Nested Bundling*

Frank Yang†

November 15, 2023

Click here for the latest version.

Abstract

A *nested bundling* strategy creates menus in which more expensive bundles include all the goods of the less expensive ones. We study when nested bundling is optimal and determine 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)* sales quantity when sold alone. We show that, under quasiconcavity assumptions, if the undominated bundles with respect to this partial order are nested, then nested bundling 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 constructive monotone comparative statics theorem. We present partial converses. Additionally, we provide distributionally robust characterizations of nested bundling. We also show that under suitable conditions it is possible to extend our analysis to allow multidimensional heterogeneity.

Keywords: Optimal bundling, nested bundling, multidimensional screening, mechanism design, monotone comparative statics.

*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 thank 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.

†Graduate School of Business, Stanford University. Email: shuny@stanford.edu.

1 Introduction

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), and e-commerce platforms (e.g., Shopify).[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 or enterprise complexity in enterprise pricing. With that simplifying assumption, we are able to allow the consumers to have general non-additive 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 types are one-dimensional, the problem of bundling is inherently multidimensional, because screening can be done using multiple instruments.

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 both are sold alone at their respective monopoly prices. 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 quasiconcavity 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

[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).

[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).

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, 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.

Our results rely on distributional assumptions such as quasiconcavity of profit functions. However, we also present a robust nesting condition (Theorem 3), which determines when a nested menu is optimal for *all* type distributions (see Section 5.1). This result requires no quasiconcavity assumptions but instead it requires richer information about the demand system, e.g., pointwise elasticity comparisons of demand curves.

Our results also rely on the one-dimensional type assumption. However, it turns out to be possible to extend our analysis to allow for multidimensional heterogeneity (see Section 5.2). We show that if the additional dimension represents horizontal preferences that are orthogonal to and separable from the one-dimensional vertical type, then the optimal mechanism *collapses* the multidimensional types to be one-dimensional — by not pricing the horizontal attributes — and hence can be characterized by our results (Theorem 4).[3] For example, in the case of Netflix, the additional heterogeneity may represent preferences for different content and the vertical types may represent distastes for ads.

In practice, the sold-alone quantities might not always be feasible to estimate if the seller needs to offer a base bundle such as a "freemium" tier to all consumers. Our nesting condition generalizes naturally to allow more comparisons of bundles beyond the sold-alone quantities (see Section 6). In particular, we can compare 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. We show that regardless of in which order the bundles are excluded by such comparisons, as soon as the remaining bundles become nested, they form an optimal menu (Theorem 5).

On the technical side, a key contribution of this paper is a new condition for monotone comparative statics with partially ordered choice sets. A crucial step toward our results

[3] Additionally, we show that even in the fully general setting, our results are locally robust in the sense that the menu we identify is approximately optimal when the bundle values are sufficiently positively correlated (Proposition 7).

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 generalizes Milgrom and Shannon (1994) by providing a new condition that is agnostic to whether the choice variables exhibit complementarity or substitutability (see Section 4.4).

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 7.1). 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 9). 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 10). 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 7.2). Even in this well-studied case, our results yield new insights by providing a new characterization of the optimal menu (Proposition 11). 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 12). 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 7.3). We obtain (see Proposition 13) 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.

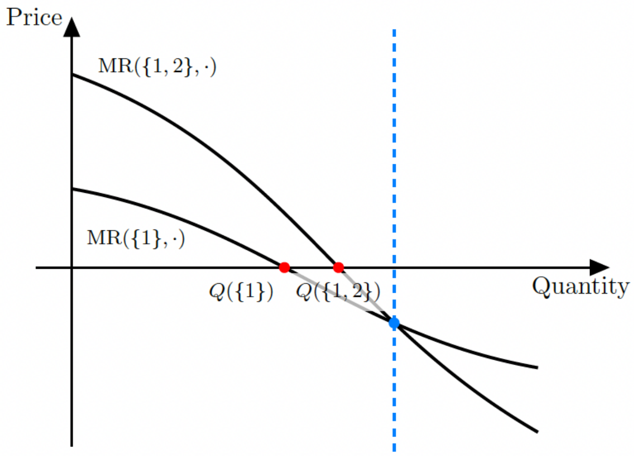
Discussion of Intuition. We now present the key intuition behind our main results (see Section 3.4 for further discussion).

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.[4] To see this, let $P(b,q)$ be the demand curve for bundle b when bundle b is sold alone, and $\mathrm{MR}(b,q)$ be the corresponding sold-alone marginal revenue curve for bundle b.

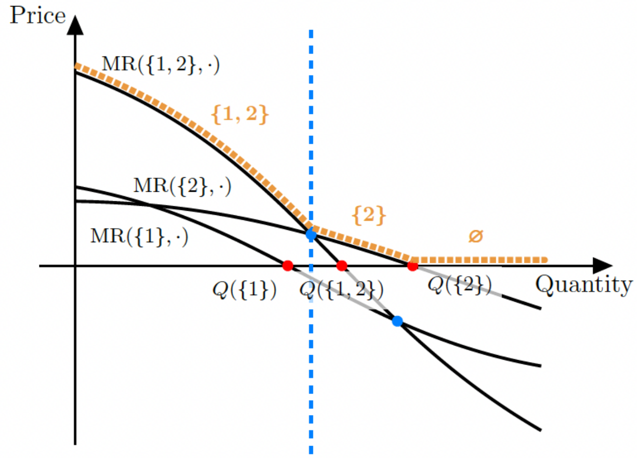
Because consumers are totally ordered, we can arrange them along a *single* quantity axis, with consumers positioned toward the right end having lower values for all the bundles. For a given menu of bundles and prices, the consumers optimally choose their favorite options, resulting in an allocation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

Fact 1. $$\text{Total Revenue} = \int_0^1 P(b(q),q)\,\mathrm{d}q - \int_0^1 \mathrm{CS}(q)\,\mathrm{d}q \qquad (\textbf{Value} - \textbf{Consumer Surplus})$$

[4] 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).



(a) MR curves under $\{1\} \preceq \{1,2\}$



(b) MR curves under $\{1\} \preceq \{1,2\}$, $\{2\} \npreceq \{1,2\}$

Figure 1: Illustration of the marginal revenue curves

$$
\begin{aligned}
&= \int_0^1 P(b(q),q)\,\mathrm{d}q + \int_0^1 q \cdot \mathrm{CS}'(q)\,\mathrm{d}q && \textbf{(Integration by Parts)} \\
&= \int_0^1 P(b(q),q)\,\mathrm{d}q + \int_0^1 q \cdot P_q(b(q),q)\,\mathrm{d}q && \textbf{(Envelope Theorem)} \\
&= \int_0^1 \mathrm{MR}(b(q),q)\,\mathrm{d}q\,.
\end{aligned}
$$

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 quasiconcave.[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 1a). This then implies that if it is profitable to sell a consumer the

[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 quasiconcave. This quasiconcavity assumption is stronger than what we actually assume in the model (see Section 2.1).

smaller bundle $\{1\}$, which happens before 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 1b). 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 1b). 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 1b) is indifferent between 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 1b. 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).

1.1 Related Literature

We study nested bundling when consumers have one-dimensional heterogeneity and non-additive preferences. Our model builds on a recent literature studying the optimality of pure bundling (i.e., selling only the grand bundle) with non-additive values (Ghili 2023, Haghpanah and Hartline 2021).[7] The closest paper is Ghili (2023) who intro-

[7] There is a long-standing literature on the profitability of price discrimination (Stokey 1979, Deneckere and McAfee 1996, Johnson and Myatt 2003, Anderson and Dana Jr 2009) which can be seen as studying the optimality of pure bundling under more restrictive assumptions. There is also an extensive literature on quality discrimination and nonlinear pricing (a la Mussa and Rosen 1978) which is a special case of our model.

duces the sold-alone quantities and shows that, under quasiconcavity assumptions, pure bundling is optimal if and only if the grand bundle has the highest sold-alone quantity. Ghili (2023)'s 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). Haghpanah and Hartline (2021) provide a ratio-monotonicity condition for the optimality of pure bundling, which motivates our robust nesting condition: under their condition, the grand bundle is again the unique undominated bundle (Corollary 5).

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.[10] The 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 com-

[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).

[10] For our robust nesting condition (Theorem 3), we use a different set of proof techniques introduced in Yang (2022) that studies multidimensional screening with costly instruments.

plementarity or substitutability.

The remainder of the paper proceeds as follows. Section 2 presents the model. Section 3 presents the main results. Section 4 sketches the main proof. Section 5 studies the robustness of the main results. Section 6 presents a generalization of the main results. Section 7 presents the applications. Section 8 concludes. Appendix A provides omitted proofs.

2 Model

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

Consumers have types $t \in \mathcal{T} := [\underline{t}, \overline{t}]$. 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 $\overline{t}$ to consume all the items: $\mathrm{argmax}_b\{v(b,\overline{t}) - C(b)\} = \overline{b}$ where $\overline{b} := \{1, \ldots, n\}$ 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. 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. ***(Incremental Monotonicity)***

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:

$$\begin{aligned} &v(a(t),t) - p(t) \geqslant v(a(\hat{t}),t) - p(\hat{t}) && \text{for all } t, \hat{t} \text{ in } \mathcal{T}; \\ &v(a(t),t) - p(t) \geqslant 0 && \text{for all } t \text{ in } \mathcal{T}. \end{aligned}$$

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 $\varnothing$).[11] A menu B is ***optimal*** if there exists an optimal mechanism (a,p) such that $a(t) \in B$ for all t.[12] Note that an

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

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

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, i.e.,

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

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

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

We assume that $\pi(b,q)$ is strictly quasiconcave in $q \in [0,1]$ with an interior maximum.[13] The ***sold-alone quantity*** $Q(b)$ is defined as the unique quantity at which the marginal profit equals zero, i.e.,

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

where $\mathrm{MR}(b,q)$ 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 quasiconcavity assumption on this profit function:

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

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.

2.1 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.

[13] For expositional simplicity, whenever we impose strict quasiconcavity 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 $g(x) \geqslant g(x')$ for all $x' \in [x_1, x_2]$ (i.e., we assume that the FOC is satisfied only at the maximum).

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 Quasiconcavity. We impose only a local quasiconcavity 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. In other words, 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 quasiconcavity is weaker than global quasiconcavity, 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 that $\pi(b_2, q) - \pi(b_1, q)$ is concave. In Section 5.1, we show that all quasiconcavity assumptions can be completely removed if we strengthen our notion of comparison from the comparison of the bundles' sold-alone quantities to be a pointwise elasticity comparison of their demand curves.

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.

However, in Section 5.2.1, we show that under suitable orthogonality and separability conditions, it is possible to extend our results to settings with multidimensional heterogeneity where the additional dimension of heterogeneity describes only horizontal preferences. In that setting, the optimal mechanism collapses the multidimensional heterogeneity to be one-dimensional and hence can be characterized by our results. In Section 5.2.2, we also show that, even in a fully general setting with multidimensional heterogeneity, our results are at least locally robust in the sense that holding the marginal distributions of bundle values fixed, the menu that we identify is approximately optimal when the bundle values are sufficiently positively correlated.

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 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.

3 Main Results

Our main results characterize *(i)* when nested bundling is optimal and *(ii)* which nested menu is optimal. In Section 3.1, we introduce a partial order that answers both questions. In Section 3.2, we provide some partial converses. In Section 3.3, we present a parametric example. In Section 3.4, we discuss the key intuition behind our results.

3.1 Optimality of Nested Bundling

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

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

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',

$$\text{both } b \text{ and } b' \text{ are undominated} \implies \text{either } b \subseteq b' \text{ or } b' \subseteq b \,. \qquad (\textbf{\textit{Nesting Condition}})$$

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.

(ii) A menu of undominated bundles is optimal.

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

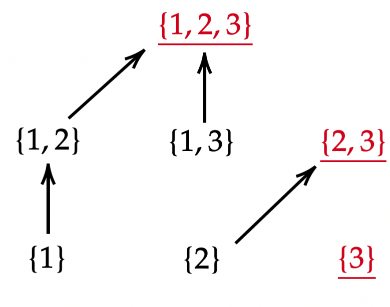


Figure 2: Illustration of the nesting condition for a three-item example. An upward arrow from b_1 to b_2 means $b_1 \preceq b_2$. The undominated bundles are nested: $\{3\} \subseteq \{2,3\} \subseteq \{1,2,3\}$.

The proof is in the appendix. We sketch the proof in Section 4. 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

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

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***, i.e., 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 below).

Theorem 1 is also agnostic to whether the items are complements or substitutes. To illustrate, consider two items and zero costs. Suppose that $v(\{1,2\},t) = \kappa \cdot \big(v(\{1\},t) + v(\{2\},t)\big)$ 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*** $\overline{b}$). 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 $Q(b_2 \mid b_1)$ denote the ***incremental quantity***, i.e., the quantity at which the incremental profit function $\pi(b_2, q) - \pi(b_1, q)$ reaches its maximum in the interval $[0, \max\{Q(b_1), Q(b_2)\}]$.[14]

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*

$$D := \left\{ b_j \in B : Q(b_{j+1} \mid b_j) \geqslant Q(b_j \mid b_{j-1}) \right\}. \tag{5}$$

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

In Section 6, we generalize both the nesting condition and the sieve algorithm to provide a more general procedure to find the minimal optimal menu.

[14] Strict quasiconcavity of $\pi(b_2, q) - \pi(b_1, q)$ on $[0, \min\{Q(b_1), Q(b_2)\}]$ implies that it is strictly quasiconcave on $[0, \max\{Q(b_1), Q(b_2)\}]$, and hence $Q(b_2 \mid b_1)$ is well-defined. Simply defining $Q(b_2 \mid b_1)$ 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 quasiconcave, then $Q(b_2 \mid b_1)$ can also be defined as the quantity maximizing the incremental profit globally.

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:

$$\text{For all } b_1 \text{ and } b_2, \qquad Q(b_1 \cup b_2) \geqslant \min\left\{Q(b_1), Q(b_2)\right\}. \qquad (\textbf{\textit{Union Quantity Condition}})$$

The following observation is instructive:

Proposition 2. *The union quantity condition implies the nesting condition.*

The proof is in the appendix. 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).[15] 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 case that menu $\big\{\{1\},\{1,2\}\big\}$ does not increase the profit beyond the single-bundle menu $\big\{\{1,2\}\big\}$.

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

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

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

Table 1: Bundle values by types for Example 1. 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.

$\{\{1,2\},\{2,3\}\}$ 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:

- price $5/2$ for the uniform lottery of getting either $\{1\}$ or $\{1,2\}$
- price $15/2$ for $\{2,3\}$

which yields a profit

$$\frac{1}{3} \times \frac{5}{2} + \frac{2}{3} \times \frac{15}{2} = \frac{35}{6} > \frac{17}{3}.$$

3.2 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, i.e.,

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

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* $b_m = \overline{b}$

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

The proof is in the appendix. 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 $\overline{b}$, 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 $\overline{b}$ as the least-selling bundle (if sold alone) in the menu.

When the menu B consists only of the grand bundle $\overline{b}$, Corollary 2 says that for pure bundling to be optimal, the grand bundle $\overline{b}$ 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 $Q(\overline{b}) \geqslant Q(b)$ 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 Appendix B.

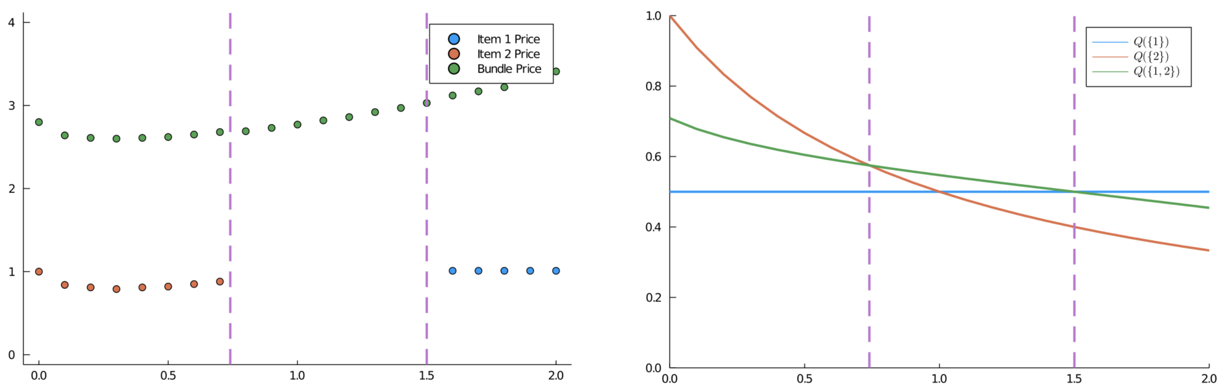
3.3 Parametric Example

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 $[0,2]$.[16] We vary parameter β from 0 to 2.

