American Economic Review 2025, 115(9): 2970–3013
https://doi.org/10.1257/aer.20230455

Nested Bundling†

By Frank Yang*

A nested bundling strategy creates menus in which more expensive bundles include all the goods of less expensive ones. We study when nested bundling is optimal and determine which nested menu is optimal, when consumers differ in one dimension. We define a partial order on bundles by (i) set inclusion and (ii) sales quantity when sold alone. We show that, under quasi-concavity assumptions, if the undominated bundles with respect to this partial order are nested, then nested bundling is optimal. We present an iterative algorithm that identifies the minimal optimal menu consisting of a subset of the undominated bundles. (*JEL* D11, D21, D42, D82, M31)

How to sell multiple products? This question, also known as optimal bundling, is of substantial economic importance to multiproduct firms. A common bundling strategy is to create *nested* menus, in which more expensive bundles include all the goods of the less expensive ones. This strategy is widely adopted across various industries, including streaming services (e.g., Netflix), software companies (e.g., Slack), e-commerce platforms (e.g., Shopify), and most recently AI services (e.g., OpenAI).[1] When is such a strategy profit maximizing? Which items to package in one tier versus another tier? How many tiers are optimal?

Even though these questions seem to be fundamental, relatively little is known because characterizations of optimal bundling are generally intractable. For instance, the optimal mechanism for selling two goods with additive and independent values remains unknown except for a few special cases (Manelli and Vincent 2006).[2]

In this paper, we answer these questions when consumers are ordered in one dimension where a higher-type consumer has higher incremental values for larger bundles. This dimension could represent, for instance, income levels in retail pricing

* Department of Economics, Harvard University (email: fyang@fas.harvard.edu). Jeff Ely was the coeditor for this article. This paper has been previously distributed with the title "The Simple Economics of Optimal Bundling." I thank especially Andy Skrzypacz, Mohammad Akbarpour, Matthew Gentzkow, Paul Milgrom, and Michael Ostrovsky for their guidance and support. I would also like to thank the coeditor and four anonymous referees, as well as Itai Ashlagi, Ben Brooks, Jeremy Bulow, Peter DeMarzo, Piotr Dworczak, Soheil Ghili, Ravi Jagadeesan, Emir Kamenica, David Kreps, Irene Lo, Alessandro Pavan, Brad Ross, Ilya Segal, Takuo Sugaya, Bob Wilson, Weijie Zhong, and participants at conferences and seminars for helpful comments and suggestions.

† Go to https://doi.org/10.1257/aer.20230455 to visit the article page for additional materials and author disclosure statement(s).

[1] For instance, Netflix offers three tiers (Netflix 2023): "Standard with ads" (ads, 1080p resolution, no downloads), "Standard" (no ads, 1080p resolution, downloads), and "Premium" (no ads, 4K + HDR, downloads). See Supplemental Appendix B.1 for details of these examples.

[2] With correlated values, it is known that restricting menus to any bounded size can lead to an arbitrarily small fraction of the optimal revenue (Hart and Nisan 2019). Even finding the optimal mechanism computationally is intractable (Daskalakis, Deckelbaum, and Tzamos 2014).

or organizational complexity in enterprise pricing. With that simplifying assumption, we are able to allow the consumers to have general nonadditive values (in particular, heterogeneous preferences over different items and complementary or substitutable preferences across different items) and sellers to have arbitrary costs for producing different bundles. While we assume that types are one-dimensional, the problem of bundling is inherently multidimensional because screening can be done using multiple instruments.

Building on Ghili (2023), we consider the following partial order on the set of bundles: A bundle b_1 is *dominated* ($\preceq$) by another bundle b_2 if (i) b_1 is a subset of b_2 and (ii) b_1 has a lower sales quantity than b_2 when each bundle is sold alone at its monopoly price. This partial order can be readily determined by examining the demand curve for each bundle separately. However, it turns out that this simple partial order, under quasi-concavity assumptions, characterizes the optimal bundling strategy.

Our first main result (Theorem 1) shows that if the undominated bundles can be totally ordered by set inclusion (the *nesting condition*), then nested bundling, in particular a menu of undominated bundles, is optimal. The proof is constructive: It presents an iterative procedure (the *sieve algorithm*) to find the minimal optimal menu that consists of a subset of the undominated bundles (Proposition 1).

In the absence of our nesting condition, even though the consumers are ordered in one dimension, the optimal mechanism need not be a nested menu, can involve dominated bundles, and may require randomization (see Example 1). We provide a sufficient condition (Proposition 2) for the nesting condition to hold, the *union quantity condition*, which simply asks the sold-alone quantity for the union of two bundles to be above the minimum of their individual sold-alone quantities. We also provide a partial converse (Proposition 3): If nested bundling is optimal, then the minimal optimal menu must include the two extremal bundles under our partial order—the grand bundle and the bundle with the highest sold-alone quantity—and must exclude any bundle dominated by another bundle in the menu.

On the technical side, our proof technique involves a new monotone comparative statics result. A crucial step toward our results is to ensure the monotonicity of the solution to a relaxed problem. Our main technical result (Theorem 2), which we call the *monotone construction theorem*, states that for any objective function on $\mathcal{X} \times \mathbb{R}$ satisfying the single-crossing property, where the choice set $\mathcal{X}$ is a partially ordered set, monotone comparative statics hold if the *chain-essential* elements in $\mathcal{X}$—the elements that cannot be removed from any chain (totally ordered subset) without decreasing the objective at some parameter—form a chain themselves. Unlike existing monotone comparative statics results, our theorem is constructive: It characterizes the range of the maximizers across parameters. Besides this constructive property, our comparative statics result also generalizes Milgrom and Shannon (1994) by providing a new condition that is agnostic to whether the choice variables exhibit complementarity or substitutability (see Section IIID).

Applications.—Besides the direct implications on optimal bundling, we present three applications of our main results.

In the first application, we provide a sufficient condition for our nesting condition using price elasticity—the *union elasticity condition*—which states that if the demand curves for two different bundles are both elastic at a certain quantity, then the demand curve for their union is also elastic at that quantity (see Section IVA). With zero marginal costs, this condition implies the nesting condition and hence the optimality of nested bundling. We also show how the optimal menu can be iteratively constructed by using items with more elastic demand curves as the *basic* items and items with more inelastic demand curves as the *upgrade* items (Proposition 6). In this case, a large bundle, if sold alone, has a sales quantity lower than its elastic items but higher than its inelastic items.

The full characterization of optimal mechanisms enables comparative statics analysis. We find that as the dispersion of values for one item increases, the monopolist switches the tiers of different items and adopts a menu size that is *U*-shaped in the dispersion parameter (Proposition 7). These comparative statics results differ significantly from those in the standard quality-differentiated goods model, such as in Johnson and Myatt (2006), as our model allows for a much richer set of preferences.

Our second application is to the quality-differentiated goods model (a la Mussa and Rosen 1978), which is a special case of our model in which there are no heterogeneous relative preferences (see Section IVB). Even in this well-studied case, our results yield new insights by providing a new characterization of the optimal menu (Proposition 8). Using this characterization, we can hold the price elasticities of different qualities constant and study the effects of cost structures on product line design. We show that it is always profitable to prune the region of a product line where the average cost curve is above its lower increasing envelope (Proposition 9). This result generalizes a finding by Johnson and Myatt (2003) and refines their intuition about when segmenting markets is profitable. Our results are new even for one-dimensional screening problems because we impose much weaker regularity assumptions compared to the textbook treatment, owing to our monotone construction theorem. This generality allows rich forms of *bunching*, which is ruled out by standard assumptions but can be characterized by our dominance order.

Our third application shows how our bundling results can provide insights into other multidimensional screening problems. Building on a connection between bundling and costly screening from Yang (2022), we use our main results to characterize when costly screening is optimal for a principal who can use both price and nonprice instruments, such as waiting time (see Section IVC). We obtain (see Proposition 10) a necessary and sufficient condition for the optimality of costly screening when the agent has *negatively correlated preferences* (higher types have higher disutilities), complementing Yang (2022), which shows that costly screening is always suboptimal when the agent has *positively correlated preferences* (higher types have lower disutilities). Our result shows that when higher types have higher disutilities, a key metric that determines the optimality of costly screening is the elasticity of disutility with respect to the agent's types.

Discussion of Intuition.—We now present the key intuition behind our main results (see Section IID for further discussion).

A key feature of the one-dimensional type space is that we can compute the total revenue from any feasible allocation, induced by some menu, from the *sold-alone* *marginal revenue curves*, regardless of how complex the allocation might be.[3] To see this, let $P(b,q)$ be the demand curve for bundle b when bundle b is sold alone, and $MR(b,q)$ be the corresponding sold-alone marginal revenue curve for bundle b.

Because consumers are totally ordered, we can arrange them along a *single* quantity axis, with consumers positioned toward the right end having lower values for all the bundles.[4] For a given menu of bundles and prices, the consumers optimally choose their favorite options, resulting in an allocation rule $b(q)$, describing the bundle choice for the consumer located at quantity q. Let $CS(q)$ be the surplus of the consumer located at quantity q (and suppose $CS(1) = 0$). The total revenue can be computed as follows:

Fact 1: Total Revenue

$$= \int_0^1 P(b(q),q)dq - \int_0^1 CS(q)dq \text{ (Value} - \text{Consumer Surplus)}$$

$$= \int_0^1 P(b(q),q)dq + \int_0^1 q \cdot CS'(q)dq \text{ (Integration by Parts)}$$

$$= \int_0^1 P(b(q),q)dq + \int_0^1 q \cdot P_q(b(q),q)dq \text{ (Envelope Theorem)}$$

$$= \int_0^1 MR(b(q),q)dq.$$

Now, suppose that we have two items $\{1,2\}$ and that costs are zero.[5] Suppose that the sold-alone quantities for the three possible bundles are $Q(\{1\}) < Q(\{1,2\}) < Q(\{2\})$. In this case, bundle $\{1\}$ is dominated by bundle $\{1,2\}$, while bundle $\{2\}$ is not dominated by bundle $\{1,2\}$. Thus, the undominated bundles are nested. Suppose that the revenue function for selling any *incremental bundle*—the option to upgrade from a smaller bundle to a larger bundle—is strictly quasi-concave.[6] This implies that (i) the MR curves cross zero once from above and (ii) the MR curve of a larger bundle also crosses the MR curve of a smaller bundle at most once from above.

There are three key observations. First, because of the ordering $Q(\{1\}) < Q(\{1,2\})$, these two quantities must be located in the region where the marginal revenue of upgrading consumers from bundle $\{1\}$ to bundle $\{1,2\}$ is *positive* (i.e., to the *left* of the vertical dashed line in Figure 1, panel A. This then implies that if it is profitable to sell a consumer the smaller bundle $\{1\}$, which happens before

[3] The use of marginal revenue curves in mechanism design has a long tradition, beginning with Bulow and Roberts (1989) in auction settings, and has been recently applied to bundling settings in Ghili (2023).

[4] Formally, the one-dimensional type space is the unique comonotonic joint distribution over bundle values that induces the single-bundle demand curves $P(b,q)$ for each bundle b.

[5] When there are positive costs, the same intuition discussed here applies by replacing the marginal revenue curves to be the marginal profit curves.

[6] For the sake of this example, we assume that these incremental revenue functions are globally quasi-concave. This quasi-concavity assumption is stronger than what we actually assume in the model (see Section I).

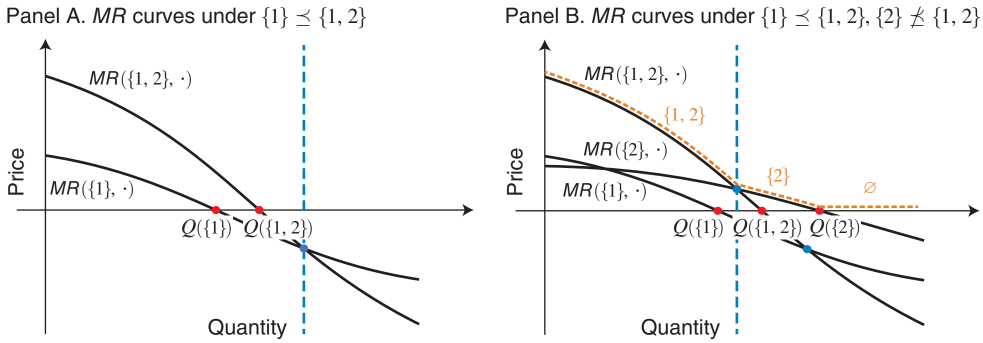

Figure 1. Illustration of the Marginal Revenue Curves

quantity $Q(\{1\})$, it is even more profitable to *upgrade* the consumer to the larger bundle $\{1,2\}$.

Second, because of the opposite ordering $Q(\{2\}) > Q(\{1,2\})$, these two quantities must be located in the region where the marginal revenue of upgrading consumers from bundle $\{2\}$ to bundle $\{1,2\}$ is *negative* (i.e., to the *right* of the vertical dashed line in Figure 1, panel B). This then implies that, after a certain quantity threshold, there always exists a region in which it is more profitable to *downgrade* the consumers from the larger bundle $\{1,2\}$ to the smaller bundle $\{2\}$.

Third, with the upgrade and downgrade operations, we can attain the *upper envelope* of the MR curves by allocating bundles to consumers in a *monotone* fashion such that a higher-type consumer receives a larger bundle in the set-inclusion order (as depicted by the bundle assignments in Figure 1, panel B). Now, because a higher-type consumer also has higher incremental values for larger bundles, we can implement this monotone allocation using upgrade prices as follows: Set the price of bundle $\{2\}$ to be its usual monopoly price, and set the price of upgrading from bundle $\{2\}$ to bundle $\{1,2\}$ to be such that the consumer located at the threshold quantity (indicated by the vertical dashed line in Figure 1, panel B) is indifferent to whether to upgrade.

The monotonicity of the allocation is crucial to guarantee that we can in fact "climb up" the MR curves. When the upper envelope of the MR curves cannot be attained by a monotone allocation rule, the optimal mechanism may require selling dominated bundles and may even do so with randomization (see Example 1). For our running example, this monotonicity is self-evident once we recognize that the configuration of MR curves must resemble Figure 1, panel B. However, in the general many-item case, it is impossible to exhaustively list all possible configurations of the MR curves. The proof relies on the new constructive monotone comparative statics result (Theorem 2).

Related Literature.—We study nested bundling when consumers have one-dimensional heterogeneity and nonadditive preferences. Our model builds on a recent literature studying the optimality of pure bundling (i.e., selling only the grand bundle) with nonadditive values (Ghili 2023; Haghpanah and Hartline 2021).[7] The closest paper is Ghili (2023), who introduces the sold-alone quantities and shows that, under quasi-concavity assumptions, pure bundling is optimal if and only if the grand bundle has the highest sold-alone quantity. Ghili's (2023) result motivates our partial order. Under his condition, the grand bundle is the unique undominated bundle, and hence, our nesting condition is trivially satisfied (Corollary 3).

There is a substantial literature on multidimensional screening and optimal bundling (beginning with Stigler 1963; Adams and Yellen 1976; McAfee, McMillan, and Whinston 1989). A general lesson is that some form of bundling is generically profitable, but characterizing optimal bundling strategies turns out to be very difficult (Armstrong 1996; Rochet and Chone 1998; Carroll 2017). Because of this difficulty, relatively little is known about how optimal bundling strategies depend on economic primitives such as price elasticities and cost structures. This paper departs from most of the bundling literature, which assumes additive values and multidimensional heterogeneity (McAfee and McMillan 1988; Manelli and Vincent 2007; Pavlov 2011; Daskalakis, Deckelbaum, and Tzamos 2017). In particular, Bergemann et al. (2022) study nested bundling with additive values and obtain conditions that are not directly comparable to ours.[8] Compared to the literature, we propose an alternative set of assumptions that might explain the popularity of nested bundling.[9] By doing so, we are also able to connect the empirically relevant economic primitives to the structure of optimal bundling strategies.

