldsymbol{Q}$ with $\boldsymbol{\gamma}(\theta) \in \operatorname{Argmax}_{q \in Q} K(q, \theta)$ a.e., we have $\boldsymbol{q}'(\theta) \neq \boldsymbol{q}^*(\theta)$ on (θ', θ''), for $\boldsymbol{q}' \colon \theta \mapsto \max\{\min\{\boldsymbol{\gamma}(\theta), x\}, y\}$). Define $S = \{\theta \in (\theta', \theta'') : \boldsymbol{q}'(\theta) = x\}$, $T = \{\theta \in (\theta', \theta'') : \boldsymbol{q}'(\theta) = y\}$, and $\Delta = \int_{(\theta', \theta'')} K(\boldsymbol{q}'(\theta), \theta) - K(\boldsymbol{q}^*(\theta), \theta) \,\mathrm{d}F(\theta)$; it holds that

$$\begin{aligned}\Delta = \int_T \int_{[\boldsymbol{q}^*(\theta), y]} K_1(\tilde{q}, \theta) \,\mathrm{d}\tilde{q} \,\mathrm{d}F(\theta) + \int_{(\theta', \theta'') \setminus (S \cup T)} K(\boldsymbol{q}'(\theta), \theta) - K(\boldsymbol{q}^*(\theta), \theta) \,\mathrm{d}F(\theta) \\ + \int_S \int_{[\boldsymbol{q}^*(\theta), x]} K_1(\tilde{q}, \theta) \,\mathrm{d}\tilde{q} \,\mathrm{d}F(\theta)\end{aligned}$$

By definition of $\boldsymbol{\gamma}$, we have $K(\boldsymbol{q}'(\theta), \theta) - K(\boldsymbol{q}^*(\theta), \theta) \geq 0$ a.e. on $(\theta', \theta'') \setminus (S \cup T)$. Additionally, we have: $\int_{[\boldsymbol{q}^*(\theta), x]} K_1(\tilde{q}, \theta) \, \mathrm{d}\tilde{q} \geq 0$ a.e. on S by definition of $\boldsymbol{q}'$, $\boldsymbol{\gamma}$, and concavity of $K(\cdot, \theta)$; similarly, we have: $\int_{[\boldsymbol{q}^*(\theta), y]} K_1(\tilde{q}, \theta) \, \mathrm{d}\tilde{q} \geq 0$ a.e. on T by definition of $\boldsymbol{q}'$, $\boldsymbol{\gamma}$, and concavity of $K(\cdot, \theta)$. It follows that $\Delta \geq 0$. To show that $\Delta > 0$, we use the fact that $\boldsymbol{q}^*$ differs from $\theta \mapsto \min\{\boldsymbol{\gamma}(\theta), q\}$ on (θ', θ''), for every allocation $\boldsymbol{\gamma}$ with $\boldsymbol{\gamma}(\theta) \in \operatorname{Argmax}_{q \in Q} K(q, \theta)$ almost everywhere.

□

Lemma B.2. *The allocation $\boldsymbol{q}^* \colon \theta \mapsto \min\{\overline{\boldsymbol{\beta}}(\theta), q\}$ has the "pooling property" (Toikka, 2011).*

Proof. First, we note that $G(F(\theta), q)$ and $H(F(\theta), q)$ do not depend on q, so we omit this argument (Remark B.1), and that J is "separable" (Toikka, 2011, Definition 3.1). Fix $q \in Q$, $(\theta', \theta'') \subseteq \Theta$, and suppose $G(F(\theta)) < H(F(\theta))$ for all $\theta \in (\theta', \theta'')$. There are three cases. First, if $\theta'' \leq T^-(q)$, then $\boldsymbol{q}^*$ is constant on (θ', θ'') by the pooling property of $\overline{\boldsymbol{\beta}}$ (Toikka, 2011, Definition 3.5, Theorem 3.7, Corollary 3.8). Second, if $\theta' \geq T^-(q)$, then $\boldsymbol{q}^*(\theta) = q$ for all $\theta \in (\theta', \theta'')$. Lastly, we consider the case in which $\theta' < T^-(q) < \theta''$. In this case, $\overline{\boldsymbol{\beta}}(\theta) < q$ for $\theta \in (\theta', a')$ and $\overline{\boldsymbol{\beta}}(\theta) \geq q$ for $\theta \in (a'', \theta'')$, for some $(a', a'') \in (\theta', a'') \times (a', \theta'')$, which contradicts the pooling property of $\overline{\boldsymbol{\beta}}$. So, there exists no such interval, and $\boldsymbol{q}^*$ has the pooling property. □