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



(a) Optimal mechanisms vs. β (b) Sold-alone quantities $Q(b)$ vs. β

Figure 3: Optimal mechanisms and sold-alone quantities $Q(b)$ for Example 2

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

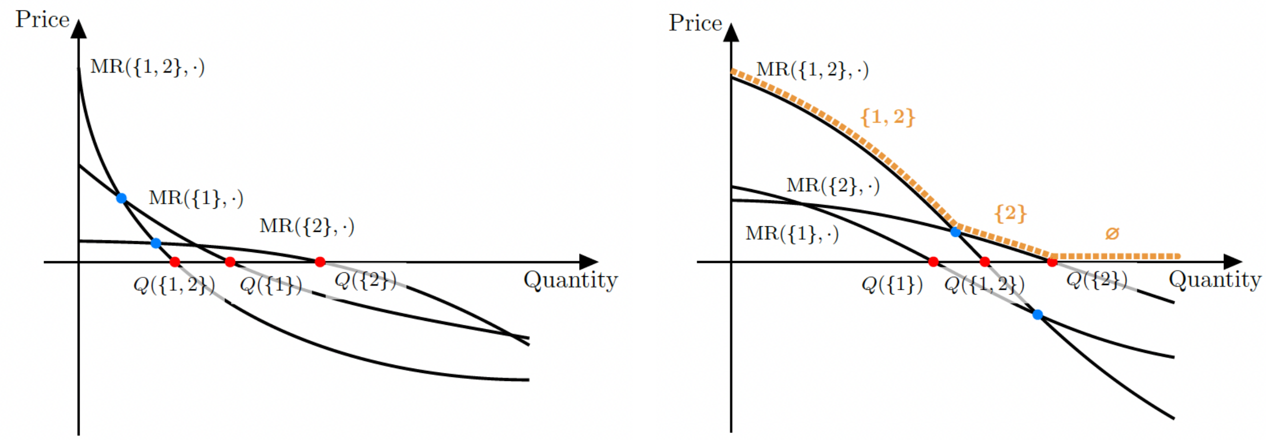
- $\{\{2\},\{1,2\}\}$ when $\beta \in [0, 0.74)$;
- $\{\{1,2\}\}$ when $\beta \in [0.74, 1.5]$;
- $\{\{1\},\{1,2\}\}$ when $\beta \in (1.5, 2]$.

The critical parameter values $\beta = 0.74$ and $\beta = 1.5$ are highlighted by the two vertical dashed lines in Figure 3a. These transitions are characterized by Theorem 1. Figure 3b plots the sold-alone quantities $Q(b)$ for the three bundles as parameter β varies. As Figure 3b 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, $\{\{1,2\}\}$ in the second region, and $\{\{1\},\{1,2\}\}$ in the third region, coinciding with the optimal menu.

3.4 Discussion of Intuition for Nested Bundling

3.4.1 Intuition Based on Marginal Revenue Curves

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



(a) MR curves under $\{1\} \not\preceq \{1,2\}$, $\{2\} \not\preceq \{1,2\}$ (b) MR curves under $\{1\} \preceq \{1,2\}$, $\{2\} \not\preceq \{1,2\}$

Figure 4: Further illustration of the marginal revenue curves

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 4a illustrates the MR curves under this case. In contrast to the case discussed in the Introduction (see Figure 4b), 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 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.

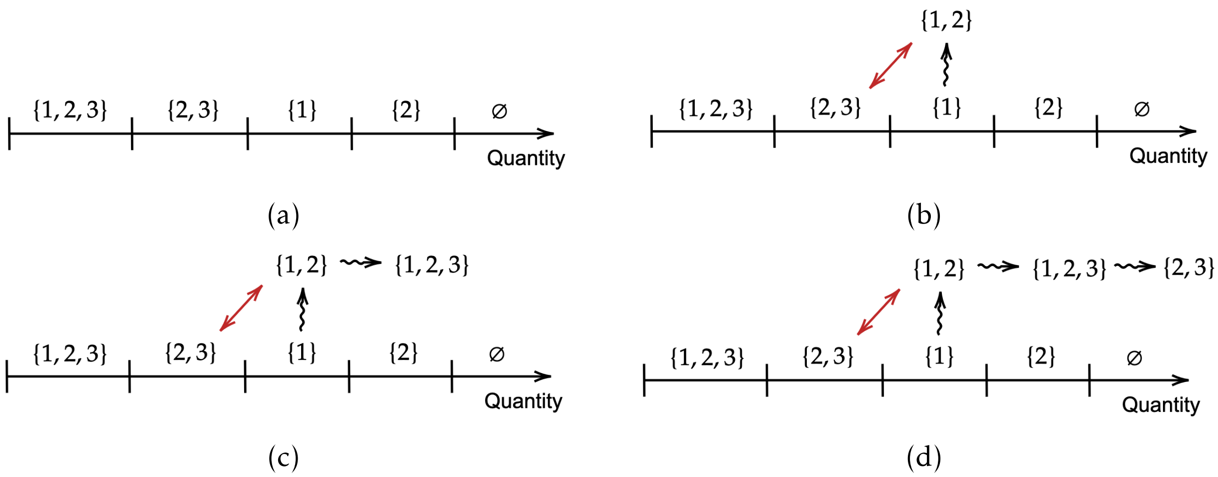


Figure 5: Illustration of the improvement argument for a three-item example

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 5a. 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 4b). 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 5b (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.[17]

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 $\{1,2\}$

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

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 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 5c; 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 5d. 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. In addition, the proof considers stochastic mechanisms and shows that, under the nesting condition, randomization cannot increase profit. Our key technical result, the monotone construction theorem, provides a weakening of the nesting condition that is both *necessary* and sufficient for these improvement arguments to yield a monotone allocation rule.

3.4.2 Alternative Intuition Based on Price Elasticities

In this section, we provide an alternative, price-theoretical intuition for nested bundling based on price elasticities. Let $\eta(b,q)$ be the usual price elasticity for bundle b evaluated at quantity q.[18] Suppose that there are two items and zero costs, and that $\eta(\{2\},q) < \eta(\{1,2\},q) < \eta(\{1\},q)$ for all quantities q. That is, bundle $\{2\}$ has a pointwise more elastic demand curve than bundle $\{1,2\}$, which in turn has a pointwise more elastic demand curve than bundle $\{1\}$. This assumption implies that $Q(\{2\}) > Q(\{1,2\}) > Q(\{1\})$, but it is much stronger than our nesting condition.

[18] That is, $\eta(b,q) := \left[\frac{\mathrm{d}\log P(b,q)}{\mathrm{d}\log q}\right]^{-1}$.

Let $\eta(\{1,2\},q \mid \{1\})$ be the price elasticity of the demand curve for the *incremental values* for bundle $\{1,2\}$ given bundle $\{1\}$. We can write

$$\eta(\{1,2\},q) = \frac{P(\{1,2\},q)}{q \cdot \frac{\mathrm{d}}{\mathrm{d}q} P(\{1,2\},q)} = \frac{\big(P(\{1,2\},q) - P(\{1\},q)\big) + P(\{1\},q)}{q \cdot \frac{\mathrm{d}}{\mathrm{d}q}\big(P(\{1,2\},q) - P(\{1\},q)\big) + q \cdot \frac{\mathrm{d}}{\mathrm{d}q} P(\{1\},q)}.$$

By the mediant inequality,[19] this implies that

$$\eta(\{1,2\},q \mid \{1\}) < \eta(\{1,2\},q) < \eta(\{1\},q).$$

That is, the demand curve for upgrading from bundle $\{1\}$ to bundle $\{1,2\}$ is even more *elastic*. In particular, it is profitable to charge a *low* enough price to sell the upgrade option to all consumers in the elastic region of the demand curve for bundle $\{1\}$. Thus, it is profitable to *exclude* bundle $\{1\}$, which is a dominated bundle, as an option from the menu. Symmetrically, for bundle $\{2\}$, we have

$$\eta(\{1,2\},q \mid \{2\}) > \eta(\{1,2\},q) > \eta(\{2\},q).$$

That is, the demand curve for upgrading from bundle $\{2\}$ to bundle $\{1,2\}$ is even more *inelastic*. In particular, it is profitable to charge a *high* enough price that leaves at least some consumers unserved in terms of the upgrade option. Thus, it is profitable to *include* bundle $\{2\}$, which is an undominated bundle, as an option in the menu.

Remark 2. Note however that, even in this two-item case, these price-theoretical arguments are incomplete. This is because they do not take into account the presence of the other item in the market when pricing the upgrade from one item to the bundle. In addition, the arguments assume that the elasticities can be pointwise ranked, which is much stronger than our nesting condition. The actual *joint* pricing problem is much more complex and cannot be simply reduced to two separate pricing problems. We emphasize that it is a *consequence* of our results that, under the nesting condition, one can pairwise compare the bundles. As explained in the Introduction and Section 3.4.1, the proof deals with the joint pricing problem by working in the quantity space rather than in the price space.

[19] That is, for any $\frac{a}{c} < \frac{b}{d}$ such that $c \cdot d > 0$, we have $\frac{a}{c} < \frac{a+b}{c+d} < \frac{b}{d}$.

4 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 quasiconcave for any two nested bundles $b_1 \subset b_2$. In the appendix, we complete the proof by weakening global quasiconcavity to local quasiconcavity, i.e., 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:

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

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*

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

We solve a relaxed problem by maximizing (8) over all measurable maps $a : \mathcal{T} \to \Delta(\mathcal{B})$, and then show that the solution to this relaxed problem is implementable. 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

- $b(t) \in \operatorname{argmax}_{b\in\mathcal{B}} \phi(b,t)$ for all t ;
- $b(t)$ is monotone in t in the set-inclusion order ;

- $b(t)$ is an undominated bundle for all t.

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 the appendix), 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).

4.1 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 $(\mathcal{X},\leqslant)$ 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***, i.e., for any $x_1 < x_2$ and $t < t'$, $g(x_1,t)\leqslant g(x_2,t) \implies g(x_1,t') < g(x_2,t')$. For any $x_1 < x_2$, let $t(x_2 \mid x_1)$ be the unique ***crossing point*** of $g(x_1,t)$ and $g(x_2,t)$:

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

Put $t(x_2 \mid x_1) = 1$ 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

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

where, in the above requirement, we also put $t(x_2 \mid x) = 1$ if no $x_2 > x$ exists, and $t(x \mid x_1) = 0$ 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 $(\mathcal{X},\leqslant)$ 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). 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*

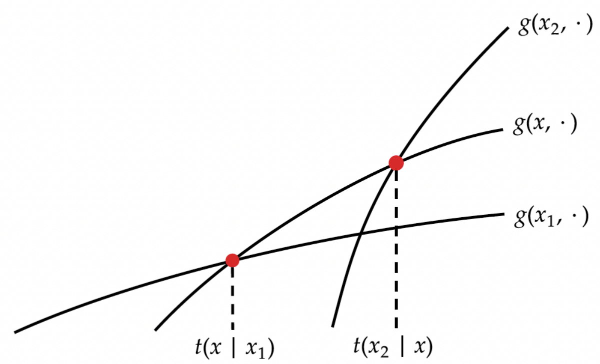


Figure 6: Illustration of the definition of a chain-essential element

(i) $x(t) \in \operatorname{argmax}_{x \in \mathcal{X}} g(x,t)$ *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 non-empty. 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

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

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

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

and let $x(s) = x_1$ if $s < t(x_2 \mid x_1)$ and $x(s) = x_n$ if $s \geqslant t(x_n \mid x_{n-1})$. 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 $s \notin \{t(x'' \mid x')\}_{x' < x''}$ (which is a finite set). Suppose for contradiction that there exists some $s \notin \{t(x'' \mid x')\}_{x' < x''}$ 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 $s > t(x'' \mid x) = 0$ and hence $g(x'', s) > g(x, s)$ by the definition of $t(x'' \mid x)$. Similarly, in case *(ii)*, there exists some $x' < x$ such that $s < t(x \mid x') = 1$ and hence $g(x', s) > g(x, s)$ by the definition of $t(x \mid x')$.

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

$$t(x \mid x') \geqslant t(x'' \mid x).$$

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

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

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

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

by the definition of $t(x \mid x')$.

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

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

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

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

and hence $g(x_j, s) \geqslant g(x', s)$ 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 $s \notin \{t(x'' \mid x')\}_{x' < x''}$, the above inequalities are all strict, and hence $g(x_j, s) > g(x', s)$ for all $x' \neq x_j \in \mathcal{Y}$.

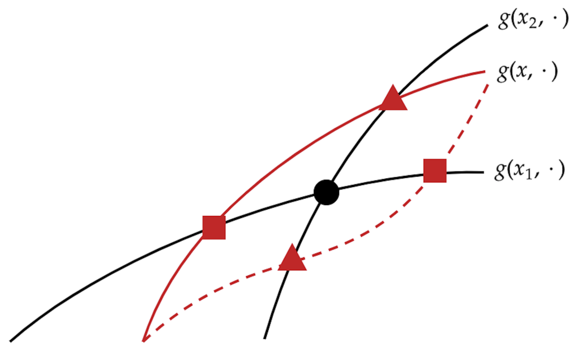


Figure 7: Illustration of the switching lemma. We hold $g(x_1,t), g(x_2,t)$ fixed and vary $g(x,t)$. Under the single-crossing property of g, note that comparing the horizontal positions of ■ and ▲ is equivalent to comparing the horizontal positions of ■ and ●.

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

4.2 Switching Lemma

Our proof of Theorem 1 will also make use of the following lemma:

Lemma 2 (Switching Lemma). *Let $(\mathcal{X}, \leqslant)$ 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 $t(x_2 \mid x_1) > 0$, we have*

$$t(x \mid x_1) < t(x_2 \mid x) \iff t(x \mid x_1) < t(x_2 \mid x_1). \tag{15}$$

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

4.3 Completion of the Proof Sketch

We apply Theorem 2 to the partially ordered set $(\mathcal{B}, \subseteq)$ and the virtual surplus function $\phi(b,t) : \mathcal{B} \times \mathcal{T} \to \mathbb{R}$. Note that, by (7), the strict global quasiconcavity 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 $t(b_2 \mid b_1)$ be the unique crossing point of $\phi(b_1,t)$ and $\phi(b_2,t)$. In particular, $t(b \mid \varnothing)$ 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 $t(b \mid \varnothing)$ is strictly between $\underline{t}$ and $\overline{t}$.

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

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

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 $\varnothing \subset b \subset b''$,

$$t(b \mid \varnothing) < t(b'' \mid \varnothing)$$

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:

- $b(t) \in \operatorname{argmax}_{b \in \mathcal{B}} \phi(b,t)$ 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 the appendix 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.

4.4 Discussion

Undominance and Chain Essential. As the proof shows, the 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 is developed with the notion of conditional dominance (see Theorem 5).

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 quasisupermodularity condition. Recall that a function $g : \mathcal{X} \to \mathbb{R}$ is ***quasisupermodular*** if for all x and $x' \in \mathcal{X}$,

$$g(x) \geqslant (>) g(x \wedge x') \implies g(x \vee x') \geqslant (>) g(x').$$

The following observation is instructive:

Proposition 4. *Let $(\mathcal{X}, \leqslant)$ 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 quasisupermodular in x for all t, then the chain-essential elements for g are totally ordered.*

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

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

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

$$t(x \vee x' \mid x \wedge x') \leqslant \max\Big\{t(x \mid x \wedge x'), t(x' \mid x \wedge x')\Big\},$$

and hence the chain-essential elements always form a chain, regardless of the value of κ.

Equivalence to Improvement Path. 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). That is, the iterative improvement arguments provided in Section 3.4.1 succeed *if and only if* the chain-essential elements are totally ordered.

Proposition 5. *Let $(\mathcal{X}, \leqslant)$ be a finite partially ordered set and $g : \mathcal{X} \times [0,1] \to \mathbb{R}$ be a function that is continuous in t and satisfies the strict single-crossing property in (x,t). The chain-essential elements for g are totally ordered if and only if:*

(i) There exists a monotone selection $x(\cdot)$ such that $x(t) \in \operatorname{argmax}_{x \in \mathcal{X}} g(x,t)$ for all t.

(ii) For all t and x_0, there exists a sequence $(x_0, \ldots, x_n)$ such that $g(x_i, t) \leqslant g(x_{i+1}, t)$ for all i, $x_n = x(t)$, and each pair (x_i, x_{i+1}) satisfies either $x_i > x_{i+1}$ or $x_i < x_{i+1}$.

The proof is in the appendix. The power of the monotone construction theorem is exactly that we require only the comparisons of ordered pairs by the definition of chain-essential elements, and yet the global solution can be constructed at all parameter values. As Proposition 5 shows, one cannot weaken our condition for this property to hold.

Weakening Single-Crossing Property. In the appendix, we show that our monotone construction theorem holds even with a *local* single-crossing property (Theorem 6), which enables us to weaken the global quasiconcavity of incremental profit functions to our local quasiconcavity condition, assumption (A2), in the proof of Theorem 1. When $\mathcal{X}$ is totally ordered, Milgrom and Shannon (1994) show that the single-crossing property holds if and only if monotone comparative statics hold for all $\mathcal{X}' \subseteq \mathcal{X}$. In comparison, when $\mathcal{X}$ is totally ordered, our local single-crossing property holds if and only if monotone comparative statics hold for all $\mathcal{X}' \subseteq \mathcal{X}$ that include the *minimum* element $\min\{\mathcal{X}\}$.[20]

5 Robustness of Nested Bundling

In this section, we study the robustness of our main results under alternative assumptions. In Section 5.1, we provide a robust nesting condition that ensures a given nested menu is optimal for all type distributions. In Section 5.2, we show that under suitable conditions it is possible to extend our analysis to allow multidimensional heterogeneity.

5.1 Distributional Robustness

Theorem 1 relies on the local quasiconcavity assumption (A2). However, we now show that we can fully remove this assumption if we strengthen our notion of comparisons and focus on deterministic mechanisms. In fact, the following result provides a robust nesting condition that ensures a given nested menu is optimal for *all* type distributions.

[20] Our local single-crossing property neither implies nor is implied by the interval dominance order of Quah and Strulovici (2009), which is equivalent to that monotone comparative statics hold for all intervals $\mathcal{X}' \subseteq \mathcal{X}$ (when $\mathcal{X}$ is totally ordered).

Theorem 3 (Robust Nesting). *Under zero costs, for any nested menu B, if:*

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

$$\frac{\mathrm{d}}{\mathrm{d}t} \log v(b_1, t) < \frac{\mathrm{d}}{\mathrm{d}t} \log v(b_2, t) \text{ for all } t \tag{16}$$

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

$$\frac{\mathrm{d}}{\mathrm{d}t} \log v(b_1, t) \geqslant \frac{\mathrm{d}}{\mathrm{d}t} \log v(b_2, t) \text{ for all } t \tag{17}$$

then menu B is optimal among deterministic mechanisms for all type distributions F.

