When Is Pure Bundling Optimal?*

Nima Haghpanah† Jason Hartline‡

July 19, 2019

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 each type has a distinct value 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, Ron Siegel, Dan Vincent, Rakesh Vohra, Jidong Zhou, and various seminar participants. We thank Berk Idem, Xiao Lin, and Garima Singal for excellent research assistance. 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, Penn State University, nima.haghpanah@gmail.com.

‡EECS Department, Northwestern University, hartline@eecs.northwestern.edu.

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 $b \subseteq \{1,\ldots,n\}$, 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 $gb = \{1,\ldots,n\}$? Or should it 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 $v(b,t)/v(gb,t)$ for each bundle b, change with the buyer's value for the grand bundle $v(gb,t)$. 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 take the discounted offer. The loss is from the spillover of types who would have paid the full price if the discounted offer were not available, but now choose 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 spillover is large, and thus mixed bundling is not profitable, if types with higher value for the grand bundle are more likely to have high relative values. Conversely, the spillover is small, and thus mixed bundling is profitable, if types with higher value for the bundle are more likely to have low relative values.

The literature on multi-product bundling mostly assumes that values are additive, i.e., the value for any bundle is the sum of the values for its constituting products.[1] Pure bundling

[1]Some exceptions are Long (1984), Armstrong (2013), and Armstrong (2016), where the focus is identifying when it is profitable for the seller to offer the bundle at a price that is less that the sum of the prices for individual products.

is generally not optimal with additive values. Adams and Yellen (1976) and McAfee et al. (1989) show that pure bundling is generally strictly dominated by mixed bundling, i.e., offering all bundles. Furthermore, mixed bundling is itself dominated by offering randomized bundles (Thanassoulis, 2004; Daskalakis et al., 2017). Nevertheless, a conventional view in the literature is that bundling is profitable, compared to selling the products separately, if the values for individual products are negatively correlated (Church and Ware, 2000). This view is mostly based on examples provided by Stigler (1963) and Adams and Yellen (1976) where values are perfectly negatively correlated, i.e., the sum of values is the same for all types. With perfect negative correlation, pure bundling is indeed optimal since it extracts the full surplus. The optimality of pure bundling with perfect negative correlation is a straightforward corollary of our result. If all types have the same value for the grand bundle, then relative values are trivially stochastically non-decreasing in the value for the grand bundle.

We do not assume that values are additive. Products may be partial substitutes or partial complements. They may even be partial substitutes for some types but partial complements for others. In fact, a key insight of our result, which we provide using an example with two products, is that the optimality of pure bundling depends on how the "relative synergy" between the products changes across types. Relative synergy is the ratio of the value for the grand bundle over the sum of the values for individual products, and measures the complementarity between products.[2] It is larger than one for a type for which the products are partial complements, and is smaller than one for a type for which the products are partial substitutes. Our condition for the optimality of pure bundling requires that the relative synergy is lower for types with higher values, that is, they consider the products to be less complementary. Notably, pure bundling may be optimal even if the products are partial substitutes for all types.

We impose little structure on the set of bundles. More generally than the setting discussed above, a bundle may be any set of divisible or indivisible products, may contain multiple units of each product, and may be randomized. A special case is when bundles are vertically differentiated, i.e., bundles are ranked in a way that each type has a higher value for a higher ranked bundle. Vertically differentiated bundles may represent different quantities or quality levels of a single product. The grand bundle is the most desirable bundle and represents the highest quantity or quality level. Thus our result provides a condition for the optimality of selling only the highest quantity or quality level. In general, however, we do not require

[2]We thank an anonymous referee for suggesting this terminology.

types to agree on the ranking of bundles, but only to agree that the grand bundle is the most desirable bundle.

We mostly focus on the case where production costs are zero. Thus our results mainly concerns markets for information goods (cable TV, software, movies, and music). The assumption of zero costs allows us to highlight the tradeoffs involved in screening types, i.e., market coverage versus spillover, and abstract away from economies of scale and scope. In an extension, we relax the assumption of zero costs and provide a condition for selling all bundles at a uniform markup above cost.

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 for 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 Cai et al. (2016) and Carroll (2017). The construction is novel and shows that ironed virtual valuations can be constructed from only downward incentive constraints (that is, other 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. Wilson (1993) and Armstrong (1996) first use this approach where, translated to our setting, each path is a ray from the origin in the value space. Eső and Szentes (2007) and Pavan et al. (2014) significantly advance this idea by allowing the decomposition to depend on the distribution of types. In particular, they decompose the type space into a base parameter and independently distributed "shocks". 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.

Related Work. Our condition is related to those of Salant (1989), Johnson and Myatt (2003), and Anderson and Dana Jr (2009). These papers study price discrimination with vertically differentiated products (Johnson and Myatt, 2003 further study price discrimination in a duopoly). They show that selling the highest quality product is optimal if there are increasing returns to quantity, a condition related to our ranking of relative values. However, since it is assumed that products are vertically differentiated, these models do not naturally apply to selling bundles of products. In addition, these papers assume that types are ranked, which maps in our setting to the case where types are on a path, and thus the results are more restrictive than our general stochastic monotonicity condition. Finally, these papers assume regularity conditions that are relaxed in this paper via ironing.

Numerous papers study optimal mechanisms for a multi-product monopolist. Bakos and Brynjolfsson (1999) show that pure bundling is optimal for selling a large number of products with independently distributed values. The reason is that bundling reduces the dispersion of consumer values. Without a large number of products, closed form solutions are known only for special cases. For instance, Daskalakis et al. (2017) identify optimal mechanisms (pure bundling and otherwise) for certain uniform distributions. Pavlov (2011) and Menicucci et al. (2015) provide sufficient conditions for the optimality of pure bundling when selling two products with additive and independently distributed values. These 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) considers two products with a bivariate Gaussian distribution of values, and shows mainly via numerical results that negative correlation implies optimality of pure bundling. On the other hand, Schmalensee (1984) also shows that pure bundling may be profitable even with positively correlated values. Fang and Norman (2006) analytically confirm the numerical results of Schmalensee (1984) for any number of products, but only compare pure bundling with separate sales (offering each individual product at a price).

1 Model and Main Result

Consider a screening problem with a single seller and a single buyer. There is a compact set of bundles B. An example is when $B = \{b \mid b \subseteq \{1, \dots, n\}\}$ is the set of all subsets of n products, although in general we impose no such structure. The cost of producing a bundle $b \in B$ is zero. There is a compact set of buyer types T. The utility of a type $t \in T$ for a bundle b and payment p to the seller is $v(b, t) - p$. Assume that v is non-negative and continuous. Note that any function v is continuous if B and T are finite. The buyer's type is its private information. The seller has a prior belief in the form of a distribution $\mu \in \Delta(T)$ over the types of the buyer. Assume that there exists a bundle $gb \in B$ such that $v(gb, t) \geq v(b, t)$ for all b and t. We refer to bundle gb as the grand bundle. Let $\emptyset \in B$ denote the outside option and normalize $v(\emptyset, t) = 0$ for all t. We refer to a tuple (B, T, v) as an environment. Some applications are discussed in Section 4.1.

We invoke the revelation principle and focus on direct mechanisms. A mechanism is a pair of functions, a bundle assignment rule $b : T \to B$ and a payment rule $p : T \to \mathbb{R}$. The mechanism (b, p) is incentive compatible (IC) if no type increases its utility by misreporting,

$$v(b(t), t) - p(t) \geq v(b(t'), t) - p(t'), \qquad \forall t, t' \in T.$$

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

$$v(b(t), t) - p(t) \geq 0, \qquad \forall t \in T.$$

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

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

over all IC and IR mechanisms.

A mechanism is a pure bundling mechanism if it offers only the grand bundle at some price p. That is, if $v(gb, t) \geq p$ then $b(t) = gb$ and $p(t) = p$, and otherwise $b(t) = \emptyset$ and $p(t) = 0$. Such a mechanism is IC and IR. We say that pure bundling is optimal if there exists a price p such that the pure bundling mechanism with 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 for the grand bundle. Below we define these terms and formally give the main theorem.

For a type t, let $r(b, t) = v(b, t)/v(gb, t) \in [0, 1]$ be the relative value for a bundle b to the value for 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.

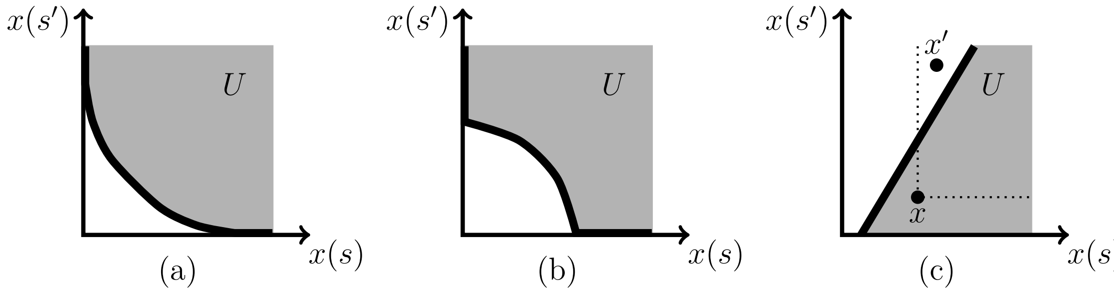

Figure 1: Sets $U \subseteq \mathbb{R}^{\{s,s'\}}$ in (a) and (b) are upper sets. The set $U \subseteq \mathbb{R}^{\{s,s'\}}$ in (c) is not an upper set.