Our main proof technique uses a Myersonian approach by maximizing a suitably defined virtual surplus function pointwise. A key technical contribution is to provide conditions under which the solution to this relaxed problem is implementable. Our main technical result, the monotone construction theorem, delivers such conditions and furthermore constructs the optimal solution. The monotone construction theorem connects to the literature on monotone comparative statics. Unlike the existing monotone comparative statics results (Milgrom and Shannon 1994; Athey 2002; Quah 2007; Quah and Strulovici 2009), our theorem is constructive and does not require a lattice structure; when the choice set is a lattice, our theorem generalizes Milgrom and Shannon (1994) by providing a new condition that is agnostic to whether the choice variables exhibit complementarity or substitutability.

The remainder of the paper proceeds as follows. Section I presents the model. Section II presents the main results. Section III sketches the main proof. Section IV presents the applications. Section V concludes. Appendix A provides omitted proofs. The Supplemental Appendix provides additional results and proofs.

[7] There is a long-standing literature on the profitability of price discrimination (Deneckere and McAfee 1996; Johnson and Myatt 2003; Anderson and Dana 2009), which can be seen as studying the optimality of pure bundling under more restrictive assumptions. There is also an extensive literature on nonlinear pricing (a la Mussa and Rosen 1978), which is a special case of our model.

[8] In a different context, Gomes and Pavan (2016) obtain conditions for a two-sided monopolistic platform to use a nested matching rule.

[9] Of course, as in the literature, our model is based on the standard theory of rational choices. Nested bundling might also arise as firms' responses to behavioral or boundedly rational consumers, e.g., to avoid choice overloading (Iyengar and Lepper 2000) or to influence sales through context effects (Simonson 1989).

I. Model

A monopolist sells n different goods $\{1, \ldots, n\}$ to a unit mass of consumers.

Consumers have types $t \in \mathcal{T} := [\underline{t}, \bar{t}]$. Types are drawn from a distribution F with a continuous, positive density f. Type t has value $v(b,t)$ for *bundle* $b \in \mathcal{B} := 2^{\{1,\ldots,n\}}$, with $v(\varnothing, t) = 0$. For any stochastic assignment $a \in \Delta(\mathcal{B})$, we define $v(a,t) := \mathbb{E}_{b\sim a}[v(b,t)]$. The monopolist incurs cost $C(b)$ to produce bundle b, with $C(\varnothing) = 0$. We assume that it is efficient for the highest type $\bar{t}$ to consume all the items: $\arg\max_b \{v(b,\bar{t}) - C(b)\} = \bar{b}$, where $\bar{b} := \{1, \ldots, n\}$ is the grand bundle.

The value function $v(b,t)$ is (i) nondecreasing in b (in the set-inclusion order), (ii) continuously differentiable in t, and (iii) strictly increasing in t whenever $v(b,t) > 0$. In addition, we will make the following monotonicity assumption:

A1 (Incremental Monotonicity): For any two nested bundles $b_1 \subset b_2$, the incremental value $v(b_2,t) - v(b_1,t)$ is strictly increasing in t whenever it is strictly positive.

The seller wants to maximize expected profits over all stochastic mechanisms. By the revelation principle, it is without loss of generality to restrict attention to direct mechanisms. Specifically, a (*stochastic*, *direct*) *mechanism* is a measurable map $(a,p) : \mathcal{T} \to \Delta(\mathcal{B}) \times \mathbb{R}$ that satisfies the usual incentive compatibility (IC) and individual rationality (IR) conditions:

$$v(a(t),t) - p(t) \geq v(a(\hat{t}),t) - p(\hat{t}) \text{ for all } t, \hat{t} \text{ in } \mathcal{T};$$

$$v(a(t),t) - p(t) \geq 0 \text{ for all } t \text{ in } \mathcal{T}.$$

Two mechanisms are *equivalent* if they differ on a zero-measure set of types.

A *menu* is a set of bundles (which we assume includes $\varnothing$).[10] A menu B is *optimal* if there exists an optimal mechanism (a,p) such that $a(t) \in B$ for all t.[11] Note that an optimal menu need not exist since the optimal mechanism can be stochastic. A menu B is *minimal optimal* if menu B is optimal and any menu $B' \subset B$ is not optimal. A menu B is *nested* if the bundles in B can be totally ordered by set inclusion. We say that *nested bundling is optimal* if there exists an optimal and nested menu.

For any bundle b, consider the *single-bundle market* in which only bundle b can be sold. Let $P(b,q)$ be the *demand curve* in this auxiliary market, that is,

$$P(b,q) := F_b^{-1}(1-q),$$

[10] To simplify notation, we omit the inclusion of $\varnothing$ in a menu whenever it is clear from the context.

[11] When an assignment $a(t) \in \Delta(\mathcal{B})$ is deterministic, we also let $a(t)$ denote the assigned bundle.

where F_b is the distribution of $v(b,t)$. Let $\pi(b,q)$ be the *profit function* for bundle b, that is,

$$\pi(b,q) := (P(b,q) - C(b))q.$$

We assume that $\pi(b,q)$ is strictly quasi-concave in $q \in [0,1]$ with an interior maximum.[12] The *sold-alone quantity* $Q(b)$ is defined as the unique quantity at which the marginal profit equals zero, that is,

$$MR(b,Q(b)) = C(b), \tag{1}$$

where $MR(b,q)$ is the usual *marginal revenue curve* for bundle b.

Under assumption (A1), note that for any two nested bundles $b_1 \subset b_2$, the difference $P(b_2,q) - P(b_1,q)$ is the demand curve generated by the incremental values for bundle b_2 given bundle b_1. Thus, $\pi(b_2,q) - \pi(b_1,q)$ is the profit function of a monopolist optimizing the quantity of the incremental bundle $b_2 \backslash b_1$, given the plan of selling every consumer bundle b_1. We will make the following quasi-concavity assumption on this profit function:

A2 (Local Quasi-Concavity): For any two nested bundles $b_1 \subset b_2$, the incremental profit $\pi(b_2,q) - \pi(b_1,q)$ is strictly quasi-concave in $q \in [0, \min\{Q(b_1), Q(b_2)\}]$.

The interval $[0, \min\{Q(b_1), Q(b_2)\}]$ is exactly the region where both individual profit functions $\pi(b_1,q)$ and $\pi(b_2,q)$ are increasing.

A. *Discussion of Assumptions*

Incremental Monotonicity: The incremental monotonicity assumption is only imposed on nested bundles $b_1 \subset b_2$. Restricting to a nested menu $\{b_1, b_2\}$, this assumption reduces to the standard *increasing differences* condition for one-dimensional screening problems. However, unlike one-dimensional screening problems, our model does not impose a total order on the allocations, which allows a much richer set of preferences.

Local Quasi-Concavity: We impose only a local quasi-concavity condition on the incremental profit function for any two nested bundles $b_1 \subset b_2$. It states that, within the interval where both $\pi(b_1,q)$ and $\pi(b_2,q)$ are increasing, the difference $\pi(b_2,q) - \pi(b_1,q)$ has at most one peak. That is, the condition assumes that, within this interval, the sum of an increasing function $\pi(b_2,q)$ and a decreasing function $-\pi(b_1,q)$ is single peaked. Local quasi-concavity is weaker than global quasi-concavity, which always holds if the incremental demand curve $P(b_2,q) - P(b_1,q)$ is log-concave (Quah and Strulovici 2012). An even stronger condition is that the incremental value $v(b_2,t) - v(b_1,t)$ follows a regular distribution in the sense of Myerson (1981), which implies $\pi(b_2,q) - \pi(b_1,q)$ is concave.

[12] For expositional simplicity, whenever we impose strict quasi-concavity of a function g on $[x_1,x_2]$, we assume in addition that $\nabla g(\cdot) = 0$ at $x \in [x_1,x_2]$ implies $g(x) \geq g(x')$ for all $x' \in [x_1,x_2]$ (i.e., we assume that the FOC is satisfied only at the maximum).

One-Dimensional Types: While we assume that the bundle values are increasing in the types, we make no restriction on how consumers' relative preferences for any non-nested bundles change across types. In particular, we allow different consumers to have different ordinal rankings over items (see Example 2). Moreover, across different consumers, the preferences for any two items can switch multiple times in arbitrary ways. The main restriction of one-dimensional types in our model is that such horizontal preferences are fixed for a given one-dimensional type t. Thus, our model is best suited for capturing settings in which some vertical attribute (such as income) is a good predictor of horizontal preferences for different items.

Complements and Substitutes: The assumptions made here are orthogonal to whether the items are complements or substitutes. To illustrate, consider a simple example where the value for a bundle b is given by $v(b,t) = v(b) \cdot t$. Note that all of the above assumptions hold if (i) types t follow a regular distribution in the sense of Myerson (1981) and (ii) $v(b)$ is monotone in the set-inclusion order, regardless of whether the value function $v(b)$ or the monopolist's cost function $C(b)$ exhibit supermodularity or submodularity.

II. Main Results

Our main results characterize (i) when nested bundling is optimal and (ii) which nested menu is optimal. In Section IIA, we introduce a partial order that answers both questions. In Section IIB, we provide some partial converses. In Section IIC, we present a parametric example. In Section IID, we discuss the key intuition behind our results.

A. *Optimality of Nested Bundling*

We define a partial order on the set of bundles $\mathcal{B}$ as follows:

$$b_1 \preceq b_2 : b_1 \subseteq b_2 \text{ and } Q(b_1) \leq Q(b_2) . \tag{2}$$

A bundle b is *dominated* if there exists $b' \neq b$ such that $b \preceq b'$ and *undominated* otherwise. We say that the *nesting condition* holds if the undominated bundles can be totally ordered by set inclusion; that is, for any two bundles b and b',

(Nesting Condition) both b and b' are undominated $\Rightarrow$ either $b \subseteq b'$ or $b' \subseteq b$.

Figure 2 illustrates this condition for a three-item example using a diagram, where an upward arrow from b_1 to b_2 represents $b_1 \preceq b_2$.

Our first main result shows that under the nesting condition, nested bundling, in particular a menu of undominated bundles, is optimal.

THEOREM 1: *Suppose that assumptions* (*A1*) *and* (*A2*) *hold. Then*, *under the nesting condition*, *we have*:

(*i*) *Nested bundling is optimal.*

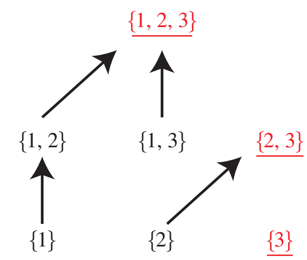

FIGURE 2. ILLUSTRATION OF THE NESTING CONDITION FOR A THREE-ITEM EXAMPLE

Notes: An upward arrow from b_1 to b_2 means $b_1 \preceq b_2$. The undominated bundles are nested: $\{3\} \subseteq \{2,3\} \subseteq \{1,2,3\}$.

(*ii*) *A menu of undominated bundles is optimal.*

(*iii*) *Every optimal mechanism is equivalent to nested bundling.*

The proof is in Appendix A. We sketch the proof in Section III. An immediate consequence of Theorem 1 is the following result.

COROLLARY 1: *Suppose that assumptions* (*A1*) *and* (*A2*) *hold. For any nested menu B*, *if*

(*i*) *for any* $b_1 \subset b_2 \in B$,

$$Q(b_1) > Q(b_2), \tag{3}$$

(*ii*) *for any* $b_1 \notin B$, *there exists* $b_2 \in B$ *such that* $b_1 \subset b_2$, *and*

$$Q(b_1) \leq Q(b_2), \tag{4}$$

then menu B is optimal.

In the special case of zero marginal costs, note that the sold-alone quantity $Q(b)$ is simply the *unit-elastic quantity*, that is, the quantity at which the demand curve $P(b,q)$ has price elasticity equal to -1. In this case, Theorem 1 shows that under suitable conditions, the optimality of a menu can be determined by simply comparing the unit-elastic quantities of different bundles.

Theorem 1 is agnostic to cost structures. In fact, Theorem 1 holds even when the socially efficient allocations require bundles that are not nested. That is, the nesting condition implies the optimality of nested bundling regardless of whether it is efficient. Theorem 1 also implies that the optimal mechanism is deterministic. This need not be true when the nesting condition is not satisfied (see Example 1).

Theorem 1 is also agnostic to whether the items are complements or substitutes. To illustrate, consider two items and zero costs. Suppose that $v(\{1,2\},t) = \kappa \cdot (v(\{1\},t) + v(\{2\},t))$, where κ is a positive constant. Depending on the value of κ, the two items can be complements ($\kappa > 1$), substitutes ($\kappa < 1$), or additive ($\kappa = 1$). However, one can verify that the nesting condition always holds in this case, regardless of the value of κ.

The undominated bundles in Theorem 1 always exist and must include the two extremal bundles under our partial order, the bundle with the highest sold-alone quantity (the *best-selling bundle* $b^\star$) and the bundle with the largest size (the *grand bundle* $\bar{b}$). If these two bundles coincide, then there is a unique undominated bundle, which by Theorem 1 implies that *pure* bundling is optimal (this recovers a result of Ghili 2023; see Corollary 3). If these two bundles do not coincide but there are no other undominated bundles, then the minimal optimal menu is a *two-tier* menu.

However, in general, a menu of undominated bundles need *not* be minimal optimal. Nevertheless, the Proof of Theorem 1 provides an iterative procedure to determine the minimal optimal menu (and its associate prices). To describe the procedure, for any $b_1 \subset b_2$, let $Q(b_2 | b_1)$ denote the *incremental quantity*, that is, the quantity at which the incremental profit function $\pi(b_2,q) - \pi(b_1,q)$ reaches its maximum in the interval $[0, \max\{Q(b_1), Q(b_2)\}]$.[13]

PROPOSITION 1 (Minimal Optimal Menu): *Suppose that assumptions (A1) and (A2) hold. For any optimal and nested menu* $B = \{b_1, \ldots, b_m\}$, *where* $b_1 \subset \cdots \subset b_m$, *let*

$$D := \{b_j \in B : Q(b_{j+1} | b_j) \geq Q(b_j | b_{j-1})\}. \tag{5}$$

Then menu $\tilde{B} := B \backslash D$ *is also an optimal menu. If* $D = \varnothing$ *and* $Q(b_m | b_{m-1}) > 0$, *then menu B is minimal optimal.*

Under the nesting condition, Theorem 1 and Proposition 1 together then describe the following algorithm, which we call the *sieve algorithm*, to determine the minimal optimal menu:

Step 1: Remove all dominated bundles.

Step 2: Remove all bundles satisfying condition (5).

Step 3: Repeat Step 2 until no such bundle exists.

Supplemental Appendix B.2 generalizes both the nesting condition and the sieve algorithm to a more general procedure to find the minimal optimal menu, which allows for comparing two bundles conditional on selling any base bundle included by the two bundles. This notion of conditional dominance offers more ways to exclude bundles from consideration.

[13] Strict quasi-concavity of $\pi(b_2,q) - \pi(b_1,q)$ on $[0, \min\{Q(b_1), Q(b_2)\}]$ implies that it is strictly quasi-concave on $[0, \max\{Q(b_1), Q(b_2)\}]$, and hence, $Q(b_2 | b_1)$ is well defined. Simply defining $Q(b_2 | b_1)$ in the interval $[0, \min\{Q(b_1), Q(b_2)\}]$ would not work for our purposes. If $\pi(b_2,q) - \pi(b_1,q)$ is globally quasi-concave, then $Q(b_2 | b_1)$ can also be defined as the quantity maximizing the incremental profit globally.

TABLE 1—BUNDLE VALUES BY TYPES FOR EXAMPLE 1

	t_1	t_2	t_3
{2,3}	4	8	9
{1,2}	4	5	7
{1}	1	1	1

Notes: The nesting condition fails here since both bundles $\{1,2\}$ and $\{2,3\}$ are undominated. In this example, bundle $\{1\}$ is dominated by bundle $\{1,2\}$ but must be included in the optimal mechanism.

Sufficient Condition for Nesting.—The nesting condition does *not* require a larger bundle to have a higher or lower sold-alone quantity. For instance, for the nesting condition to hold, it is sufficient for the union of two bundles to have a sold-alone quantity in between their individual sold-alone quantities. More generally, we say that the *union quantity condition* holds if the union of any two bundles has a sold-alone quantity above the minimum of their individual sold-alone quantities:

$$\text{(Union Quantity Condition) For all } b_1 \text{ and } b_2,\ Q(b_1 \cup b_2) \ \geq \ \min\{Q(b_1), Q(b_2)\}.$$

The following observation is instructive.