The proof is in the appendix. Before sketching the proof, we make a few remarks. Theorem 3 does not even require any single-bundle profit function $\pi(b, q)$ to be quasiconcave. Instead, Theorem 3 requires a pointwise elasticity comparison of the demand curves, as the following observation shows:

Proposition 6. *Let $\eta(b, q)$ be the price elasticity for bundle b at quantity q.*[21] *For any b and b', $\frac{\mathrm{d}}{\mathrm{d}t} \log v(b, t) \leqslant \frac{\mathrm{d}}{\mathrm{d}t} \log v(b', t)$ for all $t \in \mathcal{T}$ if and only if $\eta(b, q) \leqslant \eta(b', q)$ for all $q \in [0, 1]$.*

The proof is in the appendix. In light of Proposition 6, the conditions in Theorem 3 can be seen as a global analog of the conditions in Corollary 1, by strengthening the comparison of unit-elastic quantities to a pointwise elasticity comparison.

In the special case where the menu $B = \{\overline{b}\}$ consists only of the grand bundle, an immediate consequence of Theorem 3 is the following pure bundling result:

Corollary 5 (Haghpanah and Hartline 2021). *If $v(b, t)/v(\overline{b}, t)$ is nondecreasing in t for all bundles b, then pure bundling is optimal among deterministic mechanisms.*

In the special case where the menu $B = \{b_1, \overline{b}\}$ consists of two bundles, Theorem 3 also immediately yields the following characterization:

Corollary 6 (Two-tier Menu). *If $v(b_1, t)/v(\overline{b}, t)$ is strictly decreasing in t for some b_1, and $v(b, t)/v(\overline{b}, t)$ is nondecreasing in t for all bundles $b \neq b_1$, then menu $\{b_1, \overline{b}\}$ is optimal among deterministic mechanisms.*

Proof Sketch for Theorem 3. The proof of Theorem 3 uses a strategy different from our main proof. Without quasiconcavity assumptions, the solution to the relaxed problem in Section 4 would not be implementable. Thus, instead of using the Myersonian approach,

[21] That is, $\eta(b, q) := \left[\frac{\mathrm{d} \log P(b,q)}{\mathrm{d} \log q}\right]^{-1}$.

we use a different proof approach introduced in Yang (2022). First, for any deterministic mechanism M, the proof reconstructs a stochastic mechanism $\widetilde{M}$ that improves on the original one and uses only bundles in the nested menu B, but satisfies only a subset of the IC constraints — the *downward* IC constraints. Second, it applies the ***downward sufficiency theorem*** in Yang (2022) to argue that there exists another weakly improving, fully incentive compatible stochastic mechanism $\widehat{M}$. Third, it shows that the stochastic mechanism $\widehat{M}$, which uses lotteries that are totally ordered by stochastic dominance, can be further improved by a deterministic mechanism that uses only bundles in menu B.

5.2 Multidimensional Types

5.2.1 Collapsing Multidimensional Types

Our main model requires the bundle values to be monotone in the vertical type t. In practice, consumers may have additional heterogeneity in horizontal preferences that cannot be captured by the one-dimensional types such as preferences for different colors, genres of music, or types of movies.

We incorporate this additional dimension of horizontal preferences into our main model as follows. Each allocation now consists of a pair (b, z) where bundle $b \in \mathcal{B}$ is a subset of items and $z \in \mathcal{Z}$ is a horizontal attribute. The set $\mathcal{Z}$ is an arbitrary finite set. Horizontal attributes do not affect production costs. For any assignment (b, z), the production cost is given by $C(b)$. A type now consists of $(t, \xi) \in \mathcal{T} \times \Xi$ where $\mathcal{T}$ is a compact interval on the real line and Ξ is an arbitrary measurable space.

A type (t, ξ) consumer's payoff is given by

$$v(b, t) \cdot u(z, \xi) - p$$

where v is the bundle value function as in the main model and $u : \mathcal{B} \times \Xi \to [0, 1]$ is a horizontal utility function. In addition to the multiplicative form, we also impose the following separability property: for all $\xi \in \Xi$

$$\max_{z \in \mathcal{Z}} u(z, \xi) = 1 . \qquad (\textbf{\textit{Separability}})$$

This property assumes that for every type (t, ξ), there exists a *favorite* horizontal attribute such that when assigned, the consumer's preferences over bundles of vertical attributes can be fully described by t. For example, in the context of streaming services, the vertical attribute could be whether the content is ad-free, and the horizontal attribute could be whether the content is a documentary or a comedy. Vertical types t represent consumers' distastes for ads and horizontal types ξ represent consumers' relative tastes for different genres. The separability condition holds if when viewing their favorite genres, all type t consumers are affected by ads in the same way regardless of the genres and their relative tastes ξ.

We will also assume that the vertical type t and horizontal type ξ are statistically independent:

$$t \perp\!\!\!\perp \xi \,. \qquad (\textbf{\textit{Orthogonality}})$$

In the example of streaming services, this means that knowing a consumer's relative taste for different content reveals no information about the consumer's distaste for ads.

A mechanism in this extended model is defined as

$$(a, p) : \mathcal{T} \times \Xi \to \Delta\big(\mathcal{B} \times \mathcal{Z}\big) \times \mathbb{R}$$

that satisfies the usual IC and IR constraints as in the main model. Let $\mathcal{M}$ be the set of all mechanisms. A mechanism (a, p) ***involves no horizontal distortion*** if for all t and ξ, the assignment of the horizontal attribute is deterministic and efficient (i.e., in $\operatorname{argmax}_z u(z, \xi)$). Let $\widetilde{\mathcal{M}}$ be the set of all mechanisms involving no horizontal distortion.

Theorem 4 (Multidimensional Types)**.** *Suppose that the separability and orthogonality conditions hold. Then:*

(i) The optimal profit under $\widetilde{\mathcal{M}}$ equals to the optimal profit under $\mathcal{M}$.

(ii) If assumptions (A1), (A2), and the nesting condition hold for $\{v(b, t), C(b), F(t)\}$, then the optimal mechanism can be implemented by offering a nested menu of undominated bundles with consumers freely choosing their favorite horizontal attributes.

Theorem 4 shows that under suitable separability and orthogonality conditions, the optimal mechanism collapses multidimensional type space $\mathcal{T} \times \Xi$ to the one-dimensional type space $\mathcal{T}$ — by not distorting the consumers' horizontal choices — and hence can be characterized by our results. In the running example of streaming services, this means that the company may find it optimal to offer a two-tier menu that prices whether the content is ad-free but allows the consumers to freely choose their favorite content.

Proof Sketch for Theorem 4. The proof of Theorem 4 is in the appendix. The basic intuition behind Theorem 4 can be understood as follows. Consider a relaxed problem in which the seller can *observe* the consumers' horizontal type ξ. Now, for every (t,ξ), because of the separability condition, if the seller assigns a non-favorite horizontal attribute z and some bundle b, then it can be replicated by assigning the favorite attribute and a lottery over bundles $\{\varnothing, b\}$. Because of the orthogonality condition, observing ξ reveals no information about t, and hence there exists a *single* optimal solution to the relaxed problem, for all observed horizontal types ξ, that involves no horizontal distortion. But then it must be optimal in the original problem when the seller cannot observe ξ. Part *(ii)* of Theorem 4 follows immediately from this argument and Theorem 1 because Theorem 1 shows the optimality of nested bundling even within all stochastic mechanisms.

5.2.2 Local Robustness to Multidimensional Types

Section 5.2.1 imposes a specific structure on the consumers' preferences to separate the allocations into vertical and horizontal attributes. Now, we describe a fully general model and show that our results are at least locally robust to multidimensional heterogeneity in this fully general setting.

Each consumer has a type $v \in \mathbb{R}_+^{2^n}$ describing the private value v^b for bundle b, with $v^\varnothing = 0$. A mechanism (a,p) in this setting is defined as

$$(a,p) : \mathbb{R}_+^{2^n} \to \Delta(\mathcal{B}) \times \mathbb{R}$$

that satisfies the usual IC and IR constraints. The seller has cost $C(b)$ for producing bundle b. The type distribution is given by $\gamma \in \Delta(\mathbb{R}_+^{2^n})$. Suppose that γ has bounded, continuous marginal distributions. For every joint distribution γ, there exists a unique ***comonotonic distribution*** $\gamma^{mon} \in \Delta(\mathbb{R}_+^{2^n})$ that shares the same marginal distributions with γ but is maximally positively correlated (i.e., the Fréchet-Hoeffding copula).

To describe the sense in which our results are locally robust, we apply the misspecification framework of Madarász and Prat (2017). For two distributions $\gamma_1, \gamma_2 \in \Delta(\mathbb{R}_+^{2^n})$, γ_1 and γ_2 are said to be δ-***close*** if $\mathbb{R}_+^{2^n}$ can be partitioned into disjoint measurable sets $S_1, \ldots, S_k$ such that $||v - v'||_{\text{sup}} < \delta$ for any v, v' in the same cell S_j and $\gamma_1(S_j) = \gamma_2(S_j)$ for each S_j (Madarász and Prat 2017). We say that a distribution γ is δ-***positively-correlated*** if it is δ-close to the comonotonic distribution γ^{mon}.

Our main model can be viewed as studying the optimal bundling problem for the comonotonic distribution γ^{mon}. Our conditions can be viewed as using only the information about the marginal distributions. In particular, under comonotonicity of the bundle values, assumption (A1) can be equivalently stated as:

A1'. $P(b_2,q)-P(b_1,q)$ is strictly decreasing in q for all $b_1 \subset b_2$, where $P(b,q)$ is the single-bundle demand curve for bundle b.

Moreover, assumption (A2) and our nesting condition also depend only on the marginal distributions.

Proposition 7. *Suppose that assumptions (A1'), (A2), and the nesting condition hold for a collection of continuous marginal distributions $\{\mu^b\}_b$. Then, for any $\varepsilon > 0$, there exists $\delta > 0$ such that, for any δ-positively-correlated distribution γ with marginals given by $\{\mu^b\}_b$, the menu of undominated bundles yields a profit that is at most ε away from the optimal profit.*

The proof is in the appendix. Proposition 7 follows from Theorem 1 and the main result of Madarász and Prat (2017). In particular, Madarász and Prat (2017) provide a method of discounting the prices in a menu (without changing the bundles in the menu) to make the menu robust to local misspecifications of the type space.

6 General Procedure to Find the Optimal Menu

In practice, it might not be feasible to estimate the sold-alone quantities for all bundles when the seller must offer some base bundle (e.g., a "freemium" tier) to all consumers. In this section, we generalize our nesting condition to allow more ways to exclude bundles from consideration when finding the optimal menu.

For three bundles $b_0 \subset b_1 \subset b_2$, we say that b_1 ***is dominated by*** b_2 ***conditional on*** b_0 if

$$Q(b_1 \mid b_0) \leqslant Q(b_2 \mid b_0),$$

where for any $b \subset b'$, recall that $Q(b' \mid b)$ denotes the incremental quantity of b' given b, i.e., the quantity at which the incremental profit function $\pi(b',q) - \pi(b,q)$ reaches its maximum in the interval $[0, \max\{Q(b'), Q(b)\}]$.

A bundle b is ***strongly undominated*** if for all b', b'' such that $b' \subset b \subset b''$, we have

$$Q(b \mid b') > Q(b'' \mid b').$$

Clearly, a strongly undominated bundle must be an undominated bundle.

Theorem 5 (General Nesting)**.** *Suppose that assumptions (A1) and (A2) hold. If the menu of strongly undominated bundles is nested, then it is a minimal optimal menu.*

Theorem 5 provides the following ***conditional sieve algorithm***:

Step 1. Pick any three bundles $b_0 \subset b_1 \subset b_2$.

Step 2. Remove b_1 from consideration if $Q(b_1 \mid b_0) \leqslant Q(b_2 \mid b_0)$.

Step 3. Repeat **Steps 1**-**2** until the remaining bundles are nested.

Theorem 5 implies that when **Step 3** stops, the remaining bundles always form an optimal menu regardless of how the bundles are chosen in **Step 1**. One can further apply Proposition 1 to the remaining bundles to find the minimal optimal menu.

Proof Sketch for Theorem 5. The proof of Theorem 5 is in the appendix. Theorem 5 generalizes Theorem 1 by allowing the removal of more bundles when checking the nesting condition. The proof again relies on Theorem 2 (monotone construction theorem) and Lemma 2 (switching lemma). In fact, the strongly undominated bundles are exactly the chain-essential elements in Theorem 2, when the partially ordered choice set is $(\mathcal{B}, \subseteq)$, and the objective function is the virtual surplus function.

7 Applications

In this section, we present three applications. In Section 7.1, we further connect optimal bundling strategies to demand structures. In Section 7.2, we apply our results to quality discrimination models and study how product line design depends on cost structures. In the last application, in Section 7.3, we connect costly screening to optimal bundling.

7.1 Bundling and Elasticity

We introduce a sufficient condition for the nesting condition in Theorem 1 in terms of price elasticities. Let

$$\eta(b, q) := \left[\frac{\mathrm{d} \log P(b, q)}{\mathrm{d} \log q}\right]^{-1}$$

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

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

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 8. *Under zero costs, the union elasticity condition implies the nesting condition.*

Proposition 8 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 3.1. Note that when costs are present, we can modify the price elasticity $\eta(b,q)$ to be $\tilde{\eta}(b,q) := \left[\frac{\mathrm{d}\log(P(b,q)-C(b))}{\mathrm{d}\log q}\right]^{-1}$ 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 9. *Suppose that assumptions (A1) and (A2) hold. Under the union elasticity condition and zero costs, the following nested menu is optimal:*

$$\left\{b_1^\star,\ b_1^\star \cup b_2^\star,\ b_1^\star \cup b_2^\star \cup b_3^\star,\ \dots,\ \overline{b}\right\}.$$

The proof is in the appendix. Under the union elasticity condition, Proposition 9 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.[22] Proposition 9 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. This construction fully accounts for the complementarity or substitutability patterns across different items.

7.1.1 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 9, 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 item i if for all $s < s'$

$$Q(\{i\};s') \leqslant Q(\{i\};s), \qquad Q(\{j\};s') = Q(\{j\};s), \qquad Q(\{1,2\};s') \leqslant Q(\{1,2\};s),$$

[22] Moreover, starting from this optimal menu in Proposition 9, we can always determine the minimal optimal menu by applying Proposition 1.

and

$$Q(\{i\};s) \leqslant Q(\{1,2\};s) \implies Q(\{i\};s') \leqslant Q(\{1,2\};s').$$

That is, as parameter s increases, the demand curve for item i and the demand curve for bundle $\{1,2\}$ become more inelastic in the sense of a smaller elastic region.[23] 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 i. 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 10. *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 quasiconvex in s.

The proof is in the appendix. Proposition 10 says that if there is a sequence of demand rotations for item i, i.e., 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 in Section 3.3. 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 10 — it shifts from $\{\{2\},\{1,2\}\}$ to $\{\{1,2\}\}$, and then to $\{\{1\},\{1,2\}\}$, as parameter β increases.

7.2 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 $\mathcal{X} :=$

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

$\{0, x_1, \ldots, x_n\} \subset \mathbb{R}_+$ 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

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

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.[24]

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 11. *Suppose that assumptions (A1) and (A2) hold. Let $\widehat{Q}$ be the the upper decreasing envelope of $Q : \mathcal{X} \to [0,1]$, i.e.,*

$$\widehat{Q}(x) := \inf\Big\{g(x) : \ g \textit{ is nonincreasing and } g \geqslant Q\Big\}.$$

Let

$$X^\star := \Big\{x : \widehat{Q}(x) = Q(x)\Big\}.$$

Then $X^\star$ is an optimal menu.

The proof is in the appendix. It applies Theorem 1 to this special case. Proposition 11 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.

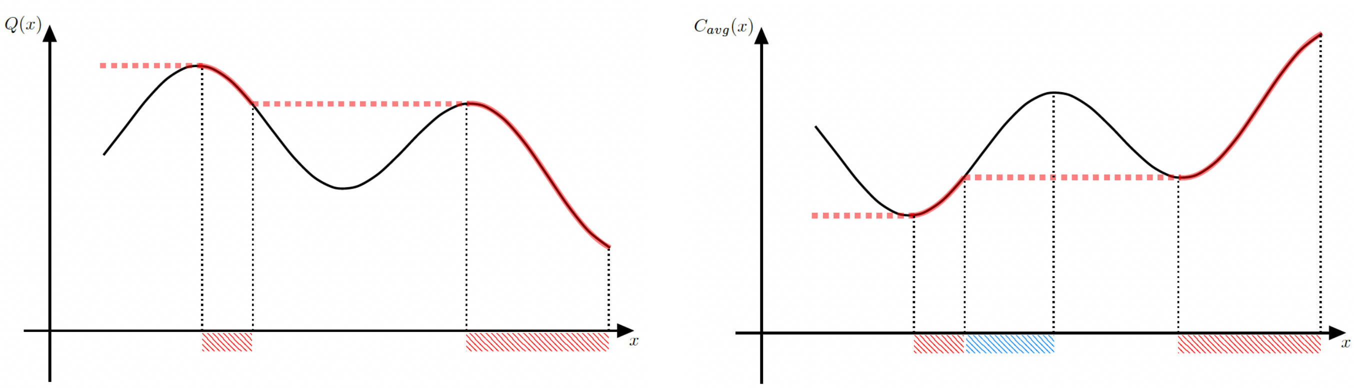
7.2.1 Product Line Design and Cost Structures

Applying Proposition 11, we can also characterize how cost structures affect the product line design under multiplicative utility functions:

Proposition 12. *Suppose that $v(x,t) = x \cdot t$ and type distribution F is regular.*[25] *Let $C_{avg}(x) :=$*

[24] For standard regularity conditions, see e.g., pp. 262–268 of Fudenberg and Tirole (1991).