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, if an upper set includes a function, then it also includes all larger functions. Some examples are shown in Figure 1. Upper sets are defined formally below, where we write $x \leq x'$ for functions $x, x' \in \mathbb{R}^S$ if $x(s) \leq x'(s)$ for all $s \in S$.

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, that is, a 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 $\mathbf{Pr}[x \in U \mid y = \hat{y}]$ 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 non-zero values for the grand bundle; i.e., $\mathbf{Pr}[r(\cdot, t) \in U \mid v(gb, t) = \hat{v}]$ 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. Section 2 specializes Theorem 1 to the case where there are only two bundles and provides two examples that illustrate the theorem and its proof. Section 3 proves the main result and provides some partial converses. Section 4 discusses some applications and interpretations. We conclude the paper in Section 5. Missing proofs are deferred to the appendix.

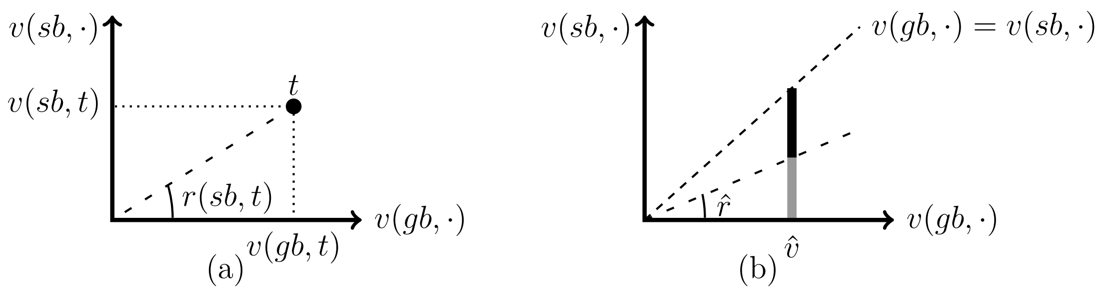

Figure 2: (a) Each type can be visualized by its values for the two bundles. The relative value is the slope of the line from the origin to the type. (b) $\mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$ is the probability that a type is in the dark-shaded set, conditioned on it being in the shaded set. Corollary 1 requires that this probability is non-decreasing in $\hat{v}$ for all $\hat{r}$.

2 Illustrative Examples

In this section we specialize Theorem 1 to the case of only two bundles and provide two examples that illustrate and the main idea behind the proof of the theorem. A reader who is only interested in our general treatment can skip ahead to Section 3.

Assume that B consists of only two bundles (other than the outside option): a "small bundle" sb and the grand bundle gb. A type t can be visualized by its values for the two bundles, as shown in Figure 2, (a). Each type t has only one non-trivial relative value $r(sb,t) = v(sb,t)/v(gb,t)$, which is the slope of the line from the origin to type t.

Theorem 1 specializes as follows. Pure bundling is optimal if the relative value $r(sb,t)$ is stochastically non-decreasing in the value for the grand bundle $v(gb,t)$. That is, the probability of $r(sb,t) \in U$ conditioned on $v(gb,t) = \hat{v}$ is non-decreasing in $\hat{v}$ for all upper sets $U \subseteq \mathbb{R}$. A set U of real numbers is an upper set if it contains all numbers greater or equal to some threshold $\hat{r}$. For such a set, $r(sb,t) \in U$ is equivalent to $r(sb,t) \geq \hat{r}$. Therefore, pure bundling is optimal if the probability of $r(sb,t) \geq \hat{r}$ conditioned on $v(gb,t) = \hat{v}$ is non-decreasing in $\hat{v}$ for all $\hat{r}$. The conditional probability $\mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$ is depicted in Figure 2, (b).

Corollary 1. *Assume that* $B = \{\emptyset, sb, gb\}$. *Pure bundling is optimal if* $\mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$ *is non-decreasing in* $\hat{v} > 0$ *for all* $\hat{r} \in \mathbb{R}$.

The two examples below illustrate the ideas behind the proof of Theorem 1. The first example considers a case where types have distinct values for the grand bundle. The second example relaxes this assumption. We solve the second example by decomposing it into problems where types do in fact have distinct values for the grand bundle. For each of these

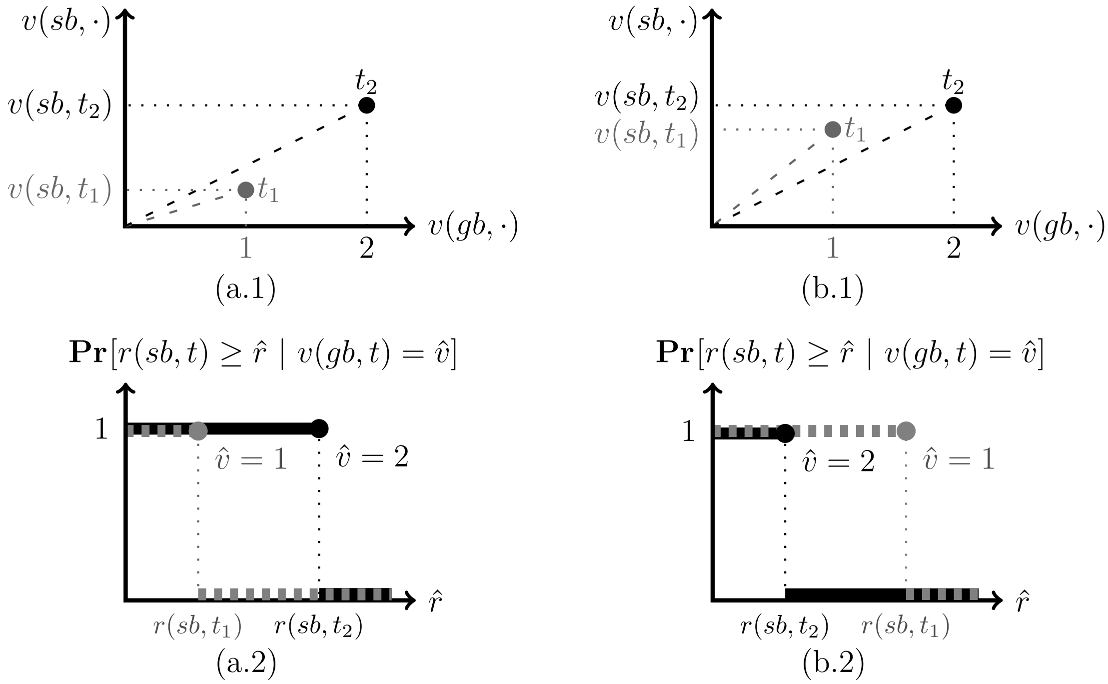

Figure 3: Panels (a.1) and (a.2) correspond to $r(sb,t_1) \leq r(sb,t_2)$. Panels (b.1) and (b.2) correspond to $r(sb,t_1) > r(sb,t_2)$. The top panels (a.1) and (b.1) depict the values of types. The bottom panels (a.2) and (b.2) show the probability of $r(sb,t) \geq \hat{r}$ conditioned on $v(gb,t) = \hat{v}$ for all $\hat{r}, \hat{v}$.

examples, we first interpret our main theorem, and then give an analysis that corroborates the theorem.

2.1 Two Types: Distinct Values for Grand Bundle

Our first example considers two types $T = \{t_1, t_2\}$ with distinct values for the grand bundle. Type t_1 (the low type) has value 1 and type t_2 (the high type) has value 2 for the grand bundle. A critical condition is whether (a) the lower type has a weakly lower relative value, i.e., $r(sb,t_1) \leq r(sb,t_2)$ as shown in panel (a.1) of Figure 3, or (b) the lower type has a higher relative value, i.e., $r(sb,t_1) > r(sb,t_2)$ as shown in panel (b.1) of Figure 3.

Corollary 1 states that pure bundling is optimal in case (a). Panel (a.2) of Figure 3 shows the conditional probabilities $\mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$ in case (a), and panel (b.2) shows these conditional probabilities in case (b). As the values for the grand bundle are distinct, these conditional distributions are point masses that do not depend on the probabilities of types. In case (a) where $r(sb,t_1) \leq r(sb,t_2)$, we have $\mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = 1] \leq \mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = 2]$ for all $\hat{r}$, and the condition of Corollary 1 is met.

The following algebraic analysis corroborates the result of Corollary 1. Let $q = \mathbf{Pr}[t = t_2]$ specify the probability of the high type. The optimal revenue among pure bundling mechanisms is $\max(1, 2q)$. 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 type t_2 buys the grand bundle, and the revenue is $2q$. Any other price results in revenue less than 1 or $2q$.