B.2 Non-regular distribution

In this section, we maintain the additional assumption that $u(q, \theta) = g(q) + \theta q$, for a strictly concave and increasing $g \colon \mathbb{R} \to \mathbb{R}$, and note that $a(q) = 0$. We consider the cumulative virtual value $H(q, \cdot) \colon \theta \mapsto \int_{[0,\theta]} J(q, F^{-1}(\tilde{\theta})) \, \mathrm{d}\tilde{\theta}$ (in the quantile space,) its lower convex envelope, $\operatorname{conv} H(q, \cdot) \colon \theta \mapsto \min\{\lambda H(q, \theta_1) + (1 - \lambda) H(q, \theta_2) \mid (\theta_1, \theta_2, \lambda) \in [0,1]^3 \text{ and } \lambda\theta_1 + (1-\lambda)\theta_2 = \theta\}$, and the right derivative of $\operatorname{conv} H(q, \cdot)$, $g(q, \cdot)$. The *ironed virtual surplus maximizer* is the largest selection $\overline{\boldsymbol{\beta}}$ from $\theta \mapsto \operatorname{Argmax}_{q \in Q} \overline{J}(q, \theta)$, in which $\overline{J}(q, \theta) \coloneqq J(0, \theta) + \int_{[0,q]} g(\tilde{q}, F(\theta)) \, \mathrm{d}\tilde{q}$; we note that $\overline{\boldsymbol{\beta}}$ is nondecreasing (Topkis, 1978; Toikka, 2011). We define the left-continuous inverse of $\overline{\boldsymbol{\beta}}$ by: $T^-(q) \coloneqq \inf\{\theta \in \Theta : \overline{\boldsymbol{\beta}}(\theta) \geq q\}$.

Remark B.1. *We note that $\overline{J}(q, \theta) = g(q) + \overline{k}(\theta) q$ for some nondecreasing $\overline{k}$. So, $\overline{J}(\cdot, \theta)$ is strictly quasiconcave for almost every $\theta \in \Theta$ because g is strictly concave. The proof of Proposition B.1 uses the fact that $\operatorname{Argmax}_{q \in Q} \overline{J}(q, \theta)$ is a singleton almost everywhere in the first claim. The remaining steps in the proof* *hold essentially as stated for more general single-crossing and concave–in-quality u.*

Proposition B.1. *Let q^M be the unique element of* $\operatorname{Argmax}_{q\in Q} V(q) - c(q)$*. Under the assumptions of this section, the allocation $\boldsymbol{q}$ is monopolist if and only if: there exists a nondecreasing allocation $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma}(\theta) = \overline{\boldsymbol{\beta}}(\theta)$ almost everywhere and $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q^M\}$ for all θ. Moreover, it holds that $q^M < q^\star$.*

Proof. The proof has three main steps. First, we characterize the solutions to $\mathcal{P}(q)$ for all $q \in Q$ by adapting the argument of Toikka (2011, proof of Theorem 3.7); second, we show that V is concave; third, we show that $q^M < q^\star$. The following preliminary observations hold by known arguments. The preliminary observation in the proof of Lemma A.2 holds, and the allocation $\boldsymbol{q}$ is monopolist iff: $\boldsymbol{q}$ solves $\mathcal{P}(q^M)$ for a quality $q^M \in \operatorname{Argmax}_{q\in Q} V(q) - c(q)$. Lastly, $\overline{J}(\cdot, \theta)$ is concave, continuously differentiable, and

$$V(q) = \max_{\boldsymbol{q}\in\boldsymbol{Q}} \int_\Theta \overline{J}(\boldsymbol{q}(\theta), \theta)\, \mathrm{d}F(\theta) \text{ subject to: } \boldsymbol{q}(\theta) \le q \text{ for all } \theta \in \Theta,$$

by Toikka (2011, respectively, by Lemma 4.10, Lemma 4.11, and Theorem 4.5.)

Claim: for fixed $q \in Q$, $\boldsymbol{q}$ solves $\mathcal{P}(q)$ iff $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q^M\}$ for all θ and some nondecreasing $\boldsymbol{\gamma} \in \boldsymbol{Q}$ with $\boldsymbol{\gamma}(\theta) = \overline{\boldsymbol{\beta}}(\theta)$ almost everywhere.

Fix $q \in Q$ and an allocation $\boldsymbol{q}^*$ such that $\boldsymbol{q}^*(\theta) = \min\{\boldsymbol{\gamma}(\theta), q\}$ for all θ and some nondecreasing $\boldsymbol{\gamma} \in \boldsymbol{Q}$ such that $\boldsymbol{\gamma}(\theta) = \overline{\boldsymbol{\beta}}(\theta)$ almost everywhere. First, we establish that $\int_\Theta G(F(\theta), q) - H(F(\theta), q)\, \mathrm{d}\boldsymbol{q}^*(\theta) = 0$. By Remark B.1, the maximizers of $\overline{J}(\cdot, \theta)$ agree almost everywhere, so the statement holds by the "pooling property" of $\theta \mapsto \min\{\overline{\boldsymbol{\beta}}(\theta), q\}$ (Lemma B.2). By Lemma B.1 and the preliminary observations, $\boldsymbol{q}^*$ maximizes $\int_\Theta \overline{J}(\boldsymbol{q}(\theta), \theta)\, \mathrm{d}F(\theta)$ subject to $\boldsymbol{q}(\theta) \le q$ for all $\theta \in \Theta$. Lastly, the objective in $\mathcal{P}(q)$ can be expressed as $\int_\Theta \overline{J}(\boldsymbol{q}(\theta), \theta)\, \mathrm{d}F(\theta) + \int_\Theta G(F(\theta), q) - H(F(\theta), q)\, \mathrm{d}\boldsymbol{q}(\theta)$, by Toikka (2011, proof of Theorem 3.7,) so we conclude that $\boldsymbol{q}^*$ solves $\mathcal{P}(q)$. Hence, it follows that: if $\boldsymbol{q} \in \boldsymbol{Q}$ is such that $\boldsymbol{q}(\theta) = \min\{\boldsymbol{\gamma}(\theta), q\}$ for all θ and some nondecreasing $\boldsymbol{\gamma} \in \boldsymbol{Q}$ with $\boldsymbol{\gamma}(\theta) = \boldsymbol{\beta}(\theta)$ almost everywhere, then $\boldsymbol{q}$ solves $\mathcal{P}(q)$.

