

When Is Pure Bundling Optimal?*

Nima Haghpanah[†]

Jason Hartline[‡]

August 28, 2020

Abstract

We study when pure bundling, i.e., offering only the grand bundle of all products, is optimal for a multi-product monopolist. Pure bundling is optimal if consumers with higher values for the grand bundle have higher relative values for smaller bundles compared to the grand bundle. Conversely, pure bundling is not optimal if consumers with higher values for the grand bundle have lower relative values. We prove the results by decomposing the problem into simpler ones in which types can be ranked according to their values for the grand bundle.

*We thank Nageeb Ali, Mark Armstrong, Ben Brooks, Gabriel Carroll, Robert Kleinberg, Vijay Krishna, Alexey Kushnir, Preston McAfee, Stephen Morris, Henrique Oliveira, Marco Scarcini, Ilya Segal, Dan Vincent, Rakesh Vohra, Jidong Zhou, and various seminar participants. We thank Berk Idem, Xiao Lin, and Garima Singal for excellent research assistance. We are particularly grateful to Ron Siegel for numerous discussions about the paper. This paper reinterprets and improves on some of the results that were previously in manuscripts titled “Reverse Mechanism Design” and “Multi-dimensional Virtual Values and Second-degree Price Discrimination”.

[†]Department of Economics, Pennsylvania State University, nima.haghpanah@gmail.com.

[‡]Department of Computer Science, Northwestern University, hartline@eecs.northwestern.edu.

1 Introduction

What is a multi-product monopolist’s optimal selling strategy? This is a classical economic question of importance for both positive and normative analysis, dating back to Stigler (1963) and Adams and Yellen (1976). We characterize when pure bundling, i.e., offering only the grand bundle of all products, is the optimal selling strategy. Our characterization is easy to state and has a straightforward intuition.

Consider a monopolistic seller of products 1 to n , and a buyer who needs at most one unit of each product. Assume that production costs are zero. The buyer’s privately known type t identifies a value $v(b, t)$ for each bundle of products $b \subseteq \{1, \dots, n\}$, and is drawn from a distribution. To maximize expected profit, should the seller use pure bundling and offer only the grand bundle of all products $b^* = \{1, \dots, n\}$? Or should she use more complex strategies such as offering a menu that includes multiple bundles at possibly different prices?

The optimality of pure bundling depends on how the buyer’s *relative values*, the ratio $v(b, t)/v(b^*, t)$ for each bundle b , change with the buyer’s value for the grand bundle $v(b^*, t)$. Pure bundling is optimal if relative values are first-order stochastically non-decreasing in the value for the grand bundle; i.e., types with higher values for the grand bundle are more likely to have higher relative values. Conversely, pure bundling is not optimal if relative values are first-order stochastically decreasing in the value for the grand bundle.

The characterization has a straightforward economic intuition. Let us compare the profit from selling only the grand bundle at some price p , to the profit from a “mixed bundling” strategy of selling a smaller bundle at a discounted price in addition to the grand bundle at the full price p . Mixed bundling has a gain and a loss compared to pure bundling. The gain is from selling to more types, i.e., types who are unwilling to pay the full price for the grand bundle but are willing to take the discounted offer. The loss is from types whose demand is diverted from the full price to the discounted offer. These types have high value for the grand bundle and high relative value for the smaller bundle (so that they find the discounted offer attractive). The loss is larger than the gain if types with higher values for the grand bundle are more likely to have high relative values, and is smaller if such types are more likely to have low relative values.

To see the interpretation of our results, consider selling a car that can be customized with an audio system. Mapped into our model, there are two products, the “basic” car and the audio system, and three non-empty bundles. The value of the bundle that contains only the audio system is zero. Suppose that consumers differ across two dimensions. First, whereas some consumers need cars to get to work, others can take public transit. The former

group have higher values for all bundles than the latter. Second, some consumers care more about listening to music while driving than do others. For the former group, the bundle that contains only the basic car has a low value relative to the grand bundle. Our results say that pure bundling is optimal (respectively not optimal) if consumers who need cars to get to work tend to care less (respectively more) about listening to music while driving.

The literature on multi-product bundling mostly assumes that values are additive, i.e., the value of any bundle is the sum of the values of its constituting products. A conventional view in the literature is that bundling is profitable if the values of individual products are negatively correlated. This view is mostly based on examples first provided by Stigler (1963) and Adams and Yellen (1976) where values are perfectly negatively correlated, i.e., the sum of values is constant for all types. When the sum of values is constant, pure bundling is optimal since it extracts the full surplus. Based on this intuition, Bakos and Brynjolfsson (1999) and Armstrong (1999) show that pure bundling is optimal if there is a large number of products with independently distributed values, since the value of the grand bundle concentrates by the law of large numbers. It is unclear from these results whether there are general conditions for the optimality of pure bundling beyond settings where the value of the grand bundle is concentrated and pure bundling extracts the full surplus. We provide such a condition. Our condition is orthogonal to how the value of the grand bundle is distributed, and instead only concerns the conditional distribution of relative values. Pure bundling may be optimal regardless of whether or not it extracts the full surplus.

With non-additive values, products may be partial substitutes or partial complements. They may even be partial substitutes for some types but partial complements for others. In fact, relative values measure the degree of complementarity between products for a given type. A type with higher relative values for all bundles considers the products to be less complementary. Our result can be re-interpreted to say that pure bundling is optimal (respectively not optimal) if types with a higher value for the grand bundle consider the products to be less (respectively more) complementary. Thus the optimality of pure bundling depends on how the degree of complementarity changes across types. Pure bundling may be optimal even if all types consider the products to be partial substitutes. We provide detailed discussion and examples in Section 6.

Our results extend to accommodate multi-unit demands and production costs. If the buyer demands multiple units of a product, we can add multiple copies of the product to the set of products and let the values of these copies be identical. With positive production costs, and assuming that the grand bundle maximizes the surplus for each type, pure bundling is

optimal if the surplus of all bundles relative to the surplus of the grand bundle stochastically increase in the surplus of the grand bundle.

1.1 Our Methodology

Our approach for proving the optimality of pure bundling consists of two components. The first component is to prove the result assuming that types are on a “path”, that is, types have distinct values for the grand bundle. In this case, pure bundling is optimal if relative values are non-decreasing in the value of the grand bundle. The proof is based on a formulation of virtual valuations that generalizes that of Myerson (1981). Assuming usual regularity conditions, the analysis is a simple extension of the standard envelope analysis. The proof without regularity assumptions constructs ironed virtual valuations, building on a duality approach from Carroll (2017) and Cai et al. (2019). The construction is novel and shows that ironed virtual valuations can be constructed from only downward incentive constraints (that is, upward constraints do not bind).

The second component of our approach is to extend the result to general type spaces. The idea is to decompose the type space into paths, and to show that pure bundling is the solution to a relaxed problem in which the seller can observe the path on which the type lies, and can design a mechanism accordingly. Since the seller can ignore this information, the revenue in the relaxed problem provides an upper bound to the optimal revenue in the original problem. We invoke a classical characterization from the statistics literature (Strassen, 1965; Kamae et al., 1977). In our setting, the characterization states that a decomposition into paths with monotone relative values exists if and only if the stochastic monotonicity condition of our main result holds. The first component of our approach then implies that pure bundling is optimal for each path.

1.2 Related Work

Optimal strategies for selling multiple products are complex. This is a common theme in the multi-product pricing literature. Pure bundling and selling products separately, i.e., offering each product at a price, are generally strictly dominated by mixed bundling, i.e., offering prices for all bundles (Adams and Yellen, 1976; McAfee et al., 1989). Mixed bundling is itself dominated by offering randomized bundles (Thanassoulis, 2004; Daskalakis et al., 2017). In fact, the optimal menu may offer infinitely many randomized bundles (Vincent and Manelli, 2007; Hart and Nisan, 2019). Finally, a seller may profit from selling damaged products

(Deneckere and McAfee, 1996).

In practice, however, sellers often use simple strategies such as offering only a few bundles. Our work is a step towards rationalizing such simple selling strategies. Pavlov (2011) and Menicucci et al. (2015) also provide sufficient conditions for the optimality of pure bundling. They study selling two products with additive and independently distributed values. Their conditions require the virtual valuation of each product to be positive on the entire support of values, and are not comparable to our conditions. Schmalensee (1984) shows mainly via numerical results that pure bundling is more profitable than separate sales when values have a Gaussian distribution and are negatively correlated. Fang and Norman (2006) provide a partial characterization of when pure bundling is more profitable than separate sales, analytically confirming the numerical results of Schmalensee (1984).

Relatively few papers on multi-product bundling allow for non-additive values. Some exceptions are Long (1984), Armstrong (2013), and Armstrong (2016), where the focus is on identifying when it is profitable for the seller to offer the bundle at a price that is less than the sum of the prices for individual products. Krishna and Rosenthal (1996) and Rosenthal and Wang (1996) characterize equilibria of simultaneous auctions for products that are partial complements.

Decomposition approaches related to ours appear in the mechanism design literature. Wilson (1993) and Armstrong (1996) use a fixed decomposition. Translated to our setting, each path is a ray from the origin in the value space. Esó and Szentes (2007) and Pavan et al. (2014) significantly advance this idea by allowing the decomposition to depend on the distribution of types. Esó and Szentes (2007) study an auctioneer who can provide information about a product to buyers. Their decomposition is the same as the one provided by Strassen (1965) and Kamae et al. (1977) in the case of a single dimension. Pavan et al. (2014) study optimal dynamic mechanism design. They decompose the type space into a base parameter and independently distributed “shocks”, and provide conditions for the solution to the decomposed problem to satisfy the global incentive constraints.

2 Model and Main Result

Consider a screening problem with a single seller and a single buyer. There are n products. A bundle $b \subseteq \{1, \dots, n\}$ is a subset of products, and the set of all bundles is $B = \{b \mid b \subseteq \{1, \dots, n\}\}$. The cost of producing any bundle is zero. There is a compact set of buyer types T . The value of a bundle b for a type $t \in T$ is $v(b, t) \geq 0$. The value of the empty bundle \emptyset

for each type is normalized to zero. The grand bundle $b^* = \{1, \dots, n\}$ maximizes each type's value over all bundles. Type t has utility $v(b, t) - p$ from consuming a bundle b and paying p to the seller. We denote the expected value of a randomized bundle $a \in \Delta(B)$ for a type t by $v(a, t) = \mathbf{E}_{b \sim a}[v(b, t)]$. The buyer's type is his private information. The seller has a prior belief in the form of a distribution $\mu \in \Delta(T)$ over the types of the buyer.

We invoke the revelation principle and focus on direct mechanisms. A mechanism is a pair of functions, a (possibly randomized) bundle assignment rule $a : T \rightarrow \Delta(B)$ and a payment rule $p : T \rightarrow \mathbb{R}$. The mechanism (a, p) is incentive compatible (IC) if misreporting does not increase the utility of any type,

$$v(a(t), t) - p(t) \geq v(a(t'), t) - p(t'), \quad \forall t, t' \in T.$$

The mechanism is individually rational (IR) if it ensures voluntary participation

$$v(a(t), t) - p(t) \geq 0, \quad \forall t \in T.$$

An IC and IR mechanism is optimal if it maximizes the expected revenue

$$\mathbf{E} [p(t)]$$

over all IC and IR mechanisms.

A pure bundling mechanism offers only the grand bundle at some price p . That is, if $v(b^*, t) \geq p$ then $a(t) = b^*$ with probability 1 and $p(t) = p$, and otherwise $a(t) = \emptyset$ with probability 1 and $p(t) = 0$. Such a mechanism is IC and IR. We say that pure bundling is optimal if a pure bundling mechanism with some price p is optimal.

The Statement of the Main Result

Our main result specifies a condition for the optimality of pure bundling. Pure bundling is optimal if relative values are stochastically non-decreasing in the value of the grand bundle. Below we define these terms and formally give the main theorem.

For a bundle b and a type t , let $r(b, t) = v(b, t)/v(b^*, t) \in [0, 1]$ be the relative value of the bundle b to the grand bundle. For a set S , let \mathbb{R}^S denote the set of functions from S to \mathbb{R} . The *profile of relative values* for a type t is a function $r(\cdot, t) \in \mathbb{R}^B$ that maps each bundle to its relative value.

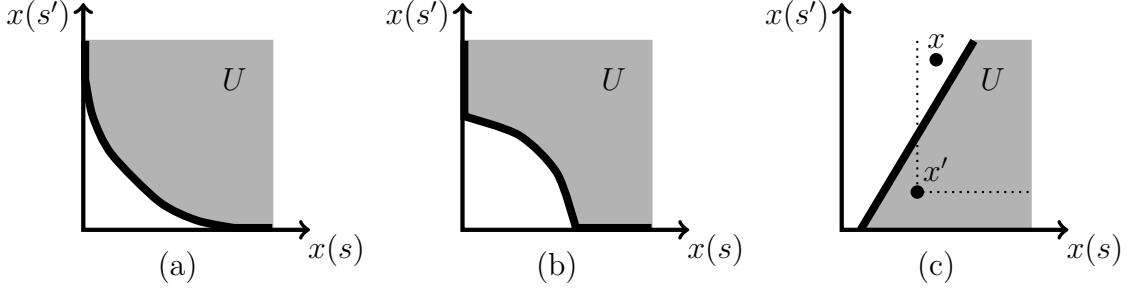


Figure 1: Sets $U \subseteq \mathbb{R}^{\{s,s'\}}$ in (a) and (b) are upper sets. The set $U \subseteq \mathbb{R}^{\{s,s'\}}$ in (c) is not an upper set.

The standard multivariate notion of first-order stochastic monotonicity (see, for example, Shaked and Shanthikumar, 2007) is stated in terms of upper sets of functions, defined next. In words, an upper set includes all functions that are larger than any function that it also includes. A function $x \in \mathbb{R}^S$ is larger than a function $x' \in \mathbb{R}^S$, denoted $x' \leq x$, if $x'(s) \leq x(s)$ for all $s \in S$. Some examples are shown in Figure 1.

Definition 1. A set $U \subseteq \mathbb{R}^S$ is an upper set if $x' \leq x$ and $x' \in U$ imply that $x \in U$.

A random variable x is stochastically non-decreasing in $y \in \mathbb{R}$ if conditioned on larger y , x is more likely to take on large values, specifically, any value in U for any upper set U .

Definition 2. A random variable $x \in \mathbb{R}^S$ is stochastically non-decreasing in a random variable $y \in \mathbb{R}$ if $\Pr[x \in U \mid y = \hat{y}]$ is non-decreasing in \hat{y} for all upper sets $U \subseteq \mathbb{R}^S$.