Now consider a "mixed bundling" mechanism that offers the small bundle sb at price $v(sb, t_1)$ and the grand bundle at price $2 - \Big(v(sb, t_2) - v(sb, t_1)\Big)$. Type t_1 is indifferent between the outside option and bundle sb, and type t_2 is indifferent between bundles sb and gb (with utility $v(sb, t_2) - v(sb, t_1)$ from either option, which we assume to be non-negative to avoid an extra case that yields the same conclusion). Breaking ties to maximize revenue, type t_1 chooses bundle sb, and type t_2 chooses bundle gb. The revenue is

$$(1-q)v(sb, t_1) + q\Big(2 - \Big(v(sb, t_2) - v(sb, t_1)\Big)\Big) = v(sb, t_1) + q\Big(2 - v(sb, t_2)\Big).$$

We now show that the revenue of the mixed bundling mechanism is at most the optimal revenue among pure bundling mechanisms if $r(sb, t_1) \leq r(sb, t_2)$, that is, $v(sb, t_1) \leq \frac{1}{2}v(sb, t_2)$. First suppose $q \geq \frac{1}{2}$. The revenue of the mixed bundling mechanism is

$$\begin{aligned} v(sb, t_1) + q\Big(2 - v(sb, t_2)\Big) &= 2q + v(sb, t_1) - qv(sb, t_2) \\ &\leq 2q + v(sb, t_1) - \tfrac{1}{2}v(sb, t_2) \\ &\leq 2q, \end{aligned}$$

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

$$\begin{aligned} v(sb, t_1) + q\Big(2 - v(sb, t_2)\Big) &\leq v(sb, t_1) + \tfrac{1}{2}\Big(2 - v(sb, t_2)\Big) \\ &= 1 + v(sb, t_1) - \tfrac{1}{2}v(sb, t_2) \\ &\leq 1, \end{aligned}$$

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

The condition of Corollary 1 is also partially necessary for the optimality of pure bundling. In particular, if $q = \frac{1}{2}$, then pure bundling is optimal if and only if $r(sb, t_1) \leq r(sb, t_2)$. If $q = \frac{1}{2}$, the revenue of the mixed bundling mechanism is $1 + v(sb, t_1) - \frac{1}{2}v(sb, t_2)$, which is no larger than the optimal revenue among pure bundling mechanisms, i.e., 1, if and only

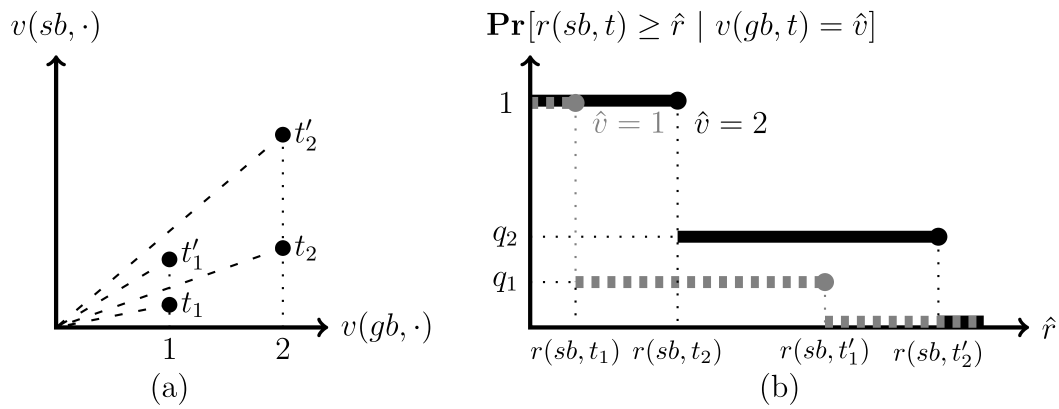

Figure 4: (a) Relative values are ordered as $r(sb, t_1) \leq r(sb, t_2) \leq r(sb, t_1') \leq r(sb, t_2')$. (b) The probability of event $r(sb, t) \geq \hat{r}$ conditioned on $v(gb, t) = 2$ is weakly higher than that conditioned on $v(gb, t) = 1$ for all $\hat{r}$ if and only if $q_1 \leq q_2$.

if $r(sb, t_1) \leq r(sb, t_2)$. We generalize this observation in Section 3.1 to provide a partial converse to Theorem 1. For $q \neq \frac{1}{2}$, $r(sb, t_1) \leq r(sb, t_2)$ is not necessary for the optimality of pure bundling.[3]

2.2 Four Types: Non-distinct Values for Grand Bundle

In the second example there are four types $T = \{t_1, t_1', t_2, t_2'\}$ with non-distinct values for the grand bundle. Types t_1 and t_1' have value 1 and types t_2 and t_2' have value 2 for the grand bundle. Assume, as depicted in Figure 4, (a), that the relative values are ordered as

$$r(sb, t_1) \leq r(sb, t_2) \leq r(sb, t_1') \leq r(sb, t_2'). \qquad (1)$$

We first interpret the condition of Corollary 1. We then corroborate Corollary 1 by decomposing the distribution into distributions over pairs of types, and applying our two type analysis.

The condition of Corollary 1 holds when the probability $q_1 = \mathbf{Pr}[t = t_1' \mid t \in \{t_1, t_1'\}]$ of high value for the small bundle conditioned on low value for the grand bundle, is at most the probability $q_2 = \mathbf{Pr}[t = t_2' \mid t \in \{t_2, t_2'\}]$ of high value for the small bundle conditioned on high value for the grand bundle. Indeed, the conditional distributions of $r(sb, t)$ can be

[3]In our example, pure bundling is optimal if q is sufficiently small or sufficiently large. In particular, if $r(sb, t_1) > r(sb, t_2)$, then pure bundling is optimal if $q \in [0, (1-v(sb, t_1))/(2-v(sb, t_2))] \cup [v(sb, t_1)/v(sb, t_2), 1]$.

written as follows

$$r(sb,t) \mid (v(gb,t)=1) = \begin{cases} r(sb,t_1) & \text{with probability } 1-q_1, \\ r(sb,t_1') & \text{with probability } q_1, \end{cases}$$

and,

$$r(sb,t) \mid (v(gb,t)=2) = \begin{cases} r(sb,t_2) & \text{with probability } 1-q_2, \\ r(sb,t_2') & \text{with probability } q_2. \end{cases}$$

The conditional probabilities $\mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$ are shown in Figure 4, (b). If $q_1 \leq q_2$, then we have $\mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = 1] \leq \mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = 2]$ for all $\hat{r}$.

Below we give an analysis that corroborates Corollary 1. With four types, verifying the optimality of pure bundling is no longer a straightforward algebraic exercise (unlike Section 2.1), since the number of possible mechanisms is quite large. Nevertheless, we can prove the optimality of pure bundling by decomposing the distribution into distributions over pairs of types. By our two type analysis above, pure bundling is optimal for any distribution supported on $\{t_1, t_2\}$, $\{t_1, t_2'\}$, or $\{t_1', t_2'\}$, where the higher type has a higher relative value. On the other hand, pure bundling is not generally optimal for a distribution over the pair of types $\{t_1', t_2\}$. The assumption that $q_1 \leq q_2$ enables us to appropriately decompose the distribution into distributions supported on $\{t_1, t_2\}$, $\{t_1, t_2'\}$, and $\{t_1', t_2'\}$.

Formally, let $q = \mathbf{Pr}[t \in \{t_2, t_2'\}]$ denote the probability of high value for the grand bundle. The distribution of types μ can be parameterized using q, q_1, and q_2,

$$\begin{aligned} &\begin{pmatrix} \mathbf{Pr}[t_1] & \mathbf{Pr}[t_1'] & \mathbf{Pr}[t_2] & \mathbf{Pr}[t_2'] \end{pmatrix} \\ &= \begin{pmatrix} (1-q)(1-q_1) & (1-q)q_1 & q(1-q_2) & qq_2 \end{pmatrix}, \end{aligned}$$

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

$$= (1-q_2)\begin{pmatrix} 1-q & 0 & q & 0 \end{pmatrix} + (q_2 - q_1)\begin{pmatrix} 1-q & 0 & 0 & q \end{pmatrix} + q_1 \begin{pmatrix} 0 & 1-q & 0 & q \end{pmatrix}.$$

Let μ^1, μ^2, and μ^3 denote the above three distributions. The supports of the three distributions is shown in Figure 5. A random type from μ can be drawn by first selecting one of μ^1, μ^2, and μ^3 with probabilities $1-q_2$, $q_2 - q_1$, and q_1, respectively, and then drawing a type from the selected distribution. Notice that $q_1 \leq q_2$ is needed to ensure that the

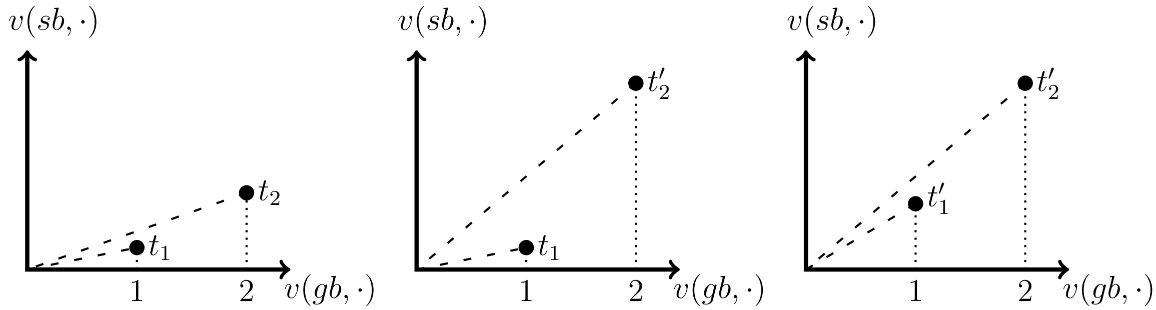

Figure 5: The supports of the three distributions μ^1, μ^2, and μ^3.

probability $q_2 - q_1$ of distribution μ^2 is non-negative. When $q_1 > q_2$ the decomposition must assign positive probability to the pair of types $\{t_1', t_2\}$ (where pure bundling is not generally optimal).