For the other direction, suppose $\boldsymbol{q}^M$ solves $\mathcal{P}(q)$ and that there does exist a nondecreasing $\boldsymbol{\gamma} \in \boldsymbol{Q}$ such that $\boldsymbol{q}^M(\theta) = \min\{\boldsymbol{\gamma}(\theta), q\}$ for all $\theta \in \Theta$ and $\boldsymbol{\gamma}(\theta) = \overline{\boldsymbol{\beta}}(\theta)$ almost everywhere. We consider the following two cases. First, suppose that $\boldsymbol{q}^M$ maximizes $\int_\Theta \overline{J}(\boldsymbol{q}(\theta), \theta)\, \mathrm{d}F(\theta)$ subject to $\boldsymbol{q}(\theta) \le q$ for all $\theta \in \Theta$; then $\boldsymbol{q}^M(\theta) = \min\{\overline{\boldsymbol{\beta}}(\theta), q\}$ a.e. by Lemma B.1. Second, suppose that $\boldsymbol{q}^M$ does not maximize $\int_\Theta \overline{J}(\boldsymbol{q}(\theta), \theta)\, \mathrm{d}F(\theta)$ subject to $\boldsymbol{q}(\theta) \le q$ for all $\theta \in \Theta$; then $\boldsymbol{q}^M$ does not solve $\mathcal{P}(q)$ by the properties of $\boldsymbol{q}^*$ in the preceding paragraph. So, the claim holds.

Claim: V is concave and its left-derivative at $q \in (0, \overline{q}]$ is $\int_{[T^-(q),1]} \overline{J}_1(q,\theta)\,\mathrm{d}F(\theta)$. As a first step, we establish that V is continuous. It holds that $|J(q,\theta)| \leq \max\{\max J(Q \times \Theta), |\min J(Q \times \Theta)|\}$. Consider a sequence $q_n \to q \in Q$, for $q_n \in Q$, $n \in \mathbb{N}$. By a dominated-convergence argument, $\lim_{n\to\infty} V(q_n) = V(q)$. So, V is continuous.

Let $q_1 < q_2$, $q_i \in Q$, $i \in \{1,2\}$, and consider the difference quotient $d := \frac{V(q_2)-V(q_1)}{q_2-q_1}$. We have

$$d = \frac{1}{q_2 - q_1} \int_{[T^-(q_1),T^-(q_2)]} \overline{J}(\overline{\boldsymbol{\beta}}(\theta),\theta) - \overline{J}(q_1,\theta)\,\mathrm{d}F(\theta) + \int_{(T^-(q_2),1]} \frac{\overline{J}(q_2,\theta) - \overline{J}(q_1,\theta)}{q_2 - q_1}\,\mathrm{d}F(\theta).$$

We construct a lower bound for d. In particular, because we have $\overline{J}(\overline{\boldsymbol{\beta}}(\theta),\theta) \geq \overline{J}(q_1,\theta)$ for all $\theta \in \Theta$, we obtain $d \geq \int_{(T^-(q_2),1]} \frac{\overline{J}(q_2,\theta)-\overline{J}(q_1,\theta)}{q_2-q_1}\,\mathrm{d}F(\theta)$. We construct an upper bound for d. In particular, because we have $\overline{J}(\overline{\boldsymbol{\beta}}(\theta),\theta) \geq \overline{J}(q_2,\theta)$ for all $\theta \in \Theta$, we obtain $d \leq \int_{[T^-(q_1),1]} \frac{\overline{J}(q_2,\theta)-\overline{J}(q_1,\theta)}{q_2-q_1}\,\mathrm{d}F(\theta)$. By left continuity of T^-, we have $\lim_{q_1\to q^-} \frac{V(q)-V(q_1)}{q-q_1} = \int_{[T^-(q),1]} \overline{J}_1(q,\theta)\,\mathrm{d}F(\theta)$, implying that V is left differentiable and the left derivative takes the stated form. By known results in convex analysis (Hiriart-Urruty and Lemaréchal, 2001, Theorem 6.4), we obtain concavity of V by continuity of V and the monotonicity of the left-derivative (monotonicity follows from the same argument as concavity of V in the proof of Proposition A.2.)

There exists a unique maximizer q^M of $q \mapsto V(q) - c(q)$ by concavity of V, and $q^\star >\in (0, \overline{q})$ by the same argument as in Proposition A.1. To complete the proof, we argue that $q^M < q^\star$. We fix $q \in (0, \overline{q})$ in what follows and define the left-derivative of V at q as $\partial_- V(q)$. Because $\boldsymbol{a}(q) = 0$ for all $q \in Q$, it suffices to show that there exists $s \in \Theta$ such that $\partial_- V(q) = \int_{[s,1]} J_1(q,\theta)\,\mathrm{d}F(\theta)$; in fact, it holds that $u_1(q,\theta) - k'(q) - J_1(q,\theta) = \frac{1-F(\theta)}{F'(\theta)} u_{12}(q,\theta)$. First, if $T^-(q) = 1$, then $\partial_- V(q) = 0$, so the claim holds. Second, if $T^-(q) = 0$, then

$$\int_{[T^-(q),1]} \overline{J}_1(q,\theta)\,\mathrm{d}F(\theta) = \operatorname{conv} H(q,1) - \operatorname{conv} H(q,0) = H(q,1) - H(q,0).$$

in which the first equality holds by definition of $\overline{J}$, the second holds because $H(q,\cdot)$ equals $\operatorname{conv} H(q,\cdot)$ at the endpoints of $[0,1]$. Hence, the claim holds.