We now state our main result.

Theorem 1. Pure bundling is optimal if the profile of relative values is stochastically non-decreasing in positive values of the grand bundle; i.e., $\Pr[r(\cdot, t) \in U \mid v(b^*, t) = \hat{v}]$ is non-decreasing in $\hat{v} > 0$ for all upper sets $U \subseteq \mathbb{R}^B$.

Organization of the paper. The rest of the paper is organized as follows. To develop some intuition, Section 3 specializes the main result to the case of two types and corroborates it algebraically. Section 4 proves the main result and provides some partial converses. Section 5 and Section 6 discuss some extensions and interpretations. Section 7 concludes. Missing proofs are deferred to the appendix.

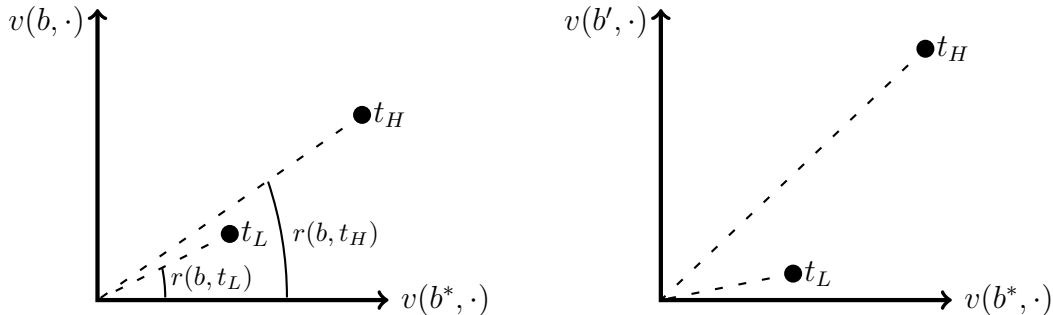


Figure 2: For a bundle b , $r(b, t)$ is the slope of the line that connects the origin to t in the graph that depicts the value for the grand bundle against the value for bundle b . The low type prefers b to b' , but the opposite holds for the high type.

3 Two Types

Suppose that there are only two types. If the two types have identical values for the grand bundle, then pure bundling extracts the full surplus and is optimal. In this case the stochastic monotonicity condition of the main result is trivially satisfied. Thus suppose that the types have distinct values for the grand bundle. Label the types t_L and t_H where $v(b^*, t_L) < v(b^*, t_H)$. Pure bundling is optimal if the low type t_L has lower relative values than the high type t_H does, that is, $r(b, t_L) \leq r(b, t_H)$ for all bundles b . This is illustrated in Figure 2. Notice that the two types may have different preferences over bundles. For instance, the low type may prefer bundle b to bundle b' even though the opposite may hold for the high type, as shown in Figure 2.

The following algebraic analysis corroborates the main result with two types, where it is assumed for simplicity that $v(b, t_H) - 1 \leq v(b, t_L) \leq v(b, t_H)$ for each bundle b , $v(b^*, t_L) = 1$, and $v(b^*, t_H) = 2$. Let $q_H = \Pr[t = t_H]$ denote the probability of the high type. The highest revenue among pure bundling mechanisms is $\max(1, 2q_H)$. Indeed, if the price of the grand bundle is 1, then both types buy the grand bundle and the revenue is 1. If the price of the grand bundle is 2, then only the high type buys the grand bundle, and the revenue is $2q_H$. Any other price results in revenue less than 1 or $2q_H$.

Now consider a “mixed bundling” mechanism that offers some bundle b at price $v(b, t_L)$ and the grand bundle at price $2 - (v(b, t_H) - v(b, t_L))$. Given such offers, and breaking ties to maximize revenue, the low type chooses bundle b , and the high type chooses the grand

bundle b^* . The revenue is

$$(1 - q_H)v(b, t_L) + q_H\left(2 - \left(v(b, t_H) - v(b, t_L)\right)\right) = v(b, t_L) + q_H\left(2 - v(b, t_H)\right).$$

We now show that the revenue of the mixed bundling mechanism is at most the highest revenue among pure bundling mechanisms if the relative values satisfy $r(b, t_L) \leq r(b, t_H)$, that is, the values satisfy $v(b, t_L) \leq \frac{1}{2}v(b, t_H)$. First suppose that $q_H \geq \frac{1}{2}$. The revenue of the mixed bundling mechanism is

$$\begin{aligned} v(b, t_L) + q_H\left(2 - v(b, t_H)\right) &= 2q_H + v(b, t_L) - q_Hv(b, t_H) \\ &\leq 2q_H + v(b, t_L) - \frac{1}{2}v(b, t_H) \\ &\leq 2q_H, \end{aligned}$$

which is the revenue of selling the grand bundle at price 2. Now suppose that $q_H \leq \frac{1}{2}$. The revenue of the mixed bundling mechanism is

$$\begin{aligned} v(b, t_L) + q_H\left(2 - v(b, t_H)\right) &\leq v(b, t_L) + \frac{1}{2}\left(2 - v(b, t_H)\right) \\ &= 1 + v(b, t_L) - \frac{1}{2}v(b, t_H) \\ &\leq 1, \end{aligned}$$

which is the revenue of selling the grand bundle at price 1.

The monotonicity of relative values is also partially necessary for the optimality of pure bundling. In particular, if the monotonicity condition is violated, then pure bundling is not optimal for *some* distribution over the two types. In fact, if the probability of the high type is $q_H = \frac{1}{2}$, then the revenue of the mixed bundling mechanism is $1 + v(b, t_L) - \frac{1}{2}v(b, t_H)$, which is larger than the optimal revenue among pure bundling mechanisms, i.e., 1, if $r(b, t_L) > r(b, t_H)$. We later generalize this observation to provide a partial converse to Theorem 1.

Notice that our algebraic analysis does not completely prove the optimality of pure bundling since it does not consider more complicated mechanisms that offer different sets of bundles or even randomizations over them. It nevertheless highlights the central role of relative values in the optimality of pure bundling.

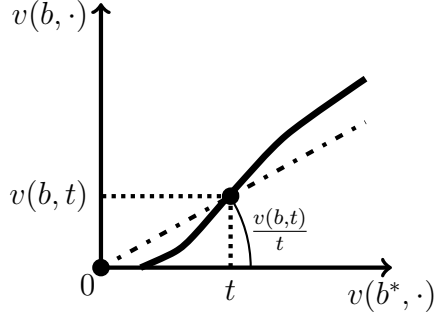


Figure 3: The function $v(b, \cdot)$ maps each type's value of the grand bundle to that type's value of bundle b . In this figure, the relative value $v(b, t)/t$ is non-decreasing in t .

4 Proof of Theorem 1 and Converses

In this section we prove Theorem 1, and provide two partial converses. We start by proving a special case of the theorem where types have distinct values for the grand bundle. We then show how the special case can be generalized to all distributions via a decomposition approach.

4.1 Paths: Distinct Values for Grand Bundle

Assume that types have distinct values for the grand bundle, that is $v(b^*, t) \neq v(b^*, t')$ for all $t \neq t'$ in the set of types T . Thus we assume without loss of generality that t equals the value of the grand bundle, that is $T \subseteq \mathbb{R}^+$ and $v(b^*, t) = t$.¹ The function $v(b, \cdot)$ maps each type's value of the grand bundle to that type's value of bundle b , as shown in Figure 3. We say that types are on “path” v .

The following proposition consists of two statements. The “if” statement is a special case of Theorem 1 when types are on a path. The special case of the condition of the main theorem is that $r(\cdot, t) = v(\cdot, t)/t$ is monotone non-decreasing in t , that is, $v(b, t)/t \leq v(b, t')/t'$ for all $t \leq t'$ and b . The “only if” statement is a partial converse to Theorem 1. It states that if types are on a path v but $v(b, t)/t$ is not monotone non-decreasing for some b , then pure bundling is not optimal for *some* distribution over types.

Proposition 1. *Assume that $T \subseteq \mathbb{R}^+$ and $v(b^*, t) = t$. Pure bundling is optimal for all distributions $\mu \in \Delta(T)$ if and only if $v(\cdot, t)/t$ is monotone non-decreasing in $t > 0$.*

¹In particular, given v and T , let $\hat{T} = \{v(b^*, t) \mid t \in T\}$, and $\hat{v}(\cdot, \hat{t}) = v(\cdot, v^{-1}(b^*, \hat{t}))$, where $v^{-1}(b^*, \hat{t})$ is a type t such that $v(b^*, t) = \hat{t}$. The inverse is well defined by the assumption values for the grand bundle are distinct. Notice that $\hat{T} \subseteq \mathbb{R}$ and $\hat{v}(b^*, \hat{t}) = \hat{t}$.

If $v(\cdot, t)/t$ is monotone non-decreasing in $t > 0$, we say that the path v is *ratio-monotone*. Geometrically, ratio-monotonicity requires that in the graph that plots the value of the grand bundle against the value of any bundle b , the slope of a ray from the origin to a type is non-decreasing along the support, as in Figure 3. Alternatively, a ray from the origin to any type is (weakly) above all lower types and (weakly) below all higher types. Ratio-monotonicity implies that types can be ranked so that the value of any bundle is lower for a lower type than for a higher type. Ratio-monotonicity does not imply any ranking on bundles. That is, it is possible that two types have opposite preferences over two bundles.

The proof of the “only if” statement is a direct extension of the two-type analysis in Section 3. Suppose that there exists a bundle b such that $v(b, t)/t$ is not monotone non-decreasing in $t > 0$. That is, there exist t, t' such that $0 < t < t'$ and $v(b, t)/t > v(b, t')/t'$. We show that there exists a distribution with support over types t and t' for which pure bundling is not optimal. In particular, let the probability of the low type t be $1 - t/t'$, and the probability of the high type t' be t/t' . The distribution is such that the seller is indifferent between selling the grand bundle at price t or t' . A “mixed bundling” mechanism that offers bundle b at a low price targeted at type t and the grand bundle at a high price targeted at type t' has higher revenue than any pure bundling mechanism.

We defer the full proof of the “if” statement to the appendix. Instead, we here provide a proof that follows the standard first order analysis and uses several assumptions. First, the marginal distribution of the value of the grand bundle is supported over an interval $[\underline{t}, \bar{t}]$, $\underline{t} > 0$ with strictly positive density. Second, $v(b, t)$ is differentiable in t for each bundle b . Third, the marginal distribution of the value of the grand bundle is regular, as defined next.

For each bundle b define $\partial_2 v(b, t) := \frac{d}{dt}v(b, t)$. Let F denote the cumulative marginal distribution of the value of the grand bundle t , and f its density. Define the virtual value of a type t for a bundle b as follows.

$$\phi(b, t) = v(b, t) - \partial_2 v(b, t) \times \frac{1 - F(t)}{f(t)}. \quad (1)$$

Define the virtual value of a randomized bundle $a \in \Delta(B)$ to be $\phi(a, t) = \mathbf{E}_{b \sim a}[\phi(b, t)]$. Recall that $v(b^*, t) = t$, which implies that $\phi(b^*, t) = t - \frac{1 - F(t)}{f(t)}$.² Note that $\phi(\emptyset, t) = 0$. We

²This is identical to the virtual value of Myerson (1981). In fact, $\phi(b, t) = v(b, t) - \frac{1 - F_b(v(b, t))}{f_b(v(b, t))}$ for any bundle b , where F_b is the cumulative marginal distribution of value of bundle b , and f_b its density. That is, the virtual value of each bundle b is equal to the virtual value of the projected distribution of values of b . This follows from Myerson (1981), since the special case of our setting where the seller can only produce bundle b is equivalent to the setting of Myerson, and his analysis applies. We use Equation (1) in our proof

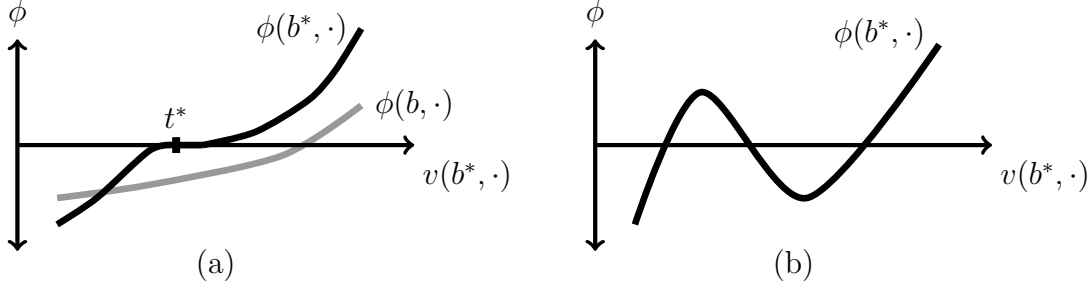


Figure 4: (a) If v is ratio-monotone, then $\phi(b, t) \leq \max(0, \phi(b^*, t))$ for all b, t . (b) The revenue of selling only the grand bundle at any price is strictly less than $\mathbf{E}[\max(0, \phi(b^*, t))]$.

say that the marginal distribution of the value of the grand bundle is regular if $\phi(b^*, t)$ is monotone non-decreasing in t .

The proof of Proposition 1 uses two lemmas. The first lemma is standard (e.g., Myerson, 1981) and applies the envelope theorem to relate revenue with virtual surplus. The expected revenue of any IC mechanism (a, p) is equal to its expected virtual surplus $\mathbf{E}[\phi(a(t), t)]$, up to a constant which is the utility of type \underline{t} .

Lemma 1. *For any incentive compatible mechanism (a, p) ,*

$$\mathbf{E}[p(t)] = \mathbf{E}[\phi(a(t), t)] - (v(a(\underline{t}), \underline{t}) - p(\underline{t})).$$

The second lemma follows directly from the definition of virtual values in Equation (1), and allows us to compare the virtual values of bundles b and b^* . If v is ratio-monotone, then for any type t , the virtual value of any bundle b is at most either 0 or the virtual value of the grand bundle, as shown in Figure 4, (a). Therefore, the virtual value of any randomized bundle a is also at most either 0 or the virtual value of the grand bundle.