Now consider a relaxed problem in which the seller can observe which of the three distributions μ^1, μ^2, or μ^3 is selected, and can design a mechanism accordingly. Since the seller can simply ignore this information, the optimal revenue in the relaxed problem upper bounds the optimal revenue in the original problem. Our two type analysis implies that the optimal revenue for each of the three distributions is $\max(1, 2q)$. Indeed, distributions μ^1, μ^2, and μ^3 are each supported on two types with values 1 and 2 for the grand bundle, where the higher type also has a higher relative value. Therefore, the optimal revenue in the relaxed problem is $\max(1, 2q)$. Since the optimal revenue among pure bundling mechanisms for distribution μ achieves this upper bound of $\max(1, 2q)$, pure bundling is optimal for distribution μ.

Randomized bundles. Our model can easily allow for randomization, once the set of bundles and the values are appropriately defined. In particular, consider the set of distributions $\tilde{B} = \Delta(B)$ over bundles $\emptyset$, sb, and gb, and let $\tilde{v}(\tilde{b}, t) = \mathbf{E}_{b \sim \tilde{b}}[v(b, t)]$ be the expectation of v for any distribution $\tilde{b} \in \tilde{B}$. In environment $(\tilde{B}, T, \tilde{v})$, the seller can sell distributions over bundles B. Pure bundling is optimal under the conditions discussed above. That is, pure bundling is optimal in environment $(\tilde{B}, T, \tilde{v})$ if $q_1 \leq q_2$. This claim requires a proof, to which we return after proving our main result.

3 Proof of Theorem 1 and Converses

In this section we prove Theorem 1, and provide two partial converses. Following the outline of Section 2, we start by proving a special case of the result where types have distinct values for the grand bundle, generalizing the two type analysis of Section 2.1. We then prove Theorem 1 by generalizing the decomposition approach of Section 2.2.

3.1 Paths: Distinct Values for Grand Bundle

Assume that types have distinct values for the grand bundle, that is $v(gb, t) \neq v(gb, t')$ for all $t \neq t'$. Thus we assume without loss of generality that t equals the value for the grand bundle, that is $T \subseteq \mathbb{R}^+$ and $v(gb, t) = t$.[4] 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. When types are on a path v, the profile of relative values is stochastically non-decreasing in the value for the grand bundle if $r(\cdot, t\,) = v(\cdot, t\,)/t$ 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 proposition shows that pure bundling is indeed optimal if $v(\cdot, t)/t$ is monotone non-decreasing in t. 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 $v(gb, t) = t$. Pure bundling is optimal for all distributions $\mu \in \Delta(T)$ if and only if $v(\cdot, t)/t$ is monotone non-decreasing in $t > 0$.*

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 for the grand bundle against the value for any bundle b, the slope of a ray from the origin to a type is non-decreasing along the support, as in Figure 6, (a). Alternatively, a ray from the origin intersects the support from above and continues below.

We start by proving the "only if" statement. The proof is a generalization of the two type argument provided in Section 2.1.

Proof of Proposition 1, "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

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

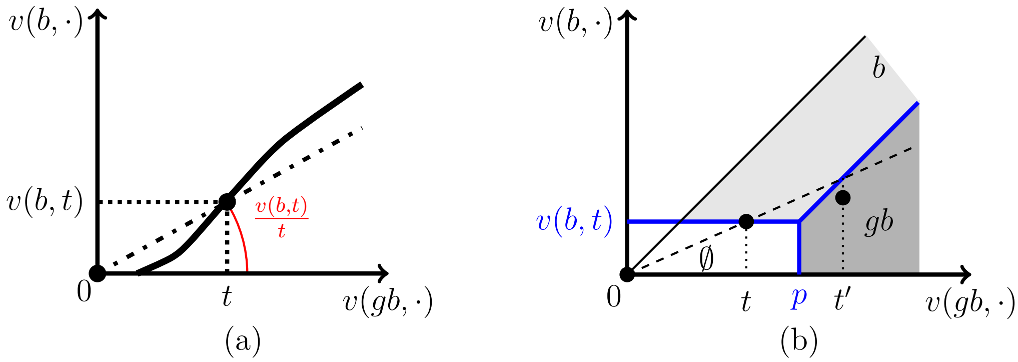

Figure 6: (a) The relative value $v(b,t)/t$ is non-decreasing along the support. $v(b,t)$ need not be convex in t. (b) The mixed bundling mechanism of Proposition 1. If $\epsilon = 0$, the price of the grand bundle $p = t' - v(b,t)(\frac{t'}{t} - 1)$ is set such that the high type chooses gb if and only if $v(b,t')/t' \leq v(b,t)/t$.

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. See Figure 6, (b). 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

$$t - \Big(t' - v(b,t)(\tfrac{t'}{t} - 1) + \epsilon\Big)$$
$$= (t - t')(1 - \tfrac{v(b,t)}{t}) - \epsilon \leq 0,$$

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

$$t' - \Big(t' - v(b,t)(\tfrac{t'}{t} - 1) + \epsilon\Big) = v(b,t)(\tfrac{t'}{t} - 1) - \epsilon. \tag{2}$$

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

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

The inequality is strict, and thus for $\epsilon > 0$ small enough, the utility of the high type calculated in equation (2) 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

$$(1 - \tfrac{t}{t'})v(b,t) + \tfrac{t}{t'}\Big(t' - v(b,t)(\tfrac{t'}{t} - 1) + \epsilon\Big)$$
$$= t + \tfrac{t}{t'}\epsilon,$$

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

Proof of the "if" Statement with Regularity Assumptions. 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 for 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 $b \in B$. Third, the marginal distribution of the value for the grand bundle is regular, as defined next.

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

$$\phi(b,t) = v(b,t) - \partial_2 v(b,t) \times \frac{1 - F_{gb}(t)}{f_{gb}(t)}. \tag{3}$$

Recall that $v(gb,t) = t$, which implies that $\phi(gb,t) = t - \frac{1-F_{gb}(t)}{f_{gb}(t)}$.[5] Note that $\phi(\emptyset,t) = 0$. We say that the marginal distribution of the value for the grand bundle is regular if $\phi(gb,t)$ is monotone non-decreasing in t.

The proof 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 (b,p) is equal to its expected virtual surplus $\mathbf{E}[\phi(b(t),t)]$, up to a constant which is the utility of type $\underline{t}$.

[5]This 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 for bundle b, and f_b its density. That is, the virtual value for each bundle b is equal to the virtual value for 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 (3) in our proof since it facilitates the comparison of virtual values based on the curvature of v.

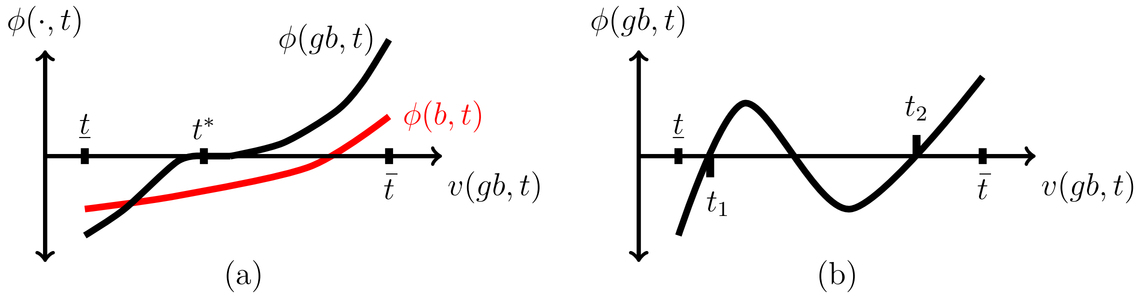

Figure 7: (a) If v is ratio-monotone, then $\phi(b,t) \leq \max(0, \phi(gb,t))$ for all b, t. If $\phi(gb,t)$ is non-decreasing, then there exists a threshold t^* such $\phi(gb,t)$ is non-positive below t^* and non-negative above t^*. (b) The revenue of selling only the grand bundle at any price is strictly less than $\mathbf{E}[\max(0, \phi(gb,t))]$.

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

$$\mathbf{E}\left[p(t)\right] = \mathbf{E}\left[\phi(b(t),t)\right] - \Big(v(b(\underline{t}),\underline{t}) - p(\underline{t})\Big).$$

The second lemma follows directly from the definition of virtual values in Equation (3), and allows us to compare the virtual values for bundles b and gb. If v is ratio-monotone, then for any type t, the virtual value for any bundle b is at most either 0 or the virtual value for the grand bundle, as shown in Figure 7, (a).