[25] That is, $\phi(t) = t - \frac{1-F(t)}{f(t)}$ is strictly increasing.



(a) $Q(x)$ and $\widehat{Q}(x)$ (b) $C_{avg}(x)$ and $\check{C}_{avg}(x)$

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

$C(x)/x$ be the average cost function. Let $\check{C}_{avg}$ be the lower increasing envelope of C_{avg}, i.e.,

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

Let

$$X^\star := \left\{ x : \check{C}_{avg}(x) = C_{avg}(x) \right\}.$$

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.[26] 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). However, Proposition 12 shows that this conclusion is incomplete when the cost structure is more complex.[27] The optimal mechanism need *not* segment the market even when average cost and marginal cost are increasing. Figure 8 illustrates. Specifically, the marginal and average costs can both increase within the blue region highlighted in Figure 8b, yet the optimal mechanism does not segment the market using these qualities. Instead, as illustrated in Figure 8a, optimal quality choices are characterized by our notion of dominance.

At first glance, Proposition 11 and Proposition 12 may seem related to the *ironing* procedures in Mussa and Rosen (1978) and Myerson (1981). However, this connection is su-

[26] 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.

[27] 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 $\alpha_1 > \alpha_2$.

perficial. Even though Proposition 11 and Proposition 12 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 (see pp. 262–268 of Fudenberg and Tirole 1991). Our assumptions are much weaker, allowing for rich forms of bunching. This generality relies on our new constructive monotone comparative statics result (Theorem 2).

7.3 Costly Screening

Consider a monopolist selling a set of quality-differentiated goods, as described in Section 7.2. In addition to setting prices for these goods, the monopolist can also use non-price 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 such costly screening optimal?

We consider a special case of the model introduced 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 $(x,y,p) : \mathcal{T} \to \mathcal{X} \times \mathcal{Y} \times \mathbb{R}$. 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 7.2. 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:

$$\pi(y,q) := c\big(y, F^{-1}(1-q)\big) \cdot q\,.$$

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 quasiconcave 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:

$$\frac{\mathrm{d}}{\mathrm{d}t}\log c(y_1,t) \leqslant \frac{\mathrm{d}}{\mathrm{d}t}\log c(y_2,t) \text{ for all } t \implies Q(y_1) \geqslant Q(y_2).$$

Our next result shows that, under quasiconcavity 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 13. *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 quasiconcave 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 appendix. The intuition behind this characterization can be understood as follows. In the absence of costly screening, as discussed in Section 7.2, a menu of different qualities can be viewed as a nested menu. Moreover, we can view that 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 $x^\star := \max\{\operatorname{argmax}_x Q(x)\}$ be the best-selling quality (if sold alone). In the absence of costly screening, by Proposition 11, 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 $Q(y) \geqslant Q(x^\star)$, then we have $Q(y) \geqslant Q(x^\star) \geqslant Q(x)$ 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 of only different qualities. Therefore, by Theorem 1, the remaining menu is an optimal menu, and hence costly screening is suboptimal.

8 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. Additionally, we provide distributionally robust characterizations of nested bundling. We further show that under suitable conditions it is possible to extend our analysis to allow multidimensional heterogeneity. We apply our results to connect empirically relevant economic primitives to optimal menu design.

References

Adams, W. J. and J. L. Yellen (1976): "Commodity Bundling and the Burden of Monopoly," *Quarterly Journal of Economics*, 90(3), 475–498.

Anderson, E. T. and J. D. Dana Jr (2009): "When is Price Discrimination Profitable?" *Management Science*, 55(6), 980–989.

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

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

Bergemann, D., A. Bonatti, A. Haupt, and A. Smolin (2022): "The Optimality of Upgrade Pricing," arXiv:2107.10323 [econ.TH].

Bulow, J. and J. Roberts (1989): "The Simple Economics of Optimal Auctions," *Journal of Political Economy*, 97(5), 1060–1090.

Carroll, G. (2017): "Robustness and Separation in Multidimensional Screening," *Econometrica*, 85(2), 453–488.

Daskalakis, C., A. Deckelbaum, and C. Tzamos (2014): "The Complexity of Optimal Mmechanism Design," in *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, SIAM, 1302–1318.

——— (2017): "Strong Duality for a Multiple-Good Monopolist," *Econometrica*, 85(3), 735–767.

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

Fuchino, S. and S. Plewik (1999): "On a Theorem of E. Helly," *Proceedings of the American Mathematical Society*, 127(2), 491–497.

Fudenberg, D. and J. Tirole (1991): *Game Theory*, MIT Press.

Ghili, S. (2023): "A Characterization for Optimal Bundling of Products with Non-Additive Values," *American Economic Review: Insights*.

Gomes, R. and A. Pavan (2016): "Many-to-Many Matching and Price Discrimination," *Theoretical Economics*, 11(3), 1005–1052.

Haghpanah, N. and J. Hartline (2021): "When is Pure Bundling Optimal?" *Review of Economic Studies*, 88(3), 1127–1156.

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

Iyengar, S. S. and M. 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.

Johnson, J. P. and D. P. Myatt (2003): "Multiproduct Quality Competition: Fighting Brands and Product Line Pruning," *American Economic Review*, 93(3), 748–774.

——— (2006): "On the Simple Economics of Advertising, Marketing, and Product Design," *American Economic Review*, 96(3), 756–784.

Madarász, K. and A. Prat (2017): "Sellers with Misspecified Models," *Review of Economic Studies*, 84(2), 790–815.

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

——— (2007): "Multidimensional Mechanism Design: Revenue Maximization and the Multiple-good Monopoly," *Journal of Economic Theory*, 137(1), 153–185.

McAfee, R. P. and J. McMillan (1988): "Multidimensional Incentive Compatibility and Mechanism Design," *Journal of Economic Theory*, 46(2), 335–354.

McAfee, R. P., J. McMillan, and M. D. Whinston (1989): "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values," *Quarterly Journal of Economics*, 104.

Milgrom, P. and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62(1), 157–180.

Mussa, M. and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18(2), 301–317.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6(1), 58–73.

Netflix (2023): “Pricing,” Accessed at https://www.netflix.com/signup/planform on October 7, 2023.

Pavlov, G. (2011): “Optimal Mechanism for Selling Two Goods,” *B.E. Journal of Theoretical Economics*, 11(1).

Quah, J. K.-H. (2007): “The Comparative Statics of Constrained Optimization Problems,” *Econometrica*, 75(2), 401–431.

Quah, J. K.-H. and B. Strulovici (2009): “Comparative Statics, Informativeness, and the Interval Dominance Order,” *Econometrica*, 77(6), 1949–1992.

——— (2012): “Aggregating the Single Crossing Property,” *Econometrica*, 80(5), 2333–2348.

Rochet, J.-C. (1987): “A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context,” *Journal of Mathematical Economics*, 16(2), 191–200.

Rochet, J.-C. and P. Chone (1998): “Ironing, Sweeping, and Multidimensional Screening,” *Econometrica*, 66(4), 783–826.

Simonson, I. (1989): “Choice Based on Reasons: The Case of Attraction and Compromise Effects,” *Journal of Consumer Research*, 16(2), 158–174.

Stigler, G. J. (1963): “United States v. Loew’s Inc.: A Note on Block-booking,” *Supreme Court Review*, 1963, 152–157.

Stokey, N. L. (1979): “Intertemporal Price Discrimination,” *Quarterly Journal of Economics*, 93(3), 355–371.

Strassen, V. (1965): “The Existence of Probability Measures with Given Marginals,” *Annals of Mathematical Statistics*, 36(2), 423–439.

Yang, F. (2022): “Costly Multidimensional Screening,” *Available at SSRN 3915700*.

A Proofs

A.1 Proof of Lemma 1

The claim follows by the same proof of **Fact 1** in the Introduction (see Myerson 1981 and Bulow and Roberts 1989).

A.2 Proof of Lemma 2

For the ($\Longleftarrow$) direction, fix any $x_1 < x < x_2$ and suppose that $t(x \mid x_1) < t(x_2 \mid x_1)$. Then, since $t(x_2 \mid x_1) > 0$, we have

$$g(x_2, t(x_2 \mid x_1)) \leqslant g(x_1, t(x_2 \mid x_1)) < g(x, t(x_2 \mid x_1))$$

and hence $t(x_2 \mid x_1) \leqslant t(x_2 \mid x)$. Thus, $t(x \mid x_1) < t(x_2 \mid x_1) \leqslant t(x_2 \mid x)$.

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

$$g(x_2, t(x \mid x_1)) < g(x, t(x \mid x_1)) \leqslant g(x_1, t(x \mid x_1))$$

and hence $t(x \mid x_1) < t(x_2 \mid x_1)$.

A.3 Proof of Theorem 1

Section 4 proves parts *(i)* and *(ii)* of Theorem 1 under a stronger assumption that the incremental profit functions are globally quasiconcave. We complete the proof by weakening global quasiconcavity to local quasiconcavity, i.e., assumption (A2). We will also show part *(iii)* of Theorem 1 in the end.

The proof strategy is the same as in Section 4, except that we generalize both Theorem 2 and Lemma 2 to hold for functions g that only has a local single-crossing property.

A.3.1 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, i.e., $x_0 < x$ 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 $x_0 \leqslant x < x'$ and all $t < t'$,

$$g(x', t) \geqslant \max\{g(x,t), g(x_0,t)\} \implies g(x',t') > \max\{g(x,t'), g(x_0,t')\}.$$

Let

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

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

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 $g(x,s) \geqslant g(x_0,s) \implies g(x,s') > g(x_0.s')$ for all $x > x_0$ and all $s' > s$. For any $x' > x > x_0$, the following three statements are equivalent:*

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

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

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

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

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

Proof. We show that *(i)* $\implies$ *(ii)* $\implies$ *(iii)* $\implies$ *(i)*.

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

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

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

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

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

Finally, we show that

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

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

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

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

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

proving the result. □

As in Section 4, an element x is ***chain essential*** for g if for all $x_1 < x < x_2$

$$t(x \mid x_1) < t(x_2 \mid x)$$

where, in the above requirement, we also put $t(x_2 \mid x) = 1$ if no $x_2 > x$ exists, and $t(x \mid x_1) = 0$ 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 4 (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*

$$t(x \mid x_1) \geqslant t(x_2 \mid x)$$

we have

$$g(x, s) \leqslant \max\left\{g(x_1, s), g(x_2, s), g(x_0, s)\right\},$$

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

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

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

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

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

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

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*

$$t(x \mid x_1) < s < t(x_2 \mid x)$$

we have

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

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

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

Moreover, by the definition of $t(x_2 \mid x)$, 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. $\square$

Theorem 6 (Monotone Construction with Local Single Crossing)**.** *Let $(\mathcal{X}, \leqslant)$ 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) $x(t) \in \operatorname{argmax}_{x \in \mathcal{X}} g(x,t)$ *for all t, and $x(t)$ is the unique maximizer for almost all t;*

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

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

Proof. We use the same construction and the same proof strategy as in Section 4. By definition $\mathcal{Y}$ must be non-empty. 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

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

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

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

and let $x(s) = x_1$ if $s < t(x_2 \mid x_1)$ and $x(s) = x_n$ if $s \geqslant t(x_n \mid x_{n-1})$. Note that by construction, $x(\cdot)$ is well-defined, and satisfies properties *(ii)* and *(iii)* in Theorem 6. 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 $s \notin \{t(x'' \mid x')\}_{x'<x''}$. Suppose for contradiction that there exists some $s \notin \{t(x'' \mid x')\}_{x'<x''}$ 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 $s > t(x'' \mid x) = 0$ and hence $g(x'',s) > g(x,s)$ by the definition of $t(x'' \mid x)$. Similarly, in case *(ii)*, there exists some $x' < x$ such that $s < t(x \mid x') = 1$ and hence $\max\{g(x',s), g(x_0,s)\} > g(x,s)$ by the definition of $t(x \mid x')$.

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

$$t(x \mid x') \geqslant t(x'' \mid x) \,.$$

But then by Lemma 4, we have

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

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

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

which by the proof of Lemma 5 implies that

$$\begin{aligned}
g(x_j, s) \geqslant \max\{g(x_{j-1}, s), g(x_0, s)\} &\text{ and } \max\{g(x_j, s), g(x_0, s)\} \geqslant g(x_{j+1}, s) \\
g(x_{j-1}, s) \geqslant \max\{g(x_{j-2}, s), g(x_0, s)\} &\text{ and } \max\{g(x_{j+1}, s), g(x_0, s)\} \geqslant g(x_{j+2}, s) \\
\vdots \qquad &\text{ and } \qquad \vdots \\
g(x_2, s) \geqslant \max\{g(x_1, s), g(x_0, s)\} &\text{ and } \max\{g(x_{n-1}, s), g(x_0, s)\} \geqslant g(x_n, s).
\end{aligned}$$

There are two cases.

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

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

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

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

for all $k > j$. Thus, $g(x_j, s) \geqslant g(x_k, s)$ for all $k > j$. Therefore, $g(x_j, s) \geqslant g(x', s)$ for all $x' \in \mathcal{Y}$. Moreover, for all $s \notin \{t(x'' \mid x')\}_{x' < x''}$, 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 $0 \leqslant s < t(x_2 \mid x_1)$. 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 $g(x_1, s) \geqslant g(x_0, s)$. 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 $g(x_1, s) \geqslant g(x_0, s)$, 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 $s \notin \{t(x'' \mid x')\}_{x' < x''}$, $x(s)$ is the unique maximizer in $\mathcal{Y}$.

Now, combining **Step 1** and **Step 2**, we immediately have that property *(i)* of Theorem 6 must hold for our construction $x(t)$, proving the result. □

A.3.2 Switching Lemma with Local Single-Crossing Property

Lemma 6 (Switching Lemma with Local Single Crossing)**.** *Let $(\mathcal{X}, \leqslant)$ 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 $t(x_2 \mid x_1) > 0$, we have*

$$t(x \mid x_1) < t(x_2 \mid x) \iff t(x \mid x_1) < t(x_2 \mid x_1).$$

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

$$g(x, s) \leqslant \max\left\{g(x_1, s), g(x_2, s), g(x_0, s)\right\},$$

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

$$t(x \mid x_1) < s < t(x_2 \mid x_1).$$

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

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

Contradiction.

For the ($\Longrightarrow$) direction, fix any $x_1 < x < x_2$ and suppose that $t(x \mid x_1) < t(x_2 \mid x)$. Suppose for contradiction that $t(x \mid x_1) \geqslant t(x_2 \mid x_1)$. Then

$$0 < t(x_2 \mid x_1) \leqslant t(x \mid x_1) < t(x_2 \mid x) \leqslant 1\,.$$

Let $s = t(x \mid x_1)$. Since

$$t(x_2 \mid x_1) \leqslant s < t(x_2 \mid x)\,,$$

we have

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

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

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

Contradiction. $\square$

A.3.3 Completion of the Proof

Parts *(i)* and *(ii)*. Parts *(i)* and *(ii)* of Theorem 1 follow by the same proof as in Section 4, with Theorem 6 replacing Theorem 2 and Lemma 6 replacing Lemma 2.

Specifically, we apply Theorem 6 to the partially ordered set $(\mathcal{B}, \subseteq)$ and the virtual surplus function $\phi(b, t) : \mathcal{B} \times \mathcal{T} \to \mathbb{R}$. By (7) in Section 4, the assumption that $\pi(b, q)$ is strictly quasiconcave 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 quasiconcavity of $\pi(b_2, q) - \pi(b_1, q)$, i.e. assumption (A2), we have

$$\phi(b_2, s) \geqslant \phi(b_1, s) \implies \phi(b_2, s') > \phi(b_1, s')$$

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

Now, to apply Theorem 6, it remains to verify that the chain-essential elements in $\mathcal{B}$ form a chain; that is, we want to show that

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

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 $\varnothing \subset b \subset b''$,

$$t(b \mid \varnothing) < t(b'' \mid \varnothing)$$

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 4. For completeness, we also prove the following lemma about implementability which is used in Section 4.