PROPOSITION 2: *The union quantity condition implies the nesting condition.*

The proof is in Appendix A. In light of Proposition 2, under zero marginal costs, Theorem 1 can be interpreted as that nested bundling is optimal if bundling results in a demand curve that is relatively elastic in the sense that the size of its elastic region is larger than at least one of the individual demand curves.

Dominated Bundle Can Be Optimal.—When the nesting condition is not satisfied, however, the optimal mechanism need not be a nested menu, can involve dominated bundles, and may even require randomization. That is, if the undominated bundles cannot be totally ordered by set inclusion, the optimal mechanism may involve selling a dominated bundle to a positive mass of consumers and may even do so with randomization. We provide such an example below. For simplicity, this counterexample is discrete, but it can be made continuous by approximation.

Example 1 (**Without Nesting Condition**)**:** Suppose that there are three items $\{1,2,3\}$ and three types of consumers $\{t_1, t_2, t_3\}$ with mass $1/3$ each. Suppose that we restrict attention to bundles $\{1\}$, $\{1,2\}$, and $\{2,3\}$ (i.e., the costs for other bundles are prohibitively high).[14] The costs for these bundles are 0. The values are given by Table 1. One may verify that the sold-alone quantities are $Q(\{1\}) = 1$ (price 1), $Q(\{1,2\}) = 1$ (price 4), and $Q(\{2,3\}) = 2/3$ (price 8). Thus, bundle $\{1,2\}$ dominates bundle $\{1\}$. Moreover, one may verify that it is indeed the

[14] The example can be extended to allow the grand bundle to be efficient for the highest type or to allow all the other bundles.

case that menu $\{\{1\},\{1,2\}\}$ does not increase the profit beyond the single-bundle menu $\{\{1,2\}\}$.

However, if the other non-nested, undominated bundle $\{2,3\}$ is allowed to be sold, then the dominated bundle $\{1\}$ becomes profitable to include. To see it, note that menu $\{\{1,2\},\{2,3\}\}$ cannot increase the profit beyond the single-bundle menu $\{\{2,3\}\}$. In particular, pricing $\{1,2\}$ at 4 and $\{2,3\}$ at 7 such that t_2 buyer is indifferent will not increase profit because t_3 buyer will choose $\{1,2\}$ instead of $\{2,3\}$. This is because preferences for $\{1,2\}$ and $\{2,3\}$ do not satisfy any single-crossing property. Now, note that the menu $\{\{1\},\{2,3\}\}$ yields a strictly higher profit than menu $\{\{2,3\}\}$: $\frac{1}{3} \times 1 + \frac{2}{3} \times 8 = \frac{17}{3} > \frac{16}{3}$. Hence, the dominated bundle $\{1\}$ is profitable to include. In fact, the optimal mechanism is stochastic: (i) price $5/2$ for the uniform lottery of getting either $\{1\}$ or $\{1,2\}$ and (ii) price $15/2$ for $\{2,3\}$. This yields a profit $\frac{1}{3} \times \frac{5}{2} + \frac{2}{3} \times \frac{15}{2} = \frac{35}{6} > \frac{17}{3}$.

B. *Partial Converse*

We provide a partial converse to Theorem 1. Recall that the *best-selling bundle* $b^\star$ is the bundle with the highest sold-alone quantity; that is,

$$Q(b^\star) \geq Q(b) \text{ for all } b,$$

which, for simplicity, is assumed to be unique.

PROPOSITION 3 (Partial Converse): *Suppose that assumptions (A1) and (A2) hold. For every minimal optimal and nested menu* $B := \{b_1, \ldots, b_m\}$, *where* $b_1 \subset \cdots \subset b_m$, *we have*

(*i*) $b_1 = b^\star$ *and* $b_m = \bar{b}$;

(*ii*) $Q(b_i) > Q(b_j)$ *for all* $b_i \subset b_j \in B$.

The proof is in Appendix A. Proposition 3 states that if nested bundling is optimal, then the minimal optimal menu must (i) include the two extremal bundles under our partial order, the bundle with the highest quantity $b^\star$ and the bundle with the largest size $\bar{b}$, and (ii) exclude any bundle dominated by some bundle in the menu. An immediate consequence of Proposition 3 is the following result.

COROLLARY 2: *Suppose that assumptions (A1) and (A2) hold. Every minimal optimal and nested menu B includes*

(*i*) *the best-selling bundle (if sold alone)* $b^\star$ *as the smallest bundle in the menu*;

(*ii*) *the grand bundle* $\bar{b}$ *as the least-selling bundle (if sold alone) in the menu.*

When the menu B consists only of the grand bundle $\bar{b}$, Corollary 2 says that for pure bundling to be optimal, the grand bundle $\bar{b}$ and the best-selling bundle $b^\star$

must coincide. Conversely, if these two bundles coincide, then the grand bundle is the unique undominated bundle, and hence, the nesting condition trivially holds. Thus, an immediate consequence of Theorem 1 and Proposition 3 is the following characterization.

COROLLARY 3 (Ghili 2023): *Suppose that assumptions (A1) and (A2) hold. Pure bundling is optimal if and only if* $Q(\bar{b}) \geq Q(b)$ *for all bundles b.*

Proposition 3 can also be used to provide sufficient conditions for nested bundling to be suboptimal. For example, a consequence of Proposition 3 is the following result.

COROLLARY 4 (Suboptimality of Nested Bundling): *Suppose that assumptions (A1) and (A2) hold. Suppose that there are two items and that the best-selling bundle* $b^\star = \{2\}$. *If the optimal profit under menu* $\{\{2\},\{1,2\}\}$ *is strictly less than the optimal profit under menu* $\{\{1\},\{1,2\}\}$, *then nested bundling is suboptimal.*

For an illustration of this corollary, see Example 3 in Supplemental Appendix B.3.

C. *Parametric Example*

Example 2: Suppose that there are two items $\{1,2\}$ and zero costs. The valuations for each bundle are given by

$$v(\{1\},t) = t, \quad v(\{2\},t) = t^\beta, \quad v(\{1,2\},t) = t + t^\beta + \sqrt{t}.$$

Types t follow a uniform distribution on $[0,2]$.[15] We vary parameter β from 0 to 2.

Figure 3, panel A plots the numerically computed optimal mechanism in terms of prices, as parameter β varies in 0.1 increments. As Figure 3, panel A shows, the optimal mechanism takes different forms as parameter β varies. Specifically, the optimal menu is given by

- $\{\{2\},\{1,2\}\}$ when $\beta \in [0, 0.74)$;
- $\{\{1,2\}\}$ when $\beta \in [0.74, 1.5]$;
- $\{\{1\},\{1,2\}\}$ when $\beta \in (1.5, 2]$.

The critical parameter values $\beta = 0.74$ and $\beta = 1.5$ are highlighted by the two vertical dashed lines in Figure 3, panel A. These transitions are characterized by Theorem 1. Figure 3, panel B plots the sold-alone quantities $Q(b)$ for the three bundles as parameter β varies. As Figure 3, panel B shows, the nesting condition holds for all values of parameter β: The undominated bundles are always nested. Specifically, the plot can be partitioned into three regions $[0, 0.74)$, $[0.74, 1.5]$, and $(1.5, 2]$. The menu of undominated bundles is $\{\{2\},\{1,2\}\}$ in the first region,

[15] Note that types $t < 1$ and types $t > 1$ have different ordinal rankings for items 1 and 2 whenever $\beta \neq 1$.

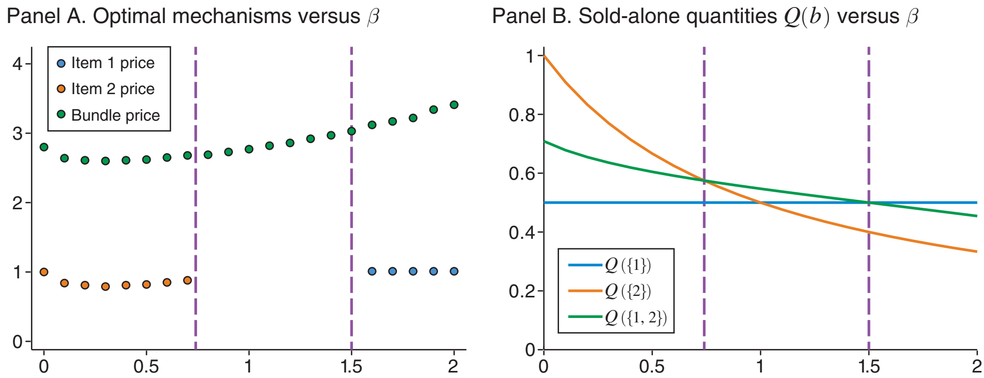

Figure 3. Optimal Mechanisms and Sold-Alone Quantities $Q(b)$ for Example 2

$\{\{1,2\}\}$ in the second region, and $\{\{1\},\{1,2\}\}$ in the third region, coinciding with the optimal menu.

D. *Discussion of Intuition for Nested Bundling*

The basic intuition behind our results is discussed in the introduction using the sold-alone MR curves. In this section, based on the MR curves, we further discuss the intuition behind (i) when nested bundling is suboptimal and (ii) why our nesting condition is sufficient when there are more than two items. As in the introduction, we present the intuition under zero marginal costs, but positive costs can be immediately incorporated by redefining the marginal revenue curves to be the marginal profit curves.

Suboptimality of Nested Bundling.—We first consider a case where nested bundling is suboptimal with two items (see Corollary 4). By the arguments in the introduction, this must be the case where all three bundles are undominated. Without loss of generality, suppose that $Q(\{1,2\}) < Q(\{1\}) < Q(\{2\})$. Suppose further that the revenue under menu $\{\{2\},\{1,2\}\}$ is less than the revenue under menu $\{\{1\},\{1,2\}\}$. Figure 4, panel A illustrates the MR curves under this case. In contrast to the case discussed in the introduction (see Figure 4, panel B), the upper envelope of the MR curves cannot be attained by a nested menu, so we cannot use the argument of "climbing up" the MR curves to find the optimal mechanism.

There are two opposing forces in this case. On the one hand, it is more profitable to attract the "medium-type" consumers than attract the "low-type" consumers since the marginal revenue of selling bundle $\{1\}$ to the "medium-type" consumers is high enough. On the other hand, it is always possible to attract a small fraction of the "low-type" consumers using bundle $\{2\}$, which can bring in a positive marginal revenue. It turns out that the second force always wins if the monopolist can ration and sell bundle $\{2\}$ with a small probability ε. This is because, roughly speaking, the gain from expanding the market this way is on the order of $O(\varepsilon)$, whereas the loss from the consumers who no longer purchase bundle $\{1\}$ is on the order of $O(\varepsilon^2)$. Intuitively, the reason why the loss is on the higher order is that before introducing

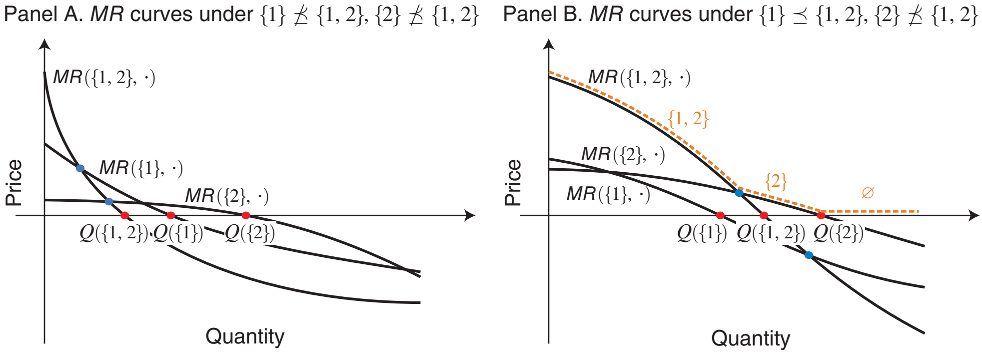

Figure 4. Further Illustration of the Marginal Revenue Curves

bundle $\{2\}$, the monopolist would have already *optimized* the prices for the menu $\{\{1\},\{1,2\}\}$ and hence suffers only a second-order loss for a small perturbation. Thus, nested bundling is suboptimal in this case.

Nested Bundling beyond Two-Item Cases.—We now explain the key insight that helps understand our results beyond the two-item cases. The intuition as discussed in the introduction still holds when there are more than two items, but we may run into issues with both the upgrade and downgrade improvements because these improvements may not be implementable in the price space.

To illustrate, suppose that there are three items and that bundle $\{1\}$ is dominated by bundle $\{1,2\}$. Suppose that we are given an initial allocation rule in the quantity space as depicted in Figure 5, panel A. By the discussion in the introduction, we know that if we can upgrade the consumers who are currently consuming bundle $\{1\}$ to bundle $\{1,2\}$, then we would achieve an improvement (see Figure 4, panel B). However, this upgrade may not be feasible because there may not be prices that can support this change in allocations, given that there are higher types who are currently purchasing bundle $\{2,3\}$, as depicted in Figure 5, panel B (highlighted by the double-headed arrow). This is because our model makes no restriction on how the consumers' relative preferences for any two non-nested bundles change across different types, which leads to a key difference between our bundling problem and the standard one-dimensional screening problem: The set of implementable allocation rules is both much richer and much more complex.[16]

The key insight that resolves this problem is the following: Such a potential conflict can only arise for *non-nested* bundles b and b', but then the nesting condition implies that at least one of them must be *dominated* (recall that the nesting condition requires the undominated bundles to be nested). In our example, this means that either bundle $\{1,2\}$ or bundle $\{2,3\}$ must be dominated. This gives us a way out because we can apply this argument again by going one layer up and further upgrading the consumers from either bundle b or bundle b' to the bundle

[16] Implementability in multidimensional settings is characterized by *cyclic monotonicity* (see Rochet 1987), which is much more complex than standard monotonicity conditions.

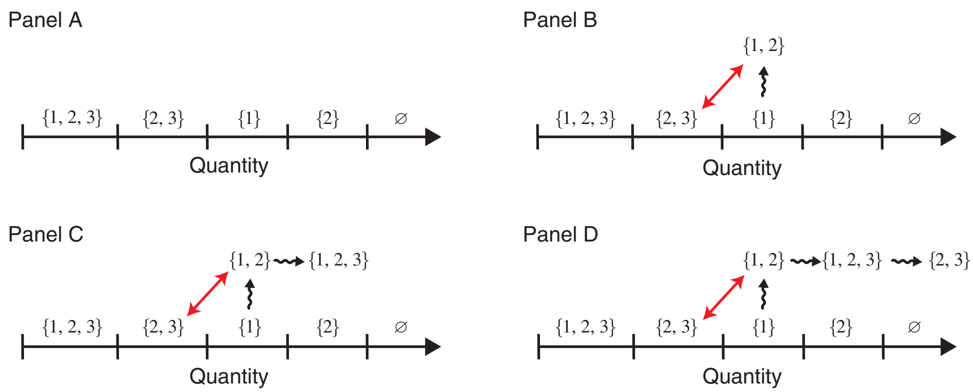

Figure 5. Illustration of the Improvement Argument for a Three-Item Example

that dominates one of them. Repeating this process would always result in a pair of nested bundles.

For our running example, suppose that bundle $\{1,2\}$ is dominated by bundle $\{1,2,3\}$ and bundle $\{2,3\}$ is undominated, as depicted in Figure 5, panel C; so the process in this example terminates in one iteration. Of course, the resulting pair of bundles can be in the "wrong" order in the sense that the higher types are assigned the smaller bundle, which we know cannot be implemented by prices. However, when that happens, since the smaller bundle is undominated, we know that if it is ever profitable to downgrade from the larger bundle to the smaller bundle at some quantity, then it is always profitable to downgrade after that quantity. For our running example, suppose that we can further profitably downgrade bundle $\{1,2,3\}$ to bundle $\{2,3\}$, as depicted in Figure 5, panel D. The allocation rule is now monotone and hence implementable. Moreover, it is more profitable than the initial allocation rule by construction. The proof shows that these arguments can be applied to any initial allocation rule, and hence, the upper envelope of the MR curves, under the nesting condition, must be attained by an implementable allocation rule.