Lemma 2. *If* $v(\cdot,t)/t$ *is monotone non-decreasing in* t*, then* $\phi(b,t) \leq \max(0, \phi(gb,t))$ *for all* b *and* 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 (b, p) is

$$\begin{aligned}\mathbf{E}\left[\phi(b(t),t)\right] - \Big(v(b(\underline{t}),\underline{t}) - p(\underline{t})\Big) &\leq \mathbf{E}\left[\phi(b(t),t)\right] \\ &\leq \mathbf{E}\left[\max(0, \phi(gb,t))\right], \qquad (4)\end{aligned}$$

where the first equality followed from IR, and the second equality followed from Lemma 2. Since $\phi(gb,t)$ is monotone, there exists a threshold t^* such that $\phi(gb,t) \leq 0$ for all $t \leq t^*$, and $\phi(gb,t) \geq 0$ for all $t \geq t^*$, as shown in Figure 7, (a). The revenue of selling only the grand bundle at price t^* is equal to $\mathbf{E}[\max(0, \phi(gb,t))]$ by Lemma 1. Thus by Inequality (4), 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 for the grand bundle is not regular. Without regularity, $\mathbf{E}[\max(0, \phi(gb,t))]$ is still an upper bound on the revenue of any mechanism since Lemma 2 and Inequality (4) require ratio-monotonicity of v but not monotonicity of $\phi(gb, t)$. However, if $\phi(gb, t)$ is not monotone, as in Figure 7, (b), then $\mathbf{E}[\max(0, \phi(gb, t))]$ 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 negative virtual value for the grand bundle, or exclude some type with positive virtual value. Thus $\mathbf{E}[\max(0, \phi(gb, t))]$ 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 proof and here discuss its geometric interpretation.

Geometric Interpretation of the Proof Without Regularity Assumptions. The general proof of Proposition 1 is based on two observations, and can be found in the appendix.

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 based on binding IC constraints from higher types. In particular, assuming that the type space is finite, the virtual value of each type t for each bundle b is defined as

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

where $\lambda(t', t)$ is the Lagrangian multiplier corresponding to the possible deviation of t' to t. Viewed as vectors, the virtual value of t is defined by shifting $v(\cdot, t)$ in the direction of the vector $v(\cdot, t) - v(\cdot, t')$, for each t' with binding IC constraints to t. Since any type t' with binding IC constraint is "above" $v(\cdot, t)$ (above the ray from the origin to $v(\cdot, t)$), the result is a virtual value vector $\hat{\phi}(\cdot, t)$ that lies "below" $v(\cdot, t)$, that is, $\hat{\phi}(b, t) \leq \frac{v(b,t)}{t}\hat{\phi}(gb, t)$. See Figure 8. The fact that $\hat{\phi}(b, t) \leq \frac{v(b,t)}{t}\hat{\phi}(gb, t)$ implies that $\hat{\phi}(b, t) \leq \max(0, \hat{\phi}(gb, t))$, and therefore virtual surplus is maximized by assigning either the outside option $\emptyset$ or the grand bundle gb to each type. We choose Lagrangian multipliers such that $\hat{\phi}(gb, t)$ is monotone non-decreasing, and therefore pure bundling maximizes virtual surplus.

3.2 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 non-zero values for the grand bundle; i.e.,* $\mathbf{Pr}[r(\cdot, t) \in U \mid v(gb, t) = \hat{v}]$ *is non-decreasing in* $\hat{v} > 0$ *for all upper sets* $U \subseteq \mathbb{R}^B$.

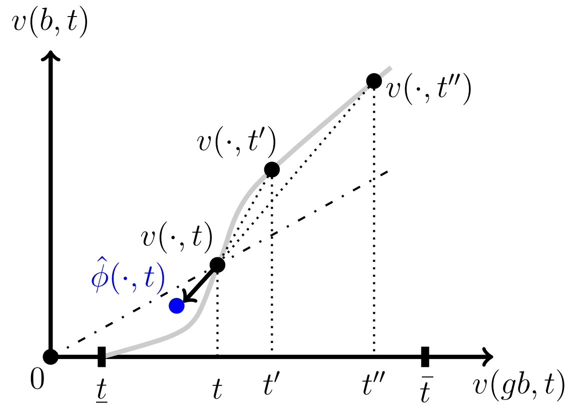

Figure 8: For each t' such that the IC constraint from t' to t binds, $\hat{\phi}$ is shifted below the ray that connects the origin to the vector $v(\cdot,t)$.

The proof generalizes the decomposition approach from Section 2.2. Suppose that the distribution of types μ can be written as a distribution over distributions $\nu \in \Delta(\Delta(T))$. A random type from μ can be drawn by first selecting a distribution $\mu' \in \Delta(T)$ from ν, and then drawing a random type from μ'. Selling the grand bundle at price p is optimal for distribution μ if it is optimal for every distribution μ' in the support of ν. Indeed, if so, then selling the grand bundle at price p is an optimal solution to the relaxed problem in which the seller can observe μ', and can select a mechanism accordingly.

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. A distribution $\nu \in \Delta(\Delta(T))$ is a *decomposition* of μ if $\mu = \mathbf{E}_{\mu'\sim\nu}[\mu']$.

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

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

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

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

To prove Theorem 1 using Proposition 1 and Lemma 3, we construct a decomposition ν 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 same* pure bundling mechanism is optimal for all μ'. That is, a single price p is an optimal price to sell the grand bundle for every μ'. A sufficient condition is that the marginal distribution of the value for the grand bundle is identical for all distributions μ' in the decomposition. We say that a decomposition ν is a ratio-monotone decomposition if it satisfies both conditions. That is, all distributions μ' in the support of ν are supported on ratio-monotone paths and have identical marginal distributions of the value for the grand bundle.

We characterize distributions μ for which a ratio-monotone decomposition exists. 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 monotone coupling. The result states that 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, let us revisit the setup of Section 2.2, where the set of bundles is $B = \{\emptyset, sb, gb\}$. Assume that $v(gb, t) \in \{1, 2\}$ for all types t. We show that a ratio-monotone decomposition of μ exists if the stochastic monotonicity condition of Theorem 1 holds, namely

$$\mathbf{Pr}\left[r(sb,t) \geq \hat{r} \mid v(gb,t) = 1\right] \leq \mathbf{Pr}\left[r(sb,t) \geq \hat{r} \mid v(gb,t) = 2\right] \tag{5}$$

for all $\hat{r} \in \mathbb{R}$ (see Corollary 1). We first construct a decomposition ν of μ, and then show that ν is a ratio-monotone decomposition if Inequality (5) holds. Assume for simplicity that the distribution of the relative value $r(sb, t)$ conditioned on the value for the grand bundle $v(sb, t)$ is continuous. The conditional distributions are shown in Figure 9.

The decomposition ν of μ is constructed as follows. Define the "quantile" $q(t) \in [0, 1]$ of a type t to be probability that a random type t' has lower value for the small bundle sb than does t, conditioned on t' having the same value for the grand bundle gb as does t. Let $\mu^{\hat{q}}$ be the distribution of types conditioned on $q(t) = \hat{q}$. Let F_q be the marginal distribution of $q(t)$. The decomposition ν first draws q from F_q, and then draws a random type from μ^q. Since $\mu = \mathbf{E}_{q\sim F_q}[\mu|q(t) = q] = \mathbf{E}_{q\sim F_q}[\mu^q]$, ν is a decomposition of μ. Notice two properties of ν. First, the quantile q is independently distributed from the value for the grand bundle. Indeed, conditioned on any $v(gb, t) = \hat{v}$, the quantile $q(t)$ is distributed uniformly on the interval $[0, 1]$. Second, a single type with value 1 for the grand bundle has quantile q, and similarly a single type with value 2 for the grand bundle has quantile q. Therefore, each

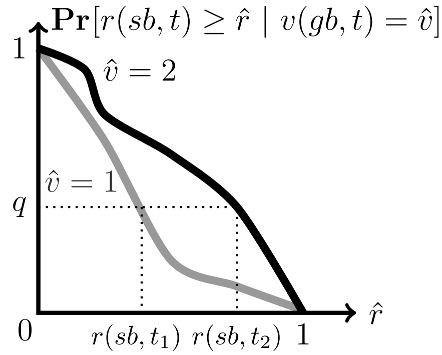

Figure 9: The dark curve is $\mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = 1]$, and the light curve is $\mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = 2]$. Since the dark curve is above the light curve, $r(sb,t_1) \leq r(sb,t_2)$.

distribution μ^q is supported on a path.

The decomposition ν constructed above is ratio-monotone if Inequality (5) holds. In fact, consider two types t_1 and t_2 with values 1 and 2 for the grand bundle and with the same quantile q. As shown in Figure 9, the relative value for type t_1 is at most the relative value for type t_2. Therefore, μ^q is supported on a ratio-monotone path, and ν is a ratio-monotone decomposition.

To summarize, we have argued that if $r(sb,t) \in \mathbb{R}$ is stochastically non-decreasing in $v(gb,t)$, then there exists a random variable $q(t)$ satisfying three properties. First, $q(t)$ and $v(gb,t)$ are independently distributed. Second, conditioned on $q(t) = q$, types have distinct values for the grand bundle. Third, conditioned on $q(t) = q$, the relative value is non-decreasing in the value for the grand bundle. In other words, the random variable q can be used to couple types with distinct values for the grand bundle, in such a way that a higher type has a higher relative value. The following lemma generalizes the construction to arbitrary sets of bundles.

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 $h^{\hat{q}} : S \times \mathbb{R} \to \mathbb{R}$ for each $\hat{q} \in Q$ such that*

(I) q and x 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 $s \in S$.

The first property of Lemma 4 is independence. The second property identifies the coupling function h that uniquely specifies x given q and y. The third property states that the coupling is monotone.