Lemma 7. *Suppose that assumption (A1) holds. Then, for any deterministic, monotone allocation rule $b(t)$, there exists a payment rule $p(t)$ such that (b,p) satisfies all IC and IR constraints and that the lowest type $\underline{t}$ receives zero payoff under (b,p).*

Proof. Let $B = \{b(t)\}_{t \in \mathcal{T}}$ which is a nested menu. Without loss of generality, let $B = \{\varnothing, b_1, \ldots, b_m\}$ where $b_1 \subset \cdots \subset b_m$. For all $i = 1, \ldots, m$, let

$$s(b_i) := \inf\left\{t \in \mathcal{T} : b(t) \supseteq b_i\right\}.$$

We construct the bundle prices $\{p^\dagger(b)\}_{b \in B}$ by the following difference equation: for all $i = 1, \ldots, m$,

$$p^\dagger(b_i) - p^\dagger(b_{i-1}) := v(b_i, s(b_i)) - v(b_{i-1}, s(b_i)),$$

where we put $b_0 = \varnothing$ and $p^\dagger(b_0) = v(b_0, t) = 0$. To prove the result, it suffices to show that for all $t \in \mathcal{T}$, we have

$$b(t) \in \operatorname*{argmax}_{b' \in B} \left\{v(b', t) - p^\dagger(b')\right\}.$$

By assumption (A1), note that

$$U(b,t) := v(b,t) - p^\dagger(b)$$

has increasing differences, and hence single-crossing property, in (b,t). Moreover, by construction, $s(b_i)$ is a crossing point of $U(b_i, t)$ and $U(b_{i-1}, t)$. Since $b(\cdot)$ is monotone, we also have

$$s(b_1) \leqslant s(b_2) \leqslant \cdots \leqslant s(b_m).$$

Now fix any $i = 0, \ldots, m$ and any $t \in (s(b_i), s(b_{i+1}))$. For the edge cases, put $s(b_0) = \underline{t}$ and

$s(b_{m+1}) = \bar{t}$. Now, observe that we have

$$U(b(t),t) = U(b_i,t) \geqslant U(b_{i+1},t) \geqslant U(b_{i+2},t) \geqslant \cdots \geqslant U(b_m,t),$$

and

$$U(b(t),t) = U(b_i,t) \geqslant U(b_{i-1},t) \geqslant U(b_{i-2},t) \geqslant \cdots \geqslant U(b_0,t).$$

Hence, for any such t, we have

$$U(b(t),t) = \max_{b' \in \mathcal{B}} \{U(b',t)\}.$$

Now, if $t = s(b_i)$ for some $i = 1,\ldots,m+1$, then by definition, we have that *(i)* $b(t) = b_i$ or $b(t) = b_{i-1}$, and *(ii)* $U(b_i,t) = U(b_{i-1},t)$. Hence, the above argument also implies that

$$U(b(t),t) = \max_{b' \in \mathcal{B}} \{U(b',t)\}.$$

Finally, suppose $t = s(b_0)$. If $s(b_0) < s(b_1)$, then the above argument holds for t. If $s(b_0) = s(b_1)$, then the above argument also holds for t when applying to $i = 1$.

Thus, for all $t \in \mathcal{T}$, we have $U(b(t),t) = \max_{b' \in \mathcal{B}}\{U(b',t)\}$. Then, the payment rule defined by $p(t) := p^\dagger(b(t))$ implements the allocation rule $b(t)$, proving the result. □

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 6. Now, fix any implementable, potentially stochastic allocation rule $a(\cdot)$. Since

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

we have that for all t,

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

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 6). Therefore, by the above argument, for almost all $t \in \mathcal{T}'$, we have

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

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

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

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.

A.4 Proof of Proposition 1

The proof strategy is the same as in Section 4. Let $B = \{b_1, \ldots, b_m\}$, where $b_1 \subset \cdots \subset b_m$, be any optimal and nested menu. We apply Theorem 6 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 6 and the proof of Theorem 1.

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

$$Q(b_2 \mid b_1) := \underset{q \in [0, \max\{Q(b_1), Q(b_2)\}]}{\operatorname{argmax}} \pi(b_2, q) - \pi(b_1, q).$$

By (7) in Section 4, we can write

$$Q(b_2 \mid b_1) = 1 - F(\tilde{t}(b_2 \mid b_1))$$

where $\tilde{t}(b_2 \mid b_1)$ is defined as

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

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

$$\tilde{t}(b_2 \mid b_1) = t(b_2 \mid b_1)$$

where $t(b_2 \mid b_1)$ 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

$$Q(b_{j+1} \mid b_j) \geqslant Q(b_j \mid b_{j-1})$$

we have

$$t(b_j \mid b_{j-1}) \geqslant t(b_{j+1} \mid b_j)$$

and hence b_j cannot be chain essential for $\phi(b,t)$ by definition.

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

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

which, by the above argument, implies that

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

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 6. Suppose for contradiction that there exists some non-empty $b \in B$ 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,

$$\phi(b,t) \leqslant \max\left\{\phi(b',t),\phi(b'',t),\phi(\varnothing,t)\right\}.$$

At the same time, by **Step 2** in the proof of Theorem 6, 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.

A.5 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.

A.6 Proof of Proposition 3

Part *(ii)* of Proposition 3 follows from the monotone construction theorem. Specifically, we apply Theorem 6 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 $\overline{b}$ is not in the menu B. We apply the monotone construction theorem, Theorem 6, to $\phi(b,t)$ and $(B \cup \overline{b}, \subseteq)$. Since B is a nested menu, $B \cup \overline{b}$ must also be a nested menu. We claim that $\overline{b}$ is always a chain-essential element. To see it, note that because for all $b \subset \overline{b}$, we have

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

where the strict inequality is due to our assumption that the grand bundle is the unique surplus-maximizing bundle for the highest type $\overline{t}$. This implies that $t(\overline{b} \mid b) < \overline{t}$ for all $b \subset \overline{b}$, where $t(b_2 \mid b_1)$ 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, $\overline{b}$ must be chain essential. Thus, by Theorem 6 and the proof of Theorem 1, we have that menu B yields a strictly lower profit than menu $B \cup \overline{b}$. 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 4 is implementable. We prove this claim using a perturbation argument.

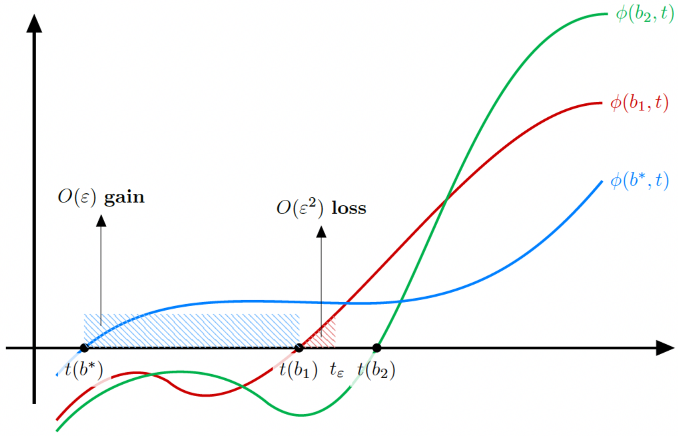


Figure 9: Illustration of the perturbation argument

A.6.1 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

$$\mathbb{E}\Big[\sum_b a'_b(t)\phi(b,t)\Big] - \mathbb{E}\Big[\sum_b a_b(t)\phi(b,t)\Big]$$

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 4. For any bundle b, let $t(b)$ be the unique crossing point of $\phi(b,t)$ and 0. Let b_1 be the smallest non-empty bundle in B. Then, upon the new option offered, we have:

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

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

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

The monopolist makes a gain from the types $t \in [t(b^\star), t(b_1))$ and suffers a loss from the types $t \in [t(b_1), t_\varepsilon)$. 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

$$\mathrm{Gain}(\varepsilon) > \mathrm{Loss}(\varepsilon).$$

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 (i.e., $O(\varepsilon)$ gain). The total loss from the types in $[t(b_1), t_\varepsilon)$ forms a triangle whose area varies in ε quadratically (i.e., $O(\varepsilon^2)$ loss). But then menu B cannot be optimal. Contradiction.

A.6.2 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

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

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

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

Let b_2 be the second smallest non-empty 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 $g(\delta) := U(t(b_1) + \delta)$. 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 $g'(\delta) \geqslant M$ for all $\delta \in [0, \bar{\delta}]$. Let $\bar{\varepsilon} := g(\bar{\delta}) > 0$. Note that for all $\varepsilon \in (0, \bar{\varepsilon})$, we have

$$g^{-1}(\varepsilon) = \int_0^\varepsilon (g^{-1})'(s)\,\mathrm{d}s = \int_0^\varepsilon \frac{1}{g'(g^{-1}(s))}\,\mathrm{d}s \leqslant \frac{1}{M}\varepsilon.$$

Note also that any type $t \in [t(b_1), \bar{t}]$ switches to this new option only if

$$U(t) \leqslant h(\varepsilon).$$

Let

$$\delta(\varepsilon) := g^{-1}(h(\varepsilon)).$$

Then, observe that for all $\varepsilon \in (0, \frac{1}{Z}\overline{\varepsilon})$, the loss in the virtual surplus is at most

$$\text{Loss}(\varepsilon) := \int_{t(b_1)}^{t(b_1)+\delta(\varepsilon)} \phi(b_1, t) f(t) \, \mathrm{d}t$$

$$\leqslant \delta(\varepsilon) \times \underbrace{\max_{t \in [t(b_1), t(b_1)+\delta(\varepsilon)]} \left\{ f(t)\phi(b_1, t) \right\}}_{=:\Phi(\varepsilon)} = \delta(\varepsilon) \times \Phi(\varepsilon) \leqslant \frac{Z}{M}\varepsilon \times \Phi(\varepsilon).$$

Observe that *(i)* $\Phi(\cdot)$ is a continuous function by Berge's theorem, and *(ii)* $\Phi(0) = 0$ since

$$\phi(b_1, t(b_1)) = 0.$$

Therefore, there exists $\overline{\varepsilon}' > 0$ such that for all $\varepsilon \in (0, \overline{\varepsilon}')$, we have

$$\Phi(\varepsilon) < \frac{MK}{Z}.$$

Now, pick any $\varepsilon \in \left(0, \min\{\frac{1}{Z}\overline{\varepsilon}, \overline{\varepsilon}'\}\right)$. We must have

$$\text{Loss}(\varepsilon) \leqslant \frac{Z}{M}\varepsilon\Phi(\varepsilon) < \varepsilon K = \text{Gain}(\varepsilon).$$

So menu B is suboptimal. Contradiction.

A.7 Proof of Proposition 4

Suppose for contradiction that there are two elements x_1 and x_2 that are both chain essential but cannot be ordered. Then, we have

$$x_1 \wedge x_2 < x_1, x_2 < x_1 \vee x_2.$$

Since x_1, x_2 are chain essential, we have

$$t(x_1 \mid x_1 \wedge x_2) < t(x_1 \vee x_2 \mid x_1),$$

$$t(x_2 \mid x_1 \wedge x_2) < t(x_1 \vee x_2 \mid x_2).$$

Suppose without loss of generality that

$$t(x_1 \mid x_1 \wedge x_2) \leqslant t(x_2 \mid x_1 \wedge x_2).$$

Fix any s such that

$$t(x_2 \mid x_1 \wedge x_2) < s < t(x_1 \vee x_2 \mid x_2).$$

Then we have

$$t(x_1 \mid x_1 \wedge x_2) \leqslant t(x_2 \mid x_1 \wedge x_2) < s < t(x_1 \vee x_2 \mid x_2),$$

which implies that

$$g(x_1, s) \geqslant g(x_1 \wedge x_2, s)$$

and

$$g(x_1 \vee x_2, s) < g(x_2, s),$$

contradicting that $g(\cdot, s)$ is quasisupermodular in x.

A.8 Proof of Proposition 5

($\Longrightarrow$) Suppose that the chain-essential elements are totally ordered. By Theorem 2, the existence of a monotone selection $x(\cdot)$ is immediate. We now show that for every $t \in [0,1]$ and every $x_0 \in \mathcal{X}$, there exists an ***improvement sequence*** $(x_0, \ldots, x_n)$ such that *(i)*

$$g(x_0, t) \leqslant g(x_1, t) \leqslant \cdots \leqslant g(x_n, t)$$

where $x_n = x(t)$ and that *(ii)* every pair (x_i, x_{i+1}) satisfies either $x_i < x_{i+1}$ or $x_i > x_{i+1}$.

We first prove this claim for all $s \notin \{t(x'' \mid x')\}_{x' < x''}$. Fix any such s. Let $\mathcal{Y}$ denote the set of chain-essential elements. Recall that $\mathcal{Y}$ is non-empty by definition. By Step 2 in the proof of Theorem 2, we know that if $x_0 \in \mathcal{Y}$, then there exists such an improvement sequence (moreover, the objective value is strictly increasing along the sequence). Now suppose $x_0 \notin \mathcal{Y}$. By Step 1 in the proof of Theorem 2, there exists $x_1 \in \mathcal{X}$ such that *(i)*

$$g(x_0, t) < g(x_1, t).$$

and that *(ii)* either $x_1 > x_0$ or $x_1 < x_0$.

If x_1 is in $\mathcal{Y}$, then we have found an improvement sequence by concatenating (x_0, x_1) with an improvement sequence that starts with x_1 (which always exists since $x_1 \in \mathcal{Y}$).

If x_1 is not in $\mathcal{Y}$, then by Step 1 in the proof of Theorem 2 again, there exists $x_2 \in \mathcal{X}$

such that *(i)*

$$g(x_1, t) < g(x_2, t).$$

and that *(ii)* either $x_2 > x_1$ or $x_2 < x_1$.

Because $\mathcal{X}$ is a finite set, this process can be repeated at most $|\mathcal{X}|$ number of times until we find a sequence $(x_0, x_1, \dots, x_n)$ such that *(i)*

$$g(x_0, t) < g(x_1, t) < \cdots < g(x_n, t).$$

and that *(ii)* every pair (x_i, x_{i+1}) satisfies either $x_i < x_{i+1}$ or $x_i > x_{i+1}$, and that *(iii)* $x_n \in \mathcal{Y}$. But then we may concatenate this improvement sequence with an improvement sequence that starts with x_n (which always exists since $x_n \in \mathcal{Y}$). Moreover, the improvement sequence that starts with x_n always ends with $x(t)$ by Step 2 in the proof of Theorem 2. Hence, our claim holds for all $s \notin \{t(x'' \mid x')\}_{x' < x''}$.

To show that this claim holds for all $s \in [0,1]$, we use a convergence argument as follows. Let $N = |\mathcal{X}|$ which is a finite number. Each improvement sequence can be viewed as a point in the finite-dimensional space $\{1, \dots, N\}^N$ which is a compact subset of $\mathbb{R}^N$.

Fix any $s \in [0,1]$ and any $x_0 \in \mathcal{X}$. First, there exists a sequence s_k converging to s where $s_k \notin \{t(x'' \mid x')\}_{x' < x''}$ (since $\{t(x'' \mid x')\}_{x' < x''}$ has measure 0 in $[0,1]$). For each s_k, there exists an improvement sequence $Z_k \in \{1, \dots, N\}^N$ by our previous step. Now, by the Bolzano-Weierstrass theorem, we know that there exists a converging subsequence Z_{k_j} such that $Z_{k_j} \to Z \in \{1, \dots, N\}^N$ as $j \to \infty$, where the convergence is with respect to the usual distance metric of $\mathbb{R}^N$. This implies that there exists some J such that for all $j \geqslant J$, $Z_{k_j} = Z$. Therefore, for all $j \geqslant J$, each s_{k_j} has the same improvement sequence Z. Denote the improvement sequence by $(x_0, x_1, \dots, x_n)$. Then, for all $j \geqslant J$, we have

$$g(x_0, s_{k_j}) \leqslant g(x_1, s_{k_j}) \leqslant \cdots \leqslant g(x_n, s_{k_j})$$

and $x_n = x(s_{k_j})$. By continuity of g in t, we have

$$g(x_0, s) \leqslant g(x_1, s) \leqslant \cdots \leqslant g(x_n, s).$$

To ensure that $x_n = x(s)$, note that $x(\cdot)$ by construction is right-continuous at all $t \in [0,1)$ and left-continuous at $t = 1$. Hence, we may choose the approximating sequence s_k to approximate s from the right if $s < 1$ and to approximate s from the left if $s = 1$. Then, we have $x_n = x(s)$, and hence $(x_0, x_1, \dots, x_n)$ is a desired improvement sequence.

($\Longleftarrow$) Suppose for contradiction that the chain-essential elements cannot be totally ordered. Let

$$\mathcal{W} := \{x(t)\}_{t\in[0,1]}$$

which is totally ordered since $x(\cdot)$ is monotone. This implies that there must exist some chain-essential element $x^\dagger$ that is not in $\mathcal{W}$. By the definition of chain-essential elements, we have

$$\max_{x':x'<x^\dagger} t(x^\dagger \mid x') < \min_{x'':x''>x^\dagger} t(x'' \mid x^\dagger),$$

where we put the left-hand side to be 0 if no $x' < x^\dagger$ exists and the right-hand side to be 1 if no $x'' > x^\dagger$ exists. Therefore, there exists some $s \in [0,1]$ such that for all $x' < x^\dagger$, we have

$$g(x^\dagger, s) > g(x', s)$$

and for all $x'' > x^\dagger$, we have

$$g(x^\dagger, s) > g(x'', s).$$

However, for such s and $x_0 = x^\dagger$, we also know that there exists an improvement sequence. In particular, there exists some x_1 such that *(i)* either $x_1 > x^\dagger$ or $x_1 < x^\dagger$ and that *(ii)*

$$g(x^\dagger, s) \leqslant g(x_1, s).$$

But that is a contradiction.

A.9 Proof of Theorem 3

The proof follows the sketch provided in Section 5.1. It builds on the techniques introduced in Yang (2022). We divide the proof into five steps. Appendix A.9.1 derives some preliminary inequalities. Appendix A.9.2 reconstructs an alternative, weakly improving mechanism that satisfies only downward IC constraints. Appendix A.9.3 applies the downward sufficiency theorem from Yang (2022) to modify the downward-IC mechanism to be fully IC. Appendix A.9.4 further modifies the fully IC, stochastic mechanism to be deterministic. Appendix A.9.5 completes the proof.

Throughout the proof, let $B = \{b_1, \ldots, b_m\}$, where $b_1 \subset \cdots \subset b_m$, denote the nested menu that satisfies the condition in Theorem 3 (the ***robust nesting condition***).

A.9.1 Preliminary Inequalities

Lemma 8. *Let $B = \{b_1, \ldots, b_m\}$, where $b_1 \subset \cdots \subset b_m$, satisfy the robust nesting condition. For any bundle $b \notin B$, let $b_k \in B$ be any bundle such that*

$$\frac{\mathrm{d}}{\mathrm{d}t} \log v(b,t) \geqslant \frac{\mathrm{d}}{\mathrm{d}t} \log v(b_k,t) \text{ for all } t.$$

Then, for all $i < j \leqslant k$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \log(v(b,t) - v(b_i,t)) > \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t) - v(b_i,t))$$

for all

$$t \in \mathcal{T}' := \left\{s : v(b_j,s) > v(b,s) > v(b_i,s)\right\}.$$

Moreover, if $\mathcal{T}' \neq \emptyset$, then it must be an interval.