Third, we claim that: if $T^-(q) \in (0,1)$, then $\operatorname{conv} H(q, T^-(q)) = H(q, T^-(q))$. First, suppose that $T^-(q) \in (0,1)$ and $\operatorname{conv} H(q, T^-(q)) < H(q, T^-(q))$. Then, by continuity of $H(q, \cdot)$ and $\operatorname{conv} H(q, \cdot)$, there exists an open interval $I \subseteq \Theta$ such that $H(q, \theta) > \operatorname{conv} H(q, \theta)$ for all $\theta \in I$. Then, by the results in Toikka (2011, Footnote 12), $\overline{\boldsymbol{\beta}}$ is constant on I, which contradicts the definition of $T^-(q)$ and the fact that $T^-(q) \in I$. Hence, if $T^-(q) \in (0,1)$, then $\operatorname{conv} H(T^-(q), q) = H(T^-(q), q)$. As a result, $\int_{[T^-(q),1]} \overline{J}_1(q, \theta) \, \mathrm{d}F(\theta) = H(q, 1) - H(q, T^-(q))$, so, the proof is complete. □

B.3 Additional results for Section 4.2

For the proofs of the following results, we denote welfare by $W \colon \boldsymbol{q} \mapsto \int_\Theta u(\boldsymbol{q}(\theta), \theta) - k(\boldsymbol{q}(\theta)) \, \mathrm{d}F(\theta) - c(\sup \boldsymbol{q})$, we use $W^\star \colon q \mapsto \int_\Theta u(q, \theta) - k(q) \, \mathrm{d}F(\theta) - c(q)$, and $\mathbb{E}_n$ refers to the distribution induced by the mixed strategies defined in Proposition 6 for an equilibrium with n active firms.

Proposition B.2. *The following welfare comparisons hold.*

1. *Welfare is decreasing in the intensity of competition among symmetric competitive equilibria, that is,*

$$\mathbb{E}_{n+1}\{W(\boldsymbol{q}[\hat{x}, \hat{y}])\} \le \mathbb{E}_n\{W(\boldsymbol{q}[\hat{x}, \hat{y}])\} \quad \textit{for all } n \ge 2.$$

2. *If the monopoly fully bunches, then monopoly dominates duopoly, that is,*

$$\textit{if } q^M \le \boldsymbol{\beta}(0), \quad \textit{then} \quad W(\boldsymbol{q}^M) > \mathbb{E}_2\{W(\boldsymbol{q}[\hat{x}, \hat{y}])\};$$

instead, duopoly dominates monopoly if: costs are sufficiently close to fixed, type θ is uniformly distributed, and full bunching does not occur, that is,

$$\textit{if } c = \left(\frac{q}{a}\right)^\alpha \textit{ and } a > \boldsymbol{\beta}(0), \quad \textit{then} \quad \lim_{\alpha \to \infty} W(\boldsymbol{q}^M) < \lim_{\alpha \to \infty} \mathbb{E}_2\{W(\boldsymbol{q}[\hat{x}, \hat{y}])\}.$$

The result is a consequence of the following ones.

Proposition B.3. *Welfare is decreasing in the intensity of competition among competitive n equilibria, that is,*

$$\mathbb{E}_{n+1}\{W(\boldsymbol{q}[\hat{x}, \hat{y}])\} \le \mathbb{E}_n\{W(\boldsymbol{q}[\hat{x}, \hat{y}])\} \quad \textit{for all } n \ge 2.$$

Proof. Step 1: preliminary observations about information rents and order statistics. Fix a competitive n equilibrium that is symmetric. Every principal makes zero profits. The utility of type θ at the pricing stage is $R(x,y,\theta) := \int_{[0,\theta]} \max\{\min\{\boldsymbol{\beta}(\tilde{\theta}), x\}, y\}\,\mathrm{d}\tilde{\theta}$, given the realizations of x and y of, resp., the first- and second-order statistics from n i.i.d. draws distributed according to the distribution function G (Proposition 6). Note that $R(x,y,\theta)$ is nondecreasing in x and y, and is increasing in x if $\theta \in (b(x), \overline{\theta}]$ and increasing in y if $\theta \in [\underline{\theta}, b(y))$.

The conditional distribution of the second-order statistic $\hat{y}$ given that the first-order statistic's realization is x is given by $G(y \mid x) := (G(y)/G(x))^{n-1}$, $y \in Q$. So, by Proposition 6, the conditional distribution is given by $G(y \mid x) = (c'(y)/V'(y))/(c'(x)/V'(x))$ and is constant in n. The conditional distribution given by $G(y \mid x)$ is also increasing in the FOSD order as x increases. Specifically, it holds that $G(y \mid x) \leq G(y \mid x')$ for all $y, x, x' \in Q$ with $x \geq x'$, and with strict inequality if, in addition, we consider the interior of the relevant supports; i.e., for all $y, x, x' \in (0, \overline{q}^M)$ with $x < x'$ (Proposition 6).