We now use Lemma 4 to prove Theorem 1. Since Lemma 4 is stated for finite S, we first prove the theorem for a finite set of bundles B. We then extend the proof in the appendix to infinite B using the continuity of v.

Proof of Theorem 1 for finite B. Consider any maximizer p^* of $p \times (1 - F_{gb}(p))$. We show that pure bundling with price p^* is optimal.

Assume 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 $h^{\hat{q}} : B \times \mathbb{R} \to \mathbb{R}$ such that

(I) q and $v(gb, t)$ are independently distributed.

(II) Conditioned on $q = \hat{q}$, $r(b, t) = h^{\hat{q}}(b, v(gb, t))$ for all $b \in B$ with probability one.

(III) $h^{\hat{q}}(b, v(gb, t))$ is monotone non-decreasing in $v(gb, t)$ for all $\hat{q}$ and b.

Let $\mu^{\hat{q}}$ be the distribution of types conditioned on $q = \hat{q}$. Note that $\mu = \mathbf{E}_{\hat{q}}[\mu | q = \hat{q}] = \mathbf{E}_{\hat{q}}[\mu^{\hat{q}}]$. Therefore, by Lemma 3, it is sufficient to prove optimality of pure bundling with price p^* for any $\mu^{\hat{q}}$.

Pure bundling with price p^* is optimal for $\mu^{\hat{q}}$ since $\mu^{\hat{q}}$ is supported on a ratio-monotone path, and the marginal distribution of the value for the grand bundle is F_{gb}. First, notice that by property (II), conditioned on $q = \hat{q}$, types have distinct values for the grand bundle. Formally, define $v^{\hat{q}}(b, t) = t h^{\hat{q}}(b, t)$. By property (II), conditioned on $q = \hat{q}$,

$$v(b, t) = v(gb, t) r(b, t) = v(gb, t) h^{\hat{q}}(b, v(gb, t)) = v^{\hat{q}}(b, v(gb, t))$$

for all bundles b with probability one. That is, types in the support of $\mu^{\hat{q}}$ are on a path $v^{\hat{q}}$ as in Section 3.1. Second, $v^{\hat{q}}$ is ratio-monotone by property (III), since $v^{\hat{q}}(b, t)/t = h^{\hat{q}}(b, t)$ is monotone non-decreasing in t for all b. Therefore Proposition 1 applies to $v^{\hat{q}}$ and implies that pure bundling is optimal for $\mu^{\hat{q}}$. By property (I), the marginal distribution of the value for the grand bundle is F_{gb}. Therefore, pure bundling with price p^* is optimal for distribution $\mu^{\hat{q}}$. □

3.3 A Partial Converse to Theorem 1

The condition of Theorem 1 is not necessary for the optimality of pure bundling. Proposition 1 provided a partial converse to Theorem 1 for the case where types are on a path. We here present a second partial converse.

Pure bundling is not optimal if the largest value in the support of $r(b,t)$ conditioned on $v(gb,t) = \hat{v}$ is decreasing in $\hat{v}$, and if some additional minor assumptions hold. Formally, let $\bar{r}(b,\hat{v})$ denote the largest value in the support of $r(b,t)$ conditioned on $v(gb,t) = \hat{v}$. We say that the distribution of values is (i) *continuous* if for all bundles b, the joint distribution of the value for the grand bundle and the value for bundle b has a probability density function $f_{gb,b}$, and (ii) *interior* if some $v > \underline{v}$ maximizes $v \times (1 - F_{gb}(v))$, where F_{gb} denotes the cumulative density function of the value for the grand bundle, and $\underline{v}$ is the lowest value in its support.

Proposition 2. *Pure bundling is not optimal if $\bar{r}(b,\hat{v})$ is monotone decreasing in $\hat{v}$ for some b, and the distribution of values is continuous and interior.*

We say that $r(b,t)$ is *stochastically decreasing at the top* in the value for the grand bundle if $\bar{r}(b,\hat{v})$ is monotone decreasing in $\hat{v}$. Notice that if types are on a path v, then $\bar{r}(b,v) = v(b,v)/v$. Therefore, a corollary of Proposition 2 is that pure bundling is not optimal if types are on a path v such that $v(b,v)/v$ is decreasing in v, and the distribution of types is continuous and interior, providing a converse to Proposition 1.

We now state a corollary of Proposition 2 that allows for a more direct comparison with Theorem 1. Pure bundling is optimal if $\mathbf{Pr}[r(b,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$ is monotone decreasing in $\hat{v}$ unless it is zero, and the distribution of values is continuous and interior.

Corollary 2. *Pure bundling is not optimal if for some bundle b, all $\hat{v} < \hat{v}'$ and all $\hat{r} \leq \bar{r}(b,\hat{v}')$,*

$$\mathbf{Pr}\left[r(b,t) \geq \hat{r} \mid v(gb,t) = \hat{v}\right] > \mathbf{Pr}\left[r(b,t) \geq \hat{r} \mid v(gb,t) = \hat{v}'\right],$$

and the distribution of values is continuous and interior.

Proof. By assumption, $\mathbf{Pr}[r(b,t) \geq \hat{r} \mid v(gb,t) = \hat{v}] > 0$ for all $\hat{r} \leq \bar{r}(b,\hat{v}')$. Since the distribution of values is continuous, $\bar{r}(b,\hat{v}') > \hat{r} = \bar{r}(b,\hat{v}')$, and Proposition 2 applies. □

Let us now compare Corollary 2 with Theorem 1. If $r(\cdot,t)$ is stochastically non-decreasing in the value for the grand bundle as Theorem 1 demands, then $\mathbf{Pr}[r(b,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$ is stochastically non-decreasing in $\hat{v}$ for all b and all $\hat{r}$. To see this, set $U = \{x \in \mathbb{R}^B \mid x(b) \geq \hat{r}\}$, and notice that $\mathbf{Pr}[r(b,t) \in U \mid v(gb,t) = \hat{v}] = \mathbf{Pr}[r(b,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$. In contrast, Corollary 2 shows that pure bundling is not optimal if $\mathbf{Pr}[r(b,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$ is monotone decreasing (unless it is zero) in $\hat{v}$ for some b and all $\hat{r}$.

4 Applications, Interpretations, and Extensions

In this section we discuss some applications, interpretations, and extensions.

We present some applications of our model in Section 4.1, and provide a closure property that simplifies verifying the stochastic monotonicity condition of our main result. We then provide some examples that allow us to interpret the stochastic monotonicity condition in Section 4.2. In Section 4.3 we relax the assumptions that costs are zero and that a grand bundle exists, and identify a condition for the optimality of selling all bundles at a uniform markup above cost, generalizing Theorem 1.

4.1 Applications of the Model

We discuss a few applications of our model below. The set of bundles B can represent any set of alternatives that the seller can assign to the buyer. Bundles may represent subsets of products, may or may not be vertically differentiated, and may be randomized.

4.1.1 Multiple Products with Multi-unit Demands

It is most natural to think of a bundle as a set of products. A bundle may contain multiple units of some products. Consider a multi-product seller with indivisible products $1, \ldots, n$, and a buyer who demands at most $\bar{u}_i \in \mathbb{Z}_{\geq 0}$ units of each product i. To model such a setting, let the set of bundles be $B = \{(b_1, \ldots, b_n) \mid b_i \in \{0, \ldots, \bar{u}_i\}\}$. A bundle b contains b_i units of each product i. The grand bundle is $gb = (\bar{u}_1, \ldots, \bar{u}_n)$. The assumption of finite $\bar{u}_i$ is required so that the grand bundle is well defined. The case where $\bar{u}_i = 1$ for all i corresponds to when the buyer demands at most a single unit of each product.

A related application is bundling with divisible products. Consider a seller with divisible products $1, \ldots, n$, and a buyer who demands at most one unit of each product i. To model such a setting, let the set of bundles be $B = \{(b_1, \ldots, b_n) \in [0,1]^n\}$. A bundle b contains a fraction $b_i \in [0,1]$ of each product i. The grand bundle is $gb = (1, \ldots, 1)$.

4.1.2 Vertically Differentiated Bundles Representing Quantities or Qualities

A special case of our setting is when bundles can be ranked such that each type has a higher value for a higher ranked bundle. A bundle in such a case may represent quantity or quality. For example, bundles are ranked in examples in Section 2 where $B = \{\emptyset, sb, gb\}$ since $v(sb, t) \leq v(gb, t)$ for all types t. With vertically differentiated bundles, a bundle can be represented with a real number, $b \in \mathbb{R}^+$, such that $v(b, t) \leq v(b', t)$ for all $b \leq b'$ and t.

A pure bundling mechanism is one that sells only the highest quality product $\max_{b \in B} b$. For the special case where $T \subseteq \mathbb{R}^+$ and values are linear $v(b, t) = b \cdot t$, pure bundling is known to be optimal from Stokey (1979), Riley and Zeckhauser (1983), and Myerson (1981). For this special case, optimality of pure bundling follows also from Proposition 1 since $v(b, t)/t = b$ is constant in t. The conditions of Salant (1989), Johnson and Myatt (2003), and Anderson and Dana Jr (2009) roughly correspond to the case in our setting where bundles are vertically differentiated, types are on a ratio-monotone path v, and the distribution of the value for the grand bundle is regular.