Proof. Since B satisfies the robust nesting condition, for all $i < j \leqslant k$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \log(v(b,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_k,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t)) > \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_i,t)) \text{ for all } t.$$

By the mediant inequality, this implies the following two inequalities: *(i)* for all $i < j \leqslant k$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t) - v(b_i,t)) > \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t)) > \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_i,t)) \geqslant 0 \text{ for all } t.$$

and *(ii)* for all $i < j \leqslant k$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \log(v(b,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t) - v(b,t)) \text{ for all } t \text{ s.t. } v(b_j,t) > v(b,t).$$

Combining these two inequalities gives *(iii)* for all $i < j \leqslant k$

$$\frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t) - v(b_i,t)) > \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t) - v(b,t)) \text{ for all } t \text{ s.t. } v(b_j,t) > v(b,t).$$

Finally, by combining the mediant inequality and inequality *(iii)*, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \log(v(b,t) - v(b_i,t)) > \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t) - v(b_i,t)) > \frac{\mathrm{d}}{\mathrm{d}t} \log(v(b_j,t) - v(b,t))$$

for all

$$t \in \mathcal{T}' = \left\{s : v(b_j,s) > v(b,s) > v(b_i,s)\right\},$$

proving the inequality. Now, note that

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(v(b,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t}\log(v(b_j,t)) \text{ for all } t$$

implies that for all $t < t'$,

$$v(b,t) \geqslant v(b_j,t) \implies v(b,t') \geqslant v(b_j,t').$$

Therefore, the set $\{s : v(b_j,s) > v(b,s)\}$ is of the form $[\underline{t},t')$ or $[\underline{t},\overline{t}]$. Similarly, the set $\{s : v(b,s) > v(b_i,s)\}$ is of the form $(t'',\overline{t}]$ or $[\underline{t},\overline{t}]$. Therefore, if $\mathcal{T}' \neq \varnothing$, then $\mathcal{T}'$ must be an interval. $\square$

A.9.2 Reconstruction

Fix any deterministic mechanism denoted by (b,p). We now reconstruct an alternative stochastic mechanism that uses only bundles in the nested menu B and satisfies downward IC constraints.

For all t, let

$$b^+(t) := \inf\left\{b' \in B : v(b',t) \geqslant v(b(t),t)\right\}, \quad b^-(t) := \sup\left\{b' \in B : v(b',t) < v(b(t),t)\right\}.$$

By construction, if $b(t) \in B$, then $b^+(t) = b(t)$. Moreover, for all t, we have

$$v(b^+(t),t) \geqslant v(b(t),t) > v(b^-(t),t).$$

Now, for all t, let

$$\alpha(t) := \frac{v(b(t),t) - v(b^-(t),t)}{v(b^+(t),t) - v(b^-(t),t)} \in (0,1].$$

The (B,b,p)-***reconstruction*** (a,p) is defined by:

- assigning each reported type t a lottery over $\{b^+(t), b^-(t)\}$ with probability $\alpha(t)$ to bundle $b^+(t)$ and probability $1-\alpha(t)$ to bundle $b^-(t)$;
- keeping the payment $p(t)$ for each reported type t unchanged.

Lemma 9. *Let $B = \{b_1,\dots,b_m\}$, where $b_1 \subset \cdots \subset b_m$, satisfy the robust nesting condition. Let (b,p) be any deterministic mechanism. Let (a,p) be the (B,b,p)-reconstruction. Then (a,p) satisfies all downward IC constraints, i.e., for all $\hat{t} < t \in \mathcal{T}$, we have*

$$v(a(t),t) - p(t) \geqslant v(a(\hat{t}),t) - p(\hat{t}).$$

Proof. Note that by construction, for all t, we have

$$\begin{aligned} v(a(t),t) &= \frac{v(b(t),t)-v(b^-(t),t)}{v(b^+(t),t)-v(b^-(t),t)} \cdot v(b^+(t),t) + \left[1-\frac{v(b(t),t)-v(b^-(t),t)}{v(b^+(t),t)-v(b^-(t),t)}\right] \cdot v(b^-(t),t) \\ &= v(b(t),t). \end{aligned}$$

Thus, $v(a(t),t)-p(t) = v(b(t),t)-p(t)$. Hence, if type t reports truthfully, then its payoff is unchanged in the reconstruction compared to that in (b,p). Because the mechanism (b,p) satisfies the downward IC constraints, to show that (a,p) satisfies the downward IC constraints, it suffices to show that the deviating payoff of type t misreporting to be $\hat{t} < t$ is weakly lower under (a,p) compared to the deviating payoff under (b,p).

Since the payment rule is unchanged, it suffices to show that for $\hat{t} < t$,

$$v(a(\hat{t}),t) = \alpha(\hat{t})v(b^+(\hat{t}),t) + (1-\alpha(\hat{t}))v(b^-(\hat{t}),t) \leqslant v(b(\hat{t}),t). \tag{A.1}$$

To ease notation, fix $\hat{t} < t$ and write $\hat{b}$, $\hat{b}^+$, and $\hat{b}^-$ for $b(\hat{t})$, $b^+(\hat{t})$, and $b^-(\hat{t})$ respectively. By construction, $\hat{b}^- \subset \hat{b}^+$. We can write the deviating payoff as

$$\begin{aligned} \alpha(\hat{t})v(b^+(\hat{t}),t) + (1-\alpha(\hat{t}))v(b^-(\hat{t}),t) &= \alpha(\hat{t})\big(v(\hat{b}^+,t) - v(\hat{b}^-,t)\big) + v(\hat{b}^-,t) \\ &= \frac{v(\hat{b},\hat{t}) - v(\hat{b}^-,\hat{t})}{v(\hat{b}^+,\hat{t}) - v(\hat{b}^-,\hat{t})}\big(v(\hat{b}^+,t) - v(\hat{b}^-,t)\big) + v(\hat{b}^-,t). \end{aligned}$$

Note that if $\hat{b} \in B$, then $\hat{b} = \hat{b}^+$ and hence (A.1) clearly holds. From now on, suppose $\hat{b} \notin B$. Then, since menu B satisfies the robust nesting condition, by the definition of $\hat{b}^+$, we must have $\hat{b}^+ \subseteq b_k \in B$ for some $b_k \supset \hat{b}$ such that

$$\frac{\mathrm{d}}{\mathrm{d}s}\log v(\hat{b},s) \geqslant \frac{\mathrm{d}}{\mathrm{d}s}\log v(b_k,s) \geqslant \frac{\mathrm{d}}{\mathrm{d}s}\log v(\hat{b}^+,s) > \frac{\mathrm{d}}{\mathrm{d}s}\log v(\hat{b}^-,s) \text{ for all } s.$$

This implies that for all $s < s'$,

$$v(\hat{b},s) \geqslant v(\hat{b}^+,s) \implies v(\hat{b},s') \geqslant v(\hat{b}^+,s'). \tag{A.2}$$

Note that if $v(\hat{b},t) \geqslant v(\hat{b}^+,t)$, then (A.1) clearly holds as

$$\alpha(\hat{t})\big(v(\hat{b}^+,t) - v(\hat{b}^-,t)\big) + v(\hat{b}^-,t) \leqslant v(\hat{b}^+,t) \leqslant v(\hat{b},t).$$

Henceforth, suppose $v(\hat{b},t) < v(\hat{b}^+,t)$. Then, by (A.2), we have $v(\hat{b},\hat{t}) < v(\hat{b}^+,\hat{t})$. By definition of $\hat{b}^-$, we have $v(\hat{b}^-,\hat{t}) < v(\hat{b},\hat{t}) < v(\hat{b}^+,\hat{t})$. But for any $s < s'$ we also have

$$v(\hat{b},s) > v(\hat{b}^-,s) \implies v(\hat{b},s') > v(\hat{b}^-,s'). \tag{A.3}$$

Thus, we have $v(\hat{b}^-,t) < v(\hat{b},t) < v(\hat{b}^+,t)$. Therefore, we have

$$\{\hat{t},t\} \subseteq \mathcal{T}' := \left\{s \in \mathcal{T} : v(\hat{b}^-,s) < v(\hat{b},s) < v(\hat{b}^+,s)\right\}.$$

Since $\mathcal{T}' \neq \varnothing$, by Lemma 8, $\mathcal{T}'$ must be an interval. Moreover, by Lemma 8, we have

$$\frac{\mathrm{d}}{\mathrm{d}s}\log(v(\hat{b},s) - v(\hat{b}^-,s)) > \frac{\mathrm{d}}{\mathrm{d}s}\log(v(\hat{b}^+,s) - v(\hat{b}^-,s))$$

for all $s \in \mathcal{T}'$, which implies that

$$\frac{\mathrm{d}}{\mathrm{d}s}\log\left[\frac{v(\hat{b},s) - v(\hat{b}^-,s)}{v(\hat{b}^+,s) - v(\hat{b}^-,s)}\right] > 0$$

for all $s \in \mathcal{T}'$. Thus, we have

$$g(s) := \frac{v(\hat{b},s) - v(\hat{b}^-,s)}{v(\hat{b}^+,s) - v(\hat{b}^-,s)}$$

is strictly increasing on the interval $\mathcal{T}'$. But then since $\hat{t} < t \in \mathcal{T}'$, we have

$$\begin{aligned}
\alpha(\hat{t})v(b^+(\hat{t}),t) + (1-\alpha(\hat{t}))v(b^-(\hat{t}),t) &= \frac{v(\hat{b},\hat{t}) - v(\hat{b}^-,\hat{t})}{v(\hat{b}^+,\hat{t}) - v(\hat{b}^-,\hat{t})}\big(v(\hat{b}^+,t) - v(\hat{b}^-,t)\big) + v(\hat{b}^-,t). \\
&\leqslant \frac{v(\hat{b},t) - v(\hat{b}^-,t)}{v(\hat{b}^+,t) - v(\hat{b}^-,t)}\big(v(\hat{b}^+,t) - v(\hat{b}^-,t)\big) + v(\hat{b}^-,t). \\
&= v(\hat{b},t) - v(\hat{b}^-,t) + v(\hat{b}^-,t) = v(\hat{b},t),
\end{aligned}$$

which proves (A.1). The claim follows. □

A.9.3 Downward Sufficiency Theorem

The (B,b,p)-reconstruction (a,p) has the property that $a(t)$ assigns a lottery only over $\{b^-(t), b^+(t)\}$ which by construction are two adjacent bundles in the chain $b_1 \subset \cdots \subset b_m$ (which we also include the empty set).

Let

$$\mathcal{A} := \left\{a \in \Delta(B) : a \in \Delta(\{b_{j-1}, b_j\}) \text{ for some } b_j \in B\right\}$$

Then, note that on the set $\mathcal{A}$, the usual stochastic dominance order $\preceq_{st}$ is a *total* order. In fact, $\mathcal{A}$ can be identified as a compact subset of $\mathbb{R}$ (with $\preceq_{st}$ being identified as the usual order $\leqslant$ on $\mathbb{R}$). For any $a \prec_{st} a'$, note that

$$v_t(a,t) = \mathbb{E}_{b\sim a}[v_t(b,t)] < \mathbb{E}_{b\sim a'}[v_t(b,t)] = v_t(a',t)$$

since for any $b_i \subset b_j$, we have

$$v_t(b_i,t) = \frac{v_t(b_i,t)}{v(b_i,t)} \cdot v(b_i,t) < \frac{v_t(b_j,t)}{v(b_j,t)} \cdot v(b_j,t) = v_t(b_j,t),$$

where the strict inequality is due to that menu B satisfies the robust nesting condition and that $0 < v(b_i,t) \leqslant v(b_j,t)$. Therefore, the preferences satisfy the ***strict increasing differences*** property: for all $a \prec_{st} a'$ and all $t < t'$,

$$v(a,t') - v(a,t) < v(a',t') - v(a',t).$$

Then, by Yang (2022), when the allocation space is given by the totally ordered set $\mathcal{A}$ and the agent's utility function is given by $v(a,t) - p$, the downward IC constraints are sufficient for optimality. Specifically, we use the following theorem:

Theorem 7 (Downward Sufficiency Theorem, Yang 2022)**.** *Suppose that $v(a,t)$ is continuous and has strict increasing differences on $\mathcal{A} \times \mathcal{T}$, where $\mathcal{A}$ and $\mathcal{T}$ are two compact subsets of $\mathbb{R}$. Then, for any $(a,p) : \mathcal{T} \to \mathcal{A} \times \mathbb{R}$ that satisfies the IR constraints and the downward IC constraints, there exists $(a',p') : \mathcal{T} \to \mathcal{A} \times \mathbb{R}$ such that*

(i) *(a',p') satisfies the IR constraints;*

(ii) *(a',p') satisfies both the upward and downward IC constraints;*

(iii) *(a',p') yields a weakly higher profit than (a,p).*

A.9.4 Purification Lemma

Now, we reduce the fully IC but stochastic mechanism given in Theorem 7 to a deterministic mechanism. This step relies on a new purification lemma:

Lemma 10 (Purification Lemma)**.** *Let B be any nested menu. Consider any mechanism (a,p) : $\mathcal{T} \to \Delta(B) \times \mathbb{R}$ such that the allocation rule is stochastically monotone, i.e., for all $t < t'$,*

$$a(t) \preceq_{st} a(t').$$

Then there exists a deterministic mechanism $(b,p'):\mathcal{T}\to B\times\mathbb{R}$ *such that* (b,p') *yields a weakly higher profit than* (a,p).

Proof. Since $a(\cdot)$ is stochastically monotone, by Strassen (1965) (see Lemma 7 in Yang 2022), there exists some random variable $\varepsilon\in\mathcal{E}$ that is independent of t, where $\mathcal{E}$ is some measurable space, and some function $b^\dagger:\mathcal{T}\times\mathcal{E}\to B$ such that *(i)* $b^\dagger(\cdot,\varepsilon)$ is monotone in the set-inclusion order for all ε, and *(ii)* for every $t\in\mathcal{T}$, we have

$$b^\dagger(t,\varepsilon)\overset{d}{=}b_{a(t)}\,,$$

where b_a denotes the B-valued random variable sampled from lottery $a\in\Delta(B)$.

Then, by Lemma 1, under the mechanism (a,p) the seller's profit is bounded from the above by

$$\begin{aligned}\mathbb{E}_t\left[\phi(a(t),t)\right]&=\mathbb{E}_t\left[\mathbb{E}_{b\sim a(t)}[\phi(b,t)]\right]\\&=\mathbb{E}_t\left[\mathbb{E}_\varepsilon\left[\phi(b^\dagger(t,\varepsilon),t)\right]\right]\\&=\mathbb{E}_\varepsilon\left[\mathbb{E}_t\left[\phi(b^\dagger(t,\varepsilon),t)\right]\right]\\&\leqslant\sup_{\varepsilon\in\mathcal{E}}\mathbb{E}_t\left[\phi(b^\dagger(t,\varepsilon),t)\right]\end{aligned}$$

Let $K:=\sup_{\varepsilon\in\mathcal{E}}\mathbb{E}_t\left[\phi(b^\dagger(t,\varepsilon),t)\right]$, and $K_\varepsilon:=\mathbb{E}_t\left[\phi(b^\dagger(t,\varepsilon),t)\right]$. There exists a sequence ε_j such that as $j\to\infty$, we have

$$K_{\varepsilon_j}\to K\,.$$

Now, consider the sequence $b^\dagger(\cdot,\varepsilon_j)$. This is a sequence of monotone functions. By Helly's selection theorem for monotone functions on linearly ordered sets (Fuchino and Plewik 1999, Theorem 7), there exists a subsequence $\{b^\dagger(\cdot,\varepsilon_{j_k})\}_k$ that converge pointwise. For all t, let

$$b(t):=\lim_{k\to\infty}b^\dagger(t,\varepsilon_{j_k})\,.$$

Clearly, we have that $b(\cdot)$ is monotone. Moreover, we have

$$\mathbb{E}_t\left[\phi(b(t),t)\right]=\mathbb{E}_t\left[\lim_{k\to\infty}\phi(b^\dagger(t,\varepsilon_{j_k}),t)\right]=\lim_{k\to\infty}K_{\varepsilon_{j_k}}=\sup_{\varepsilon\in\mathcal{E}}\mathbb{E}_t\left[\phi(b^\dagger(t,\varepsilon),t)\right]\geqslant\mathbb{E}_t\left[\phi(a(t),t)\right].$$

Since $b(\cdot)$ is monotone, by Lemma 1 and Lemma 7, there exists some payment rule $p'(\cdot)$ such that (b,p') is a mechanism and yields the seller a profit of $\mathbb{E}_t\left[\phi(b(t),t)\right]$. The result follows. □

A.9.5 Completion of the Proof

We complete the proof of Theorem 3. Let $B = \{b_1, \dots, b_m\}$ where $b_1 \subset \cdots \subset b_m$ be the nested menu that satisfies the robust nesting condition. Let

$$\mathcal{A} := \Big\{ a \in \Delta(B) : a \in \Delta(\{b_{j-1}, b_j\}) \text{ for some } b_j \in B \Big\}.$$

Let $\mathcal{M}_D$ be the space of deterministic mechanisms. Let $\mathcal{M}_B$ be the space of deterministic mechanisms (b, p) such that $b(t) \in B$ for all t. Let $\mathcal{M}_{\mathcal{A}}$ be the space of stochastic mechanisms (a, p) such that $a(t) \in \mathcal{A}$ for all t. Let $\widehat{\mathcal{M}}_{\mathcal{A}}$ be the space of measurable maps $(a, p) : \mathcal{T} \to \mathcal{A} \times \mathbb{R}$ such that (a, p) satisfy all IR constraints and all *downward* IC constraints.