Remark 1: As the above discussion shows, our nesting condition is essential in two ways: (i) It facilitates the comparison of marginal revenues, and (ii) it provides a way out from complex implementability constraints by guiding us toward an even more profitable allocation rule that we know is implementable. The actual proof follows these intuitions. The proof also considers stochastic mechanisms and shows that, under the nesting condition, randomization cannot increase profit.[17] Our key technical result, the monotone construction theorem, gives a weakening of the nesting condition that is both necessary and sufficient for these improvement arguments to yield a monotone allocation rule.

[17] Consistent with Strausz (2006), our proof shows that, under the nesting condition, only the local downward incentive constraints bind, which is crucial to ensure the optimality of deterministic mechanisms.

III. Proof Sketch for the Main Results

In this section, we sketch the joint proof of Theorem 1 and Proposition 1. For simplicity, we assume in this section that the incremental profit function is globally quasi-concave for any two nested bundles $b_1 \subset b_2$. In Appendix A, we complete the proof by weakening global quasi-concavity to local quasi-concavity, that is, assumption (A2).

Following Myerson (1981), let

$$\phi(b,t) := v(b,t) - C(b) - \frac{1 - F(t)}{f(t)} v_t(b,t) \tag{6}$$

be the *virtual surplus* function. Following Bulow and Roberts (1989), we note that this function can be equivalently interpreted as the *sold-alone marginal profit* for bundle b evaluated at the quantity such that the marginal consumer is of type t:

$$\phi(b,t) = MR(b,q)|_{q=1-F(t)} - C(b). \tag{7}$$

A key difference between our problem and one-dimensional mechanism design problems is that we do not have access to a simple characterization of implementable allocation rules. However, as shown in the introduction, we can compute the total profit from any implementable allocation rule using the sold-alone marginal profit functions.

LEMMA 1: *Consider any mechanism (a,p) that gives the lowest type $\underline{t}$ zero payoff. Then, the seller's expected profit under the mechanism (a,p) is given by*

$$\mathbb{E}\left[\sum_{b\in\mathcal{B}} a_b(t)\phi(b,t)\right]. \tag{8}$$

The proof is similar to that of Fact 1 and follows from Myerson (1981) and Bulow and Roberts (1989) (see Appendix A). Now, we solve a relaxed problem by maximizing (8) over all measurable maps $a : \mathcal{T} \to \Delta(\mathcal{B})$ and show that the solution to this relaxed problem is implementable.[18] Note that by linearity, we have

$$\max_{a:\mathcal{T}\to\Delta(\mathcal{B})} \mathbb{E}\left[\sum_{b\in\mathcal{B}} a_b(t)\phi(b,t)\right] = \max_{a:\mathcal{T}\to\mathcal{B}} \mathbb{E}\left[\sum_{b\in\mathcal{B}} a_b(t)\phi(b,t)\right] = \mathbb{E}\left[\max_{b\in\mathcal{B}} \phi(b,t)\right]. \tag{9}$$

We will show that there exists a pointwise solution $b(t)$ that satisfies

- $b(t) \in \arg\max_{b\in\mathcal{B}} \phi(b,t)$ for all t;
- $b(t)$ is monotone in t in the set-inclusion order;
- $b(t)$ is an undominated bundle for all t.

[18] That is, we relax the implementability constraints on the allocation rules and use the fact that any optimal mechanism must give the lowest type zero payoff.

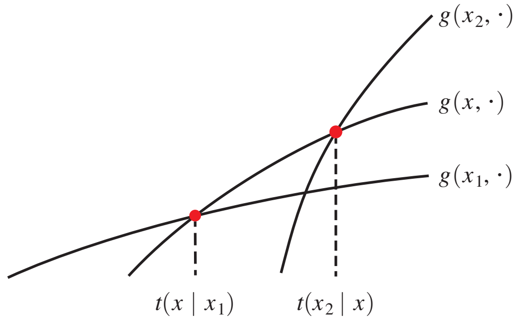

Figure 6. Illustration of the Definition of a Chain-Essential Element

If we can show the above, then we obtain Theorem 1 (parts (i) and (ii)) as follows. By assumption (A1) (incremental monotonicity), we know that $v(b,t)$ has the increasing differences property when restricted to a nested menu. By the standard argument (see Lemma 7 in Supplemental Appendix B.4), the monotone allocation rule $b(t)$ solving the relaxed problem would then be implementable and hence optimal. Therefore, nested bundling, in particular a menu of undominated bundles, is optimal. In fact, the proof will explicitly construct this solution $b(t)$ and show that the set of assigned bundles $\{b(t)\}_{t \in \mathcal{T}}$ coincides with the menu given by our sieve algorithm (see Proposition 1). Note that the existence of such a monotone solution $b(t)$ is precisely a monotone comparative statics question where the choice set is given by $\mathcal{B}$, endowed with the partial order $\subseteq$, and the parameter space is $\mathcal{T}$, with the objective being $\phi(b,t)$.

A. *Monotone Construction Theorem*

To establish the existence of such an allocation rule $b(t)$, our key technical result is an abstract monotone comparative statics theorem. To state it, let $(\mathcal{X}, \leq)$ be a finite partially ordered set and $g : \mathcal{X} \times [0,1] \to \mathbb{R}$ be a function satisfying the *strict single-crossing property*; that is, for any $x_1 < x_2$ and $t < t'$, $g(x_1,t) \leq g(x_2,t) \Rightarrow g(x_1,t') < g(x_2,t')$. For any $x_1 < x_2$, let $t(x_2|x_1)$ be the unique *crossing point* of $g(x_1,t)$ and $g(x_2,t)$:

$$t(x_2|x_1) := \inf\{t \in [0,1] : g(x_2,t) > g(x_1,t)\}.$$

Put $t(x_2|x_1) = 1$ if the above set is empty. An element $x \in \mathcal{X}$ is called *chain essential* for the function g if for all $x_1, x_2 \in \mathcal{X}$ such that $x_1 < x < x_2$, we have

$$t(x|x_1) < t(x_2|x), \tag{10}$$

where, in the above requirement, we also put $t(x_2|x) = 1$ if no $x_2 > x$ exists, and $t(x|x_1) = 0$ if no $x_1 < x$ exists. Figure 6 illustrates this definition.

A chain-essential element x maximizes the objective over all possible *chains* (totally ordered subsets) that contain x, for at least some parameter. The following result asserts that if the chain-essential elements form a chain themselves (the *chain condition*), then each chain-essential element maximizes the objective over the *entire* choice set for at least some parameter and does so monotonically.

THEOREM 2 (Monotone Construction Theorem): *Let $(\mathcal{X}, \leq)$ be a finite partially ordered set. Suppose that $g : \mathcal{X} \times [0,1] \rightarrow \mathbb{R}$ is continuous in t and satisfies the strict single-crossing property in (x,t). Let $\mathcal{Y} \subseteq \mathcal{X}$ be the set of chain-essential elements for g. If $\mathcal{Y}$ is totally ordered, then there exists $x(t)$ such that*

(i) $x(t) \in \arg\max_{x \in \mathcal{X}} g(x,t)$ for all t, and $x(t)$ is the unique maximizer for almost all t;

(ii) $x(t)$ is monotone in t;

(iii) $\{x(t)\}_{t \in [0,1]} = \mathcal{Y}$.

The proof of Theorem 2 is constructive. Note that by definition, $\mathcal{Y}$ must be nonempty. If $\mathcal{Y}$ has only one element, let $x(s)$ be that element for all $s \in [0,1]$. Otherwise, because $\mathcal{Y}$ is totally ordered, we can let the elements in $\mathcal{Y}$ be $x_1 < x_2 < \cdots < x_n$. Since the elements in $\mathcal{Y}$ are chain essential, by (10), we must have

$$0 < t(x_2 | x_1) < \cdots < t(x_n | x_{n-1}) < 1. \tag{11}$$

For any $s \in [0,1]$, let

$$x(s) = x_j \text{ if } s \in \big[t(x_j | x_{j-1}), t(x_{j+1} | x_j)\big), \tag{12}$$

and let $x(s) = x_1$ if $s < t(x_2 | x_1)$ and $x(s) = x_n$ if $s \geq t(x_n | x_{n-1})$. Note that by construction, $x(\cdot)$ is well defined and satisfies properties (ii) and (iii) in Theorem 2. We now show that $x(t)$ maximizes $g(x,t)$ for all t and uniquely so for almost all t.

Step 1: First, we claim that for all $s \in [0,1]$, we have

$$\max_{x \in \mathcal{X}} g(x,s) = \max_{x \in \mathcal{Y}} g(x,s). \tag{13}$$

Because $\mathcal{X}$ is finite, note that by continuity of g in s, it suffices to show the above holds for almost all $s \in [0,1]$. We claim that (13) holds for all $s \notin \{t(x'' | x')\}_{x' < x''}$ (which is a finite set). Suppose for contradiction that there exists some $s \notin \{t(x'' | x')\}_{x' < x''}$ such that (13) does not hold. Then, there must exist some $x \notin \mathcal{Y}$ that maximizes $g(\cdot, s)$ over $\mathcal{X}$.

First, suppose that there is either (i) no $x' < x$ or (ii) no $x'' > x$. Because $x \notin \mathcal{Y}$, in case (i), there exists some $x'' > x$ such that $s > t(x'' | x) = 0$, and hence, $g(x'', s) > g(x, s)$ by the definition of $t(x'' | x)$. Similarly, in case (ii), there exists some $x' < x$ such that $s < t(x|x') = 1$, and hence, $g(x', s) > g(x, s)$ by the definition of $t(x|x')$.

Now, suppose otherwise. Then, because $x \notin \mathcal{Y}$, there exist some $x' < x < x''$ such that

$$t(x|x') \geq t(x''|x).$$

There are again two cases. Case (iii): If $s > t(x|x')$, then we have $s > t(x''|x)$, and hence,

$$g(x'', s) > g(x, s)$$

by the definition of $t(x''|x)$. Case (iv): If $s < t(x|x')$, then we have

$$g(x', s) > g(x, s)$$

by the definition of $t(x|x')$.

In all of the four cases, the element x cannot maximize $g(\cdot, s)$ over $\mathcal{X}$. Contradiction.

Step 2: Second, we claim that for all $s \in [0,1]$, we have

$$g(x(s), s) = \max_{x' \in \mathcal{Y}} g(x', s), \tag{14}$$

where $x(\cdot)$ is constructed in (12). Fix any $s \in [0,1]$. Let $x_j = x(s)$. By construction, we have

$$0 < t(x_2|x_1) < \cdots < t(x_j|x_{j-1}) \leq s < t(x_{j+1}|x_j) < \cdots < t(x_n|x_{n-1}) < 1,$$

which by the definition of $t(\cdot|\cdot)$ implies that

$$\begin{gathered} g(x_j, s) \geq g(x_{j-1}, s) \text{ and } g(x_j, s) \geq g(x_{j+1}, s) \\ g(x_{j-1}, s) \geq g(x_{j-2}, s) \text{ and } g(x_{j+1}, s) \geq g(x_{j+2}, s) \\ \vdots \qquad \text{and} \qquad \vdots \\ g(x_2, s) \geq g(x_1, s) \text{ and } g(x_{n-1}, s) \geq g(x_n, s), \end{gathered}$$

and hence, $g(x_j, s) \geq g(x', s)$ for all $x' \in \mathcal{Y}$. (The same reasoning works for the edge cases of $j = 1$ and $j = n$ as well.) Moreover, note that for all $s \notin \{t(x''|x')\}_{x'<x''}$, the above inequalities are all strict, and hence, $g(x_j, s) > g(x', s)$ for all $x' \neq x_j \in \mathcal{Y}$.

Now, combining Step 1 and Step 2, we immediately have that property (i) of Theorem 2 must hold for our construction $x(t)$.

B. *Switching Lemma*

Our Proof of Theorem 1 will also make use of the following lemma.

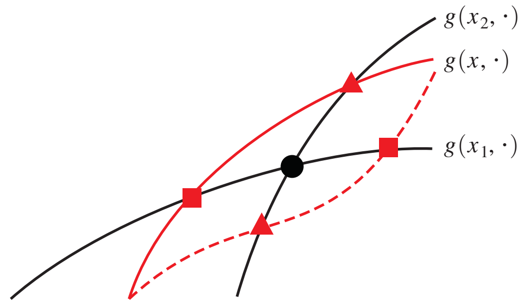

FIGURE 7. ILLUSTRATION OF THE SWITCHING LEMMA

Notes: We hold $g(x_1,t), g(x_2,t)$ fixed and vary $g(x,t)$. Under the single-crossing property of g, note that comparing the horizontal positions of ■ and ▲ is equivalent to comparing the horizontal positions of ■ and ●.

LEMMA 2 (Switching Lemma): *Let $(\mathcal{X}, \leq)$ be a finite partially ordered set. Suppose that $g : \mathcal{X} \times [0,1] \to \mathbb{R}$ is continuous in t and satisfies the strict single-crossing property in (x,t). For any $x_1 < x < x_2$ where $t(x_2|x_1) > 0$, we have*

$$t(x|x_1) < t(x_2|x) \Leftrightarrow t(x|x_1) < t(x_2|x_1). \tag{15}$$

The proof is in Appendix A. Lemma 2 allows us to switch $t(x_2|x)$ in the definition of a chain-essential element to be $t(x_2|x_1)$. Figure 7 illustrates.

C. *Completion of the Proof Sketch*

We apply Theorem 2 to the partially ordered set $(\mathcal{B}, \subseteq)$ and the virtual surplus function $\phi : \mathcal{B} \times \mathcal{T} \to \mathbb{R}$. Note that, by (7), the strict global quasi-concavity of $\pi(b_2,q) - \pi(b_1,q)$ for any two nested bundles $b_1 \subset b_2$ implies that $\phi(b,t)$ has the strict single-crossing property in (b,t). For any $b_1 \subset b_2$, let $t(b_2|b_1)$ be the unique crossing point of $\phi(b_1,t)$ and $\phi(b_2,t)$. In particular, $t(b|\varnothing)$ is the crossing point of $\phi(\varnothing,t) \equiv 0$ and $\phi(b,t)$. By the assumption $Q(b) \in (0,1)$, we have that $t(b|\varnothing)$ is strictly between $\underline{t}$ and $\bar{t}$.

To apply Theorem 2, we need to show that the chain-essential elements in $\mathcal{B}$ form a chain; that is, we need to show that

$$\mathcal{Y} := \left\{b \in \mathcal{B} : t(b|b') < t(b''|b) \text{ for all } b' \subset b \subset b''\right\}$$

is totally ordered by set inclusion. Taking $b' = \varnothing$, by Lemma 2, we have that any bundle $b \in \mathcal{Y}$ must satisfy that for $\varnothing \subset b \subset b''$,

$$t(b|\varnothing) < t(b''|\varnothing),$$

which implies that

$$Q(b) > Q(b''),$$

where $Q(\cdot)$ is the sold-alone quantity. Hence, every $b \in \mathcal{Y}$ is an undominated bundle. But, by the nesting condition, the set of undominated bundles is totally ordered by set inclusion, and hence, $\mathcal{Y}$ is totally ordered by set inclusion. Thus, Theorem 2 applies and yields an allocation rule $b(t)$ that satisfies

- $b(t) \in \arg\max_{b \in \mathcal{B}} \phi(b,t)$ for all t, and $b(t)$ is the unique maximizer for almost all t;
- $b(t)$ is monotone in t in the set-inclusion order;
- $b(t) \in \mathcal{Y}$ is an undominated bundle for all t.

Parts (i) and (ii) of Theorem 1 thus follow immediately by the argument provided at the beginning of this section. Part (iii) of Theorem 1 also follows because $b(t)$ is the unique maximizer for $\phi(b,t)$ for almost all t (see Appendix A for details).