The distribution of the first-order statistic $\hat{x}$, instead, is given by $(G(x))^n$. So, by Proposition 6, the distribution of $\hat{x}$ is nonincreasing in the FOSD order as n increases. Specifically, it holds that $(G(x))^n = (c'(x)/V'(x))^{n/(n-1)}$, and so

$$(c'(x)/V'(x))^{\frac{n}{n-1}} \leq (c'(x)/V'(x))^{\frac{n+1}{n}}.$$

for all $x \in Q$ and $n \geq 2$, with strict inequality if $x \in (0, \overline{q}^M)$.

Step 2: The conditional expectation of welfare given x is monotone in x and constant in n. By the above observations (Step 1) and known results about first-order-dominance (FOSD) order the conditional expectation of information rents given $\hat{x} = x$ is constant in n and nondecreasing in x. Specifically, the conditional expectation of information rents given $\hat{x} = x$ is

$$R(x) := \int_{[0,\overline{q}^M]} R(x,y,\theta)\,\mathrm{d}(G(y)/G(x))^{n-1},$$

and $R(x)$ is constant in n and nondecreasing in x. Moreover, $R(x) > R(x')$ if $\underline{\theta} < b(\hat{y}) < b(x')$ and $b(x) < \overline{\theta}$ with positive probability — as induced by the distribution given by the distribution function $y \mapsto (G(y)/G(x))^{n-1}$ for $\hat{y}$ given $\hat{x} = x$.

Step 3: The expected welfare is monotone in n. By the above observations (Step 1) and known results about FOSD , the expected information rents are nonincreasing in n. Specifically, define

$$R_n := \int_{[0,\bar{q}^M]} R(x)\,\mathrm{d}(G(x))^n,$$

and observe that $R_n = \mathbb{E}_n\{W(\boldsymbol{q}[\hat{x},\hat{y}])\}$; we have that $R_n \geq R_{n+1}$ for all $n \geq 2$. Moreover, monotonicity is strict if $\underline{\theta} < b(\hat{y}) < b(x')$ and $b(x) < \overline{\theta}$ with positive probability given n and $n+1$. $\square$

Proposition B.4. *The following statement holds:*

$$\textit{if } q^M \leq \boldsymbol{\beta}(0), \quad \textit{then} \quad W(\boldsymbol{q}^M) > \mathbb{E}_2\{W(\boldsymbol{q}[\hat{x},\hat{y}])\}.$$

Proof. Note that: the utility from a free good of quality y to type 0 is $g(y)$, and the marginal revenue under full bunching is given by g'.

Step 1: Welfare under full bunching and monopoly. Under full bunching at q^M, welfare is given by $W^\star(q^M)$, that is $W^\star(q^M) = \int_{[0,q^M]} w^\star(q)\,\mathrm{d}q$, using $w^\star(q) := g'(q) + \int_\Theta \theta\,\mathrm{d}F(\theta) - c'(q)$.

Step 2: Welfare under full bunching and competition. Fix a competitive and symmetric n equilibrium and suppose $\boldsymbol{\beta}(0) \geq x$. In this case, using full bunching and $\underline{\theta} = 0$, we obtain $V'(q) = g'(q)$ for all $q \in (0, q^M]$. Welfare conditional on x and y is $R(x,y,\theta) = g(x) + \theta x - (g(x) - g(y))$. So, the conditional expectation of welfare given x is

$$\begin{aligned} R(x) &:= \int_{[0,q^M]} R(x,y,\theta)\,\mathrm{d}\big((G(y)/G(x))^{n-1}\big), \\ &= \int_{[0,q^M]} g(y)\,\mathrm{d}\big((G(y)/G(x))^{n-1}\big) + x\int_\Theta \theta\,\mathrm{d}F(\theta). \end{aligned}$$

Using the formula for G in Proposition 6 and integration by parts, we have

$$R(x) = g(x) - c(x)\frac{c'(x)}{g'(x)} + x\int_\Theta \theta\,\mathrm{d}F(\theta).$$

So, we express expected welfare under full bunching and competition as

$$\begin{aligned} R_n &:= \int_{[0,q^M]} R(x)\,\mathrm{d}(G(x))^n, \\ &= \int_{[0,q^M]} g(x) - c(x)\frac{g'(x)}{c'(x)} + x\int_\Theta \theta\,\mathrm{d}F(\theta)\,\mathrm{d}(c'(x)/V'(x))^{n/(n-1)}. \end{aligned}$$

Step 3: Welfare under full bunching and duopoly. We express expected welfare under full bunching and a 2 equilibrium that is competitive and symmetric as

$$R_2 = \int_{[0,q^M]} g(x) - c(x)\frac{g'(x)}{c'(x)} + x\int_\Theta \theta \, \mathrm{d}F(\theta) \, \mathrm{d}(c'(x)/V'(x))^2,$$
$$= g(q^M) - c(q^M)\frac{g'(q^M)}{c'(q^M)} - \int_{[0,q^M]} \left[\frac{\partial}{\partial x}\left(g(x) - c(x)\frac{g'(x)}{c'(x)}\right) + \int_\Theta \theta \, \mathrm{d}F(\theta)\right](c'(x)/V'(x))^2 \, \mathrm{d}x,$$

using integration by parts. Using $V'(x) = g'(x)$ and $\frac{c'(\overline{q}^M)}{g'(\overline{q}^M)} = 1$, we have

$$W^\star(q^M) - R_2 = \int_{[0,q^M]} \left[\frac{\partial}{\partial x}\left(g(x) - c(x)\frac{g'(x)}{c'(x)}\right) + \int_\Theta \theta \, \mathrm{d}F(\theta)\right](c'(x)/V'(x))^2 \, \mathrm{d}x.$$