Lemma 2. *If $v(\cdot, t)/t$ is monotone non-decreasing in t , then $\phi(a, t) \leq \max(0, \phi(b^*, t))$ for all randomized bundles $a \in \Delta(B)$ and types $t \in T$.*

We now use the above two lemmas to show that pure bundling is optimal if v is ratio-monotone. By Lemma 1, the revenue of any IC and IR mechanism (a, p) is

$$\begin{aligned} \mathbf{E}[\phi(a(t), t)] - (v(a(\underline{t}), \underline{t}) - p(\underline{t})) &\leq \mathbf{E}[\phi(a(t), t)] \\ &\leq \mathbf{E}[\max(0, \phi(b^*, t))], \end{aligned} \tag{2}$$

since it facilitates the comparison of virtual values based on the curvature of v .

where the first inequality follows from IR, and the second inequality follows from Lemma 2. Since $\phi(b^*, t)$ is monotone, there exists a threshold t^* such that $\phi(b^*, t) \leq 0$ for all $t \leq t^*$, and $\phi(b^*, t) \geq 0$ for all $t \geq t^*$. The revenue of selling only the grand bundle at price t^* is equal to $\mathbf{E}[\max(0, \phi(b^*, t))]$ by Lemma 1. Thus by Inequality (2), the revenue of selling only the grand bundle at price t^* is weakly higher than that of any IC and IR mechanism, and pure bundling is optimal.

The proof above does not work if the marginal distribution of the value of the grand bundle is not regular. Without regularity, $\mathbf{E}[\max(0, \phi(b^*, t))]$ is still an upper bound on the revenue of any mechanism since Lemma 2 and Inequality (2) require ratio-monotonicity of v but not monotonicity of $\phi(b^*, t)$. However, if $\phi(b^*, t)$ is not monotone, as in Figure 4, (b), then $\mathbf{E}[\max(0, \phi(b^*, t))]$ is also strictly higher than the revenue of any pure bundling mechanism. Indeed, any pure bundling mechanism must either sell the grand bundle to some type with a negative virtual value of the grand bundle, or exclude some type with positive virtual value. Thus $\mathbf{E}[\max(0, \phi(b^*, t))]$ cannot be used to argue that pure bundling obtains more revenue than all mechanisms. The general proof of Proposition 1 relies on an ironing technique. We defer the formal proof but discuss its geometric interpretation.

The proof is based on two observations. First, only “downward” IC constraints bind. In particular, if $t < t'$, then the IC constraint that corresponds to a deviation of type t to type t' does not bind. Second, each type can be assigned a virtual value function based on binding IC constraints. In particular, assuming that the type space is finite, the virtual value function $\hat{\phi}(\cdot, t)$ of each type t is

$$\hat{\phi}(\cdot, t) = v(\cdot, t) + \sum_{t': \text{IC from } t' \text{ to } t \text{ binds}} \lambda(t', t) (v(\cdot, t) - v(\cdot, t')),$$

where $\lambda(t', t) > 0$ is the Lagrangian multiplier for the IC constraint that corresponds to the deviation of type t' to type t . To see the geometric interpretation, suppose that the IC constraint of only a single type t' to type t binds. The virtual value function of type t is

$$\hat{\phi}(\cdot, t) = v(\cdot, t) + \lambda(t', t) (v(\cdot, t) - v(\cdot, t')).$$

Thus, viewed as vectors, $\hat{\phi}(\cdot, t)$ is obtained by moving $v(\cdot, t)$ away from $v(\cdot, t')$. By ratio-monotonicity, $v(\cdot, t')$ is above the ray that connects the origin to $v(\cdot, t)$, which implies that

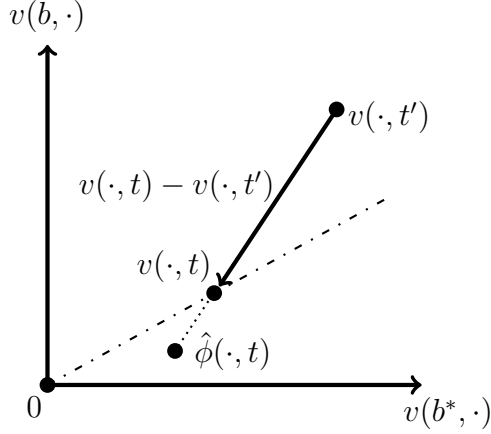


Figure 5: The IC constraint from type t' to type t binds. The virtual value profile of type t is defined by moving the values $v(\cdot, t)$ of type t away from the values $v(\cdot, t')$ of type t' .

$\hat{\phi}(\cdot, t)$ is below that ray, as shown in Figure 5. That is, for each bundle b ,

$$\hat{\phi}(b, t) \leq \frac{v(b, t)}{t} \cdot \hat{\phi}(b^*, t).$$

This property holds more generally if only downward IC constraints bind. Since the relative value of bundle b is at most 1, we have $\hat{\phi}(b, t) \leq \max(0, \hat{\phi}(b^*, t))$, and therefore virtual surplus is maximized by assigning either the outside option \emptyset or the grand bundle b^* to each type. We choose Lagrangian multipliers such that $\hat{\phi}(b^*, t)$ is monotone non-decreasing, and therefore pure bundling maximizes virtual surplus. The resulting virtual values for the grand bundle are equivalent to Myerson's ironed virtual values when values are drawn from the marginal distribution of the value for the grand bundle F .

4.2 Proof of Theorem 1: Non-distinct Values for Grand Bundle

Equipped with Proposition 1, we now prove Theorem 1, restated below.

Theorem 1. *Pure bundling is optimal if the profile of relative values is stochastically non-decreasing in positive values of the grand bundle; i.e., $\Pr[r(\cdot, t) \in U \mid v(b^*, t) = \hat{v}]$ is non-decreasing in $\hat{v} > 0$ for all upper sets $U \subseteq \mathbb{R}^B$.*

The proof is based on decomposing the distribution of types μ . Suppose that there exists a distribution over distributions of types $D \in \Delta(\Delta(T))$ such that $\mu = \mathbf{E}_{\mu' \sim D}[\mu']$. We refer to D as a *decomposition* of μ . A random type from μ can be drawn by first selecting a distribution over types $\mu' \in \Delta(T)$ from D , and then drawing a random type from μ' . Now

consider a modified problem where the seller has the power to observe μ' , and is allowed to select a mechanism accordingly. The optimal revenue in this modified problem is at least as high as in the original problem, since the seller can simply ignore the extra information and select the same mechanism for all μ' . If selling the grand bundle at price p is optimal for each μ' , then it is the optimal solution to the modified problem, and hence to the original problem.

Formally, we have the following lemma. The lemma holds for any mechanism (not necessarily pure bundling), although we only use it to prove the optimality of pure bundling.

Lemma 3. *Consider a decomposition D of μ . A mechanism (a, p) is optimal for distribution μ if it is optimal for all distributions $\mu' \in \Delta(T)$ in the support of D .*

Proof. The proof is by linearity of expectation. Suppose that a mechanism (a, p) is optimal for all μ' in the support of D . Consider any IC and IR mechanism (b', p') . We have

$$\mathbf{E}_{t \sim \mu} [p'(t)] = \mathbf{E}_{\mu' \sim D} [\mathbf{E}_{t \sim \mu'} [p'(t)]] \leq \mathbf{E}_{\mu' \sim D} [\mathbf{E}_{t \sim \mu'} [p(t)]] = \mathbf{E}_{t \sim \mu} [p(t)],$$

where the inequality followed from the optimality of (a, p) for μ' . □

To prove the main result, we construct a decomposition D of μ that satisfies two properties. First, pure bundling is optimal for every distribution μ' in the decomposition. By Proposition 1, it is sufficient that μ' is supported on a ratio-monotone path. Second, the marginal distribution of the value of the grand bundle is identical for all distributions μ' in the decomposition. This property ensures *the same* pure bundling mechanism is optimal for all μ' . We say that a decomposition D is a ratio-monotone decomposition if it satisfies both conditions. That is, all distributions μ' in the support of D are supported on ratio-monotone paths and have identical marginal distributions of the value of the grand bundle. By Lemma 3, if a ratio-monotone decomposition exists, then pure bundling is optimal.

A ratio-monotone decomposition exists if and only if μ satisfies the stochastic monotonicity condition of Theorem 1. The characterization simply invokes a classical result from the statistics literature which relates first-order stochastic dominance to the existence of a monotone coupling. That is, if a random variable stochastically dominates another, then the two random variables can be coupled such that the first one is greater than the second one with probability one.

To develop some intuition, consider the following example. There are four types $T = \{t_1, t'_1, t_2, t'_2\}$ with non-distinct values for the grand bundle, depicted in Figure 6. The value

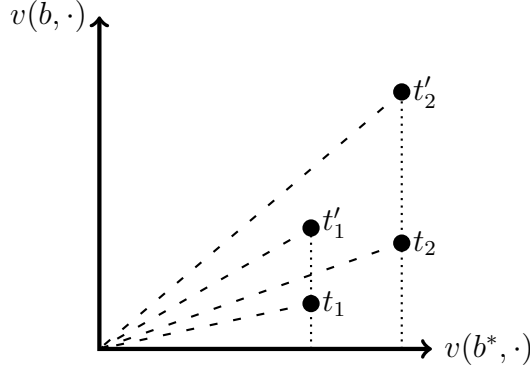


Figure 6: For each bundle b , relative values are ordered as $r(b, t_1) \leq r(b, t_2) \leq r(b, t'_1) \leq r(b, t'_2)$.

of the grand bundle is identical for types t_1 and t'_1 , and is lower than the value of the grand bundle for types t_2 and t'_2 , which is also identical. The relative values are ordered as

$$r(b, t_1) \leq r(b, t_2) \leq r(b, t'_1) \leq r(b, t'_2) \quad (3)$$

for each bundle b . Notice that the value of a bundle b may be lower for type t_2 than for type t'_1 , even though the value of the grand bundle is higher for type t_2 than for type t'_1 .

The stochastic monotonicity condition of Theorem 1 holds when the probability $q_1 = \Pr[t = t'_1 \mid t \in \{t_1, t'_1\}]$ that conditioned on a low value of the grand bundle, the value of the other bundles are high, is at most the probability $q_2 = \Pr[t = t'_2 \mid t \in \{t_2, t'_2\}]$ that conditioned on a high value of the grand bundle, the value of the other bundles are high. To see this, consider any upper set $U \subseteq \mathbb{R}^B$. Given the ranking of relative values and the definition of an upper set, if the set includes the relative values of t_1 , or respectively t'_1 , then it also includes the relative values of t_2 , respectively t'_2 . Now consider three cases. First, if the upper set includes the relative values of both t_1 and t'_1 , then it also includes the relative values of both t_2 and t'_2 . Thus, conditioned on either a low or a high value of the grand bundle, relative values are in the set with probability 1 and monotonicity is trivially satisfied. Second, if the upper set includes the relative values of t'_1 but not t_1 , then it also includes the relative values of t'_2 (and perhaps t_2). Therefore, conditioned on a low or a high value of the grand bundle, relative values are in the set with probability q_1 or at least q_2 (which is at least q_1 by assumption). Third, if the upper set includes the relative values of neither t_1 nor t'_1 , then conditioned on a low value of the grand bundle, relative values are in the set with probability 0 and monotonicity is trivially satisfied.

We now construct a ratio-monotone decomposition assuming $q_1 \leq q_2$. Let $q_H = \Pr[t \in$

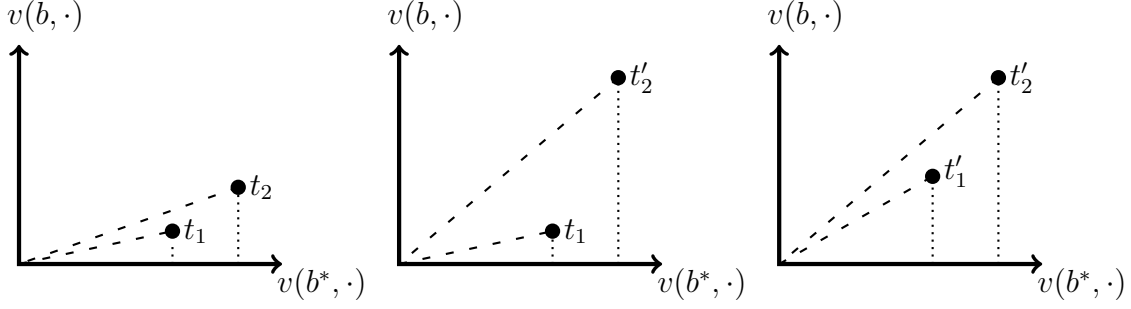


Figure 7: The supports of the three distributions μ^1, μ^2 , and μ^3 .

$\{t_2, t'_2\}$ denote the probability of high value of the grand bundle. The distribution of types μ can be parameterized using q_H, q_1 , and q_2 ,

$$\begin{aligned} & \left(\mathbf{Pr}[t_1], \mathbf{Pr}[t'_1], \mathbf{Pr}[t_2], \mathbf{Pr}[t'_2] \right) \\ &= \left((1 - q_H)(1 - q_1), (1 - q_H)q_1, q_H(1 - q_2), q_Hq_2 \right), \end{aligned}$$

and can be written as a convex combination of three distributions,

$$= (1 - q_2) \left(1 - q_H, 0, q_H, 0 \right) + (q_2 - q_1) \left(1 - q_H, 0, 0, q_H \right) + q_1 \left(0, 1 - q_H, 0, q_H \right).$$

Thus the distribution μ is decomposed into three distributions, denoted μ^1, μ^2 , and μ^3 , with probabilities $1 - q_2, q_2 - q_1$, and q_1 , respectively. The three distributions have supports on ratio-monotone paths, as shown in Figure 7, and have identical probability q_H that the value of the grand bundle is high. Thus μ has a ratio-monotone decomposition. Notice that it is necessary to have $q_1 \leq q_2$ as otherwise the probability of distribution μ^2 would become negative. In fact, if $q_1 > q_2$, then any decomposition of μ into distributions supported on paths must include a distribution supported on t'_1 and t_2 which is not ratio-monotone.

We now summarize the above discussion in a generalizable form. A type t can be drawn from μ by first selecting one of the three distributions with appropriate probabilities, and then drawing a type from the selected distribution. To select a distribution, we can draw q uniformly at random from the interval $[0, 1]$, and then select μ^1 if $q \leq 1 - q_2$, μ^2 if $1 - q_2 < q \leq 1 - q_1$, and μ^3 otherwise. The random variable q satisfies three properties.

- (I) q and $v(b^*, t)$ are independently distributed. In fact, the value of the grand bundle has the same distribution in μ^1, μ^2 , and μ^3 . Thus, conditioned on any q , the distribution

of the value of the grand bundle is the same.

- (II) q and the value of the grand bundle pin down a type, and hence relative values. This is because each distribution μ^1, μ^2 , or μ^3 is supported on a path.
- (III) For a fixed q , relative values are monotone non-decreasing in the value of the grand bundle. This is because each path is ratio-monotone.

The following lemma generalizes the construction to all distributions.