Finally, to see how Proposition 1 (minimal optimal menu) follows, note that for any optimal, nested menu B, we can apply Theorem 2 to the totally ordered set $(B, \subseteq)$ and the virtual surplus function $\phi(b,t)$. Now, the set of chain-essential elements $\mathcal{Y}_B \subseteq B$ is always totally ordered, and hence, $\mathcal{Y}_B$ must be the minimal optimal menu by the construction given in Theorem 2. If condition (5) in Proposition 1 holds for some $\varnothing \neq D \subseteq B$, then any bundle $b \in D$ cannot be in $\mathcal{Y}_B$ and hence can be removed. Otherwise, it can be shown that $\mathcal{Y}_B = B$, and hence, menu B is a minimal optimal menu.

D. *Discussion*

Undominance and Chain Essential.—As the proof shows, the set of chain-essential elements in $(\mathcal{B}, \subseteq)$ for the objective function $\phi(b,t)$ is always a subset of the undominated bundles. Thus, the nesting condition is a sufficient condition for the chain-essential elements to form a chain. However, it is not necessary. We can further generalize Theorem 1 using Theorem 2. This generalization is developed via the notions of conditional dominance and strongly undominated bundles (see Theorem 4 in Supplemental Appendix B.2).

Connection to Monotone Comparative Statics.—Unlike existing monotone comparative statics results, the monotone construction theorem does not require a lattice structure. However, when $\mathcal{X}$ is a lattice, the theorem generalizes the canonical result of Milgrom and Shannon (1994): Under the single-crossing property, our chain condition is implied by their quasi-supermodularity condition. Recall that a function $g : \mathcal{X} \to \mathbb{R}$ is *quasi-supermodular* if for all x and $x' \in \mathcal{X}$, $g(x) \geq (>) g(x \wedge x') \Rightarrow g(x \vee x') \geq (>) g(x')$. The following observation is instructive.

PROPOSITION 4: *Let $(\mathcal{X}, \leq)$ be a finite lattice and $g : \mathcal{X} \times [0,1] \to \mathbb{R}$ be a function satisfying the strict single-crossing property in (x,t). If $g(\cdot,t)$ is quasi-supermodular in x for all t, then the chain-essential elements for g are totally ordered.*

The proof is in Supplemental Appendix B.5. To see that the chain condition is strictly weaker than quasi-supermodularity, let $\mathcal{X} := \{x, x', x \vee x', x \wedge x'\}$ be a four-element lattice. Suppose that

$$g(x \vee x', t) = \kappa \cdot \big(g(x,t) + g(x',t)\big), \quad g(x \wedge x', t) = 0.$$

Note that quasi-supermodularity of $g(\cdot, t)$ requires that $\kappa \geq 1$. However, provided that g has the strict single-crossing property in (x, t), we have

$$t(x \vee x' | x \wedge x') \leq \max\big\{t(x | x \wedge x'), t(x' | x \wedge x')\big\},$$

and hence, the chain-essential elements always form a chain, regardless of the value of κ. Moreover, Supplemental Appendix B.6 shows that the chain condition is not only sufficient for monotone comparative statics but also *necessary* if one requires that the maximizer at each parameter can be found using only comparisons of the objective with ordered pairs (i.e., the pairs that satisfy the single-crossing property).

IV. Applications

In this section, we present three applications. In Section IVA, we further connect optimal bundling strategies to demand structures. In Section IVB, we apply our results to quality discrimination models and study how product line design depends on cost structures. In the last application, in Section IVC, we connect costly screening to optimal bundling. The proofs for this section can be found in Supplemental Appendix B.

A. *Bundling and Elasticity*

We introduce a sufficient condition for the nesting condition in Theorem 1 in terms of price elasticities. Let $\eta(b,q) := \left[\frac{d \log P(b,q)}{d \log q}\right]^{-1}$ be the usual *price elasticity* for bundle b evaluated at quantity q. We say that the *union elasticity condition* holds if for any bundles b_1 and b_2, we have

(Union Elasticity Condition)
$$\eta(b_1, q) < -1 \text{ and } \eta(b_2, q) < -1$$
$$\Rightarrow \eta(b_1 \cup b_2, q) < -1.$$

That is, if the demand curves for two bundles are both elastic at a certain quantity q, then the demand curve for their union is also elastic at quantity q.

PROPOSITION 5: *Under zero costs, the union elasticity condition implies the nesting condition.*

Proposition 5 follows immediately from Proposition 2 by noting that under zero costs, the union elasticity condition implies the union quantity condition that we introduced in Section IIA. Note that when costs are present, we can modify the price elasticity $\eta(b,q)$ to be $\tilde{\eta}(b,q) := \left[\frac{d\log(P(b,q) - C(b))}{d\log q}\right]^{-1}$ to incorporate costs into the elasticity measure.

Applying our main results, we can fully characterize the optimal menu under the union elasticity condition. To state the characterization, we first arrange the bundles according to their sold-alone quantities $Q(b)$ and define $b_i^\star$ as the *i-th best-selling bundle*, with ties broken arbitrarily. Then, we have the following result.

PROPOSITION 6: *Suppose that assumptions* (*A1*) *and* (*A2*) *hold. Under the union elasticity condition and zero costs*, *the following nested menu is optimal:*

$$\left\{b_1^\star,\ b_1^\star \cup b_2^\star,\ b_1^\star \cup b_2^\star \cup b_3^\star,\ \ldots,\ \bar{b}\right\}.$$

The proof is in the Supplemental Appendix. Under the union elasticity condition, Proposition 6 provides a simple recipe for constructing the optimal menu: (i) Arrange all bundles in descending order based on their sold-alone quantities, and (ii) successively merge them, excluding duplicates.[19] Proposition 6 shows that the optimal mechanism iteratively creates nests such that items with a more elastic demand curve become the *basic* items and items with a more inelastic demand curve become the *upgrade* items, with both measured by the size of their elastic regions (i.e., unit-elastic quantities). Note also that this mechanism sorts the bundles, rather than the items, by their sold-alone quantities, accounting for the complementarity or substitutability patterns across different items.

Comparative Statics of Optimal Menu for Demand Rotations.—Price elasticities can be affected by advertising and marketing, which can act as demand rotation in the sense of Johnson and Myatt (2006). Using Proposition 6, we can analyze the comparative statics of optimal bundling given a sequence of demand rotations. Suppose that there are two items and zero costs. Consider a family of demand systems indexed by parameter $s \in \mathbb{R}$, with $\eta(b,q;s)$ denoting the price elasticities and $Q(b;s)$ denoting the sales volumes. We use the following notion of demand rotations: There is a sequence of (*clockwise*, *sales-ordered*) *demand rotations* for an item, say item 1, if for all $s < s'$,

$$Q(\{1\};s') \leq Q(\{1\};s),\ Q(\{2\};s') = Q(\{2\};s),\ Q(\{1,2\};s') \leq Q(\{1,2\};s),$$

and

$$Q(\{1\};s)\ \leq\ Q(\{1,2\};s)\ \Rightarrow\ Q(\{1\};s')\ \leq\ Q(\{1,2\};s').$$

[19] Moreover, starting from this optimal menu in Proposition 6, we can always determine the minimal optimal menu by applying Proposition 1.

That is, as parameter s increases, the demand curve for item 1 and the demand curve for bundle $\{1,2\}$ become more inelastic in the sense of a smaller elastic region.[20] The last condition ensures that the indirect change in the demand curve for bundle $\{1,2\}$ is smaller than the direct change in the demand curve for item 1. To state our result, we define the *tier of item i* in a nested menu $B := \{b_1, \ldots, b_m\}$, where $b_1 \subset \cdots \subset b_m$, as the index of the smallest bundle in B that includes item i, denoted by $r_i(B)$.

PROPOSITION 7: *Suppose that assumptions (A1) and (A2) hold. Suppose that there are two items and zero costs and that the union elasticity condition holds for all s. Let $B^{OPT}(s)$ be the minimal optimal menu. Then, in a sequence of demand rotations for item i, we have that*

(i) *the tier of item i in the optimal menu $r_i(B^{OPT}(s))$ is nondecreasing in s;*

(ii) *the tier of item $j \neq i$ in the optimal menu $r_j(B^{OPT}(s))$ is nonincreasing in s;*

(iii) *the size of the optimal menu $|B^{OPT}(s)|$ is quasi-convex in s.*

The proof is in the Supplemental Appendix. Proposition 7 says that if there is a sequence of demand rotations for item i, that is, an increase in the dispersion of consumers' values for item i, then the item gets promoted to be the *upgrade* item, while the other item gets demoted to be the *basic* item, and the optimal menu first gets coarser and then gets finer. This result complements Johnson and Myatt (2006), who study the effect of value dispersion on a monopolist's quality design. They show that demand rotation always leads to an expansion of the product line. In contrast, our bundling setting involves the monopolist switching the tiers of different items and adopting a menu size that is U-shaped in the dispersion parameter. For example, consider Example 2. As parameter β increases, there is a sequence of demand rotations for item 2. The optimal menu changes in a way that is consistent with Proposition 7; it shifts from $\{\{2\},\{1,2\}\}$ to $\{\{1,2\}\}$, and then to $\{\{1\},\{1,2\}\}$, as parameter β increases.

B. *Quality Discrimination*

A special case of our model is the quality discrimination model a la Mussa and Rosen (1978). Our results provide new insights even in this well-studied setting. Let $\mathcal{X} := \{0, x_1, \ldots, x_n\} \subset \mathbb{R}_+$ be a set of *qualities*, with $0 < x_1 < \cdots < x_n$. In this model, a type-t consumer has value $v(x,t)$ for a good of quality x; the monopolist incurs cost $C(x)$ to supply a good of quality x. This can be viewed as a special case of our model, where we define the values and costs for the bundles as follows: For all $k = 1, \ldots, n$, let

$$v(\{1, \ldots, k\}, t) := v(x_k, t), \quad C(\{1, \ldots, k\}) := C(x_k).$$

[20] A sufficient (but far from necessary) condition is that, as parameter s increases, the demand curves become pointwise more inelastic.

Let $v(b,t) = 0$, $C(b) = 0$ for all bundles b that are not of the form $\{1,\ldots,k\}$. In this case, the nesting condition is always satisfied. Assumption (A1) reduces to the standard increasing differences condition, and assumption (A2) reduces to a local regularity condition that is much weaker than the standard regularity conditions.[21]

Let $Q(x)$ be the sold-alone quantity of the good of quality x; thus, $Q : \mathcal{X} \to [0,1]$. For simplicity of exposition, assume that $Q(x) \in (0,1)$ for all $x \in \mathcal{X}$. Our next result provides a new characterization of optimal quality discrimination.

PROPOSITION 8: *Suppose that assumptions (A1) and (A2) hold. Let $\hat{Q}$ be the upper decreasing envelope of $Q : \mathcal{X} \to [0,1]$; that is,*

$$\hat{Q}(x) := \inf\{g(x) : g \text{ is nonincreasing and } g \geq Q\}.$$

Let $X^\star := \{x : \hat{Q}(x) = Q(x)\}$. Then $X^\star$ is an optimal menu.

The proof is in the Supplemental Appendix. It applies Theorem 1 to this special case. Proposition 8 offers a simple way of pruning the product line using only the sold-alone quantities. To prune the product line to a minimal optimal menu, we can further apply Proposition 1 to this special case using the incremental quantities.

Product Line Design and Cost Structures.—Applying Proposition 8, we can also characterize how cost structures affect the product line design under multiplicative utility functions.

PROPOSITION 9: *Suppose that $v(x,t) = x \cdot t$ and type distribution F is regular.*[22] *Let $C_{avg}(x) := C(x)/x$ be the average cost function. Let $\check{C}_{avg}$ be the lower increasing envelope of C_{avg}; that is,*

$$\check{C}_{avg}(x) := \sup\{g(x) : g \text{ is nondecreasing and } g \leq C_{avg}\}.$$

Let $X^\star := \{x : \check{C}_{avg}(x) = C_{avg}(x)\}$. Then $X^\star$ is an optima menu.

This result generalizes Proposition 1 of Johnson and Myatt (2003), where the average cost curve is assumed to be U-shaped.[23] They conclude that "It is optimal to segment the market with multiple products exactly in the region where average cost and marginal cost are increasing" (Johnson and Myatt 2003, p. 759). However, Proposition 9 shows that this conclusion is incomplete when the cost structure is more complex.[24] The optimal mechanism need *not* segment the market even when average cost and marginal cost are increasing. Figure 8 illustrates. Specifically, the

[21] For standard regularity conditions, see, e.g., pp. 262–68 of Fudenberg and Tirole (1991).

[22] That is, $\phi(t) := t - (1 - F(t))/f(t)$ is strictly increasing.

[23] They define a cost structure to be *U-shaped* if there exists some quality threshold x_k below which the average cost is decreasing and above which the marginal and average costs are increasing. In this case, the menu of undominated qualities $X^\star$ coincides with the region of increasing marginal and average costs $\{x_k,\ldots,x_n\}$. Moreover, by Proposition 1, menu $X^\star$ in this case is the minimal optimal menu.

[24] Average costs are not U-shaped whenever there are kinks in the cost function due to a mix of production technologies, e.g., $C(x) = \min\{k_1 + x^{\alpha_1}, k_2 + x^{\alpha_2}\}$, where $k_1 < k_2$ and $\alpha_1 > \alpha_2$.

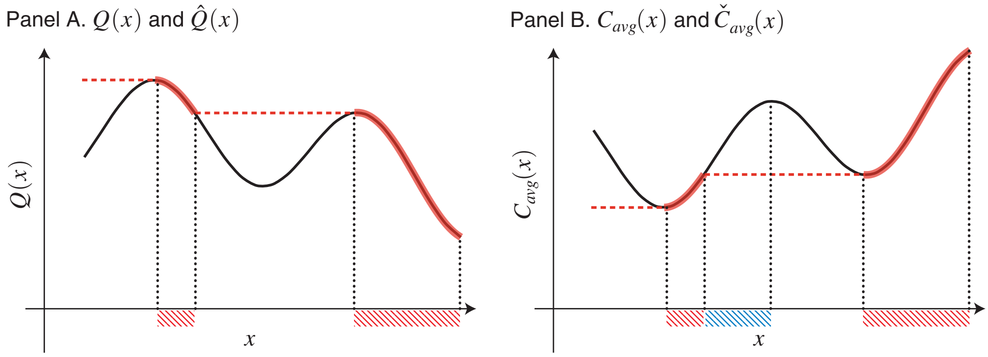

Figure 8. Illustration of $\hat{Q}(\cdot)$ and $\check{C}_{avg}(\cdot)$

marginal and average costs can both increase within the blue-shaded middle region highlighted in Figure 8, panel B, yet the optimal mechanism does not segment the market using these qualities. Instead, as illustrated in Figure 8, panel A, optimal quality choices are characterized by our notion of dominance.

At first glance, Proposition 8 and Proposition 9 may seem related to the *ironing* procedures in Mussa and Rosen (1978) and Myerson (1981). However, this connection is superficial. Even though Proposition 8 and Proposition 9 characterize various bunching regions, they operate in a setting where ironing is *not* needed. In the standard textbook treatment of one-dimensional screening, the regularity assumptions that rule out ironing also rule out the possibility of bunching.[25] Our assumptions are much weaker, allowing for rich forms of bunching. This generality relies on our new constructive monotone comparative statics result (Theorem 2).

C. *Costly Screening*

Consider a monopolist selling quality-differentiated goods, as in Section IVB. In addition to setting prices for these goods, the monopolist can also use nonprice instruments by requiring customers to perform certain costly actions, such as waiting in line or collecting coupons, in order to qualify for certain offers. When is costly screening optimal?

We consider a special case of the model in Yang (2022). A consumer's payoff is given by $u(x,t) - c(y,t) - p$, where $x \in \{0,x_1,\ldots,x_n\} =: \mathcal{X} \subset \mathbb{R}_+$ denotes the *quality* and $y \in \{0,y_1,\ldots,y_m\} =: \mathcal{Y} \subset \mathbb{R}_+$ denotes the *costly action*, with the normalization $u(0,t) = c(0,t) = 0$. The seller's payoff is given by $-C(x) + p$, where $C(\cdot)$ is the production cost. From Yang (2022), we know that if $c(y,t)$ is *nonincreasing* in t, then the optimal deterministic mechanism does not use the costly instruments (i.e., $y(t) = 0$ for all types t).