The claim holds because we have $\int \theta \, \mathrm{d}F > 0$ and

$$\frac{\partial}{\partial x}\left(g(x) - c(x)\frac{g'(x)}{c'(x)}\right) = -c(x)\frac{g''(x)c'(x) - g'(c)c''(x)}{(c'(x))^2} \geq 0.$$

□

Proposition B.5. *Assume that: preferences are linear, $c(q) = (q/a)^\alpha$, for $\alpha > 1,\ a > 0$, and that θ is uniformly distributed. The following statement holds:*

$$\mathbb{E}_2\{W(\boldsymbol{q}[\hat{x}, \hat{y}])\} - W(\boldsymbol{q}^M) \to \frac{1}{8}a \text{ as } \alpha \to \infty.$$

Proof. We establish the result using the following claims.

(0) The following equality holds:

$$\mathbb{E}_2\{W(\boldsymbol{q}[\hat{x}, \hat{y}])\} - W(\boldsymbol{q}^M) = \mathbb{E}_2\left\{\frac{1}{8}(x + y - q^M) + \frac{1}{4}y\right\} - \frac{1}{4}\int_Q q \, \mathrm{d}H(q),$$

defining $H\colon Q \mapsto \mathbb{R} : q \mapsto \min\left\{\frac{c'(q)}{V'(q)}, 1\right\}$. To see that the equality holds: consumer surplus of a monopolist is

$$U^M := \int_\Theta q^M(\theta - \varphi^{-1}(0))_+ \, \mathrm{d}F(\theta)$$
$$= q^M/8;$$

the consumer surplus in oligopoly is

$$
\begin{aligned}
U^O(x,y) :=& y\int_{[0,\varphi^{-1}(0)]} \theta \,\mathrm{d}F(\theta) + \int_{(\varphi^{-1}(0),1]} x\theta - (x-y)\varphi^{-1}(0)\,\mathrm{d}F(\theta) \\
=& y/8 + x3/8 - (x-y)/4 \\
=& x/8 + 3y/8;
\end{aligned}
$$

the change in consumer surplus is

$$
\begin{aligned}
\Delta(x,y) :=& U^O(x,y) - U^M \\
=& \frac{1}{8}(x+y-q^M) + \frac{1}{4}y;
\end{aligned}
$$

the monopoly profits are

$$
\begin{aligned}
\Pi^M &= (1-\varphi^{-1}(0))\varphi^{-1}(0)q^M - c(q^M) \\
&= \frac{1}{4}q^M - c(q^M) \\
&= \frac{1}{4}\int_{[0,q^M]} 1 - H(q)\,\mathrm{d}q \\
&= \frac{1}{4}\int_Q q\,\mathrm{d}H(q).
\end{aligned}
$$

(1) $q^M \to a$ as $\alpha \to \infty$. By direct calculation: $q^M = \left(\frac{a^\alpha}{4\alpha}\right)^{\frac{1}{\alpha-1}}$. The result follows because $\log(4\alpha)^{\frac{1}{\alpha-1}} \to 0$ as $\alpha \to \infty$.

(2) The r.v. $\hat{y} \mid \hat{x} = x$ tends to x weakly as $\alpha \to \infty$. Define the r.v. Y_x on Q with the conditional distribution of $\hat{y}$ given that $\hat{x} = x$, for $x \in Q$. The result follows because the distribution function of Y_x is $q \mapsto \left(\min\left\{\frac{c'(q)}{c'(x)}, 1\right\}\right)^{\alpha-1}$. Specifically, it holds that Y_x approaches the constant x in probability as $\alpha \to \infty$, for all $x \in Q$.

(3) The r.v. Z with distribution function H tends to a weakly as $\alpha \to \infty$. Define the r.v. Z on Q with the distribution function H. The result follows because, by direct calculation, $H\colon q \mapsto \frac{\alpha}{a}\left(\min\left\{\frac{q}{a}, 1\right\}\right)^{\alpha-1}$.

(4) The r.v. $\hat{x}$ tends to a weakly as $\alpha \to \infty$. Define the r.v. X with distribution function $q \mapsto H(q)H(q)$. The result follows from part (3).

By properties of weak convergence, and the fact that the support of the relevant random variables are contained in $[0, q^*]$, the result follows. $\square$

B.4 On the expression for marginal revenues

In this section, we provide more details for the argument that $b(q)$ is the type θ that maximizes the marginal-revenue expression $(1 - F(\theta))(g'(q) + \theta)$; most of these details appear, e.g., in (Wilson, 1993).