In the inter-temporal price discrimination setting of Stokey (1979), relative values have a natural interpretation as a buyer's discount rate for delayed consumption. Our result states that selling the product immediately is optimal if consumers with higher values for the product are more likely to have higher discount rates, that is, they are more patient.

4.1.3 Randomized Bundles and a Closure Property

Our model can incorporate randomization, once the set of bundles and values are appropriately defined. Theorem 1 applies and identifies a condition for the optimality of pure bundling. However, with randomized bundles, verifying the condition of the theorem may appear challenging due to the large size of the profile of relative values (a relative value for each randomized bundle). We "prune" the condition by showing that relative values for non-deterministic bundles can be safely ignored. In other words, the exact same condition that implies the optimality of pure bundling with deterministic bundles also implies its optimality with randomized bundles. This result employs on a simple closure property of stochastic monotonicity, which we also apply in future subsections.

Formally, for a given environment (B, T, v), let $\tilde{B} = \Delta(B)$ be the set of all distributions over B, and let $\tilde{v}(\tilde{b}, t) = \mathbf{E}_{b \sim \tilde{b}}[v(b, t)]$ be the expectation of v for all randomize bundles $\tilde{b} \in \tilde{B}$. The seller in environment $(\tilde{B}, T, \tilde{v})$ can sell any distribution over B.

Theorem 1 states that pure bundling is optimal in environment $(\tilde{B}, T, \tilde{v})$ if $\tilde{r}(\cdot, t) = \tilde{v}(\cdot, t)/v(gb, t) \in \mathbb{R}^{\tilde{B}}$ is stochastically non-decreasing in the value for the grand bundle $v(gb, t)$.

Verifying the condition may appear challenging due to the large size of the profile of relative values $\tilde{r}(\cdot, t)$. For instance, $\tilde{r}(\cdot, t) \in \mathbb{R}^{\tilde{B}}$ is an infinite dimensional profile even if B is finite. Nonetheless, the proposition below shows that it is enough to verify the condition for the lower dimensional profile $r(\cdot, t) = v(\cdot, t)/v(gb, t) \in \mathbb{R}^B$. In other words, if the stochastic monotonicity condition holds in environment (B, T, v), then it also holds in environment $(\tilde{B}, T, \tilde{v})$.

Proposition 3. *For a given environment* (B, T, v)*, let* $\tilde{B} = \Delta(B)$ *and* $\tilde{v}(\tilde{b}, t) = \mathbf{E}_{b\sim\tilde{b}}[v(b, t)]$ *for all* $\tilde{b} \in \tilde{B}$*. Pure bundling is optimal in environment* $(\tilde{B}, T, \tilde{v})$ *if* $\mathbf{Pr}[r(\cdot, t) \in U \mid v(gb, t) = \hat{v}]$ *is non-decreasing in* $\hat{v} > 0$ *for all upper sets* $U \subseteq \mathbb{R}^B$*.*

For example, recall the setup of Section 2.2, where $B = \{\emptyset, sb, gb\}$ and $T = \{t_1, t_1', t_2, t_2'\}$. Assume as before that t_1 and t_1' have value 1 and types t_2 and t_2' have value 2 for the grand bundle, and the relative values are ordered as $r(sb, t_1) \leq r(sb, t_2) \leq r(sb, t_1') \leq r(sb, t_2')$, as shown in Figure 4. We showed in Section 2 that the condition of Theorem 1 holds if $q_1 \leq q_2$. We also claimed that pure bundling is optimal even if randomization is allowed. The claim follows directly from Proposition 3.

Proposition 3 is itself a corollary of the following lemma. The lemma provides a closure property. If x is stochastically non-decreasing in y, then any monotone function of x is also stochastically non-decreasing in y. In other words, stochastic monotonicity is closed under monotone transformations. We apply the lemma also in the future subsections to prune the set of conditions that must be verified for Theorem 1 to hold.

Lemma 5. *[Shaked and Shanthikumar, 2007, Theorem 6.B.16 (a)] Consider two sets* $\mathcal{S}, \mathcal{S}'$*, random variables* $x \in \mathbb{R}^{\mathcal{S}}$ *and* $y \in \mathbb{R}$*, and a monotone non-decreasing function* $g : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{S}'}$ *(i.e.,* $g(x) \leq g(x')$ *if* $x \leq x'$*). The distribution of* $g(x) \in \mathbb{R}^{\mathcal{S}'}$ *is stochastically non-decreasing in* y *if the distribution of* x *is stochastically non-decreasing in* y*.*

Equipped with Lemma 5, we now prove Proposition 3.

Proof of Proposition 3. Notice that $\tilde{r}(\tilde{b}, t) = \mathbf{E}_{b\sim\tilde{b}}[r(b, t)]$. The function $\tilde{r}(\cdot, t)$ is monotone non-decreasing in $r(\cdot, t)$. By assumption, $r(\cdot, t)$ is stochastically non-decreasing in $v(gb, t)$. Lemma 5 implies that $\tilde{r}(\cdot, t)$ is stochastically non-decreasing in $v(gb, t)$. Theorem 1 implies that pure bundling is optimal in environment $(\tilde{B}, T, \tilde{v})$. □

As another example of Proposition 3, consider the case of two indivisible products $B = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. Theorem 1 states that pure bundling is optimal if the profile of relative values $(\frac{v(\{1\},t)}{v(gb,t)}, \frac{v(\{2\},t)}{v(gb,t)})$ is stochastically non-decreasing in the value for the grand bundle

$v(gb, t)$. More strongly, Proposition 3 implies that under the same condition, pure bundling is optimal even if randomization is allowed. We provide an interpretation of this condition later when we discuss a related example with divisible products in Section 4.2.2.

4.2 Interpretations of the Results

In this section we provide interpretations of our results using three examples. First we present an example with a single product and multi-unit demands. We interpret the value for the grand bundle as a measure of a type's wealth (similar interpretations are common in the screening literature, as in Armstrong, 1999). In this view, our result states that pure bundling is optimal if wealthier consumers tend to have lower need for higher quantities. Second we consider an example with two products. We show that pure bundling is optimal if wealthier consumers consider the products to be less complementary. Third we consider a multi-product example with additive values, where the value for a bundle is the sum of the values of its constituting products. We show that pure bundling is generally not optimal except for special cases already identified in the literature.

4.2.1 A Single Product: Wealth versus Need

In this section we consider an example with a single product and two-unit demands. We view the value for the grand bundle as a measure of a type's wealth, and the relative values as a measure of a type's need for higher quantities. In this view, the interpretation of our main result is that pure bundling is optimal if wealthier consumers have lower need for higher quantities. For instance, this condition holds if wealthier households tend to be smaller.[6]

Let the set of bundles be $B = \{\emptyset, sb, gb\}$ (similar to Section 2). Interpret the small bundle sb as low quantity, and the grand bundle gb as high quantity of a single product. Recall that the utility of a type t for a bundle b and payment p to the seller is

$$v(b, t) - p.$$

[6] The median household income in the United States does indeed decline for households beyond four persons according to the US Census Bureau (2006).

Dividing utility of type t for all bundle-payment pairs by a constant does not affect the type's choices. Therefore, the choices made by such a type is identical to the choices of a type that has utility

$$\frac{v(b,t)}{v(gb,t)} - \frac{p}{v(gb,t)}$$

for bundle b and payment p. The marginal utility for money is $1/v(gb,t)$. Assuming that the marginal utility for money is inversely related with wealth, we interpret $v(gb,t)$ as a measure of the type's wealth. We also interpret $v(sb,t)/v(gb,t)$ as an inverse measure of a type's need for high quantities, where high $v(sb,t)/v(gb,t)$ indicates low need for higher quantities.

Theorem 1 (Corollary 1) states that pure bundling is optimal if $v(sb,t)/v(gb,t)$ is stochastically non-decreasing in $v(gb,t)$. That is, wealthier consumers are more likely to have lower need for high quantities. This condition holds if wealthier households tend to be smaller.

To interpret the converse result (Proposition 2), it is more natural to view the set of bundles as the set of quality levels, where sb denotes low quality and gb denotes high quality. In this view, $v(sb,t)/v(gb,t)$ is an inverse measure of the type's sensitivity to quality, where low $v(sb,t)/v(gb,t)$ indicates high sensitivity to quality. Proposition 2 states that pure bundling is not optimal if wealthier consumers have higher sensitivity to quality.

4.2.2 Heterogeneous Products and Relative Synergies

We now provide an example with two heterogeneous divisible products. An interpretation of Theorem 1 is that pure bundling is optimal if relative synergy, the ratio of the value for the grand bundle over the sum of the values of the individual products, is lower for consumers with higher values for the grand bundle. Relative synergy measures the complementarity between products. In this view, Theorem 1 states that pure bundling is optimal if wealthier consumers consider the products to be less complementary.

Assume that there are two divisible products, $B = [0,1]^2$ and the set of types is $T \subseteq \mathbb{R}^3_+$. The value of a type $t = (t_{gb}, t_1, t_2)$ for a bundle b is

$$v(b,t) = t_{gb} \cdot \Big(t_1 b_1 + t_2 b_2 + (1 - t_1 - t_2) b_1 b_2\Big), \tag{6}$$

where $t_{gb} \geq 0$ and $t_1, t_2 \in [0,1]$. The grand bundle is $gb = (1,1)$.