Lemma 4 (Strassen, 1965; Kamae et al., 1977). *Consider jointly distributed random variables $(x, y) \in \mathbb{R}^S \times \mathbb{R}$ for some finite set S . The distribution of x is stochastically non-decreasing in y if and only if there exists a random variable $q \in Q$, jointly distributed with (x, y) , and a function $h^{\hat{q}} : S \times \mathbb{R} \rightarrow \mathbb{R}$ for each $\hat{q} \in Q$ such that*

- (I) q and y are independently distributed.
- (II) Conditioned on $q = \hat{q}$, $x(s) = h^{\hat{q}}(s, y)$ for all $s \in S$ with probability one.
- (III) $h^{\hat{q}}(s, y)$ is monotone non-decreasing in y for all \hat{q} and $s \in S$.

In terms of our discussion above, x represents relative values, y represents the value of the grand bundle, and h is the mapping that pins down relative values given q and the value of the grand bundle. The first property of Lemma 4 is independence. The second property states that q and y pin down x via function h . The third property states that for each q , the mapping from y to x is monotone.

We now use Lemma 4 to prove Theorem 1.

Proof of Theorem 1. Recall that F denotes the cumulative marginal distribution of the value of the grand bundle. Consider any maximizer p^* of $p(1 - F(p))$. We show that pure bundling with price p^* is optimal.

Suppose that $r(\cdot, t) \in \mathbb{R}^B$ is stochastically non-decreasing in $v(b^*, t)$. By Lemma 4, there exists a random variable $q \in Q$ and functions $h^{\hat{q}} : B \times \mathbb{R} \rightarrow \mathbb{R}$ such that

- (I) q and $v(b^*, t)$ are independently distributed.
- (II) Conditioned on $q = \hat{q}$, $r(b, t) = h^{\hat{q}}(b, v(b^*, t))$ for all $b \in B$ with probability one. In words, conditioned on q , a type is pinned down by his value of the grand bundle, i.e., types are on a path.

(III) $h^{\hat{q}}(b, v(b^*, t))$ is monotone non-decreasing in $v(b^*, t)$ for all \hat{q} and b . That is, conditioned on q , each path is ratio-monotone.

Let $\mu^{\hat{q}}$ be the distribution of types conditioned on $q = \hat{q}$. Each such distribution is supported on a ratio-monotone path by properties (II) and (III). By property (I), the marginal distribution of the value of the grand bundle is F . Therefore, pure bundling with price p^* is optimal for each $\mu^{\hat{q}}$. Lemma 3 implies that pure bundling is optimal for distribution μ . \square

4.3 A Partial Converse to Theorem 1

Our main result requires that the profile of relative values is stochastically non-decreasing in the value of the grand bundle, which implies that each relative value is stochastically non-decreasing in the value of the grand bundle. We now show that pure bundling is not optimal if the relative value of some bundle is stochastically decreasing in the value of the grand bundle, and if two minor additional assumptions hold. We start with definitions.

Definition 3. *A random variable $x \in \mathbb{R}$ is stochastically decreasing in a random variable $y \in \mathbb{R}$ if for all $\hat{y} < \hat{y}'$ and \bar{x} , there exists $\delta > 0$ such that $\Pr[x \geq \bar{x} \mid y = \hat{y}] \geq \Pr[x \geq \bar{x} - \delta \mid y = \hat{y}']$.*

To see the connection with Definition 2, suppose that $x \in \mathbb{R}$ is stochastically non-increasing in y , that is, $\Pr[x \geq \bar{x} \mid y = \hat{y}]$ is non-increasing in \hat{y} for any \bar{x} . This is equivalent to saying that for any $\hat{y} < \hat{y}'$, there exists $\delta \geq 0$ such that $\Pr[x \geq \bar{x} \mid y = \hat{y}] \geq \Pr[x \geq \bar{x} - \delta \mid y = \hat{y}']$. Definition 3 instead requires that $\delta > 0$, providing a *strict* notion of stochastic order.³

We say that the distribution of values is (i) *continuous* if for all bundles b , the joint distribution of the value of the grand bundle and the value of bundle b has a probability density function with continuous support, and (ii) *interior* if some $v > \underline{v}$ maximizes $v \times (1 - F(v))$, where F denotes the distribution of the value of the grand bundle, and \underline{v} is the lowest value in its support.

Proposition 2. *Pure bundling is not optimal if $r(b, t)$ is stochastically decreasing in the value of the grand bundle for some b , and the distribution of values is continuous and interior.*

³An alternative definition of strict order, which may appear more natural at first, is to require $\Pr[x \geq \bar{x} \mid y = \hat{y}]$ to be decreasing in \hat{y} . The two definitions are equivalent if x has full support over \mathbb{R} conditioned on any value of y . Otherwise, this alternative definition requires adjustments. To see this, suppose that $x = -y$. According to Definition 3, x is stochastically decreasing in y . But note that for all $\hat{y} > 0$, $\Pr[x \geq 0 \mid y = \hat{y}] = 0$, therefore the conditional probability is not decreasing, violating the alternative definition. Since we do not assume full support, we use Definition 3.

If types are on a path, then the condition of the proposition is that the relative value of bundle b strictly decreases along the path. Thus the proposition provides a partial converse to Proposition 1, which says that pure bundling is optimal if the relative value of each bundle weakly increases along the path.

5 Extensions

Multi-unit demands. Our model can accommodate multi-unit demands. If the buyer demands multiple units of some product, we can add multiple copies of the product to the set of products and let the values of these copies be identical. For example, suppose that there are two products 1 and 2, and the buyer may demand at most two units of product 1 and one unit of product 2. Define the set of products to be $\{1_a, 1_b, 2\}$, where 1_a and 1_b represent copies of product 1. The values satisfy $v(\{1_a\}, t) = v(\{1_b\}, t)$ and $v(\{1_a, 2\}, t) = v(\{1_b, 2\}, t)$ for all types t to reflect that 1_a and 1_b are identical products. More generally, the value of a bundle is measurable with respect to the number of copies of each product in the bundle.

Costs. Our model can also accommodate production costs. This is because we can transform the problem to one with zero costs by subtracting costs from values. In particular, suppose each bundle b costs $c(b)$ to produce (and the empty bundle costs zero). Consider a transformed setting with the same set of bundles and types where the value of bundle b for type t is $\tilde{v}(b, t) = v(b, t) - c(b)$, and costs are zero. For any IC an IR mechanism (a, p) in the original setting, the mechanism (a, \tilde{p}) where $\tilde{p}(t) = p(t) - \mathbf{E}_{b \sim a(t)}[c(b)]$ is IC and IR in the transformed setting, and vice versa.⁴ Further, the expected profit of mechanism (a, p) in the original setting is equal to the expected revenue of mechanism (a, \tilde{p}) in the transformed setting,

$$\mathbf{E}_t \left[p(t) - \mathbf{E}_{b \sim a(t)} [c(b)] \right] = \mathbf{E}_t [\tilde{p}(t)].$$

Thus, if pure bundling is optimal in the transformed setting, then it is also optimal in the original setting. Our main result can now be used to obtain conditions for the optimality of pure bundling in the transformed setting, and hence in the original setting with costs. To ensure that the grand bundle has the highest value for all types in the transformed setting,

⁴To see this, consider the utility of a type t from reporting t' in the two settings. In the original setting, the utility is $v(a(t'), t) - p(t')$. In the transformed setting, the utility is $\tilde{v}(a(t'), t) - \tilde{p}(t') = v(a(t'), t) - \mathbf{E}_{b \sim a(t')} [c(b)] - (p(t') - \mathbf{E}_{b \sim a(t')} [c(b)]) = v(a(t'), t) - p(t')$.

we assume that the grand bundle is the efficient non-empty bundle, that is, b^* maximizes the surplus $v(b, t) - c(b)$ over all bundles $b \neq \emptyset$ for each type t . This assumption is satisfied if bundles have nearly uniform costs, but may be violated if a smaller bundle has a significantly lower cost than the grand bundle, since if so, there may be a type for which the small bundle is more efficient than the grand bundle. The transformed profile of relative values is $\tilde{r}(\cdot, t) \in \mathbb{R}^B$ where $\tilde{r}(b, t) = \tilde{v}(b, t)/\tilde{v}(b^*, t)$.

Proposition 3. *Suppose that each bundle b costs $c(b)$ and the grand bundle is the efficient non-empty bundle. Pure bundling is optimal if the transformed profile of relative values is stochastically non-decreasing in positive transformed values of the grand bundle; i.e., $\Pr[\tilde{r}(\cdot, t) \in U \mid \tilde{v}(b^*, t) = \hat{v}]$ is non-decreasing in $\hat{v} > 0$ for all upper sets $U \subseteq \mathbb{R}^B$.*

As an application of Proposition 3, consider a “main” product that can be produced at cost $c \geq 0$ together with several “add-ons” that have zero costs. An example is a car and software add-ons such as voice command and satellite radio services. To map the application with add-ons into our model, let the set of products be equal to the set of all add-ons in addition to the main product. The value of any bundle that does not include the main product is zero. All non-empty bundles have the same cost c .⁵ The grand bundle is the efficient non-empty bundle since it has the highest value over all bundles. Now Proposition 3 applies and says that pure bundling is optimal if the transformed profile of relative values $\tilde{r}(\cdot, t)$ where $\tilde{r}(b, t) = (v(b, t) - c)/(v(b^*, t) - c)$ is stochastically non-decreasing in the transformed values of the grand bundle. A sufficient condition is that the profile of relative values r is stochastically non-decreasing in value of the grand bundle.

Corollary 1. *Suppose that each non-empty bundle costs $c \geq 0$. Pure bundling is optimal if the profile of relative values is stochastically non-decreasing in positive values of the grand bundle.*

Qualities. Our model can capture the problem of selling a product at several quality levels. For this problem, our results give optimality conditions for selling only the highest quality. To see this, consider the special case of our setting where the products are identical, that is, the value of a bundle depends only on the number of products in it. This is mathematically equivalent to a problem where there is a single product that can be produced in different quality levels, and the buyer demands at most a single unit of the product.

⁵A bundle that does not include the main product has zero cost. However, a mechanism that assigns such a bundle can be replaced with one that assigns the empty bundle instead and has the same profit. Thus we can assume that the cost of producing such a bundle is also c without affecting optimal mechanisms.

To formally state the result, suppose that there are n quality levels. Let $v(m, t)$ denote the value of quality level $m \in \{1, \dots, n\}$ for type t , and let $c(m)$ denote the cost of such a quality level. The surplus of quality level m for type t is $\tilde{v}(m, t) = v(m, t) - c(m)$ and the transformed profile of relative values is $\tilde{r}(\cdot, t) \in \mathbb{R}^n$ where $\tilde{r}(m, t) = \tilde{v}(m, t)/\tilde{v}(n, t)$.

Corollary 2. *Suppose that the highest quality $m = n$ maximizes the surplus $\tilde{v}(m, t)$ over all $m \in \{1, \dots, n\}$ for each type t . Selling only the highest quality is optimal if the transformed profile of relative values is stochastically non-decreasing in positive surplus of the highest quality; i.e., $\Pr[\tilde{r}(\cdot, t) \in U \mid \tilde{v}(n, t) = \hat{v}]$ is non-decreasing in $\hat{v} > 0$ for all upper sets $U \subseteq \mathbb{R}^n$.*

The conditions of Corollary 2 are related to the conditions in Salant (1989), Deneckere and McAfee (1996), Johnson and Myatt (2003), and Anderson and Dana Jr (2009) for the optimality of only selling the highest quality. These papers make several assumptions that we do not, including the regularity of the marginal distribution of the value of the highest quality, and most importantly, increasing differences. Increasing differences requires that types can be ranked so that a higher type has a higher marginal value for consuming a higher quality than a lower quality. Increasing differences may be violated in our setting even in the special case with identical products and ratio-monotone paths. Another assumption in these papers is increasing returns to quality. The weakest form of this assumption is from Anderson and Dana Jr (2009) which requires that types have distinct values (they are on a path) and $\tilde{v}(m, t)$ is log-submodular. This is a stronger assumption than ratio-monotonicity (which is itself stronger than the stochastic monotonicity assumption of Corollary 2).⁶

6 Interpretations

Optimality of pure bundling depends on how relative values vary with the value of the grand bundle across types. To understand our results, re-parameterize a type by his (i) relative values, and (ii) value of the grand bundle. We first interpret each parameter while holding the other fixed. The interpretation is in terms of top-up margins and price-sensitivity, which we now discuss.

Relative values can be used to measure the value of adding products to a bundle. To see this, consider a type t and a bundle b . Define the top-up margin of the bundle to be

⁶Log-submodularity requires that $\log(\tilde{v}(m+1, t)) - \log(\tilde{v}(m, t))$ is non-increasing in t for any $m < n$. Summing from m to n , $\log(\tilde{v}(n, t)) - \log(\tilde{v}(m, t))$ must be non-increasing in t . Raising the expression to the power of e , $\tilde{v}(n, t)/\tilde{v}(m, t)$ is non-increasing in t . This is the ratio-monotonicity condition of Corollary 2 when types are on a path.

the percentage increase in value from “topping up” that bundle, i.e., adding all the products that are not in b to the bundle b . This top-up margin is

$$100\% \times \left(\frac{v(b^*, t) - v(b, t)}{v(b, t)} \right) = 100\% \times \left(\frac{1}{r(b, t)} - 1 \right),$$

which is inversely related to the relative value $r(b, t)$ of bundle b . Thus relative values inversely measure top-up margins. Intuitively, consumers with higher top-up margins for all bundles consider the products to be more complementary.

The value of the grand bundle inversely measures price-sensitivity. If we keep the relative values fixed, then increasing the value of the grand bundle leads to the consumer choosing a weakly more expensive bundle from any menu. This is because the marginal value of money is constant at 1, and thus increasing the value of the grand bundle increases the marginal rate of substitution between money and any bundle.⁷ Price-sensitivity may be related to the consumer’s wealth or income.

We can now rephrase our results in terms of top-up margins and price-sensitivity. Pure bundling is optimal (respectively not optimal) if more price-sensitive consumers have higher (respectively lower) top-up margins, i.e., they consider the products to be more (respectively less) complementary. We now provide some examples to illustrate how top-up margins may vary with price-sensitivity.