In this section, we consider the opposite case where $c(y,t)$ is strictly *increasing* in t (for all $y > 0$). We also restrict attention to deterministic mechanisms

[25] For example, see pp. 262–68 of Fudenberg and Tirole (1991).

$(x,y,p) : \mathcal{T} \to \mathcal{X} \times \mathcal{Y} \times \mathbb{R}$. We say that *costly screening is optimal* if every optimal mechanism requires a positive mass of consumers to perform some costly action $y > 0$ and *suboptimal* otherwise.

Let $\pi(x,q)$ be the profit function of selling quality x alone and $Q(x)$ the corresponding sold-alone quantity as in Section IVB. For any costly action y, it is helpful to consider an auxiliary problem of selling the *pass to avoid action* y (e.g., a pass to skip the waiting line). Let $\pi(y,q)$ be the profit function for this problem: $\pi(y,q) := c\left(y, F^{-1}(1-q)\right) \cdot q$. If $u(x,t) - c(y,t)$ is increasing in t, then $\pi(x,q) - \pi(y,q)$ is exactly the profit function of selling quality x when requiring action y. Suppose that $\pi(y,q)$ is strictly quasi-concave in q. Let $Q(y)$ be the sold-alone quantity that maximizes $\pi(y,q)$.

The sold-alone quantity $Q(y)$ can be thought of as an elasticity index since the faster the cost of action y increases with types on the log scale, the lower $Q(y)$ would be.[26] Our next result shows that, under quasi-concavity assumptions, costly screening is optimal if and only if there exists a costly action with sufficiently high elasticity of disutility when measured by this index.

PROPOSITION 10: *Suppose that assumptions (A1) and (A2) hold for $\{u(x,t), C(x), F(t)\}$. Suppose that for all $x,y > 0$, $u(x,t) - c(y,t)$ is strictly increasing in t and $\pi(x,q) - \pi(y,q)$ is strictly quasi-concave in q. Then costly screening is optimal if and only if* $\min_{y>0} Q(y) < \max_{x>0} Q(x)$.

The proof is in the Supplemental Appendix. The intuition behind this characterization can be understood as follows. In the absence of costly screening, as discussed in Section IVB, a menu of different qualities can be viewed as a nested menu. Moreover, we can view requiring a costly action y to purchase quality x as a *damaged* bundle (x,y) that is a subset of the *undamaged* bundle x. Therefore, the question of whether costly screening is optimal reduces to the question of whether selling a specific nested menu is optimal.

Let $x^\star := \max\{\arg\max_x Q(x)\}$ be the best-selling quality (if sold alone). In the absence of costly screening, by Proposition 8, we know that the best-selling quality $x^\star$ would be optimally offered as the base quality level in the menu. If there exists a costly action y such that $Q(y) < Q(x^\star)$, then the damaged bundle $(x^\star, y)$ has an even higher sold-alone quantity. Intuitively, this is because the costly action y compresses the distribution of values for the damaged bundle. Thus, $(x^\star, y)$ can be profitably included in the menu to expand the market by Proposition 3.

On the other hand, if $Q(y) \geq Q(x^\star)$, then we have $Q(y) \geq Q(x^\star) \geq Q(x)$ for all qualities x since quality $x^\star$ is the best-selling quality. This implies that any damaged bundle (x,y) has a lower sold-alone quantity than the undamaged bundle x. Intuitively, this is because the costly action y now makes the distribution of values for the damaged bundle more dispersed. But then bundle (x,y) is dominated by bundle x. Removing all such dominated bundles leaves a nested menu that consists

[26] Formally, $\frac{d}{dt}\log c(y_1,t) \leq \frac{d}{dt}\log c(y_2,t)$ for all $t \Rightarrow Q(y_1) \geq Q(y_2)$.

of only different qualities. Thus, by Theorem 1, the remaining menu is optimal, and hence, costly screening is suboptimal.

V. Conclusion

This paper studies when nested bundling is optimal and determines which nested menu is optimal, when consumers differ in one dimension. We introduce a partial order on the set of bundles defined by (i) set inclusion and (ii) sold-alone quantity. We show that if the set of undominated bundles is nested, then nested bundling, in particular a menu of undominated bundles, is optimal. We provide an iterative procedure to determine the minimal optimal menu that consists of a subset of the undominated bundles. The proof technique involves a new monotone comparative statics result that is constructive and requires no lattice structure. We also provide necessary conditions for a given nested menu to be optimal. We apply our results to connect empirically relevant economic primitives to optimal menu design.

Appendix A. Proofs

A1. *Proof of Lemma 1*

Consider any mechanism (a,p) under which the lowest type $\underline{t}$ has zero payoff. By the envelope theorem (Milgrom and Segal 2002), any type t consumer has payoff given by[27]

$$v(a(t),t) - p(t) = \int_{\underline{t}}^{t} v_t(a(s),s)ds.$$

Therefore, the expected profit of the seller is given by

$$\begin{aligned}&\int_{\underline{t}}^{\bar{t}} \left[p(t) - C(a(t))\right] dF(t)\\ &= \int_{\underline{t}}^{\bar{t}} \left[\left[v(a(t),t) - \int_{\underline{t}}^{t} v_t(a(s),s)ds\right] - C(a(t))\right] dF(t).\end{aligned}$$

By the standard argument of integration by parts, we can write the consumer surplus as

$$\begin{aligned}\int_{\underline{t}}^{\bar{t}} \int_{\underline{t}}^{t} v_t(a(s),s)dsdF(t) &= \int_{\underline{t}}^{\bar{t}} v_t(a(t),t)dt - \int_{\underline{t}}^{\bar{t}} F(t)v_t(a(t),t)dt\\ &= \int_{\underline{t}}^{\bar{t}} \frac{1 - F(t)}{f(t)} v_t(a(t),t)dF(t).\end{aligned}$$

[27] Recall the notation that for any $a \in \Delta(\mathcal{B})$, we write $v(a,t) = \mathbb{E}_{b\sim a}[v(b,t)]$ and $C(a) = \mathbb{E}_{b\sim a}[C(b)]$.

Therefore, the expected profit of the seller is

$$\begin{aligned}\int_{\underline{t}}^{\bar{t}}\big[p(t) - C(a(t))\big]dF(t) &= \int_{\underline{t}}^{\bar{t}}\left[v(a(t),t) - C(a(t)) - \frac{1 - F(t)}{f(t)}v_t(a(t),t)\right]dF(t) \\ &= \int_{\underline{t}}^{\bar{t}}\left[\sum_{b\in\mathcal{B}} a_b(t)\phi(b,t)\right]dF(t) \\ &= \mathbb{E}\left[\sum_{b\in\mathcal{B}} a_b(t)\phi(b,t)\right],\end{aligned}$$

proving the claim. ∎

A2. *Proof of Lemma 2*

For the ($\Leftarrow$) direction, fix any $x_1 < x < x_2$, and suppose that $t(x|x_1) < t(x_2|x_1)$. Then, since $t(x_2|x_1) > 0$, we have $g(x_2, t(x_2|x_1)) \leq g(x_1, t(x_2|x_1)) < g(x, t(x_2|x_1))$, and hence, $t(x_2|x_1) \leq t(x_2|x)$. Thus, $t(x|x_1) < t(x_2|x_1) \leq t(x_2|x)$.

For the ($\Rightarrow$) direction, fix any $x_1 < x < x_2$, and suppose that $t(x|x_1) < t(x_2|x)$. Since $t(x_2|x_1) > 0$, if $t(x|x_1) = 0$, then we are done. Otherwise, we have $t(x|x_1) > 0$, and hence, $g(x_2, t(x|x_1)) < g(x, t(x|x_1)) \leq g(x_1, t(x|x_1))$, and hence, $t(x|x_1) < t(x_2|x_1)$. ∎

A3. *Proof of Theorem 1*

Section III proves parts (i) and (ii) of Theorem 1 under a stronger assumption that the incremental profit functions are globally quasi-concave. We complete the proof by weakening global quasi-concavity to local quasi-concavity, that is, assumption (A2). We will also show part (iii) of Theorem 1 in the end. The proof strategy is the same as in Section III, except that we generalize both Theorem 2 and Lemma 2 to hold for functions g that only have a local single-crossing property.

Monotone Construction Theorem with Local Single-Crossing Property: Let $\mathcal{X}$ be a finite partially ordered set. Suppose that $\mathcal{X}$ has a *minimum* x_0; that is, $x_0 < x$ for all $x \neq x_0$. A function $g : \mathcal{X} \times [0,1] \to \mathbb{R}$ has *strict local single-crossing property* if for all $x_0 \leq x < x'$ and all $t < t'$,

$$g(x',t) \geq \max\{g(x,t), g(x_0,t)\} \Rightarrow g(x',t') > \max\{g(x,t'), g(x_0,t')\}.$$

Let

$$t(x'|x) := \inf\{t \in [0,1] : g(x',t) > \max\{g(x,t), g(x_0,t)\}\},$$

where we put $t(x'|x) := 1$ if the above set is empty. Write $t(x)$ as a shorthand for $t(x|x_0)$.

The definition of strict local single-crossing property imposes that (i) for any $x > x_0$, $g(x,\cdot)$ single-crosses $g(x_0,\cdot)$ from below and (ii) for any $x' > x > x_0$, $g(x',\cdot)$ single-crosses $\max\{g(x,\cdot), g(x_0,\cdot)\}$ from below. The following lemma provides two equivalent characterizations of the strict local single-crossing property that will be helpful later.

LEMMA 3: *Let $\mathcal{X}$ be a finite partially ordered set with a minimum element x_0. Suppose that $g(x,\cdot)$ is continuous for all x and that $g(x,s) \geq g(x_0,s) \Rightarrow g(x,s') > g(x_0,s')$ for all $x > x_0$ and all $s' > s$. For any $x' > x > x_0$, the following three statements are equivalent:*

(*i*) $g(x',s) \geq \max\{g(x,s), g(x_0,s)\} \Rightarrow g(x',s') > \max\{g(x,s'), g(x_0,s')\}$ *for all* $s' > s$;

(*ii*) $g(x',s) \geq g(x,s) \Rightarrow g(x',s') > g(x,s')$ *for all* $s' > s \geq \min\{t(x), t(x')\}$;

(*iii*) $g(x',s) \geq g(x,s) \Rightarrow g(x',s') > g(x,s')$ *for all* $s' > s \geq \max\{t(x), t(x')\}$.

Moreover, if any of the above three conditions holds, then we also have

$$t(x'|x) = \inf\Big\{s \in \big[\min\{t(x), t(x')\}, 1\big] : g(x',s) > g(x,s)\Big\}.$$

PROOF:

We show that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i).

(i) $\Rightarrow$ (ii): Suppose that condition (i) holds and that $g(x',s) \geq g(x,s)$ for some s such that $1 > s \geq \min\{t(x), t(x')\}$. Fix any $s' > s$. If $s \geq t(x)$, then $g(x',s) \geq g(x,s) \geq g(x_0,s)$, which by condition (i), implies that $g(x',s') > g(x,s')$. Now suppose $s < t(x)$. Then we must have $t(x') \leq s < t(x)$. Hence, $g(x',s) \geq g(x_0,s) \geq g(x,s)$, which by condition (i), also implies that $g(x',s') > g(x,s')$.

(ii) $\Rightarrow$ (iii): This is immediate from the definition.

(iii) $\Rightarrow$ (i): Suppose that condition (iii) holds. Fix any $s' > s$, and suppose that $g(x',s) \geq \max\{g(x,s), g(x_0,s)\}$. Then, we have $s \geq t(x')$. Hence, we have $g(x',s') > g(x_0,s')$. Thus, it suffices to show that $g(x',s') > g(x,s')$. If $s \geq t(x)$, then by condition (iii), we have $g(x',s') > g(x,s')$. Now suppose $s < t(x)$. Then we must have $t(x') \leq s < t(x)$. There are two cases: If $s' \leq t(x)$, then we have $g(x,s') \leq g(x_0,s') < g(x',s')$; otherwise, if $s' > t(x)$, then note that since $t(x') < t(x)$, we have

$$g(x,t(x)) \leq g(x_0,t(x)) < g(x',t(x)),$$

which by condition (iii) implies that $g(x',s') > g(x,s')$.

Finally, we show that

$$t(x'|x) = \inf\left\{s \in \left[\min\{t(x),t(x')\},1\right] : g(x',s) > g(x,s)\right\}.$$

First, note that by definition, we must have $t(x'|x) \geq t(x') \geq \min\{t(x),t(x')\}$. Second, note that for any $s > t(x'|x)$, we have $g(x',s) > \max\{g(x,s),g(x_0,s)\} \geq g(x,s)$. Now, fix any s such that $\min\{t(x),t(x')\} \leq s < t(x'|x)$. We claim that $g(x',s) < g(x,s)$. To see it, note that since $s < t(x'|x)$, we have

$$g(x',s) < \max\{g(x,s),g(x_0,s)\}.$$

Since $s \geq \min\{t(x),t(x')\}$, we also have either that $g(x,s) \geq g(x_0,s)$ or that $g(x',s) \geq g(x_0,s)$. In either case, we must then have $g(x',s) < g(x,s)$. Then, it follows that

$$t(x'|x) = \inf\left\{s \in \left[\min\{t(x),t(x')\},1\right] : g(x',s) > g(x,s)\right\},$$

proving the result. ∎

As in Section III, an element x is *chain essential* for g if for all $x_1 < x < x_2$,

$$t(x|x_1) < t(x_2|x),$$

where, in the above requirement, we also put $t(x_2|x) = 1$ if no $x_2 > x$ exists, and $t(x|x_1) = 0$ if no $x_1 < x$ exists.

The next two lemmas show that this definition of chain-essential elements extends the key properties of our earlier definition in Section III (when g has the global single-crossing property). In particular, the chain-essential elements are exactly the ones that cannot be removed from any *chain* without decreasing the objective value at some parameter.

LEMMA 4: *Let $\mathcal{X}$ be a finite partially ordered set with a minimum element x_0. Suppose that g is continuous in t and has the strict local single-crossing property. For any $x_1 < x < x_2$ such that*

$$t(x|x_1) \geq t(x_2|x),$$

we have

$$g(x,s) \leq \max\{g(x_1,s),g(x_2,s),g(x_0,s)\},$$

for all s and strictly so for all $s \notin \{t(x),t(x_1),t(x|x_1),t(x_2|x)\}$.

PROOF:

Fix any $s \notin \{t(x), t(x_1), t(x|x_1), t(x_2|x)\}$. If $s > t(x|x_1)$, then we have $s > t(x_2|x)$, and hence,

$$g(x,s) < g(x_2,s) \leq \max\{g(x_1,s), g(x_2,s), g(x_0,s)\}.$$

If $s < t(x|x_1)$, then we have that

$$g(x,s) < \max\{g(x_1,s), g(x_0,s)\} \leq \max\{g(x_1,s), g(x_2,s), g(x_0,s)\}.$$

Thus, the required strict inequality holds for s. The weak inequality holds for all $s \in [0,1]$ by the continuity of g. ∎

LEMMA 5: *Let $\mathcal{X}$ be a finite partially ordered set with a minimum element x_0. Suppose that g is continuous in t and has the strict local single-crossing property. For any $x_1 < x < x_2$ and any s such that*

$$t(x|x_1) < s < t(x_2|x),$$

we have

$$g(x,s) > \max\{g(x_1,s), g(x_2,s), g(x_0,s)\}.$$

PROOF:

Fix any s such that $t(x|x_1) < s < t(x_2|x)$. By the definition of $t(x|x_1)$, we have

$$g(x,s) > \max\{g(x_1,s), g(x_0,s)\}.$$

Moreover, by the definition of $t(x_2|x)$, we then have

$$g(x_2,s) < \max\{g(x,s), g(x_0,s)\} = g(x,s).$$

Therefore, we have

$$g(x,s) > \max\{g(x_1,s), g(x_2,s), g(x_0,s)\},$$

proving the claim. ∎

THEOREM 3 (Monotone Construction with Local Single Crossing): *Let $(\mathcal{X}, \leq)$ be a finite partially ordered set with a minimum element x_0. Suppose that $g : \mathcal{X} \times [0,1] \to \mathbb{R}$ is continuous in t and satisfies the strict local single-crossing property in (x,t). Let $\mathcal{Y} \subseteq \mathcal{X}$ be the set of chain-essential elements for g. If $\mathcal{Y}$ is totally ordered, then there exists $x(t)$ such that*

(*i*) $x(t) \in \arg\max_{x \in \mathcal{X}} g(x,t)$ *for all t, and $x(t)$ is the unique maximizer for almost all t*;

(*ii*) $x(t)$ *is monotone in* t;

(*iii*) $\{x(t)\}_{t \in [0,1]} = \mathcal{Y}$.