Then we have

$$\begin{aligned}
\sup_{(b,p)\in\mathcal{M}_D} \mathbb{E}[p(t)] &\leqslant \sup_{(a,p)\in\widehat{\mathcal{M}}_{\mathcal{A}}} \mathbb{E}[p(t)] && \text{(Lemma 9)} \\
&\leqslant \sup_{(a,p)\in\mathcal{M}_{\mathcal{A}}} \mathbb{E}[p(t)] && \text{(Theorem 7)} \\
&\leqslant \sup_{(b,p)\in\mathcal{M}_B} \mathbb{E}[p(t)], && \text{(Lemma 10)}
\end{aligned}$$

where the last inequality also uses the fact that when the allocation space is constrained to be the totally ordered set $\mathcal{A}$, any fully IC mechanism must have an allocation rule that is monotone in the stochastic dominance order $\preceq_{st}$. By Lemma 6 of Yang (2022), there exists a solution to the problem $\sup_{(b,p)\in\mathcal{M}_B} \mathbb{E}[p(t)]$. Therefore, menu B is optimal among all deterministic mechanisms.

A.10 Proof of Proposition 6

Note that $\eta(b, q) \leqslant \eta(b', q)$ for all $q \in [0, 1]$ holds if and only if

$$\left[\frac{\mathrm{d}\log P(b,q)}{\mathrm{d}\log q}\right]^{-1} \leqslant \left[\frac{\mathrm{d}\log P(b',q)}{\mathrm{d}\log q}\right]^{-1} \quad \text{for all } q \in [0,1]$$

which holds if and only if

$$\frac{\mathrm{d}\log P(b,q)}{\mathrm{d}\log q} \geqslant \frac{\mathrm{d}\log P(b',q)}{\mathrm{d}\log q} \quad \text{for all } q \in [0,1]$$

which holds if and only if

$$\frac{\mathrm{d}\log P(b,q)}{\mathrm{d}q} \geqslant \frac{\mathrm{d}\log P(b',q)}{\mathrm{d}q} \quad \text{for all } q \in [0,1]$$

which holds if and only if

$$\frac{P_q(b,q)}{P(b,q)} \geqslant \frac{P_q(b,q)}{P(b,q)} \quad \text{for all } q \in [0,1]$$

which holds if and only if

$$\frac{v_t(b,F^{-1}(1-q))}{v(b,F^{-1}(1-q))} \cdot \frac{-1}{f(F^{-1}(1-q))} \geqslant \frac{v_t(b',F^{-1}(1-q))}{v(b',F^{-1}(1-q))} \cdot \frac{-1}{f(F^{-1}(1-q))} \quad \text{for all } q \in [0,1]$$

which holds if and only if

$$\frac{v_t(b,t)}{v(b,t)} \leqslant \frac{v_t(b',t)}{v(b',t)} \quad \text{for all } t \in [\underline{t},\overline{t}],$$

where the last step follows by a change of variable $t = F^{-1}(1-q)$. The claim follows.

A.11 Proof of Theorem 4

Recall that $\mathcal{M}$ is the set of all mechanisms, and $\widetilde{\mathcal{M}}$ is the set of all mechanisms involving no horizontal distortion. We show that under the separability and orthogonality conditions, the optimal profit under $\mathcal{M}$ equals that under $\widetilde{\mathcal{M}}$.

Let $\widehat{M}$ be the set of measurable maps:

$$(a,p) : \mathcal{T} \times \Xi \to \Delta(\mathcal{B} \times \mathcal{Z}) \times \mathbb{R},$$

such that *(i)* it satisfies all IR constraints, *(ii)* it involves no horizontal distortion (i.e., the assignment of the horizontal attribute is deterministic and efficient), *(iii)* no type (t,ξ) wants to misreport to be $(\hat{t},\xi)$ for all $\hat{t},t,\xi$ (i.e., satisfies all IC constraints of the form $(t,\xi) \to (\hat{t},\xi)$).

We show that

$$\sup_{(a,p)\in\mathcal{M}} \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big] \overset{(i)}{\leqslant} \sup_{(a,p)\in\widehat{\mathcal{M}}} \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big] \overset{(ii)}{\leqslant} \sup_{(a,p)\in\widetilde{\mathcal{M}}} \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big], \tag{A.4}$$

where $C(a)$ denotes $\mathbb{E}_{(b,z)\sim a}[C(b)]$. This immediately implies the result since by definition we also have

$$\sup_{(a,p)\in\mathcal{M}} \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big] \geqslant \sup_{(a,p)\in\widetilde{\mathcal{M}}} \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big].$$

We divide the proof into two steps, which respectively prove inequalities *(i)* and *(ii)*.

A.11.1 Reconstruction

First, to show

$$\sup_{(a,p)\in\mathcal{M}} \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big] \leqslant \sup_{(a,p)\in\widehat{\mathcal{M}}} \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big], \tag{A.5}$$

note that it suffices to show that for every $(a,p)\in\mathcal{M}$, we can find a weakly improving $(a',p')\in\widehat{\mathcal{M}}$.

Fix any $(a,p)\in\mathcal{M}$. Let $I := |\mathcal{B}|$ and $J := |\mathcal{Z}|$. We may write the mechanism (a,p) as assigning probabilities α_{ij} and payment p to each type (t,ξ), where

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\alpha_{ij}(t,\xi) = 1\,.$$

For each ξ and i, define

$$\beta_i(t,\xi) := \sum_j \alpha_{ij}(t,\xi)u(z_j,\xi) \in [0,1]\,,$$

where we have used that for all z and ξ,

$$0 \leqslant u(z,\xi) \leqslant 1\,.$$

Note that

$$\sum_i \beta_i(t,\xi) = \sum_i\sum_j \alpha_{ij}(t,\xi)u(z_j,\xi) \leqslant 1\,.$$

Let

$$\beta_0(t,\xi) := 1 - \sum_i \beta_i(t,\xi) \in [0,1]\,.$$

Without loss of generality, let $b_1 = \varnothing$. Consider the following (a',p'): for each type (t,ξ),

- assign deterministically $z \in \operatorname{argmax}_{\mathcal{Z}} u(z,\xi)$

- assign bundle $b_i \neq \varnothing$ with probability $\beta_i(t,\xi)$
- assign bundle $\varnothing$ with probability $\beta_0(t,\xi) + \beta_1(t,\xi)$
- keep payments unchanged.

By construction, the assignments are well-defined. We now argue that the map (a', p') satisfies all IC constraints that involve only misreporting along the t dimension, i.e.,

$$\sum_{i,j} \alpha'_{ij}(t,\xi) v(b_i,t) u(z_j,\xi) - p(t,\xi) \geqslant \sum_{i,j} \alpha'_{ij}(\hat{t},\xi) v(b_i,t) u(z_j,\xi) - p(\hat{t},\xi) \text{ for all } t, \hat{t}. \tag{A.6}$$

Note that by construction, we have for all $t, \hat{t} \in \mathcal{T}$,

$$\begin{aligned}
\sum_{i,j} \alpha'_{ij}(\hat{t},\xi) v(b_i,t) u(z_j,\xi) &= \sum_{i\neq 1} \beta_i(\hat{t},\xi) v(b_i,t) + \Big(\beta_0(\hat{t},\xi) + \beta_1(\hat{t},\xi)\Big) v(\varnothing, t) \\
&= \sum_{i\neq 1} \Big(\sum_j \alpha_{ij}(\hat{t},\xi) u(z_j,\xi) \Big) v(b_i,t) \\
&= \sum_{i,j} \alpha_{ij}(\hat{t},\xi) v(b_i,t) u(z_j,\xi),
\end{aligned}$$

which immediately implies that (A.6) since the original mechanism (a,p) satisfies all IC constraints. Moreover, note that

$$\mathbb{E}\Big[p'(t,\xi) - C(a'(t,\xi))\Big] = \mathbb{E}\Big[p(t,\xi) - C(a'(t,\xi))\Big] \geqslant \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big],$$

where the inequality is due to that for all t, ξ

$$\begin{aligned}
C(a'(t,\xi)) = \sum_i \beta_i(t,\xi) C(b_i) &= \sum_i \sum_j \alpha_{ij}(t,\xi) u(z_j,\xi) C(b_i) \\
&\leqslant \sum_i \sum_j \alpha_{ij}(t,\xi) C(b_i) = C(a(t,\xi)),
\end{aligned}$$

where we have used $C(b) \geqslant 0$ with $C(\varnothing) = 0$, and $u(z,\xi) \in [0,1]$.

Therefore, (A.5) must hold.

A.11.2 Decomposition

We now show that

$$\sup_{(a,p)\in\widehat{\mathcal{M}}} \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big] \leqslant \sup_{(a,p)\in\widetilde{\mathcal{M}}} \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big]. \tag{A.7}$$

Note that because ξ is independent of t, we have

$$\sup_{(a,p)\in\widehat{\mathcal{M}}} \mathbb{E}\Big[p(t,\xi) - C(a(t,\xi))\Big] \leqslant \mathbb{E}_\xi\left[\sup_{(a,p)\in\widehat{\mathcal{M}}} \mathbb{E}_t\Big[p(t,\xi) - C(a(t,\xi))\Big]\right].$$

Now, fix any realized ξ'. Note that the problem

$$\sup_{(a,p)\in\widehat{\mathcal{M}}} \mathbb{E}_t\Big[p(t,\xi') - C(a(t,\xi'))\Big]$$

is equivalent to

$$\sup_{(a,p):\mathcal{T}\times\{\xi'\}\to\Delta(\mathcal{B}\times\mathcal{Z})\times\mathbb{R}} \mathbb{E}_t\Big[p(t,\xi') - C(a(t,\xi'))\Big] \tag{A.8}$$

subject to that *(i)* for all $t,\hat{t}$

$$\sum_{i,j}\alpha_{ij}(t,\xi')v(b_i,t) - p(t,\xi') \geqslant \sum_{i,j}\alpha_{ij}(\hat{t},\xi')v(b_i,t) - p(\hat{t},\xi'),$$

and that *(ii)* for all t

$$\sum_{i,j}\alpha_{ij}(t,\xi')v(b_i,t) - p(t,\xi') \geqslant 0.$$

This is because any $(a,p)\in\widehat{\mathcal{M}}$ satisfies the no horizontal distortion condition, all IR constraints, and the IC constraints of the form $(t,\xi)\to(\hat{t},\xi)$.

However, note that problem (A.8) is *identical* for every ξ' up to relabeling. Therefore, for every ξ', we must have that (A.8) is equivalent to

$$\sup_{(a,p):\mathcal{T}\to\Delta(\mathcal{B}\times\mathcal{Z})\times\mathbb{R}} \mathbb{E}_t\Big[p(t) - C(a(t))\Big] \tag{A.9}$$

subject to that *(i)* for all $t,\hat{t}$

$$\sum_i\Big(\sum_j\alpha_{ij}(t)\Big)v(b_i,t) - p(t) \geqslant \sum_i\Big(\sum_j\alpha_{ij}(\hat{t})\Big)v(b_i,t) - p(\hat{t}),$$

and that *(ii)* for all t

$$\sum_i\Big(\sum_j\alpha_{ij}(t)\Big)v(b_i,t) - p(t) \geqslant 0.$$

But then we may let $\beta_i(t) = \sum_j \alpha_{ij}(t)$, which reduces (A.9) to

$$\sup_{(\beta,p):\mathcal{T}\to\Delta(\mathcal{B})\times\mathbb{R}} \mathbb{E}_t\left[p(t) - C(\beta(t))\right] \tag{A.10}$$

subject to that *(i)* for all $t, \hat{t}$

$$v(\beta(t),t) - p(t) \geqslant v(\beta(\hat{t}),t) - p(\hat{t}),$$

and that *(ii)* for all t

$$v(\beta(t),t) - p(t) \geqslant 0.$$

Note that any feasible solution to (A.10) can be implemented as posting a menu of lotteries over bundles with consumers freely choosing their favorite horizontal attributes. Therefore,

$$\sup_{(\beta,p):\mathcal{T}\to\Delta(\mathcal{B})\times\mathbb{R}} \mathbb{E}_t\left[p(t) - C(\beta(t))\right] \leqslant \sup_{(a,p)\in\widetilde{\mathcal{M}}} \mathbb{E}_{t,\xi}\left[p(t,\xi) - C(a(t,\xi))\right].$$

But then

$$\begin{aligned}\sup_{(a,p)\in\widehat{\mathcal{M}}} \mathbb{E}\left[p(t,\xi) - C(a(t,\xi))\right] &\leqslant \mathbb{E}_\xi\left[\sup_{(a,p)\in\widehat{\mathcal{M}}} \mathbb{E}_t\left[p(t,\xi) - C(a(t,\xi))\right]\right]\\ &\leqslant \sup_{(a,p)\in\widetilde{\mathcal{M}}} \mathbb{E}\left[p(t,\xi) - C(a(t,\xi))\right],\end{aligned}$$

proving (A.7).

A.11.3 Completion of the Proof

By (A.5) and (A.7), we have that (A.4) must hold. This proves part *(i)* of Theorem 4.

By the proof of part *(i)*, if there exists an optimal solution (β, p) to (A.10), then we have found an optimal mechanism, by posting a menu of lotteries over the bundles described by (β, p) with consumers freely choosing their favorite horizontal attributes.

By Theorem 1, under the conditions in part *(ii)* of Theorem 4, there exists an optimal, deterministic solution to (A.10). In particular, the optimal solution is given by a menu of undominated bundles. The result follows.

A.12 Proof of Proposition 7

Fix the collection of continuous marginal distributions $\{\mu^b\}_b$. Let $\gamma^\dagger$ denote the unique comonotonic distribution with such marginals. For every joint distribution γ that is δ-positively-correlated and has marginals given by $\{\mu^b\}_b$, we have that $\gamma^\dagger$ and γ are δ-close.

Let $\mathcal{T} = [0,1]$ and $F(t)$ be the uniform distribution. For every bundle b, let

$$v(b,t) := P(b, 1-t),$$

where $P(b,q)$ is the single-bundle demand curve generated by marginal distribution μ_b. Note that $v(b,t) \sim \mu_b$ for all b. This implies that

$$\Big(v(b_1,t), \ldots, v(b_{2^n},t)\Big) \sim \gamma^\dagger,$$

since the comonotonic distribution is unique. Note that assumptions (A1') imply that for all $b_1 \subset b_2$, $v(b_2,t) - v(b_1,t)$ is strictly increasing in t. Thus, we have that assumptions (A1), (A2), and the nesting condition hold for $\{v(b,t), F(t), C(b)\}$.

Therefore, by Theorem 1, the optimal mechanism for the joint distribution $\gamma^\dagger$ is given by a menu of undominated bundles. Let (a,p) be the optimal mechanism when the joint distribution is $\gamma^\dagger$. In particular, we know that (a,p) is deterministic and $\{a(v)\}_v \subseteq B$ where B is the menu of undominated bundles.

Now, we make use of the following result:

Lemma 11 (Madarász and Prat 2017, Carroll 2017)**.** *For any $\varepsilon > 0$, there exists $\delta > 0$ such that for any mechanism (a,p), there exists another mechanism $(\tilde{a},\tilde{p})$ such that (i) for any two δ-close distributions γ,γ',*

$$\mathbb{E}_{\gamma'}[\tilde{p}(v) - C(\tilde{a}(v))] > \mathbb{E}_{\gamma}[p(v) - C(a(v))] - \varepsilon,$$

and that (ii)

$$\{\tilde{a}(v)\}_v \subseteq \text{Closure}(\{a(v)\}_v).$$

Fix $\varepsilon' = \frac{\varepsilon}{2}$ and let $\delta > 0$ and $(\tilde{a},\tilde{p})$ be given by Lemma 11 against ε'. Note that by Lemma 11, $(\tilde{a},\tilde{p})$ is also deterministic and assigns every type an undominated bundle.

Now, fix any γ that is δ-close to $\gamma^\dagger$. By Lemma 11 and the construction of (a,p), we have

$$\mathbb{E}_{\gamma}\Big[\tilde{p}(v) - C(\tilde{a}(v))\Big] > \mathbb{E}_{\gamma^\dagger}\Big[p(v) - C(a(v))\Big] - \frac{\varepsilon}{2} = \sup_{(a',p')\in\mathcal{M}} \mathbb{E}_{\gamma^\dagger}\Big[p'(v) - C(a'(v))\Big] - \frac{\varepsilon}{2}. \quad \text{(A.11)}$$

Fix any mechanism (a'', p''). We apply Lemma 11 again to get $(\tilde{a}'', \tilde{p}'')$ such that

$$\mathbb{E}_{\gamma^\dagger}\left[\tilde{p}''(v) - C(\tilde{a}''(v))\right] > \mathbb{E}_{\gamma}\left[p''(v) - C(a''(v))\right] - \frac{\varepsilon}{2}.$$

Since this holds for all (a'', p''), we have

$$\sup_{(a',p')\in\mathcal{M}} \mathbb{E}_{\gamma^\dagger}\left[p'(v) - C(a'(v))\right] \geqslant \sup_{(a'',p'')\in\mathcal{M}} \mathbb{E}_{\gamma}\left[p''(v) - C(a''(v))\right] - \frac{\varepsilon}{2}. \tag{A.12}$$

But then combining (A.11) and (A.12), we have

$$\mathbb{E}_{\gamma}\left[\tilde{p}(v) - C(\tilde{a}(v))\right] > \sup_{(a',p')\in\mathcal{M}} \mathbb{E}_{\gamma^\dagger}\left[p'(v) - C(a'(v))\right] - \frac{\varepsilon}{2} \geqslant \sup_{(a'',p'')\in\mathcal{M}} \mathbb{E}_{\gamma}\left[p''(v) - C(a''(v))\right] - \varepsilon.$$

Since the above holds for any joint distribution γ that is δ-close to $\gamma^\dagger$, the result follows.