The first example is flights. A flight may combine the basic transportation service with several “add-on” services such as luggage services. Each of these services is a different product. The value of any bundle that does not include the transportation service is zero. Business travelers are more likely to “travel light” than leisure travelers, so they may have a lower margin for topping up the transportation service with luggage services. They are also less price-sensitive. Thus, top-up margins increase in price-sensitivity, and bundling luggage services with the basic transportation service is optimal. A related example is hotel stays. Compared to leisure travelers, business travelers have a lower margin for topping up their hotel stay with access to a child-friendly pool, but a higher margin for topping up their hotel stay with Internet access. Therefore, it is optimal to bundle pool access, but not Internet access, with the hotel stay.

As another example, consider a firm that sells entertainment products such as TV shows.

⁷Formally, consider types t and t' with identical relative values and $v(b^*, t) < v(b^*, t')$. Suppose type t chooses bundle b price at p and type t' chooses bundle b' price at p' . The choices of the two types imply that $v(b, t) - p \geq v(b', t) - p'$ and $v(b', t') - p' \geq v(b, t') - p$. Summing the two inequalities, and factoring the identical relative values, we have $(v(b^*, t) - v(b^*, t')) \cdot (r(b, t) - r(b', t)) \geq 0$. Therefore, $r(b, t) \leq r(b', t)$. Since type t chooses b even though it has a lower value, the price of b must be weakly less, $p \leq p'$.

Top-up margins depend on a consumer’s available leisure time. A consumer with limited leisure time may be able to follow only a few shows, so the value for having access to additional shows is low, leading to low top-up margins. More leisure time leads to higher top-up margins. Thus pure bundling is optimal (respectively not optimal) if more price-sensitive consumers have more (respectively less) leisure time. Leisure time may vary with a consumer’s income, which is related to the consumer’s price-sensitivity. A related example is when consumers are households, and top-up margins increase in the number of individuals in the household. More generally, top-up margins may be affected by constraints that limit consumption capacity, and these constraints may vary with price-sensitivity.

Another example is credit cards. A credit card is a bundle of products. For instance, one product may give a reward point per dollar spent. Each point can be redeemed as a cent to make certain travel purchases. Since the ability to redeem points for travel is more valuable for a consumer who travels more, the margin for topping up the credit card with this product is higher for such a consumer. The top-up margin may vary for other reasons. To redeem points, the consumer needs to identify the credit card company’s travel partners and know how to transfer the points to those partners. Thus, the top-up margin is higher for a more savvy consumer or one with a lower opportunity cost of time, both of which may be related to price-sensitivity.⁸ Bundling the product with the credit card is optimal (respectively not optimal) if more price-sensitive consumers are more (respectively less) savvy and have a lower (respectively higher) opportunity cost of time. A related example is software products that are more effective when used together. For instance, a software package may allow the user to set up a presentation program together with a data analysis program so that the charts in the presentation update automatically if the results of the data analysis change. Less savvy consumers are less likely to value this possibility, leading to lower top-up margins.

A parametric example. Top-up margins quantify the degree of complementarity between products. They are thus related to the notions of complements and substitutes. To illustrate, consider an example with two products. For type t , the value of product 1 is $w(t)$, the value of product 2 is $\alpha w(t)$, and the value of the bundle is $(1 + \alpha)w(t) + \delta$, where $\alpha \geq 0$ and $\delta \in \mathbb{R}$ are constants and $w : T \rightarrow \mathbb{R}^+$ is a function. Thus δ is the “added value” of consuming the products together. The products are partial complements if $\delta \geq 0$ and are partial substitutes

⁸These effects may work in opposite directions. Compared to a college student, a retiree may be less price-sensitive, less tech savvy, and have a lower opportunity cost of time. Comparing within an age group, a high income individual is likely to be more savvy and have a higher opportunity cost of time.

if $\delta \leq 0$. The top-up margins of the two products are

$$100\% \times \frac{\alpha w(t) + \delta}{w(t)} \text{ and } 100\% \times \frac{w(t) + \delta}{\alpha w(t)}.$$

If the products are partial complements, i.e., $\delta \geq 0$, then the top-up margins weakly decrease in the value of the grand bundle and thus pure bundling is optimal. Otherwise, top-up margins increase in the value of the grand bundle and pure bundling is not optimal (Proposition 2 also requires distribution to be continuous and interior). Krishna and Rosenthal (1996) use this model with $\alpha = 1$ and $\delta \geq 0$ to study the equilibria of spectrum auctions.

More generally, different types may have different added values. To illustrate, suppose that for some function $\delta : T \rightarrow \mathbb{R}$, the value of the grand bundle for type t is $(1 + \alpha)w(t) + \delta(t)$. Pure bundling is optimal if $\delta(t)/w(t)$ is stochastically non-increasing in the value of the grand bundle. That is, as the value of the grand bundle grows, the added value increases at a slower rate compared to the “base value” $w(t)$. As before, the interpretation is that types with a higher value for the grand bundle consider the products to be less complementary. A sufficient condition is that the base value and the added value are independently distributed and the products are partial complements for all types, i.e., $\delta(t) \geq 0$ for all t .⁹ Notably, however, pure bundling may be optimal even if products are partial substitutes for all types. For example, if $\delta(t) = \tau w(t)$ for some constant τ , then the top-up margins are constant and thus pure bundling is optimal even if $\tau \leq 0$.

7 Conclusions

We study when it is optimal to sell only the grand bundle to all consumers. Pure bundling is optimal if consumers with higher values are more likely to have higher relative values for smaller bundles compared to the grand bundle. We decompose the type space into paths, and prove the optimality of pure bundling for each path.

Our approach addresses a main challenge in multi-dimensional screening, namely identifying binding incentive constraints. We relax two sets of incentive constraints: any upward constraint along a path, and any constraint across paths. The fact that the solution to the relaxed problem is incentive compatible establishes that the relaxed constraints are not binding under our stochastic monotonicity condition. On the other hand, if the stochastic

⁹Given the conditions, $\delta(t)/w(t)$ is stochastically non-increasing in $w(t)$, and $w(t)$ itself is stochastically non-decreasing in the value of the grand bundle. By transitivity, $\delta(t)/w(t)$ is stochastically non-increasing in the value of the grand bundle.

monotonicity condition does not hold, our approach does not provide any guidance on which constraints do or do not bind.

The theoretical literature on multi-product bundling shows that optimal mechanisms are often complex. Our work is a step towards rationalizing simple selling strategies. In practice, even if firms do not offer just a single bundle, they tend to offer only a few bundles. Ideally a theory of bundling would explain why this is the case, and which small set of bundles would be offered by the seller. Although we do not study how to offer only a few bundles, we hope that our work sheds light on possible answers. For instance, to show that certain products should be sold only as a bundle, it might be useful to consider the relative values for all subsets of that bundle.

Our condition for the optimality of pure bundling is not knife edge. If the distribution of relative values is strictly stochastically increasing in the value for the grand bundle, then the same will hold for a small perturbation of the distribution. Thus pure bundling remains optimal for perturbations of the distribution. This robustness property is violated in Stokey (1979) and Riley and Zeckhauser (1983) (and Myerson, 1981, in the special case of a single buyer) where the relative values are constant across types. The analysis of Armstrong (1996), applied to our setting, shows that the same result holds when relative values are independently distributed from the value for the grand bundle. In either case, the distribution of relative values is constant in the value for the grand bundle, which is a special case of our condition. A local perturbation of such a distribution violates the stochastic monotonicity condition required for optimality of pure bundling. Thus, such instances are at the boundary between optimality and non-optimality of pure bundling.

References

- Adams, W. J. and Yellen, J. L. (1976). Commodity bundling and the burden of monopoly. *The Quarterly Journal of Economics*, pages 475–498.
- Anderson, E. T. and Dana Jr, J. D. (2009). When is price discrimination profitable? *Management Science*, 55(6):980–989.
- Armstrong, M. (1996). Multiproduct nonlinear pricing. *Econometrica*, 64(1):51–75.
- Armstrong, M. (1999). Price discrimination by a many-product firm. *The Review of Economic Studies*, 66(1):151–168.

- Armstrong, M. (2013). A more general theory of commodity bundling. *Journal of Economic Theory*, 148(2):448–472.
- Armstrong, M. (2016). Nonlinear pricing. *Annual Review of Economics*, 8:583–614.
- Bakos, Y. and Brynjolfsson, E. (1999). Bundling information goods: Pricing, profits, and efficiency. *Management Science*, 45(12):1613–1630.
- Cai, Y., Devanur, N. R., and Weinberg, S. M. (2019). A duality-based unified approach to bayesian mechanism design. *SIAM Journal on Computing*, (0):STOC16–160.
- Carroll, G. (2017). Robustness and separation in multidimensional screening. *Econometrica*, 85(2):453–488.
- Daskalakis, C., Deckelbaum, A., and Tzamos, C. (2017). Strong duality for a multiple-good monopolist. *Econometrica*, 85(3):735–767.
- Deneckere, R. J. and McAfee, P. (1996). Damaged goods. *Journal of Economics & Management Strategy*, 5(2):149–174.
- Eső, P. and Szentes, B. (2007). Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies*, 74(3):705–731.
- Fang, H. and Norman, P. (2006). To bundle or not to bundle. *The RAND Journal of Economics*, 37(4):946–963.
- Hart, S. and Nisan, N. (2019). Selling multiple correlated goods: Revenue maximization and menu-size complexity. *Journal of Economic Theory*, 183:991–1029.
- Johnson, J. P. and Myatt, D. P. (2003). Multiproduct quality competition: Fighting brands and product line pruning. *American Economic Review*, 93(3):748–774.
- Kamae, T., Krengel, U., and O’Brien, G. L. (1977). Stochastic inequalities on partially ordered spaces. *The Annals of Probability*, pages 899–912.
- Krishna, V. and Rosenthal, R. W. (1996). Simultaneous auctions with synergies. *Games and Economic Behavior*, 17(1):1–31.
- Long, J. B. (1984). Comments on “gaussian demand and commodity bundling”. *The Journal of Business*, 57(1):S235–S246.

- Madarász, K. and Prat, A. (2017). Sellers with misspecified models. *The Review of Economic Studies*, 84(2):790–815.
- McAfee, R. P., McMillan, J., and Whinston, M. D. (1989). Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics*, 104(2):371–383.
- Menicucci, D., Hurkens, S., and Jeon, D.-S. (2015). On the optimality of pure bundling for a monopolist. *Journal of Mathematical Economics*, 60:33–42.
- Milgrom, P. and Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601.
- Myerson, R. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1):pp. 58–73.
- Pavan, A., Segal, I., and Toikka, J. (2014). Dynamic mechanism design: A myersonian approach. *Econometrica*, 82(2):601–653.
- Pavlov, G. (2011). Optimal mechanism for selling two goods. *The BE Journal of Theoretical Economics*, 11(1).
- Riley, J. and Zeckhauser, R. (1983). Optimal selling strategies: When to haggle, when to hold firm. *The Quarterly Journal of Economics*, 98(2):267–289.
- Rosenthal, R. W. and Wang, R. (1996). Simultaneous auctions with synergies and common values. *Games and Economic Behavior*, 17(1):32–55.
- Salant, S. W. (1989). When is inducing self-selection suboptimal for a monopolist? *The Quarterly Journal of Economics*, 104(2):391–397.
- Schmalensee, R. (1984). Gaussian demand and commodity bundling. *Journal of Business*, pages S211–S230.
- Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer Science & Business Media.
- Stigler, G. J. (1963). United States v. Loew’s inc.: A note on block-booking. *The Supreme Court Review*, 1963:152–157.

- Stokey, N. L. (1979). Intertemporal price discrimination. *The Quarterly Journal of Economics*, pages 355–371.
- Strassen, V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, pages 423–439.
- Thanassoulis, J. (2004). Haggling over substitutes. *Journal of Economic theory*, 117(2):217–245.
- Vincent, D. and Manelli, A. (2007). Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *Journal of Economic Theory*, 137(1):153–185.
- Wilson, R. B. (1993). *Nonlinear pricing*. Oxford University Press.

A Appendix

A.1 Proof of Lemma 1

Lemma 1. *For any incentive compatible mechanism (a, p) ,*

$$\mathbf{E} [p(t)] = \mathbf{E} [\phi(a(t), t)] - (v(a(\underline{t}), \underline{t}) - p(\underline{t})).$$

Proof. Fix an IC mechanism (a, p) . Let u be the indirect utility function of mechanism (a, p) , that is $u(t) := v(a(t), t) - p(t)$. Incentive compatibility implies that $u(t) = \max_{t'} v(a(t'), t) - p(t')$. The envelope theorem (Theorem 1 and Theorem 2 in Milgrom and Segal, 2002) implies that

$$\frac{d}{dt}u(t) = \partial_2 v(a(t), t), \tag{4}$$

whenever the derivative of u exists, and that u can be represented as an integral of its derivative.

Now consider the expectation of u , and apply integration by parts to write

$$\begin{aligned}
\mathbf{E} [u(t)] &= \int_{\underline{t}}^{\bar{t}} u(t)f(t) dt \\
&= \int_{\underline{t}}^{\bar{t}} \frac{d}{dt}u(t)(1 - F(t)) dt + u(\underline{t}) \\
&= \int_{\underline{t}}^{\bar{t}} \partial_2 v(a(t), t)(1 - F(t)) dt + u(\underline{t}) \\
&= \mathbf{E} \left[\partial_2 v(a(t), t) \frac{1 - F(t)}{f(t)} \right] + u(\underline{t})
\end{aligned}$$

where the second to last equality followed from substituting (4). Finally, the expected revenue can be written as the difference of surplus and consumer rents,

$$\begin{aligned}
\mathbf{E} [p(t)] &= \mathbf{E} [v(a(t), t) - u(t)] \\
&= \mathbf{E} \left[v(a(t), t) - \partial_2 v(a(t), t) \frac{1 - F(t)}{f(t)} \right] - u(\underline{t}) \\
&= \mathbf{E} [\phi(a(t), t)] - (v(a(\underline{t}), \underline{t}) - p(\underline{t})).
\end{aligned}$$

□

A.2 Proof of Lemma 2

Lemma 2. *If $v(\cdot, t)/t$ is monotone non-decreasing in t , then $\phi(a, t) \leq \max(0, \phi(b^*, t))$ for all randomized bundles $a \in \Delta(B)$ and types $t \in T$.*

Proof. Since $v(b, t)/t$ is monotone and v is differentiable in t ,

$$\frac{d}{dt} \left(\frac{v(b, t)}{t} \right) = \frac{\partial_2 v(b, t)t - v(b, t)}{t^2} \geq 0.$$

As a result,

$$\frac{v(b, t)}{t} \leq \partial_2 v(b, t). \tag{5}$$

Directly from the definition of virtual values,

$$\begin{aligned}
\phi(b, t) &= v(b, t) - \partial_2 v(b, t) \times \frac{1 - F(t)}{f(t)} \\
&\leq v(b, t) - \frac{v(b, t)}{t} \times \frac{1 - F(t)}{f(t)} \\
&= \frac{v(b, t)}{t} \left[t - \frac{1 - F(t)}{f(t)} \right]
\end{aligned}$$

where the inequality followed from (5). Notice that the right hand side is $\frac{v(b, t)}{t}$ multiplied by the virtual value for the grand bundle $\phi(b^*, t)$. Therefore,

$$\phi(b, t) \leq \frac{v(b, t)}{t} \phi(b^*, t). \quad (6)$$

We now complete the proof by considering two cases and applying Inequality (6). If $0 \leq \phi(b^*, t)$ we have

$$\phi(b, t) \leq \frac{v(b, t)}{t} \phi(b^*, t) \leq \phi(b^*, t),$$

where the second inequality followed since $\frac{v(b, t)}{t} \leq 1$. If $\phi(b^*, t) \leq 0$ we have

$$\phi(b, t) \leq \frac{v(b, t)}{t} \phi(b^*, t) \leq 0.$$

Given the above two inequalities, we have $\phi(b, t) \leq \max(0, \phi(b^*, t))$ for all b and t . As a result, $\phi(a, t) = \mathbf{E}[\phi(b, t)] \leq \max(0, \phi(b^*, t))$ for all randomized bundles a and types t . \square

A.3 Proof of Proposition 1

In this section we prove Proposition 1. The proof of the “if” statement of Proposition 1 requires a setup, with which we start.