A different way to formalize the monopolist problem is as a choice of a tariff, i.e., a function $T\colon Q \to \mathbb{R}$. The monopolist solves the problem $\mathcal{P}'$

$$\sup_{\boldsymbol{q}, T} \int_{\Theta} T(\boldsymbol{q}(\theta)) \,\mathrm{d}F(\theta) - c(\sup \boldsymbol{q}(\Theta)) \text{ subject to:}$$
$$\text{for all } (\theta, q) \in \Theta \times Q,\ u(\boldsymbol{q}(\theta), \theta) - T(\boldsymbol{q}(\theta)) \geq u(q, \theta) - T(q),\ u(\boldsymbol{q}(\theta), \theta) - T(\boldsymbol{q}(\theta)) \geq 0.$$

This problem is equivalent to the original monopolist problem $\mathcal{P}^M$ (Guesnerie and Laffont, 1984). The unique solution $(\boldsymbol{q}^M, t^M)$ to $\mathcal{P}^M$, characterized in Proposition A.2, can be implemented by a tariff T^M such that $T(q) = t^M(b(q))$ if $q \in \boldsymbol{q}(\Theta)$. Additionally, if the hazard rate of F is increasing, then T^M is concave. In what follows, we consider $\mathcal{P}'$ under the restriction that: T is concave, and, for all θ, $u(q, \theta) - T(q)$ is convex as a function of q in a neighborhood of $\boldsymbol{q}^M(\theta)$. This condition holds if $T = T^M$, so is without loss in our setup, and is common for this direct approach to screening, see Section 8.1 in Wilson (1993) for sufficient conditions in terms of the primitives.

After a change of variables, the revenues are expressed as $\int_{[0, \sup \boldsymbol{q}(\Theta)]} T(q) \,\mathrm{d}H(q)$, letting $H(q) \coloneqq F(\partial_+ T(q) - g'(q))$, for the right-derivative $\partial_+ T$ of T. Integrating by parts, we obtain $\int_{[0, \sup \boldsymbol{q}(\Theta)]} T'(q)(1 - H(q)) \,\mathrm{d}q$ (Machina, 1982, Lemma 2, for which we use concavity of T.) Hence, we find the optimal tariff T by setting its derivative T' to solve the revenue maximization "pointwise," up to a term that is invariant in quality. The optimal price schedule T' satisfies $T'(q) \in \operatorname{Argmax}_{p \in \mathbb{R}_+} p(1 - F(p - g'(q)))$; intuitively, any marginal quality increment is priced as if it constitutes an individual market.

From the solution to the problem $\mathcal{P}(\overline{q})$ and the fact that the optimal T has $T(q) = t^M(b(q))$, we know that $T'(q)$ is equal to $g'(q) + b(q)$. Moreover, if $y^* = p^* - g'(q)$, then p^* maximizes $p \mapsto p(1 - F(p - g'(q)))$ iff y^* maximizes $y \mapsto (y + g'(q))(1 - F(y))$. So, the reason why $b(q)$ maximizes the marginal-revenue expression $(g'(q) + \theta)(1 - F(\theta))$ by choice of type θ is the fact that the marginal utility of type $b(q)$ is the optimal price of a marginal quality increment from q.

Remark B.2. *The formula for V' and the property of b described in the main* *text can be extended. Define $s(p,q) = \inf\{\theta \in \Theta \mid u_1(q,\theta) \geq p\}$ and let $(\boldsymbol{q}^*, t^*)$ solve the monopolist problem $\mathcal{P}(q)$. It holds that*

1. $V'(q) = u_1(q, b(q))(1 - F(b(q)))$;

2. $u_1(q, b(q)) \in \operatorname{Argmax}_{p \in \mathbb{R}_+} p(1 - F(s(p,q)))$.

First, by Proposition A.2 and the envelope theorem, we have that $t^*(b(q)) = u(q, b(q)) - \int_{[0,b(q)]} u_2(\boldsymbol{q}^*(\tilde{\theta}), \tilde{\theta})\, \mathrm{d}\tilde{\theta}$ (Milgrom and Segal, 2002). Hence, by continuous differentiability of b on $(\boldsymbol{\beta}(0), \infty)$, we have $(t^* \circ b)'(q) = u_1(q, b(q))$ for all $q > \boldsymbol{\beta}(0)$. Therefore, by the expression for V derived in the preceding paragraph, we have $V'(q) = (1 - F(b(q)))u_1(q, b(q))$.

References

Admati, Anat R. and Paul Pfleiderer (1986), "A monopolistic market for information." *Journal of Economic Theory*, 39(2), 400–438.

Bergemann, Dirk and Alessandro Bonatti (2019), "Markets for information: An introduction." *Annual Review of Economics*, 11, 85–107.

Bergemann, Dirk, Alessandro Bonatti, and Alex Smolin (2018), "The design and price of information." *American Economic Review*, 108(1), 1–48.

Bergemann, Dirk, Alessandro Bonatti, and Alex Smolin (2025a), "The economics of large language models: Token allocation, fine-tuning, and optimal pricing." URL https://arxiv.org/abs/2502.07736.

Bergemann, Dirk and Rahul Deb (2025), "Robust pricing for cloud computing." URL https://arxiv.org/abs/2502.07168.

Bergemann, Dirk, Tibor Heumann, and Stephen Morris (2025b), "Screening with persuasion." URL https://arxiv.org/abs/2212.03360.

Bergemann, Dirk and Marco Ottaviani (2021), "Information markets and nonmarkets." In *Handbook of Industrial Organization* (Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, eds.), volume 4, 593–672, Elsevier, Amsterdam.

Boiteux, Marcel (1960), "Peak-load pricing." *The Journal of Business*, 33(2), 157–179.

Bonatti, Alessandro, Munther Dahleh, Thibaut Horel, and Amir Nouripour (2024), "Selling information in competitive environments." *Journal of Economic Theory*, 216, 105779.