PROOF:

We use the same construction and the same proof strategy as in Section III. By definition, $\mathcal{Y}$ must be nonempty. If $\mathcal{Y}$ has only one element, let $x(s)$ be that element for all $s \in [0,1]$. Otherwise, because $\mathcal{Y}$ is totally ordered, we can let the elements in $\mathcal{Y}$ be $x_1 < x_2 < \cdots < x_n$. Since the elements in $\mathcal{Y}$ are chain essential, by definition, we must have

$$0 < t(x_2 | x_1) < \cdots < t(x_n | x_{n-1}) < 1.$$

For any $s \in [0,1]$, let

$$x(s) = x_j \text{ if } s \in [t(x_j | x_{j-1}), t(x_{j+1} | x_j)),$$

and let $x(s) = x_1$ if $s < t(x_2 | x_1)$ and $x(s) = x_n$ if $s \geq t(x_n | x_{n-1})$. Note that by construction, $x(\cdot)$ is well defined and satisfies properties (ii) and (iii) in Theorem 3. We now show that $x(t)$ maximizes $g(x,t)$ for all t and uniquely so for almost all t.

Step 1: First, we claim that for all $s \in [0,1]$, we have

$$\max_{x \in \mathcal{X}} g(x,s) = \max_{x \in \mathcal{Y}} g(x,s).$$

Because $\mathcal{X}$ is finite, note that by continuity of g in s, it suffices to show the above holds for almost all $s \in [0,1]$. We claim that the above holds for all $s \notin \{t(x'' | x')\}_{x' < x''}$. Suppose for contradiction that there exists some $s \notin \{t(x'' | x')\}_{x' < x''}$ such that it does not hold. Then, there must exist some $x \notin \mathcal{Y}$ that maximizes $g(\cdot, s)$ over $\mathcal{X}$.

First, suppose that there is either (i) no $x' < x$ or (ii) no $x'' > x$. Because $x \notin \mathcal{Y}$, in case (i), there exists some $x'' > x$ such that $s > t(x'' | x) = 0$, and hence, $g(x'', s) > g(x, s)$ by the definition of $t(x'' | x)$. Similarly, in case (ii), there exists some $x' < x$ such that $s < t(x | x') = 1$, and hence, $\max\{g(x', s), g(x_0, s)\} > g(x, s)$ by the definition of $t(x | x')$.

Now, suppose otherwise. Then, in this case, because $x \notin \mathcal{Y}$, there exist some $x' < x < x''$ such that

$$t(x | x') \geq t(x'' | x).$$

But then by Lemma 4, we have

$$g(x,s) < \max\{g(x', s), g(x'', s), g(x_0, s)\}.$$

In all of these cases, the element x cannot maximize $g(\cdot, s)$ over $\mathcal{X}$. Contradiction.

Step 2: Second, we claim that for all $s \in [0,1]$, we have

$$g(x(s),s) = \max_{x' \in \mathcal{Y}} g(x',s).$$

This holds trivially if $|\mathcal{Y}| = 1$. Hence, suppose $|\mathcal{Y}| > 1$. Fix any $s \in [0,1]$. Let $x_j = x(s)$. By construction, we have

$$0 < t(x_2|x_1) < \cdots < t(x_j|x_{j-1}) \leq s < t(x_{j+1}|x_j) < \cdots < t(x_n|x_{n-1}) < 1,$$

which by the Proof of Lemma 5 implies that

$$g(x_j,s) \geq \max\{g(x_{j-1},s),g(x_0,s)\} \text{ and } \max\{g(x_j,s),g(x_0,s)\} \geq g(x_{j+1},s)$$

$$g(x_{j-1},s) \geq \max\{g(x_{j-2},s),g(x_0,s)\} \text{ and } \max\{g(x_{j+1},s),g(x_0,s)\} \geq g(x_{j+2},s)$$

$$\vdots \qquad \text{and} \qquad \vdots$$

$$g(x_2,s) \geq \max\{g(x_1,s),g(x_0,s)\} \text{ and } \max\{g(x_{n-1},s),g(x_0,s)\} \geq g(x_n,s).$$

There are two cases.

Case (i)**:** $j \geq 2$. Note that the left column above implies that

$$g(x_j,s) \geq g(x_i,s)$$

for all $i < j$ and that $g(x_j,s) \geq g(x_0,s)$. But by the right column above, we also have

$$\max\{g(x_j,s),g(x_0,s)\} \geq g(x_k,s)$$

for all $k > j$. Thus, $g(x_j,s) \geq g(x_k,s)$ for all $k > j$. Therefore, $g(x_j,s) \geq g(x',s)$ for all $x' \in \mathcal{Y}$. Moreover, for all $s \notin \{t(x''|x')\}_{x'<x''}$, the same argument implies that $g(x_j,s) > g(x',s)$ for all $x' \neq x_j \in \mathcal{Y}$.

Case (ii)**:** $j = 1$. In this case, we have $0 \leq s < t(x_2|x_1)$. By the same reasoning as in the previous case, we have

$$\max\{g(x_1,s),g(x_0,s)\} > g(x_k,s)$$

for all $k > 1$. Now, note that we must have $g(x_1,s) \geq g(x_0,s)$. Because otherwise, by the above, we immediately have

$$g(x_0,s) > g(x_1,s) \text{ and } g(x_0,s) > g(x_k,s)$$

for all $k > 1$, and hence,

$$g(x_0,s) > \max_{x \in \mathcal{Y}} g(x,s),$$

which is impossible by Step 1. Thus, we have $g(x_1,s) \geq g(x_0,s)$, and hence,

$$g(x_1,s) = \max\{g(x_1,s), g(x_0,s)\} > g(x_k,s)$$

for all $k > 1$. Therefore, $g(x_1,s) > g(x',s)$ for all $x' \neq x_1 \in \mathcal{Y}$.

Combining these two cases, we have that for all $s \in [0,1]$, $g(x(s),s) = \max_{x' \in \mathcal{Y}} g(x',s)$, and moreover, for all $s \notin \{t(x''|x')\}_{x'<x''}$, $x(s)$ is the unique maximizer in $\mathcal{Y}$.

Now, combining Step 1 and Step 2, we immediately have that property (i) of Theorem 3 must hold for our construction $x(t)$, proving the result. ∎

Switching Lemma with Local Single-Crossing Property:

LEMMA 6 (Switching Lemma with Local Single Crossing): *Let* $(\mathcal{X}, \leq)$ *be a finite partially ordered set with a minimum element* x_0*. Suppose that* $g : \mathcal{X} \times [0,1] \to \mathbb{R}$ *is continuous in t and satisfies the strict local single-crossing property in* (x,t)*. For any* $x_1 < x < x_2$ *where* $t(x_2|x_1) > 0$*, we have*

$$t(x|x_1) < t(x_2|x) \Leftrightarrow t(x|x_1) < t(x_2|x_1).$$

PROOF:

For the $(\Leftarrow)$ direction, fix any $x_1 < x < x_2$, and suppose that $t(x|x_1) < t(x_2|x_1)$. Suppose for contradiction that $t(x|x_1) \geq t(x_2|x)$. Then by Lemma 4, we have

$$g(x,s) \leq \max\{g(x_1,s), g(x_2,s), g(x_0,s)\},$$

for all $s \in [0,1]$. Note that there exists s such that

$$t(x|x_1) < s < t(x_2|x_1).$$

For such s, we have

$$g(x,s) > \max\{g(x_1,s), g(x_0,s)\}, \quad \max\{g(x_1,s), g(x_0,s)\} > g(x_2,s).$$

Therefore, we have

$$g(x,s) > \max\{g(x_1,s), g(x_2,s), g(x_0,s)\}.$$

Contradiction.

For the $(\Rightarrow)$ direction, fix any $x_1 < x < x_2$, and suppose that $t(x|x_1) < t(x_2|x)$. Suppose for contradiction that $t(x|x_1) \geq t(x_2|x_1)$. Then

$$0 < t(x_2|x_1) \leq t(x|x_1) < t(x_2|x) \leq 1.$$

Let $s = t(x|x_1)$. Since

$$t(x_2|x_1) \leq s < t(x_2|x),$$

we have

$$g(x_2,s) \ \geq \ \max\{g(x_1,s),g(x_0,s)\} \ = \ \max\{g(x,s),g(x_0,s)\},$$

where the last equality is due to $0 < s = t(x|x_1) < 1$. But then, since $s < t(x_2|x)$, we have

$$\max\{g(x,s),g(x_0,s)\} \ > \ g(x_2,s).$$

Contradiction. ∎

Completion of the Proof:

Parts (*i*) *and* (*ii*).—Parts (i) and (ii) of Theorem 1 follow by the same proof as in Section III, with Theorem 3 replacing Theorem 2 and Lemma 6 replacing Lemma 2.

Specifically, we apply Theorem 3 to the partially ordered set $(\mathcal{B},\subseteq)$ and the virtual surplus function $\phi:\mathcal{B}\times\mathcal{T}\to\mathbb{R}$. By (7) in Section III, the assumption that $\pi(b,q)$ is strictly quasi-concave implies that $\phi(b,\cdot)$ single-crosses $\phi(\varnothing,\cdot) \equiv 0$ from below. Let $t(b)$ denote the unique crossing point. Moreover, for any $\varnothing \neq b_1 \subset b_2$, under the strict local quasi-concavity of $\pi(b_2,q) - \pi(b_1,q)$, that is, assumption (A2), we have

$$\phi(b_2,s) \ \geq \ \phi(b_1,s) \ \Rightarrow \ \phi(b_2,s') \ > \ \phi(b_1,s')$$

for all $s' > s \geq \max\{t(b_1),t(b_2)\}$. But then by Lemma 3, the function $\phi(b,t)$ must satisfy the strict local single-crossing property. Let $t(b_2|b_1)$ denote the unique crossing point of $\phi(b_2,\cdot)$ and $\max\{\phi(b_1,\cdot),\phi(\varnothing,\cdot)\}$.

Now, to apply Theorem 3, it remains to verify that the chain-essential elements in $\mathcal{B}$ form a chain; that is, we want to show that

$$\mathcal{Y} := \{b \in \mathcal{B}: t(b|b') < t(b''|b) \text{ for all } b' \subset b \subset b''\}$$

is totally ordered by set inclusion. Taking $b' = \varnothing$, by Lemma 6, we have that any bundle $b \in \mathcal{Y}$ must satisfy that for $\varnothing \subset b \subset b''$, $t(b|\varnothing) < t(b''|\varnothing)$, which implies that $Q(b) > Q(b'')$, where $Q(\cdot)$ is the sold-alone quantity. Hence, every $b \in \mathcal{Y}$ is an undominated bundle. But, by the nesting condition, the set of undominated bundles is totally ordered by set inclusion, and hence, $\mathcal{Y}$ is totally ordered by set inclusion. The rest of the proof is identical to that in Section III.

Part (*iii*).—We now prove part (iii) of Theorem 1. First, note that if $b \in \mathcal{B}$ satisfies

$$\phi(b,t) \ > \ \max_{b' \in \mathcal{B} \setminus b} \phi(b',t),$$

then by the linearity of probabilities and the finiteness of $\mathcal{B}$, we have

$$\phi(b,t) \ > \ \max_{a \in \Delta(\mathcal{B}) \setminus \delta_b} \mathbb{E}_{b' \sim a}[\phi(b',t)],$$

where δ_b denotes the Dirac measure centered on b.

Let $b(\cdot)$ denote the constructed allocation rule given by Theorem 3. Now, fix any implementable, potentially stochastic allocation rule $a(\cdot)$. Since

$$b(t) \in \operatorname*{arg\,max}_{b' \in \mathcal{B}} \phi(b', t),$$

we have that for all t,

$$\mathbb{E}_{b' \sim a(t)}\left[\phi(b', t)\right] \leq \phi(b(t), t).$$

Moreover, let $\mathcal{T}' := \left\{t \in \mathcal{T} : a(t) \neq \delta_{b(t)}\right\}$. Recall that $b(t)$ is the unique maximizer of the problem $\max_{b' \in \mathcal{B}} \phi(b', t)$ for almost all t (see Theorem 3). Therefore, by the above argument, for almost all $t \in \mathcal{T}'$, we have

$$\mathbb{E}_{b' \sim a(t)}\left[\phi(b', t)\right] < \phi(b(t), t).$$

If $a(\cdot)$ attains the optimal profit for the seller, then by Lemma 1, we must have

$$\mathbb{E}\left[\mathbb{E}_{b' \sim a(t)}\left[\phi(b', t)\right]\right] = \mathbb{E}\left[\phi(b(t), t)\right],$$

which implies that $\mathcal{T}'$ has measure 0. Therefore, $a(\cdot)$ is equivalent to $\delta_{b(\cdot)}$ almost everywhere. By the envelope theorem, we also have that the payment rules implementing $a(\cdot)$ and $\delta_{b(\cdot)}$ must coincide almost everywhere. Thus, any optimal mechanism is equivalent to the nested bundling mechanism that we constructed. ∎

A4. *Proof of Proposition 1*

The proof strategy is the same as in Section III. Let $B = \{b_1, \ldots, b_m\}$, where $b_1 \subset \cdots \subset b_m$, be any optimal and nested menu. We apply Theorem 3 to the totally ordered set $(B, \subseteq)$, with the objective function being $\phi(b, t)$ (and the minimum element being $\varnothing$). The set of chain-essential elements $\mathcal{Y}_B \subseteq B$ is always totally ordered, and hence, $\mathcal{Y}_B$ must be a minimal optimal menu by Theorem 3 and the Proof of Theorem 1.

First, suppose that condition (5) in Proposition 1 holds for some nonempty $D \subseteq B$. We claim that any bundle $b \in D$ cannot be chain essential and hence can be removed. To see this, recall that for any $\varnothing \subset b_1 \subset b_2$,

$$Q(b_2 | b_1) := \operatorname*{arg\,max}_{q \in \left[0,\, \max\{Q(b_1), Q(b_2)\}\right]} \pi(b_2, q) - \pi(b_1, q).$$

By (7) in Section III, we can write $Q(b_2 | b_1) = 1 - F\left(\tilde{t}(b_2 | b_1)\right)$, where

$$\tilde{t}(b_2 | b_1) := \inf\left\{s \in \left[\min\{t(b_1), t(b_2)\}, \bar{t}\right] : \phi(b_2, s) > \phi(b_1, s)\right\},$$

where $t(b)$ denotes the unique crossing point of $\phi(b, \cdot)$ and $\phi(\varnothing, \cdot) \equiv 0$. But by Lemma 3, we also know that $\tilde{t}(b_2 | b_1) = t(b_2 | b_1)$, where $t(b_2 | b_1)$ is defined as the unique crossing point of $\phi(b_2,\cdot)$ and $\max\{\phi(b_1,\cdot),\phi(\varnothing,\cdot)\}$. Therefore, for any bundle $b_j \in \mathcal{B}$ such that $Q(b_{j+1}|b_j) \geq Q(b_j|b_{j-1})$, we have $t(b_j|b_{j-1}) \geq t(b_{j+1}|b_j)$, and hence, b_j cannot be chain essential for $\phi(b,t)$ by definition.