A.3.1 Setting

Assume that set of types $T \subseteq \mathbb{R}_+$ is finite. That is, types are $t_0 (= \underline{t}) < t_1 < \dots < t_I$. To avoid dividing by zero, assume without loss of generality that $t_0 > 0$. Let $f(t_i)$ denote the probability of type t_i . We prove Proposition 1 for the finite case, and then extend it, similar

to Carroll (2017), to general distributions applying an approximation result from Madarász and Prat (2017).

A.3.2 Generalized Virtual Values

We now discuss a general construction of virtual values based on Lagrangian duality, identical to Carroll (2017); Cai et al. (2019). For any i and j from $\{0, \dots, I\}$, let $\lambda(j, i) \geq 0$ be the Lagrangian multiplier of the IC constraint

$$v(a(t_j), t_j) - p(t_j) \geq v(a(t_i), t_j) - p(t_i).$$

Define the Lagrangian

$$\begin{aligned} \mathcal{L}(\lambda, a, p) &= \left(\sum_i p(t_i) f(t_i) \right) + \left(\sum_{i,j} \lambda(j, i) (v(a(t_j), t_j) - p(t_j) - (v(a(t_i), t_j) - p(t_i))) \right) \\ &= \sum_i \left(p(t_i) \left(f(t_i) - \sum_j \lambda(i, j) + \sum_j \lambda(j, i) \right) \right) \\ &\quad + \sum_i \left(\left(\sum_j \lambda(i, j) v(a(t_i), t_i) \right) - \left(\sum_j \lambda(j, i) v(a(t_i), t_j) \right) \right). \end{aligned}$$

For any IC mechanism (a, p) , the Lagrangian is an upper bound on the revenue of the mechanism. Furthermore, duality implies that for any optimal mechanism (a, p) , there exist optimal Lagrangian multipliers λ such that (a, p) minimizes the Lagrangian $\mathcal{L}(\lambda, a, p)$. Note that the optimal λ must satisfy

$$f(t_i) - \sum_j \lambda(i, j) + \sum_j \lambda(j, i) = 0. \tag{7}$$

Otherwise, the dual solution is unbounded. Indeed, if the expression above is strictly positive, by decreasing $p(t_i)$ the Lagrangian approaches $-\infty$. Similarly, if the expression is strictly negative, by increasing $p(t_i)$ the Lagrangian approaches $-\infty$.

We call λ satisfying (7) feasible. We interpret (7) as a flow constraint and refer to $\lambda(j, i)$ the flow incoming to i from j . In any feasible λ , the term involving p in the Lagrangian

cancel and the Lagrangian becomes

$$\begin{aligned}\mathcal{L}(\lambda, a, p) &= \sum_i \left(\left(\sum_j \lambda(i, j) v(a(t_i), t_i) \right) - \left(\sum_j \lambda(j, i) v(a(t_i), t_j) \right) \right) \\ &= \sum_i \left(v(a(t_i), t_i) f(t_i) - \sum_j \lambda(j, i) (v(a(t_i), t_j) - v(a(t_i), t_i)) \right),\end{aligned}\quad (8)$$

where the equality followed from substituting (7). For any λ , define the *induced virtual value* of a bundle b for a type t as follows,

$$\phi(b, t_i) := v(b, t_i) - \frac{1}{f(t_i)} \sum_j \lambda(j, i) (v(b, t_j) - v(b, t_i)), \quad \forall t_i \in T, b \in B. \quad (9)$$

Extend the definition of virtual values to randomized bundles by setting, for any randomized bundle $a \in \Delta(B)$, $\phi(a, t) = \mathbf{E}_{b \sim a}[\phi(b, t)]$. Substituting 9 into 8, the Lagrangian is the expected virtual surplus of a

$$\mathcal{L}(\lambda, a, p) = \sum_{t_i} \phi(a(t_i), t_i) \times f(t_i). \quad (10)$$

We summarize the analysis above in the following lemma.

Lemma 5 (Carroll, 2017; Cai et al., 2019). *A mechanism (a, p) is optimal if and only if there exists feasible λ such that the assignment rule a maximizes the virtual surplus defined in Equations 9 and 10, and (a, p) and λ satisfy the complementary slackness condition,*

$$\lambda(j, i) \left(v(a(t_j), t_j) - p(t_j) - (v(a(t_i), t_j) - p(t_i)) \right) = 0. \quad (11)$$

A.3.3 A Construction of Ironed Virtual Values

The lemma below is similar to Lemma 1. It provides a comparison of virtual values based on the change in $v(b, t)/t$, if the Lagrangian variables are non-zero only for downward constraints, i.e., $\lambda(j, i) = 0$ for all $j < i$.

Lemma 6. *If $\lambda(j, i) = 0$ for all $j < i$ and $v(b, t_i)/t_i$ is monotone non-decreasing in t_i , then the induced virtual values, defined via Equation (9), satisfy*

$$\frac{v(b, t_i)}{t_i} \phi(b^*, t_i) \geq \phi(b, t_i), \quad \forall t_i, b \in B.$$

Proof. Recall that $v(b^*, t_i) = t_i$ for all i . Directly from the definition of virtual values,

$$\begin{aligned} \frac{v(b, t_i)}{t_i} \phi(b^*, t_i) &= \frac{v(b, t_i)}{t_i} \left(t_i - \frac{1}{f(t_i)} \sum_j \lambda(j, i) (v(b^*, t_j) - v(b^*, t_i)) \right) \\ &= v(b, t_i) - \frac{1}{f(t_i)} \sum_j \lambda(j, i) \left(\frac{v(b, t_i)}{t_i} v(b^*, t_j) - v(b, t_i) \right). \end{aligned}$$

Since $\lambda(j, i) = 0$ for all $j < i$, we have

$$\begin{aligned} \frac{v(b, t_i)}{t_i} \phi(b^*, t_i) &= v(b, t_i) - \frac{1}{f(t_i)} \sum_{j:j>i} \lambda(j, i) \left(\frac{v(b, t_i)}{t_i} v(b^*, t_j) - v(b, t_i) \right) \\ &\geq v(b, t_i) - \frac{1}{f(t_i)} \sum_{j:j>i} \lambda(j, i) (v(b, t_j) - v(b, t_i)) \\ &= v(b, t_i) - \frac{1}{f(t_i)} \sum_j \lambda(j, i) (v(b, t_j) - v(b, t_i)) \\ &= \phi(b, t_i), \end{aligned}$$

where the inequality followed from monotonicity of $v(b, t_i)/t_i$. \square

To interpret Lemma 6 geometrically, note from definition (9) that viewed as a vector, $\phi(\cdot, t_i)$ is equal to the vector $v(\cdot, t_i)$, shifted proportionally to $v(\cdot, t_j) - v(\cdot, t_i)$ for all $t_j > t_i$ with strictly positive $\lambda(j, i)$. The resulting vector is “below” the ray that connects the origin to $v(\cdot, t_i)$, as depicted in Figure 5.

Given Lemma 6, we would like to construct Lagrangian dual variables $\bar{\lambda}$ such that (1) $\bar{\lambda}$ is feasible (2) $\bar{\lambda}$ is non-zero only for downward constraints so that Lemma 6 applies, (3) the induced virtual value $\bar{\phi}$ of definition (9) for the favorite bundle $\bar{\phi}(b^*, t)$ is monotone non-decreasing in t , and (4) the assignment rule that only assigns the grand bundle b^* to types t with positive $\bar{\phi}(b^*, t)$ satisfies the complementary slackness condition with $\bar{\lambda}$. The lemma below shows that such dual variables exist (we verify property (4) later).

Lemma 7. *There exist dual variables $\bar{\lambda}$ with induced virtual value $\bar{\phi}$ such that*

- (I) $\bar{\lambda}$ is feasible, that is, it satisfies (7),
- (II) If $\bar{\lambda}(j, i) > 0$ then $i < j$,
- (III) $\bar{\phi}(b^*, t)$ is monotone non-decreasing in t ,

(IV) If $\bar{\lambda}(j, i) > 0$ then $\bar{\phi}(b^*, t_j) = \bar{\phi}(b^*, t_{j''})$ for all j', j'' such that $i \leq j', j'' < j$.

Proof. For each i , let $G_i = \sum_{j \geq i} f(t_j)$, and let $G_{I+1} = 0$. Define the revenue function R , with support $\{G_i\}_{i \in I}$ as follows,

$$R(G_i) = t_i G_i.$$

Define the ironed revenue function \bar{R} , defined over support $\{G_i\}_{i \in I}$, to be the lowest concave function that is pointwise higher than R . We now inductively construct Lagrangian variables $\bar{\lambda}$ such that its induced virtual value $\bar{\phi}$ for b^* satisfies

$$\bar{\phi}(b^*, t_i) = \frac{\bar{R}(G_i) - \bar{R}(G_{i+1})}{f(t_i)}.$$

That is, $\bar{\phi}(b^*, t_i)$ is the slope of the ironed revenue curve at t_i . By concavity of \bar{R} , $\bar{\phi}(b^*, t_i)$ is monotone non-decreasing and property (III) of the lemma is satisfied.

From $\kappa = n$ to $\kappa = 0$, we recursively define the Lagrangian λ^κ , its induced virtual value ϕ^κ , and the associated revenue function R^κ given the previous iterations. At the end of the induction, we set $\bar{\lambda} = \lambda^0$ and $\bar{\phi} = \phi^0$. The Lagrangian variables for $\kappa = n$ are defined as follows,

$$\lambda^n(i, j) = \begin{cases} G_i & \text{if } j = i - 1, \\ 0 & \text{otherwise.} \end{cases}$$

That is, the Lagrangian is non-zero only for *local* downward IC constraints. In each iteration κ (including $\kappa = n$), the virtual value ϕ^κ is defined given λ^κ via (9), and the revenue curve is,

$$R^\kappa(G_i) = \sum_{j \geq i} \phi^\kappa(b^*, t_j) f(t_j), \quad (12)$$

which is equivalent to

$$\phi^\kappa(b^*, t_i) = \frac{R^\kappa(G_i) - R^\kappa(G_{i+1})}{f(t_i)}.$$

In each iteration κ , the variables satisfy four properties: (i) λ^κ is feasible; (ii) λ^κ is positive only for downward constraints; (iii) $R^\kappa(G_i) = \bar{R}(G_i)$ if $i \geq \kappa$ and $R^\kappa(G_i) = R(G_i)$ if $i < \kappa$; and, (iv) if $\lambda^\kappa(j, i) > 0$ and $j > i + 1$, then $R(G_{j'}) < \bar{R}(G_{j'})$ for all j' such that

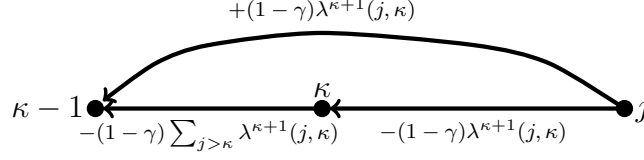


Figure 8: The change in λ^κ compared to $\lambda^{\kappa+1}$. For each $j > \kappa$, a fraction of the flow from j to κ and then to $\kappa - 1$ is rerouted to go directly from j to $\kappa - 1$.

$i < j' < j$. Properties (i), (ii), and (iv) are trivially satisfied when $\kappa = n$. To see property (iii) when $\kappa = n$, write

$$\begin{aligned}
R^n(G_i) &= \sum_{j \geq i} \phi^n(b^*, t_j) f(t_j) \\
&= \sum_{j \geq i} \left(v(b^*, t_j) f(t_j) - \sum_{j'} \lambda^n(j', j) (v(b^*, t_{j'}) - v(b^*, t_j)) \right) \\
&= \sum_{j \geq i} \left(t_j f(t_j) - \sum_{j'} \lambda^n(j', j) (t_{j'} - t_j) \right) \\
&= \sum_{j \geq i} \left(t_j f(t_j) - G_{j+1} (t_{j+1} - t_j) \right) \\
&= \sum_{j \geq i} \left(t_j G_j - G_{j+1} t_{j+1} \right) \\
&= t_i G_i = R(G_i).
\end{aligned}$$

In iteration $\kappa < n$, the Lagrangian is updated as follows. Let $\lambda^\kappa = \lambda^{\kappa+1}$, except for the following modifications,

$$\lambda^\kappa(j, \kappa) = \gamma \lambda^{\kappa+1}(j, \kappa), \quad \forall j > \kappa, \quad (13)$$

$$\lambda^\kappa(j, \kappa - 1) = \lambda^{\kappa+1}(j, \kappa - 1) + (1 - \gamma) \lambda^{\kappa+1}(j, \kappa), \quad \forall j > \kappa, \quad (14)$$

$$\lambda^\kappa(\kappa, \kappa - 1) = \lambda^{\kappa+1}(\kappa, \kappa - 1) - (1 - \gamma) \sum_{j > \kappa} \lambda^{\kappa+1}(j, \kappa), \quad (15)$$

for a parameter γ to be identified shortly. See Figure 8. Define ϕ^κ via (9) and R^κ via (12).