Carroll, Gabriel (2023), "Contract theory." In *Online and Matching-Based Market Design* (Federico Echenique, Nicole Immorlica, and Vijay V. Vazirani, eds.), 614–634, Cambridge University Press, Cambridge, UK.

Chade, Hector and Jeroen Swinkels (2021), "Screening in vertical oligopolies." *Econometrica*, 89(3), 1265–1311.

Champsaur, Paul and Jean-Charles Rochet (1989), "Multiproduct duopolists." *Econometrica*, 57(3), 533–557.

Corrao, Roberto, Joel P. Flynn, and Karthik A. Sastry (2023), "Nonlinear pricing with underutilization: A theory of multi-part tariffs." *American Economic Review*, 113(3), 836–60.

Deneckere, Raymond J. and R. Preston McAfee (1996), "Damaged goods." *Journal of Economics & Management Strategy*, 5(2), 149–174.

Gale, Ian L. and Thomas J. Holmes (1993), "Advance-purchase discounts and monopoly allocation of capacity." *The American Economic Review*, 83(1), 135–146.

Garrett, Daniel F., Renato Gomes, and Lucas Maestri (2019), "Competitive screening under heterogeneous information." *The Review of Economic Studies*, 86(4), 1590–1630.

Goldfarb, Avi and Catherine Tucker (2019), "Digital economics." *Journal of Economic Literature*, 57(1), 3–43.

Grubb, Michael D. (2009), "Selling to overconfident consumers." *American Economic Review*, 99(5), 1770–1807.

Guesnerie, Roger and Jean-Jacques Laffont (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(3), 329–369.

Hahn, Jong-Hee (2006), "Damaged durable goods." *The RAND Journal of Economics*, 37(1), 121–133.

Halac, Marina, Elliot Lipnowski, and Daniel Rappaport (2024), "Pricing for coordination."

Hiriart-Urruty, Jean-Baptiste and Claude Lemaréchal (2001), *Fundamentals of Convex Analysis*. Springer, Berlin, Heidelberg.

Ichihashi, Shota (2021), "The economics of data externalities." *Journal of Economic Theory*, 196, 105316.

Inderst, Roman (2008), "Durable goods with quality differentiation." *Economics Letters*, 100(2), 173–177.

Jehiel, Philippe, Benny Moldovanu, and Ennio Stacchetti (1999), "Multidimensional mechanism design for auctions with externalities." *Journal of Economic Theory*, 85(2), 258–293.

Johnson, Justin P. and David P. Myatt (2003), "Multiproduct quality competition: Fighting brands and product line pruning." *American Economic Review*, 93(3), 748–774.

Kreps, David M. and José A. Scheinkman (1983), "Quantity precommitment and Bertrand competition yield Cournot outcomes." *The Bell Journal of Economics*, 14(2), 326–337.

Loertscher, Simon and Ellen Muir (2022), "Monopoly pricing, optimal randomization, and resale." *Journal of Political Economy*, 130(3), 566–635.

Machina, Mark J. (1982), "Expected utility analysis without the independence axiom." *Econometrica*, 50(2), 277–324.

Maskin, Eric and John Riley (1984), "Monopoly with incomplete information." *The RAND Journal of Economics*, 15(2), 171–196.

Milgrom, Paul and Ilya Segal (2002), "Envelope theorems for arbitrary choice sets." *Econometrica*, 70(2), 583–601.

Moulin, Hervé (1994), "Serial cost-sharing of excludable public goods." *The Review of Economic Studies*, 61(2), 305–325.

Mussa, Michael and Sherwin Rosen (1978), "Monopoly and product quality." *Journal of Economic Theory*, 18(2), 301–317.

Oren, Shmuel, Stephen Smith, and Robert Wilson (1985), "Capacity pricing." *Econometrica*, 53(3), 545–566.

Rochet, Jean-Charles and Jean Tirole (2003), "Platform competition in two-sided markets." *Journal of the European Economic Association*, 1(4), 990–1029.

Rodríguez Olivera, Rosina (2024), "Strategic incentives and the optimal sale of information." *American Economic Journal: Microeconomics*, 16(2), 296–353.

Rothschild, Michael and Joseph Stiglitz (1976), "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information." *The Quarterly Journal of Economics*, 90(4), 629–649.

Sandmann, Christopher (2025), "When are sparse menus profit-maximizing?"

Segal, Ilya (1999), "Contracting with externalities." *The Quarterly Journal of Economics*, 114(2), 337–388.

Segal, Ilya and Michael D. Whinston (2003), "Robust predictions for bilateral contracting with externalities." *Econometrica*, 71(3), 757–791.

Shapiro, Carl and Hal R. Varian (1998), "Versioning: The smart way to sell information." *Harvard Business Review*, 76(6), 106–118.

Stole, Lars A. (2007), "Price discrimination and competition." In *Handbook of Industrial Organization* (M. Armstrong and R. Porter, eds.), volume 3, 2221–2299, Elsevier, Amsterdam.

Toikka, Juuso (2011), "Ironing without control." *Journal of Economic Theory*, 146(6), 2510–2526.

Topkis, Donald M. (1978), "Minimizing a submodular function on a lattice." *Operations Research*, 26(2), 305–321.

Wilson, Robert (1993), *Nonlinear pricing*. Oxford University Press, New York.