A.13 Proof of Theorem 5

We follow the same proof strategy as in Theorem 1 and Proposition 1. In particular, we apply the monotone construction theorem (with the local single-crossing property), Theorem 6, to the partially ordered set $(\mathcal{B}, \subseteq)$ and virtual surplus function $\phi(b,t)$.

The result follows if we show that the strongly undominated bundles are exactly the chain-essential elements in Theorem 6 (see the proof of Theorem 1 and Proposition 1). We follow the notation as in the proof of Theorem 1. Because by assumption $Q(b) \in (0,1)$, we have that $t(b) \in (\underline{t}, \overline{t})$ for all $b \neq \varnothing$. Therefore, for any $\varnothing \subset b_1 \subset b_2$, by Lemma 3, we have

$$t(b_2 \mid b_1) \geqslant \min\{t(b_1), t(b_2)\} > 0.$$

Thus, for any $b_1 \subset b_2$, we have $t(b_2 \mid b_1) > 0$.

Fix any strongly undominated bundle b. We show that b is chain essential. If $b = \varnothing$ or $b = \overline{b}$, then we have that b is a chain-essential element because *(i)* $\underline{t} < t(b')$ for all $b' \supset \varnothing$ and *(ii)* $t(\overline{b} \mid b') < \overline{t}$ for all $b' \subset \overline{b}$ (see the proof of Proposition 3). Thus, suppose that $\varnothing \subset b \subset \overline{b}$. Suppose for contradiction that b is not chain essential. Then there exists $b_1 \subset b \subset b_2$ such that

$$t(b \mid b_1) \geqslant t(b_2 \mid b).$$

Since $t(b_2 \mid b_1) > 0$, by Lemma 6, this implies

$$t(b \mid b_1) \geqslant t(b_2 \mid b_1).$$

Now, by Lemma 3 and the proof of Proposition 1, we have

$$Q(b \mid b_1) \leqslant Q(b_2 \mid b_1),$$

contradicting to that b is strongly undominated.

Now, fix any chain-essential element b. We show that b is strongly undominated. Note that, for all $b_1 \subset b \subset b_2$, by the definition of chain-essential elements, we have

$$t(b \mid b_1) < t(b_2 \mid b).$$

Since $t(b_2 \mid b_1) > 0$, by Lemma 6, this implies that

$$t(b \mid b_1) < t(b_2 \mid b_1).$$

Then, by Lemma 3 and the proof of Proposition 1, we have

$$Q(b \mid b_1) > Q(b_2 \mid b_1).$$

Since this holds for all $b_1 \subset b \subset b_2$, we have that b is strongly undominated.

A.14 Proof of Proposition 8

Since costs are zero, the price elasticity for any bundle b at quantity $Q(b)$ must satisfy

$$\eta(b, Q(b)) = -1.$$

Recall that

$$\mathrm{MR}(b,q) = P(b,q)\big[1 + \frac{1}{\eta(b,q)}\big].$$

This implies that the elasticity curve $\eta(b,\cdot)$ single-crosses -1 from below because the MR curve $\mathrm{MR}(b,\cdot)$ single-crosses 0 from above.

We claim that, under zero costs, the union elasticity condition implies the union quantity condition. Indeed, under zero costs and the union elasticity condition, for any b_1, b_2, because

$$\eta\big(b_1 \cup b_2, Q(b_1 \cup b_2)\big) = -1,$$

we have

$$\eta\big(b_1, Q(b_1 \cup b_2)\big) \geqslant -1 \quad \text{or} \quad \eta\big(b_2, Q(b_1 \cup b_2)\big) \geqslant -1,$$

and hence

$$Q(b_1 \cup b_2) \geqslant \min\left\{Q(b_1), Q(b_2)\right\}.$$

Thus, the union quantity condition holds. Thus, the nesting condition holds by Proposition 2.

A.15 Proof of Proposition 9

Let B be the proposed menu. By Proposition 8, the nesting condition holds. Hence, by Theorem 1, it suffices to show that any (non-empty) bundle $b \notin B$ is dominated. We start by showing that for all i, we have

$$Q(b_1^\star \cup \cdots \cup b_i^\star) \geqslant Q(b_i^\star).$$

We prove this by induction on i. The base case $i = 1$ is trivial. For the inductive step, suppose that the claim holds for $i-1$. Now, observe that

$$Q(b_1^\star \cup \cdots \cup b_i^\star) \geqslant \min\left\{Q(b_1^\star \cup \cdots \cup b_{i-1}^\star), Q(b_i^\star)\right\} \geqslant \min\left\{Q(b_{i-1}^\star), Q(b_i^\star)\right\} = Q(b_i^\star),$$

where *(i)* the first inequality follows from that the union elasticity condition implies the union quantity condition (as shown in the proof of Proposition 8), *(ii)* the second inequality follows from the inductive hypothesis, and *(iii)* the last equality follows from the definition of $b_i^\star$ and $b_{i-1}^\star$. This proves the inductive step.

Now, fix any $b \notin B$. There exists some index j such that $b = b_j^\star$. Since $b \notin B$, we have

$$b = b_j^\star \subset b_1^\star \cup \cdots \cup b_j^\star.$$

But, by the previous step, we also have

$$Q(b) = Q(b_j^\star) \leqslant Q(b_1^\star \cup \cdots \cup b_j^\star).$$

Thus, bundle b is dominated, completing the proof.

A.16 Proof of Proposition 10

Without loss of generality, suppose that there is a sequence of demand rotations for item 2. By Proposition 9, nested bundling is always optimal at any parameter s. By Proposition 1, the minimal optimal menu $B^{OPT}(s)$ equals the set of undominated bundles.

To prove claim *(i)*, observe that it suffices to show that if $B^{OPT}(s) = \{\{1\},\{1,2\}\}$, then $B^{OPT}(s')$ must be $\{\{1\},\{1,2\}\}$ for any $s < s'$. Suppose not. Then, for some $s < s'$, we have

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

which is impossible by our notion of demand rotations.

To prove claim *(ii)*, observe that it suffices to show that if $B^{OPT}(s') = \{\{2\},\{1,2\}\}$, then $B^{OPT}(s)$ must be $\{\{2\},\{1,2\}\}$ for any $s < s'$. Suppose not. Then, for some $s < s'$, we have

$$Q(\{2\};s) \leqslant Q(\{1,2\};s), \qquad Q(\{2\};s') > Q(\{1,2\};s'),$$

which is impossible by our notion of demand rotations.

To prove claim *(iii)*, observe that it suffices to show that it cannot be $|B^{OPT}(s)| = 1, |B^{OPT}(s')| = 2, |B^{OPT}(s'')| = 1$ for any $s < s' < s''$. To see why this is impossible, note that: if $B^{OPT}(s') = \{\{1\},\{1,2\}\}$, then $r_2(B^{OPT}(\cdot))$ cannot be nondecreasing, contradicting claim *(i)*; if $B^{OPT}(s') = \{\{2\},\{1,2\}\}$, then $r_1(B^{OPT}(\cdot))$ cannot be nonincreasing, contradicting claim *(ii)*.

A.17 Proof of Proposition 11

By Theorem 1, it suffices to show that if $x^\dagger \notin X^\star$, then $x^\dagger$ is dominated by another quality level $x \neq x^\dagger$. Fix any $x^\dagger \notin X^\star$. Then $\widehat{Q}(x^\dagger) > Q(x^\dagger)$. Suppose, for contradiction, that there does not exist any $x > x^\dagger$ such that $Q(x) \geqslant Q(x^\dagger)$. Then, we have

$$\max_{x > x^\dagger} \{\widehat{Q}(x)\} = \max_{x > x^\dagger} \{Q(x)\} < Q(x^\dagger).$$

Let

$$\widetilde{Q}(x) := \begin{cases} \widehat{Q}(x) & \text{if } x \neq x^\dagger\,; \\ Q(x) & \text{otherwise}\,. \end{cases}$$

Then note that $\widetilde{Q}(x)$ is also nonincreasing and everywhere above $Q(x)$. Moreover, $\widetilde{Q}(x)$ is everywhere below $\widehat{Q}(x)$ with $\widetilde{Q}(x^\dagger) < \widehat{Q}(x^\dagger)$. Contradiction.

A.18 Proof of Proposition 12

Note that since $Q(x) \in (0,1)$, we must have

$$\mathrm{MR}(Q(x)) \cdot x - C(x) = 0\,,$$

where the quality-adjusted marginal revenue curve $\mathrm{MR}(q) := \frac{\mathrm{d}}{\mathrm{d}q}(F^{-1}(1-q)\cdot q)$ is a strictly decreasing, continuous function. Therefore, we have

$$Q(x) = \mathrm{MR}^{-1}\Big(C_{avg}(x)\Big).$$

Now, observe that for any function $h : \mathcal{X} \to \mathbb{R}$ and any strictly decreasing function $\Phi : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbf{U}^{-}\big[\Phi \circ h\big] = \Phi \circ \mathbf{L}^{+}\big[h\big],$$

where $\mathbf{U}^{-}[\,\cdot\,]$ denotes the upper decreasing envelope operator and $\mathbf{L}^{+}[\,\cdot\,]$ denotes the lower increasing envelope operator. Thus, we have

$$\widehat{Q}(x) = \mathrm{MR}^{-1}\Big(\breve{C}_{avg}(x)\Big)$$

because $\mathrm{MR}^{-1}(\,\cdot\,)$ is strictly decreasing. The claim follows from Proposition 11.

A.19 Proof of Proposition 13

We map this problem into a bundling problem as follows. Consider a bundling problem with $n+m$ many items, where the first n items represent ***quality upgrades*** exactly as in Section 7.2, and the remaining m items represent the ***passes to avoid*** each of the m costly activities. Specifically, for any (x_i, y_j), we define

$$v\Big(\big\{1,\ldots,i\big\} \cup \{n+1,\ldots,n+m\big\}\backslash\big\{n+j\big\},t\Big) := u(x_i,t) - c(y_j,t),$$

with $v\big(\big\{1,\ldots,i\big\} \cup \{n+1,\ldots,n+m\big\},t\big) := u(x_i,t)$ being the value of quality x_i without any costly action. We can map the production costs accordingly and let $v(b,t) = C(b) = 0$ for bundles b that are not of the above form. With a slight abuse of notation, we also write (x,y) as the bundle of quality x and costly action y, and write $Q(x,y)$ as the corresponding sold-alone quantity for this *damaged* bundle, i.e., the unique quantity maximizing the profit function $\pi(x,q) - \pi(y,q)$.

($\Longleftarrow$) Suppose that $\min_{y>0} Q(y) < \max_{x>0} Q(x)$. Suppose for contradiction that there exists an optimal deterministic mechanism that does not use any costly instruments. Then, by Proposition 11, there exists an optimal menu B such that

$$x^{\star} := \max\Big\{\operatorname*{argmax}_{x>0} Q(x)\Big\}$$

is the base-tier quality in menu B. Let $y^{\star} := \min\big\{\operatorname{argmin}_{y>0} Q(y)\big\}$. By assumption,

$Q(y^\star) < Q(x^\star)$. Because $\pi(x^\star, q)$, $\pi(y^\star, q)$, and $\pi(x^\star, q) - \pi(y^\star, q)$ are strictly quasiconcave, we have

$$Q(y^\star) < Q(x^\star) \implies Q(y^\star) < Q(x^\star) < Q(x^\star, y^\star).$$

But this implies that

$$t(x^\star, y^\star) < t(x^\star),$$

where, as in the proof of Theorem 1, $t(\cdot)$ denotes the type at which the associated virtual surplus function crosses zero.[28] By Proposition 11 and the construction of Theorem 1, all types below $t(x^\star)$ consumes $\varnothing$ under the optimal mechanism. However, consider the perturbation of assigning the types $s \in [t(x^\star, y^\star), t(x^\star))$ the damaged bundle $(x^\star, y^\star)$. Because

$$u(x^\star, t) - (u(x^\star, t) - c(y^\star, t)) = c(y^\star, t)$$

is strictly increasing in t, there exist prices to implement this change of the allocation. This change must increase the total profit by Lemma 1 since the virtual surplus function associated with bundle $(x^\star, y^\star)$ is strictly positive for all types $s > t(x^\star, y^\star)$. Contradiction.

($\Longrightarrow$) Suppose that $\min_{y>0} Q(y) \geqslant \max_{x>0} Q(x)$. We show that there exists an optimal mechanism that does not use any costly instruments. Note that for all $x' > 0$ and $y' > 0$,

$$Q(y') \geqslant \min_{y>0} Q(y) \geqslant \max_{x>0} Q(x) \geqslant Q(x').$$

Because $\pi(x', q)$, $\pi(y', q)$, and $\pi(x', q) - \pi(y', q)$ are strictly quasiconcave, this implies that

$$Q(y') \geqslant Q(x') \geqslant Q(x', y').$$

Therefore, each damaged bundle (x', y') is dominated by the undamaged version x'. Now, note that, by the quasiconcavity assumptions, *(i)* the virtual surplus function $\phi(x', t)$ single-crosses $\phi(x', t) - \phi(y', t)$ from below, *(ii)* both virtual surplus functions $\phi(x', t)$ and $\phi(x', t) - \phi(y', t)$ single-crosses 0 from below. Therefore, by the proof of Theorem 1, we have that $\phi(x', t) - \phi(y', t) \leqslant \max\{\phi(x', t), 0\}$ for all $t \in \mathcal{T}$.

Then, by Lemma 1, the optimal value of this screening problem is bounded from above by

$$\mathbb{E}\Big[\max_{x \in \mathcal{X}} \phi(x, t)\Big].$$

Since assumptions (A1) and (A2) hold for $\{u(x,t), C(x), F(t)\}$, by the proof of Theorem 1,

[28]For any bundle (x, y), the associated virtual function is given by $u(x,t) - C(x) - c(y,t) - \frac{1-F(t)}{f(t)}\big(u_t(x,t) - c_t(y,t)\big) = \phi(x,t) - \phi(y,t)$ where $\phi(x,t) := \frac{\mathrm{d}}{\mathrm{d}q}\pi(x,q)\,|_{q=1-F(t)}$ and $\phi(y,t) := \frac{\mathrm{d}}{\mathrm{d}q}\pi(y,q)\,|_{q=1-F(t)}$.

	t_1	t_2	t_3
$\{1,2\}$	4	12	25
$\{1\}$	2	6	8
$\{2\}$	3	3	6

(a) Nesting condition fails

	t_1	t_2	t_3
$\{1,2\}$	4	12	20
$\{1\}$	2	6	8
$\{2\}$	3	3	6

(b) Nesting condition holds

Table 2: Bundle values by types for Example 3. Circled are *sold-alone* monopoly prices. In case *(a)*, the nesting condition fails and the optimal mechanism is stochastic. In case *(b)*, the nesting condition holds, and the optimal mechanism is deterministic and given by the menu of undominated bundles.

the seller can attain the above profit by selling a deterministic menu of different qualities. Thus, costly screening is suboptimal.

B Additional Example

We provide an additional example that shows totally ordered types are not sufficient for the optimality of nested bundling. This example further illustrates our nesting condition. For simplicity, the example is discrete, but it can be made continuous by approximation.

Example 3. Suppose that there are two items $\{1,2\}$ and three types of consumers $\{t_1, t_2, t_3\}$ with mass $1/3$ each. Suppose that the costs are zero. We consider two cases.

Case *(a)*. The values are given by Table 2a. One can verify that the sold-alone quantities are given by $Q(\{1,2\}) = 1/3$, $Q(\{1\}) = 2/3$, and $Q(\{2\}) = 1$ (the sold-alone prices are circled in Table 2a). Thus, none of the bundles are dominated. So the nesting condition fails. Note that the nested menu $\{\{1\},\{1,2\}\}$ yields a profit $29/3$ (by pricing $\{1,2\}$ at 23, and $\{1\}$ at 6), and the nested menu $\{\{2\},\{1,2\}\}$ yields a profit $28/3$ (by pricing $\{1,2\}$ at 22, and $\{2\}$ at 3). The optimal deterministic menu in this case is not nested: it prices the bundle $\{1,2\}$ at 22, $\{1\}$ at 6, and $\{2\}$ at 3, which results in a profit

$$\frac{1}{3} \times \big(22 + 6 + 3\big) = \frac{31}{3} > \frac{29}{3}.$$

Moreover, the fully optimal mechanism is stochastic:

- price 22 for bundle $\{1,2\}$
- price $72/11$ for a lottery that puts probability $10/11$ on bundle $\{1\}$ and probability $1/11$ on bundle $\{1,2\}$

- price 3 for bundle $\{2\}$

which yields a profit

$$\frac{1}{3} \times \left(22 + \frac{72}{11} + 3\right) = \frac{347}{33} > \frac{31}{3}.$$

The suboptimality of nested bundling can be understood using Corollary 4 since the nested menu $\big\{\{2\},\{1,2\}\big\}$ that includes the best-selling bundle $\{2\}$ yields a strictly lower profit than the other nested menu $\big\{\{1\},\{1,2\}\big\}$.

Case *(b)*. The values are given by Table 2b, which is exactly the same as Table 2a except that type t_3's value for bundle $\{1,2\}$ is lowered from 25 to 20. Given this change, when bundle $\{1,2\}$ is sold alone, the monopoly price would be 12 and the quantity $Q(\{1,2\})$ would be 2/3 rather than 1/3. Then, bundle $\{1\}$ is dominated, and the nesting condition holds. The optimal mechanism in this case is deterministic and given by the nested menu $\big\{\{2\},\{1,2\}\big\}$, which coincides with the undominated bundles. The optimality of nested bundling and the construction of optimal menu follow directly from Theorem 1 and Proposition 1.