Let us verify that the variables in iteration κ satisfy the properties (i) to (iv) mentioned above, assuming they do so in iteration $\kappa + 1$. We start with verifying feasibility of λ^κ . For any $j > \kappa$ a fraction of the outgoing flow to κ is shifted to $\kappa - 1$, and (7) remains satisfied. For κ , the incoming and the outgoing flows are each reduced by $\sum_{j > \kappa} \lambda^\kappa(j, \kappa)$. For $\kappa - 1$, a fraction of the incoming flow from κ is reduced by $\sum_{j > \kappa} \lambda^\kappa(j, \kappa)$ and the incoming flow from

all types $j > \kappa$ is increased by $\sum_{j>\kappa} \lambda^\kappa(j, \kappa)$.

Assuming that $\lambda^{\kappa+1}$ satisfies property (ii), then so will λ^κ by construction.

To verify property (iii), notice that in iteration κ the incoming flow for all $j \neq \kappa - 1, \kappa$ does not change. Therefore $R^\kappa = R^{\kappa+1}$ for all types other than κ and $\kappa - 1$. To see the equality $R^\kappa(G_{\kappa-1}) = R(G_{\kappa-1})$, notice that

$$\phi^\kappa(b^*, \kappa) = \phi^{\kappa+1}(b^*, \kappa) + \frac{1-\gamma}{f(t_\kappa)} \sum_{j>\kappa} \lambda^{\kappa+1}(j, \kappa)(t_j - t_\kappa),$$

and,

$$\begin{aligned} \phi^\kappa(b^*, \kappa - 1) &= \phi^{\kappa+1}(b^*, \kappa - 1) - \frac{1-\gamma}{f(t_{\kappa-1})} \sum_{j>\kappa} \lambda^{\kappa+1}(j, \kappa)((t_j - t_{\kappa-1}) - (t_\kappa - t_{\kappa-1})) \\ &= \phi^{\kappa+1}(b^*, \kappa - 1) - \frac{1-\gamma}{f(t_{\kappa-1})} \sum_{j>\kappa} \lambda^{\kappa+1}(j, \kappa)(t_j - t_\kappa). \end{aligned}$$

Summing up the above two equations, we get

$$\phi^\kappa(b^*, \kappa)f(t_\kappa) + \phi^\kappa(b^*, \kappa - 1)f(t_{\kappa-1}) = \phi^{\kappa+1}(b^*, \kappa)f(t_\kappa) + \phi^{\kappa+1}(b^*, \kappa - 1)f(t_{\kappa-1}).$$

From 12, we conclude that $R^\kappa(G_{\kappa-1}) = R^{\kappa+1}(G_{\kappa-1}) = R(G_{\kappa-1})$.

To finish the verification of property (iii), we argue that there exists a value of γ such that $R^\kappa(G_\kappa) = \bar{R}(G_\kappa)$. If $\gamma = 1$, then $R^\kappa(G_\kappa) = R^{\kappa+1}(G_\kappa)$, which by induction assumption is equal to $R(G_\kappa)$. Since $R(G_\kappa) \leq \bar{R}(G_\kappa)$, we have $R^\kappa(G_\kappa) \leq \bar{R}(G_\kappa)$ if $\gamma = 1$. On the other hand, we show that $R^\kappa(G_\kappa) \geq \bar{R}(G_\kappa)$ if $\gamma = 0$. Notice that

$$\bar{\phi}(b^*, \kappa) = \frac{R(G_j) - R(G_{j'})}{G_j - G_{j'}}$$

for some $j \leq \kappa < j'$ (possibly $j = \kappa$). Since $R(G_{j'}) = t_{j'}G_{j'} \geq t_jG_{j'}$, we have

$$\bar{\phi}(b^*, \kappa) \leq \frac{t_jG_j - t_jG_{j'}}{G_j - G_{j'}} = t_j.$$

Now note that if $\gamma = 0$, then $\phi^\kappa(b^*, t_\kappa) = t_\kappa \geq \bar{\phi}(b^*, \kappa)$. Thus if $\gamma = 0$,

$$\begin{aligned} R^\kappa(G_\kappa) &= \phi^\kappa(b^*, t_\kappa)f(t_\kappa) + R^\kappa(G_{\kappa+1}) \\ &\geq \bar{\phi}(b^*, \kappa) + \bar{R}(G_{\kappa+1}) \\ &= \bar{R}(G_{\kappa+1}). \end{aligned}$$

Note that $R^\kappa(G_\kappa)$ is a continuous functions of γ . Therefore there must exist a value of γ such that $R^\kappa(G_\kappa) = \bar{R}(G_\kappa)$.

Finally, we argue that property (iv) is satisfied. Note that the only positive flow that is possibly created in iteration κ is $\lambda^\kappa(j, \kappa - 1)$ for $j > \kappa$. Such a flow is created only if $R(G_\kappa) < \bar{R}(G_\kappa)$ (as otherwise $\gamma = 1$) and $\lambda^{\kappa+1}(j, \kappa) > 0$. In this case, by induction hypothesis, we must have $R(G_{j'}) < \bar{R}(G_{j'})$ for all j' such that $\kappa - 1 < j' < j$.

We complete the proof by showing that $\bar{\lambda} = \lambda^0$ satisfies properties of the lemma. Properties (I) and (II) follow directly from properties (i) and (ii) of induction. Property (III) follows since \bar{R} is concave. Property (IV) follows since if $\lambda^\kappa(j, i) > 0$ and $j > i + 1$, then $R(G_{j'}) < \bar{R}(G_{j'})$ for all j' such that $i < j' < j$. Note from the definition of $\bar{\phi}$ and \bar{R} that if $R(G_{j'}) < \bar{R}(G_{j'})$ then $\bar{\phi}(b^*, t_{j'-1}) = \bar{\phi}(b^*, t_{j'})$. As a result, $\bar{\phi}(b^*, t_{j'}) = \bar{\phi}(b^*, t_{j''})$ for all j', j'' such that $i \leq j', j'' < j$. \square

A.3.4 Approximation

To extend the proof from distributions with finite support to general distributions, we apply the following result that implies that if two distributions are close to each other, then their optimal revenue is also close. Distributions μ and μ' are δ -close if T can be partitioned into disjoint sets T_1, \dots, T_n such that $v(b, t) - v(b, t') \leq \delta$ for all $t, t' \in T_i$, and $\Pr_\mu[t \in T_i] = \Pr_{\mu'}[t \in T_i]$.

Lemma 8 (Madarász and Prat, 2017; Lemma 4.3 in Carroll, 2017). *For every $\epsilon > 0$, there exists $\delta > 0$ such that for every mechanism (a, p) , there exists a mechanism (a', p') such that every pair of distributions μ, μ' that are δ -close satisfy $\mathbf{E}_{\mu'}[p'(t)] \geq \mathbf{E}_\mu[p(t)] - \epsilon$.*

A.3.5 The Proof

We now complete the proof of Proposition 1, restated below.

Proposition 1. *Assume that $T \subseteq \mathbb{R}^+$ and $v(b^*, t) = t$. Pure bundling is optimal for all distributions $\mu \in \Delta(T)$ if and only if $v(\cdot, t)/t$ is monotone non-decreasing in $t > 0$.*

Proof. We first prove the “only if” statement. Assume that there exists a bundle b such that $v(b, t)/t$ is not monotone non-decreasing in $t > 0$. Therefore, there exist t, t' such that $0 < t < t'$ and $v(b, t)/t > v(b, t')/t'$. We show that there exists a distribution with support over types t and t' for which pure bundling is not optimal. In particular, let the probability of the low type t be $1 - t/t'$, and the probability of the high type t' be t/t' . The optimal revenue among pure bundling mechanisms is t , which is obtained by offering the grand bundle at price t or t' .

Consider a “mixed bundling” mechanism that assigns bundle b to the low type at price $v(b, t)$, and the grand bundle to the high type at price $t' - v(b, t)(t'/t - 1) + \epsilon$, for $\epsilon > 0$ to be identified shortly. We show that the mixed bundling mechanism obtains higher revenue than the optimal pure bundling revenue, t .

Let us verify the incentive constraints. The utility of the low type from truthtelling is 0, and from deviating (i.e., reporting the high type) is

$$t - \left(t' - v(b, t)\left(\frac{t'}{t} - 1\right) + \epsilon \right) = (t - t')(1 - \frac{v(b, t)}{t}) - \epsilon \leq 0,$$

where the inequality followed since $t - t' \leq 0$, $v(b, t) \leq t$, and $\epsilon \geq 0$. Therefore the IC and IR constraints for the low type are satisfied. To verify the incentive constraints for the high type, notice that the utility of the high type from truthtelling is

$$t' - \left(t' - v(b, t)\left(\frac{t'}{t} - 1\right) + \epsilon \right) = v(b, t)\left(\frac{t'}{t} - 1\right) - \epsilon. \quad (16)$$

Since $v(b, t)/t > v(b, t')/t'$ and $v' > v$,

$$v(b, t)\left(\frac{t'}{t} - 1\right) > \max(0, v(b, t') - v(b, t)).$$

The inequality is strict, and thus for $\epsilon > 0$ small enough, the utility of the high type calculated in Equation (16) is at least $\max(0, v(b, t') - v(b, t))$, which is the utility the high type can receive from the outside option or reporting to be the low type. Thus the IC and IR constraints for the high type are satisfied, and the mechanism satisfies all constraints.

The revenue of the mixed bundling mechanism is

$$\begin{aligned} \left(1 - \frac{t}{t'}\right)v(b, t) + \frac{t}{t'}\left(t' - v(b, t)\left(\frac{t'}{t} - 1\right) + \epsilon\right) \\ = t + \frac{t}{t'}\epsilon, \end{aligned}$$

which is strictly higher than the optimal pure bundling revenue, t . Thus pure bundling is not optimal.

We now prove the “if” statement, starting by assuming that T is finite. We later extend the proof to arbitrary T using Lemma 8.

Consider $\bar{\lambda}$ and $\bar{\phi}$ from Lemma 7. Since $\bar{\phi}(b^*, t)$ is monotone non-decreasing in t , there exists a threshold t^* such that $\bar{\phi}(b^*, t) < 0$ if $t \leq t^*$, and $\bar{\phi}(b^*, t) \geq 0$ otherwise. Consider a pure bundling mechanism (a, p) that only offers the grand bundle b^* for price t^* . We show that (a, p) is optimal by arguing that the mechanism and $\bar{\lambda}$ satisfy conditions of Lemma 5.

Feasibility of $\bar{\lambda}$ follows directly from property (1) of Lemma 7. By property (2) of Lemma 7, and since $v(b, t_i)/t_i$ is monotone non-decreasing in t_i , $\bar{\phi}$ satisfies

$$\frac{v(b, t_i)}{t_i} \phi(b^*, t_i) \geq \phi(b, t_i), \forall t_i, b \in B.$$

As a result, using an argument identical to the informal proof of Proposition 1 in the main body (which is omitted), a maximizes virtual surplus.

It only remains to verify the complementary slackness condition (11). If a pair of types get the same assignment in (a, p) (either b^* or \emptyset), then they are indifferent for each other's assignment and complementary slackness is satisfied. Therefore we only need to verify complementary slackness between two types t_i and t_j such that $t_i < t^* \leq t_j$. By property (2) of Lemma 7, $\bar{\lambda}(i, j) = 0$. If $t_j = t^*$, then t_j gets utility of zero and is indifferent to the assignment of t_i . If $t_j > t^*$, then $\bar{\phi}(t_{j-1}) \geq \bar{\phi}(t^*) > \bar{\phi}(t_i)$. Property (4) of Lemma 7 implies that $\bar{\lambda}(j, i) = 0$.

We now extend the proof to arbitrary T (not necessarily finite). Consider a distribution μ with arbitrary support T . We argue that if pure bundling is not optimal for μ , then there must exist a distribution μ' whose support is a finite subset of T for which pure bundling is not optimal. If $v(b, t)/t$ is monotone non-decreasing over T , then it is also monotone non-decreasing over any subset of T . The analysis above shows that pure bundling is optimal for μ' . As a result, pure bundling is also optimal for μ .

Suppose for contradiction that pure bundling is not optimal for μ . Then there exists a mechanism (b, p) whose revenue is $\sigma > 0$ more than the optimal pure bundling revenue. Let p^* denote the optimal price for selling only the grand bundle, and let OPT_{PB} denote its revenue. We construct the support $\{t'_0, t'_1, \dots\}$ of μ' inductively from T . Let $t'_0 = \inf(T)$.

Set $\epsilon = \sigma/3$, and consider δ from Lemma 8. Given t'_i , define t'_{i+1} as follows,

$$t'_{i+1} = \sup\{t \in T \mid v(b, t) \leq v(b, t'_i) + \delta, \forall b\}.$$

Let the probability of t'_i in μ' be the probability of $[t'_i, t'_{i+1})$ in μ . The support of μ' is finite by the assumption that T is compact. Notice that distributions μ and μ' are δ -close. From Lemma 8, we conclude that there exists a mechanism (b', p') such that $\mathbf{E}_{\mu'}[p'(t)] \geq \mathbf{E}_{\mu}[p(t)] - \epsilon$. The optimal revenue among pure bundling mechanisms for distribution μ' is at most $OPT_{PB} + \epsilon \leq \mathbf{E}_{\mu}[p(t)] - \sigma + \epsilon < \mathbf{E}_{\mu}[p(t)] - \epsilon \leq \mathbf{E}_{\mu'}[p'(t)]$. Therefore, pure bundling is not optimal for distribution μ' . This contradicts our analysis above for distributions μ' with finite support. \square

A.4 Proof of Proposition 2

We prove Proposition 2 by showing that a mechanism that offers a bundle b at a discounted price in addition to the grand bundle at full price obtains higher revenue than any pure bundling mechanism.

Proposition 2. *Pure bundling is not optimal if $r(b, t)$ is stochastically decreasing in the value of the grand bundle for some b , and the distribution of values is continuous and interior.*

Proof. We prove a stronger claim. Let $\bar{r}(b, \hat{v})$ denote the largest value in the support of $r(b, t)$ conditioned on $v(b^*, t) = \hat{v}$. If $r(b, t)$ is stochastically decreasing in the value of the grand bundle, then $\bar{r}(b, \hat{v})$ is monotone decreasing in \hat{v} . To see this, notice that $0 = \mathbf{Pr}[r(b, t) \geq \bar{r}(b, \hat{v}) \mid v(b^*, t) = \hat{v}]$. Therefore, by definition of stochastic monotonicity, for any $\hat{v}' > \hat{v}$ there exists $\delta > 0$ such that $0 \geq \mathbf{Pr}[r(b, t) \geq \bar{r}(b, \hat{v}) - \delta \mid v(b^*, t) = \hat{v}']$. Therefore, $\bar{r}(b, \hat{v}') < \bar{r}(b, \hat{v})$.