Now, suppose that $D = \varnothing$ and $Q(b_m|b_{m-1}) > 0$. Then we must have

$$1 > Q(b_1) > Q(b_2|b_1) > \cdots > Q(b_m|b_{m-1}) > 0,$$

which, by the above argument, implies that

$$\underline{t} < t(b_1) < t(b_2|b_1) < \cdots < t(b_m|b_{m-1}) < \bar{t}.$$

To show that menu B is minimal optimal, it suffices to show that $\mathcal{Y}_B = B$; that is, we want to show that all $b \in B$ are chain essential. This follows by the Proof of Theorem 3. Suppose for contradiction that there exists some nonempty $b \in B$ that is not a chain-essential element. Then, by the definition of chain-essential elements and Lemma 4, there exist $b' \neq b, b'' \neq b \in B$ such that for all t,

$$\phi(b,t) \leq \max\{\phi(b',t),\phi(b'',t),\phi(\varnothing,t)\}.$$

At the same time, by Step 2 in the Proof of Theorem 3, for any $b \in B$, there exists some s such that $\phi(b,s) > \max_{\hat{b}\in B\setminus\{b\}} \phi(\hat{b},s)$. Contradiction. ∎

A5. *Proof of Proposition 2*

Suppose for contradiction that there exist undominated bundles b_1 and b_2 that are not nested. Then $b_1 \subset b_1 \cup b_2$ and $b_2 \subset b_1 \cup b_2$. Because b_1 and b_2 are undominated, we have

$$Q(b_1) > Q(b_1 \cup b_2), \quad Q(b_2) > Q(b_1 \cup b_2),$$

and hence, $\min\{Q(b_1),Q(b_2)\} > Q(b_1 \cup b_2)$, contradicting to the union quantity condition. ∎

A6. *Proof of Proposition 3*

Part (ii) of Proposition 3 follows from the monotone construction theorem. Specifically, we apply Theorem 3 to the virtual surplus function $\phi(b,t)$ and the totally ordered set $(B,\subseteq)$, where B is the minimal optimal and nested menu. Since B is minimal optimal, we must have that B is the set of chain-essential elements, but that implies that any bundle in B cannot be dominated by another bundle in B, and hence, $Q(b_i) > Q(b_j)$ for all $b_i \subset b_j \in B$.

Now, suppose for contradiction that the grand bundle $\bar{b}$ is not in the menu B. We apply the monotone construction theorem, Theorem 3, to $\phi(b,t)$ and $(B \cup \bar{b},\subseteq)$. Since B is a nested menu, $B \cup \bar{b}$ must also be a nested menu. We claim that $\bar{b}$ is always a chain-essential element. To see it, note that for all $b \subset \bar{b}$, we have

$$\phi(\bar{b},\bar{t}) = v(\bar{b},\bar{t}) - C(\bar{b}) > v(b,\bar{t}) - C(b) = \phi(b,\bar{t}),$$

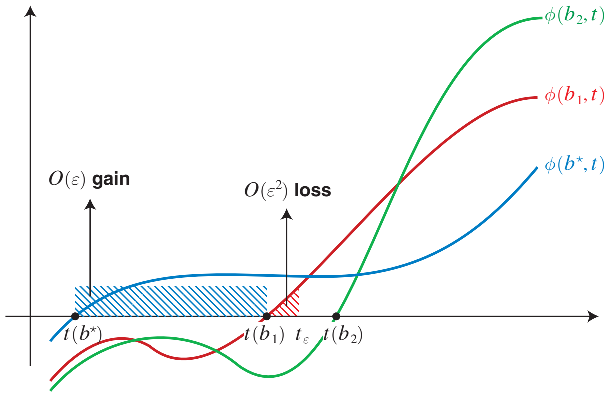

Figure 9. Illustration of the Perturbation Argument

where the strict inequality is due to our assumption that the grand bundle is the unique surplus-maximizing bundle for the highest type $\bar{t}$. This implies that $t(\bar{b}\,|\,b) < \bar{t}$ for all $b \subset \bar{b}$, where $t(b_2|b_1)$ denotes the unique crossing point of $\phi(b_2,\cdot)$ and $\max\{\phi(b_1,\cdot),\phi(\varnothing,\cdot)\}$ for any $b_1 \subset b_2$ (see the Proof of Theorem 1). Therefore, $\bar{b}$ must be chain essential. Thus, by Theorem 3 and the Proof of Theorem 1, we have that menu B yields a strictly lower profit than menu $B \cup \bar{b}$. But menu B is an optimal menu. Contradiction.

Because of part (ii) of Proposition 3, if menu B includes the best-selling bundle $b^\star$, then $b^\star$ must be the smallest bundle in menu B. Thus, it suffices to show that $b^\star \in B$. To prove this claim, we need a different proof strategy because Proposition 3 asserts that $b^\star \in B$ regardless of whether the solution to the relaxed problem in Section III is implementable. We prove this claim using a perturbation argument.

Sketch of the Perturbation Argument.—Suppose for contradiction that $b^\star \notin B$. Consider adding the following option to the original menu B: a lottery of getting bundle $b^\star$ with a small probability ε, at the price of ε multiplied by the monopoly price of $b^\star$. By Lemma 1, the net profit change to the monopolist after adding this new option can be computed as $\mathbb{E}\left[\sum_b a'_b(t)\phi(b,t)\right] - \mathbb{E}\left[\sum_b a_b(t)\phi(b,t)\right]$, where a is the original allocation rule under menu B and a' is the induced allocation rule after the consumers readjust their optimal choices given the new option.

Note that by the Proof of Theorem 1, the allocation rule a must be equivalent to our construction given in Section III when applied to menu B. For any bundle b, let $t(b)$ be the unique crossing point of $\phi(b,t)$ and 0. Let b_1 be the smallest nonempty bundle in B. Then, upon the new option offered, we have

(i) all types $t \in \left[\underline{t}, t(b^\star)\right)$ will not take this option;

(ii) all types $t \in \left[t(b^\star), t(b_1)\right)$ will switch from $\varnothing$ to this option;

(iii) all types $t \in \left[t(b_1), t_\varepsilon\right)$ will switch from b_1 to this option, for some threshold t_ε.

The monopolist makes a gain from the types $t \in \left[t(b^\star), t(b_1)\right)$ and suffers a loss from the types $t \in \left[t(b_1), t_\varepsilon\right)$. It is crucial to compute the gain and the loss in terms of the virtual surplus. Denote them by $\mathrm{Gain}(\varepsilon)$ and $\mathrm{Loss}(\varepsilon)$. The key observation is that for $\varepsilon > 0$ small enough, we have $\text{Gain}(\varepsilon) > \text{Loss}(\varepsilon)$. Figure 9 illustrates with an example where $B = \{b_1, b_2\}$. The total gain from the types in $[t(b^\star), t(b_1))$ forms a rectangle whose area varies in ε linearly. The total loss from the types in $[t(b_1), t_\varepsilon)$ forms a triangle whose area varies in ε quadratically. But then menu B cannot be optimal. Contradiction.

Details of the Perturbation Argument.—First, we provide a lower bound on the gain in the virtual surplus. Because types in $[t(b^\star), t(b_1))$ will take this new option, the gain in the virtual surplus is at least

$$\text{Gain}(\varepsilon) \coloneqq \varepsilon \times \underbrace{\int_{t(b^\star)}^{t(b_1)} \phi(b^\star, t) dF(t)}_{=: K} = \varepsilon K > 0,$$

where the inequality $K > 0$ uses the single-crossing property of $\phi(b^\star, t)$.

Now, we provide an upper bound on the loss in the virtual surplus. Note that any type t who takes this option obtains a payoff that is at most

$$h(\varepsilon) \coloneqq \varepsilon \times \underbrace{\left[v(b^\star, \bar{t}) - v(b^\star, t(b^\star))\right]}_{=: Z} = \varepsilon Z.$$

Let b_2 be the second-smallest nonempty bundle in B (if it does not exist, put $t(b_2) = 1$ in what follows). Note that $t(b^\star) < t(b_1) < t(b_2)$ (see Figure 9). By the construction of (a, p), for any $\delta \in [0, t(b_2) - t(b_1)]$, we have $U(t(b_1) + \delta) = v(b_1, t(b_1) + \delta) - v(b_1, t(b_1))$, where U denotes the indirect utility function under (a, p).

Let $g(\delta) \coloneqq U(t(b_1) + \delta)$. Note that $v(b_1, t(b_1)) > 0$ and hence $v_t(b_1, t(b_1)) > 0$ by assumption. Thus, $\partial_+ g(0) > 0$. Since g' is continuous on $[0, t(b_2) - t(b_1)]$, there exist some constants $\bar{\delta} \in (0, t(b_2) - t(b_1))$ and $M > 0$ such that $g'(\delta) \geq M$ for all $\delta \in [0, \bar{\delta}]$. Let $\bar{\varepsilon} \coloneqq g(\bar{\delta}) > 0$. Note that for all $\varepsilon \in (0, \bar{\varepsilon})$, we have

$$g^{-1}(\varepsilon) = \int_0^\varepsilon (g^{-1})'(s) ds = \int_0^\varepsilon \frac{1}{g'(g^{-1}(s))} ds \leq \frac{1}{M}\varepsilon.$$

Note that any type $t \in [t(b_1), \bar{t}]$ switches to this new option only if $U(t) \leq h(\varepsilon)$. Let $\delta(\varepsilon) \coloneqq g^{-1}(h(\varepsilon))$. Then, observe that for all $\varepsilon \in \left(0, \frac{1}{Z}\bar{\varepsilon}\right)$, the loss in the virtual surplus is at most

$$\begin{aligned}
\text{Loss}(\varepsilon) &\coloneqq \int_{t(b_1)}^{t(b_1)+\delta(\varepsilon)} \phi(b_1, t) f(t) dt \\
&\leq \delta(\varepsilon) \times \underbrace{\max_{t \in [t(b_1), t(b_1)+\delta(\varepsilon)]} \{f(t)\phi(b_1, t)\}}_{=: \Phi(\varepsilon)} \\
&\leq \frac{Z}{M}\varepsilon \times \Phi(\varepsilon).
\end{aligned}$$

Observe that (i) $\Phi(\cdot)$ is a continuous function by Berge's theorem and (ii) $\Phi(0) = 0$ since $\phi\big(b_1, t(b_1)\big) = 0$. Therefore, there exists $\bar{\varepsilon}' > 0$ such that for all $\varepsilon \in (0, \bar{\varepsilon}')$, we have $\Phi(\varepsilon) < MK/Z$. Now, pick any $\varepsilon \in \Big(0, \min\Big\{\frac{1}{Z}\bar{\varepsilon}, \bar{\varepsilon}'\Big\}\Big)$. We must have $\text{Loss}(\varepsilon) \leq \frac{Z}{M}\varepsilon\,\Phi(\varepsilon) < \varepsilon K = \text{Gain}(\varepsilon)$. So menu B is suboptimal. Contradiction. ∎

REFERENCES

Adams, William James, and Janet L. Yellen. 1976. "Commodity Bundling and the Burden of Monopoly." *Quarterly Journal of Economics* 90 (3): 475–98.

Anderson, Eric T., and James D. Dana Jr. 2009. "When Is Price Discrimination Profitable?" *Management Science* 55 (6): 980–89.

Armstrong, Mark. 1996. "Multiproduct Nonlinear Pricing." *Econometrica* 64 (1): 51–75.

Athey, Susan. 2002. "Monotone Comparative Statics under Uncertainty." *Quarterly Journal of Economics* 117 (1): 187–223.

Bergemann, Dirk, Alessandro Bonatti, Andreas Haupt, and Alex Smolin. 2022. "The Optimality of Upgrade Pricing." Cowles Foundation Discussion Paper 2290.

Bulow, Jeremy, and John Roberts. 1989. "The Simple Economics of Optimal Auctions." *Journal of Political Economy* 97 (5): 1060–90.

Carroll, Gabriel. 2017. "Robustness and Separation in Multidimensional Screening." *Econometrica* 85 (2): 453–88.

Daskalakis, Constantinos, Alan Deckelbaum, and Christos Tzamos. 2014. "The Complexity of Optimal Mmechanism Design." In *SODA '14: Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, edited by Chandra Chekuri, 1302–18. Society for Industrial and Applied Mathematics.

Daskalakis, Constantinos, Alan Deckelbaum, and Christos Tzamos. 2017. "Strong Duality for a Multiple-Good Monopolist." *Econometrica* 85 (3): 735–67.

Deneckere, Raymond J., and R. Preston McAfee. 1996. "Damaged Goods." *Journal of Economics and Management Strategy* 5 (2): 149–74.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. MIT Press.

Ghili, Soheil. 2023. "A Characterization for Optimal Bundling of Products with Nonadditive Values." *American Economic Review: Insights* 5 (3): 311–26.

Gomes, Renato, and Alessandro Pavan. 2016. "Many-to-Many Matching and Price Discrimination." *Theoretical Economics* 11 (3): 1005–52.

Haghpanah, Nima, and Jason Hartline. 2021. "When Is Pure Bundling Optimal?" *Review of Economic Studies* 88 (3): 1127–56.

Hart, Sergiu, and Noam Nisan. 2019. "Selling Multiple Correlated Goods: Revenue Maximization and Menu-Size Complexity." *Journal of Economic Theory* 183: 991–1029.

Iyengar, Sheena S., and Mark R. Lepper. 2000. "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology* 79 (6): 995–1006.

Johnson, Justin P., and David P. Myatt. 2003. "Multiproduct Quality Competition: Fighting Brands and Product Line Pruning." *American Economic Review* 93 (3): 748–74.

Johnson, Justin P., and David P. Myatt. 2006. "On the Simple Economics of Advertising, Marketing, and Product Design." *American Economic Review* 96 (3): 756–84.

Manelli, Alejandro M., and Daniel R. Vincent. 2006. "Bundling as an Optimal Selling Mechanism for a Multiple-Good Monopolist." *Journal of Economic Theory* 127 (1): 1–35.

Manelli, Alejandro M., and Daniel R. Vincent. 2007. "Multidimensional Mechanism Design: Revenue Maximization and the Multiple-Good Monopoly." *Journal of Economic Theory* 137 (1): 153–85.

McAfee, R. Preston, and John McMillan. 1988. "Multidimensional Incentive Compatibility and Mechanism Design." *Journal of Economic Theory* 46 (2): 335–54.

McAfee, R. Preston, John McMillan, and Michael D. Whinston. 1989. "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values." *Quarterly Journal of Economics* 104 (2): 371–83.

Milgrom, Paul, and Ilya Segal. 2002. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica* 70 (2): 583–601.

Milgrom, Paul, and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62 (1): 157–80.

Mussa, Michael, and Sherwin Rosen. 1978. "Monopoly and Product Quality." *Journal of Economic Theory* 18 (2): 301–17.

Myerson, Roger B. 1981. "Optimal Auction Design." *Mathematics of Operations Research* 6 (1): 58–73.
Netflix. 2023. "Pricing." https://web.archive.org/web/20231009190817/https://www.netflix.com/signup/planform (accessed November 2023).
Pavlov, Gregory. 2011. "Optimal Mechanism for Selling Two Goods." *B. E. Journal of Theoretical Economics* 11 (1): 0000102202193517041664.
Quah, John K.-H. 2007. "The Comparative Statics of Constrained Optimization Problems." *Econometrica* 75 (2): 401–31.
Quah, John K.-H, and Bruno Strulovici. 2009. "Comparative Statics, Informativeness, and the Interval Dominance Order." *Econometrica* 77 (6): 1949–92.
Quah, John K.-H, and Bruno Strulovici. 2012. "Aggregating the Single Crossing Property." *Econometrica* 80 (5): 2333–48.
Rochet, Jean-Charles. 1987. "A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context." *Journal of Mathematical Economics* 16 (2): 191–200.
Rochet, Jean-Charles, and Philippe Chone. 1998. "Ironing, Sweeping, and Multidimensional Screening." *Econometrica* 66 (4): 783–826.
Simonson, Itamar. 1989. "Choice Based on Reasons: The Case of Attraction and Compromise Effects." *Journal of Consumer Research* 16 (2): 158–74.
Stigler, George J. 1963. "United States versus Loew's Inc.: A Note on Block-Booking." *Supreme Court Review* 4: 152–57.
Strausz, Roland. 2006. "Deterministic versus Stochastic Mechanisms in Principal-Agent Models." *Journal of Economic Theory* 128 (1): 306–14.
Yang, Frank. 2022. "Costly Multidimensional Screening." SSRN. http://dx.doi.org/10.2139/ssrn.3915700.