Consider the optimal price p for only selling the grand bundle b^* . We prove the optimal pure bundling revenue can be improved by offering another bundle, contradicting the optimality of pure bundling. In particular, consider the change in revenue as a result of offering bundle b for a price $\bar{v}(b, p - \epsilon)$ in addition to b^* for price p . In this new mechanism, a type t chooses b^* if $v(b^*, t) - p \geq \max(0, v(b, t) - \bar{v}(b, p - \epsilon))$, chooses b if $v(b, t) - \bar{v}(b, p - \epsilon) \geq \max(0, v(b^*, t) - p)$, and chooses \emptyset otherwise. See Figure 9.

Now consider the change in revenue as a result of the discounted offer. The results are twofold. On one hand, a set of types $T_+(\epsilon)$ will pay $\bar{v}(b, p - \epsilon)$ for b (shaded dark in Figure 9). The gain in revenue $\nabla_+(\epsilon)$ is

$$\nabla_+(\epsilon) := \bar{v}(b, p - \epsilon) \times \mu_+(\epsilon),$$

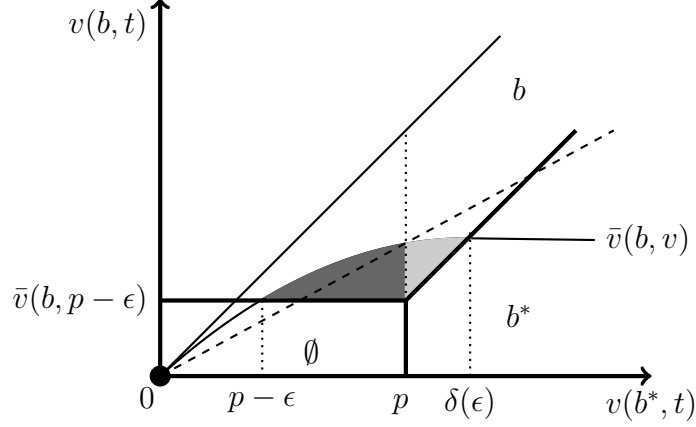


Figure 9: As a result of offering bundle b for a price $\bar{v}(b, p - \epsilon)$ in addition to b^* for a price p , there will be a gain and a loss in revenue. The gain is from the types in the region shaded dark. The loss is from the types in the region shaded light.

where $\mu_+(\epsilon)$ is the probability of the set of types $T_+(\epsilon)$,

$$\mu_+(\epsilon) := \mu(T_+(\epsilon)) = \int_{p-\epsilon}^p \int_{\bar{v}(b, p-\epsilon)}^{\bar{v}(b, v)} f_{b^*, b}(v, z) dz dv, \quad (17)$$

and $f_{b^*, b}$ denotes the density function of the joint distribution of the value for the grand bundle and value for bundle b . On the other hand, a set of types $T_-(\epsilon)$ with value slightly higher than p for b^* will change their decision from selecting bundle b^* to bundle b (shaded light in Figure 9). The loss of revenue $\nabla_-(\epsilon)$ is

$$\nabla_-(\epsilon) := (p - \bar{v}(b, p - \epsilon)) \times \mu_-(\epsilon),$$

where $\mu_-(\epsilon)$ is the probability of $T_-(\epsilon)$,

$$\mu_-(\epsilon) := \mu(T_-(\epsilon)) = \int_p^{p+\delta(\epsilon)} \int_{t-p+\bar{v}(b, p-\epsilon)}^{\bar{v}(b, v)} f_{b^*, b}(v, z) dz dv.$$

Note that $\nabla_+(0) = \nabla_-(0) = 0$. We now show that the gain is larger than the loss for small enough ϵ , $\nabla_+(\epsilon) > \nabla_-(\epsilon)$ by showing that $\nabla'_+(0) = \nabla'_-(0) = 0$, and $\nabla''_+(0) > \nabla''_-(0)$. Directly from the definitions, we have

$$\begin{aligned} \nabla'_+(\epsilon) &= -\partial_2 \bar{v}(b, p - \epsilon) \times \mu_+(\epsilon) + \bar{v}(b, p - \epsilon) \times \mu'_+(\epsilon), \\ \nabla''_+(\epsilon) &= \partial_{22} \bar{v}(b, p - \epsilon) \times \mu_+(\epsilon) + \bar{v}(b, p - \epsilon) \times \mu''_+(\epsilon). \end{aligned}$$

The derivatives of μ'_+ can be calculated as follows

$$\begin{aligned}
\mu'_+(\epsilon) &= - \int_{\bar{v}(b,p-\epsilon)}^{\bar{v}(b,p-\epsilon)} f_{b^*,b}(p-\epsilon, z) dz - \int_{p-\epsilon}^p \partial_2 \bar{v}(b, p-\epsilon) f_{b^*,b}(v, \bar{v}(b, p-\epsilon)) dv \\
&= - \int_{p-\epsilon}^p \partial_2 \bar{v}(b, p-\epsilon) f_{b^*,b}(v, \bar{v}(b, p-\epsilon)) dv, \text{ and} \\
\mu''_+(\epsilon) &= \partial_2 \bar{v}(b, p-\epsilon) f_{b^*,b}(p-\epsilon, \bar{v}(b, p-\epsilon)) - \int_{p-\epsilon}^p \frac{d}{d\epsilon} \partial_2 \bar{v}(b, p-\epsilon) f_{b^*,b}(v, \bar{v}(b, p-\epsilon)) dv.
\end{aligned} \tag{18}$$

Now notice from (17) that $\mu_+(0) = 0$ and from (18) that $\mu'_+(0) = 0$. Therefore, $\nabla'_+(0) = 0$. Now we can calculate $\nabla''_+(0)$ as follows

$$\nabla''_+(0) = \bar{v}(b, p) \times \mu''_+(0) = \bar{v}(b, p) \times \partial_2 \bar{v}(b, p) f_{b^*,b}(p, \bar{v}(b, p)). \tag{19}$$

Similarly we verify that $\nabla'_-(0) = 0$ and calculate $\nabla''_-(0)$.

$$\begin{aligned}
\nabla'_-(\epsilon) &= -\partial_2 \bar{v}(b, p-\epsilon) \times \mu_-(\epsilon) + (p - \bar{v}(b, p-\epsilon)) \times \mu'_-(\epsilon), \\
\nabla''_-(\epsilon) &= \bar{v}''_a(p-\epsilon) \times \mu_-(\epsilon) + (p - \bar{v}(b, p-\epsilon)) \times \mu''_-(\epsilon).
\end{aligned}$$

We calculate the derivatives of μ_- ,

$$\begin{aligned}
\mu'_-(\epsilon) &= \delta'(\epsilon) \int_{\delta(\epsilon)-p+\bar{v}(b,p-\epsilon)}^{\bar{v}(b,\delta(\epsilon))} f_{b^*,b}(\delta(\epsilon), z) dz \\
&\quad - \int_p^{\delta(\epsilon)} \partial_2 \bar{v}(b, p-\epsilon) f_{b^*,b}(v, v-p+\bar{v}(b, p-\epsilon)) dv \\
&= - \int_p^{\delta(\epsilon)} \partial_2 \bar{v}(b, p-\epsilon) f_{b^*,b}(v, v-p+\bar{v}(b, p-\epsilon)) dv. \\
\mu''_-(\epsilon) &= -\delta'(\epsilon) \partial_2 \bar{v}(b, p-\epsilon) f_{b^*,b}(\delta(\epsilon), \delta(\epsilon)-p+\bar{v}(b, p-\epsilon)) \\
&\quad - \int_p^{\delta(\epsilon)} \frac{d}{d\epsilon} \partial_2 \bar{v}(b, p-\epsilon) f_{b^*,b}(v, v-p+\bar{v}(b, p-\epsilon)) dv
\end{aligned} \tag{20}$$

As a result, we have that $\nabla'_-(0) = 0$, and

$$\nabla''_-(0) = (p - \bar{v}(b, p)) \times \delta'(0) \partial_2 \bar{v}(b, p) f_{b^*,b}(p, \bar{v}(b, p)).$$

So in order to complete the proof, we need to show that

$$\bar{v}(b, p) > (p - \bar{v}(b, p)) \delta'(0). \tag{21}$$

The above inequality directly follows from the monotonicity of the curve, as follows. The threshold $\delta(\epsilon)$ defines a type that is indifferent between choosing b^* and b ,

$$\delta(\epsilon) - p = \bar{v}(b, \delta(\epsilon)) - \bar{v}(b, p - \epsilon).$$

Differentiation with respect to ϵ and evaluating at $\epsilon = 0$ gives

$$\delta'(0)(1 - \partial_2 \bar{v}(b, p)) = \partial_2 \bar{v}(b, p).$$

Substituting into (21), we need to show that

$$\bar{v}(b, p) > (p - \bar{v}(b, p)) \frac{\partial_2 \bar{v}(b, p)}{1 - \partial_2 \bar{v}(b, p)}. \quad (22)$$

Since by assumption $\partial_2 \bar{v}(b, p) < \bar{v}(b, p)/p$, we have

$$\frac{\partial_2 \bar{v}(b, p)}{1 - \partial_2 \bar{v}(b, p)} < \frac{\bar{v}(b, p)}{p - \bar{v}(b, p)},$$

which is identical to (22), completing the proof. \square

A.5 Proof of Proposition 3

Proposition 3. *Suppose that each bundle b costs $c(b)$ and the grand bundle is the efficient non-empty bundle. Pure bundling is optimal if the transformed profile of relative values is stochastically non-decreasing in positive transformed values of the grand bundle; i.e., $\Pr[\tilde{r}(\cdot, t) \in U \mid \tilde{v}(b^*, t) = \hat{v}]$ is non-decreasing in $\hat{v} > 0$ for all upper sets $U \subseteq \mathbb{R}^B$.*

Proof. The proof is provided in the main body. The only subtlety is to ensure that the Theorem 1 applies to the transformed setting where values may be negative. In the transformed setting, we can assume without loss of generality that each type's value of the grand bundle is non-negative. This is because given the assumption that grand bundle is the efficient non-empty bundle, if the value of the grand bundle is negative for some type t , then value of all non-empty bundles is negative for that type. As a result, type t must be excluded in any optimal mechanism. Thus we can assume that such a type has zero value for all bundles without affecting optimal mechanisms. In the proof of Theorem 1, the values of bundles other than the grand bundle may be negative. In fact, the decomposition approach that is used to prove the theorem from Proposition 1 does not depend on the sign of the random

variable. The proof of Proposition 1 also does not require that the values of bundles other than the grand bundle are non-negative. \square

A.6 Proof of Corollary 1

Corollary 1. *Suppose that each non-empty bundle costs $c \geq 0$. Pure bundling is optimal if the profile of relative values is stochastically non-decreasing in positive values of the grand bundle.*

Proof. To apply Proposition 3, we show that the probability of the event $\tilde{r}(\cdot, t) \in U$ conditioned on $\tilde{v}(b^*, t) = \hat{v}$ is non-decreasing in $\hat{v} > c$ for any upper set U . It is convenient to rephrase the event $\tilde{r}(\cdot, t) \in U$ in terms of the relative values r . To do so, for any c and \hat{v} , let $U(c, \hat{v})$ be the set of all $x \in \mathbb{R}^B$ for which there exists some $y \in U$ such that $(x(b)(\hat{v} + c) - c)/\hat{v} = y(b)$ for all b . Conditioned on $\tilde{v}(b^*, t) = \hat{v}$, the event $\tilde{r}(\cdot, t) \in U$ is equivalent to the event $r(\cdot, t) \in U(c, \hat{v})$. In fact, suppose that $\tilde{v}(b^*, t) = \hat{v}$. Then

$$\frac{r(b, t)(\hat{v} + c) - c}{\hat{v}} = \frac{v(b, t) - c}{\tilde{v}(b^*, t)} = \frac{\tilde{v}(b, t)}{\tilde{v}(b^*, t)} = \tilde{r}(b, t),$$

for all b . Thus, by definition of $U(c, \hat{v})$, $\tilde{r}(\cdot, t) \in U$ is equivalent to the event $r(\cdot, t) \in U(c, \hat{v})$.

Notice two properties of $U(c, \hat{v})$. First, $U(c, \hat{v})$ is an upper set. To see this, suppose that $x \in U(c, \hat{v})$ and consider x' such that $x' \geq x$. By definition, there exists $y \in U$ such that $(x(b)(\hat{v} + c) - c)/\hat{v} = y(b)$ for all b . Define y' so that $y'(b) = (x'(b)(\hat{v} + c) - c)/\hat{v}$. Notice that $y' \geq y$. Therefore, since U is an upper set, we have $y' \in U$ and thus $x' \in U(c, \hat{v})$. Second, if $x \in U(c, \hat{v})$, $x \leq 1$, and $\hat{v}' \geq \hat{v}$, then $x \in U(c, \hat{v}')$. To see this, since $x \in U(c, \hat{v})$, then for some $y \in U$, we have $(x(b)(\hat{v} + c) - c)/\hat{v} = y(b)$ for all b . Define y' so that $y'(b) = (x(b)(\hat{v}' + c) - c)/\hat{v}'$. Notice that $y' \geq y$. Since U is an upper set and $y \in U$, we must also have $y' \in U$ and therefore $x \in U(c, \hat{v}')$.

Now consider $c < \hat{v} < \hat{v}'$. We have

$$\begin{aligned} \Pr \left[\tilde{r}(\cdot, t) \in U \mid \tilde{v}(b^*, t) = \hat{v}' \right] &= \Pr \left[r(\cdot, t) \in U(c, \hat{v}') \mid \tilde{v}(b^*, t) = \hat{v}' \right] \\ &\geq \Pr \left[r(\cdot, t) \in U(c, \hat{v}) \mid \tilde{v}(b^*, t) = \hat{v}' \right] \\ &\geq \Pr \left[r(\cdot, t) \in U(c, \hat{v}) \mid \tilde{v}(b^*, t) = \hat{v} \right] \\ &= \Pr \left[\tilde{r}(\cdot, t) \in U \mid \tilde{v}(b^*, t) = \hat{v} \right]. \end{aligned}$$

The first inequality followed since as argued above, $r(\cdot, t) \in U(c, \hat{v})$, $r(\cdot, t) \leq 1$, and $\hat{v}' \geq \hat{v}$

imply that $r(\cdot, t) \in U(c, \hat{v})$. The second inequality followed by the assumption that r is stochastically non-decreasing in the value of the grand bundle and the fact that $U(c, \hat{v})$ is an upper set. \square