Let us interpret the parameters. Similar to Section 4.2.1, we interpret the value for the grand bundle $v(gb,t) = t_{gb}$ as a measure of a type's wealth. Parameters t_1 and t_2 are the relative values of individual products compared to the grand bundle. The relative synergy is

$$\frac{v(gb,t)}{v\Big((1,0),t\Big)+v\Big((0,1),t\Big)} = \frac{t_{gb}}{t_{gb}t_1 + t_{gb}t_2} = \frac{1}{t_1+t_2}.$$

Therefore $t_1 + t_2$ is the inverse of the relative synergy. Types with lower relative synergy (higher $t_1 + t_2$) consider the products to be less complementary. In particular,

$$\begin{cases} t_1+t_2<1: & v\Big((b_1,0),t\Big)+v\Big((0,b_2),t\Big) < v\Big((b_1,b_2),t\Big) \quad \text{(partial complements)} \\ t_1+t_2=1: & v\Big((b_1,0),t\Big)+v\Big((0,b_2),t\Big) = v\Big((b_1,b_2),t\Big) \quad \text{(additive values)} \\ t_1+t_2>1: & v\Big((b_1,0),t\Big)+v\Big((0,b_2),t\Big) > v\Big((b_1,b_2),t\Big) \quad \text{(partial substitutes).} \end{cases}$$

Proposition 4 applies and implies that pure bundling is optimal if (t_1, t_2) is stochastically non-decreasing in t_{gb}. That is, wealthier consumers are more likely to have low relative synergy and consider the products to be less complementary.

Proposition 4. *Assume that* $B = [0, 1]^2$, $T \subseteq \mathbb{R}^3_+$, *and values are given by Equation* (6). *Pure bundling is optimal if* $(t_1, t_2) \in \mathbb{R}^2$ *is stochastically non-decreasing in* $t_{gb} > 0$.

Proof. We use Lemma 5 to show that $r(\cdot, t) \in \mathbb{R}^B$ is stochastically non-decreasing in $t_{gb} > 0$, and then the optimality of pure bundling follows from Theorem 1. To apply Lemma 5, we need to show that $r(b, t)$ is non-decreasing in (t_1, t_2) for all b. Rearranging r, we have

$$\begin{aligned} r(b,t) &= t_1 b_1 + t_2 b_2 + (1 - t_1 - t_2) b_1 b_2 \\ &= t_1(b_1 - b_1 b_2) + t_2(b_2 - b_1 b_2) + b_1 b_2, \end{aligned}$$

which is non-decreasing in (t_1, t_2) since $b_1 - b_1 b_2 \geq 0$ and $b_2 - b_1 b_2 \geq 0$. □

Conversely, Proposition 2 states that pure bundling is not optimal if t_i is stochastically decreasing at the top in t_{gb} for some i, and the distribution of values is continuous and interior.

Notably, pure bundling may be optimal even if the products are partial substitutes for all types ($t_1 + t_2 > 1$ for all types). Proposition 4 shows that for pure bundling to be optimal, the products need not be partial complements for all types, but rather products should be less complementary for consumers with higher values.

A large part of the literature on multi-product bundling studies the case where values are additive, i.e., $t_1 + t_2 = 1$. Note that the condition of Proposition 4 is quite restrictive with additive values. Indeed, for (t_1, t_2) to be stochastically non-decreasing in t_{gb} while $t_1 + t_2 = 1$, (t_1, t_2) must be independently distributed of t_{gb}.[7] Therefore, with additive values, pure bundling is not optimal except for restrictive cases. We dedicate the next section to a more in depth analysis of additive values.

4.2.3 Non-optimality of Pure Bundling With Additive Values

Mostly for tractability reasons, a large part of the literature on multi-product bundling has assumed that values are additive, i.e., the value for a bundle is the sum of the values of its constituting products. Nevertheless, the literature has not identified general conditions, beyond some special cases, for the optimality of pure bundling with additive values. An implication of Proposition 1 is that pure bundling cannot be generally optimal for a given set of additive values, in a sense we formalize below, except for the special cases already identified in the literature.

Consider selling n divisible products to a buyer who needs at most one unit of each product. That is, the set of bundles is $B = \{0, 1\}^n$, where $b_i = 1$ if product i is in bundle $(b_1, \ldots, b_n)$, and $b_i = 0$ otherwise. Values are additive if $T \subseteq \mathbb{R}^n_+$ and $v(b, t) = \sum_i b_i t_i$ for a bundle b. That is, t_i is the value for a unit of product i, and the value for a bundle is the sum of the values of its constituting products. The grand bundle is $gb = (1, \ldots, 1)$.

The prior work has identified two conditions on T that imply the optimality of pure bundling regardless of the distribution of types. First, pure bundling is optimal if all types have the same value for the grand bundle, as argued in Stigler (1963) and Adams and Yellen (1976). This shown in Figure 10, (a), where the sum of values is constant across types (only values for individual products are depicted unlike previous pictures which depict values for bundles). In this case, pure bundling extracts the full surplus. Second, it follows from Riley and Zeckhauser (1983) and Myerson (1981) that pure bundling is optimal if the relative values are constant, $t_i/t_j = t'_i/t'_j$ for all types $t, t' \in T$ and products i, j, as shown in Figure 10, (b). In other words, there exist $\delta_1, \ldots, \delta_n$ such that $t_i = t_1 \delta_i$ for all types $t \in T$ and products i. The proposition below shows that if pure bundling is optimal for all distributions over T, one of the aforementioned two conditions must hold.

Proposition 5. *Let $B = \{0, 1\}^n$, $T \subseteq \mathbb{R}^n_+$, and $v(b, t) = \sum_i b_i t_i$. Pure bundling is optimal for all distributions over T if and only if either of the following conditions holds*

[7] The stochastic monotonicity condition requires that $\mathbf{Pr}[t_1 \geq \hat{r} \mid t_{gb} = \hat{v}]$ and $\mathbf{Pr}[t_2 \geq \hat{r} \mid t_{gb} = \hat{v}] = \mathbf{Pr}[1 - t_1 \geq \hat{r} \mid t_{gb} = \hat{v}]$ are both non-decreasing in $\hat{v}$ for all $\hat{r}$. Therefore t_1 must be independently distributed of t_{gb}. Since $t_2 = 1 - t_1$, (t_1, t_2) is also independently distributed of t_{gb}.

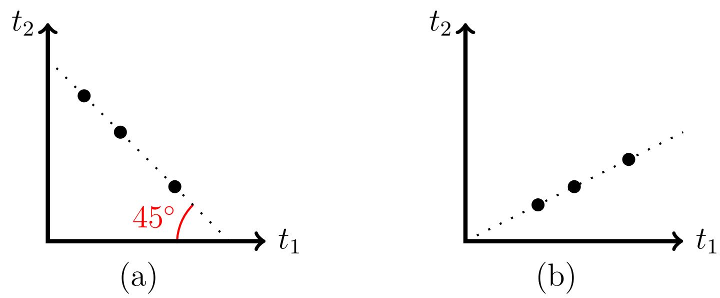

Figure 10: Values for individual products 1 and 2. (a) $\sum_i t_i$ is constant. (b) t_1/t_2 is constant.

1. *$\sum_i t_i = \sum_i t'_i$ for all $t, t' \in T$.*

2. *there exist $\delta_1, \dots, \delta_n$ such that $t_i = t_1\delta_i$ for all types t and products i.*

Proof. Suppose that the first condition holds. The stochastic monotonicity condition of Theorem 1 trivially holds since $v(gb, t)$ is constant over all types. Therefore pure bundling is optimal.

Suppose that the second condition holds. In this case, types have distinct values for the grand bundle, $\sum_i t_i \neq \sum_i t'_i$ for all $t \neq t'$. Otherwise, we have $t_1 \sum_i \delta_i = t'_1 \sum_i \delta_i$ and thus $t_1 = t'_1$ and $t = t'$. Notice also that the relative value $v(b,t)/v(gb,t) = (\sum_{i\in b} t_1\delta_i)/(\sum_i t_1\delta_i) = (\sum_{i\in b} \delta_i)/(\sum_i \delta_i)$ is constant over all types. Therefore, Proposition 1 applies and implies that pure bundling is optimal.

We now prove the necessity of the above two cases. Assume that pure bundling is optimal for all distributions over T. We show that if first condition does not hold, then the second condition must hold. Suppose therefore that there exist types t, t' such that $\sum_i t_i \neq \sum_i t'_i$. Assume without loss of generality that $\sum_i t_i < \sum_i t'_i$. Proposition 1 states that for pure bundling to be optimal for all distributions $\mu \in \Delta(\{t, t'\})$, we must have $r(b, t) \leq r(b, t')$ for all bundles b. In particular, by letting b be a bundle that includes a unit of product j and no other product, we must have

$$\frac{t_j}{\sum_i t_i} \leq \frac{t'_j}{\sum_i t'_i}. \tag{7}$$

However, summing over all j, each side of the above inequality adds up to 1. Therefore all inequalities must hold with equality, $\frac{t_j}{\sum_i t_i} = \frac{t'_j}{\sum_i t'_i}$. As a result, we have $t_i/t_j = t'_i/t'_j$ for all products i and j.

We have argued so far that for every pair of types such that $\sum_i t_i \neq \sum_i t'_i$, we must have $t_i/t_j = t'_i/t'_j$ for all products i and j. We now show that every pair of types must indeed satisfy $\sum_i t_i \neq \sum_i t'_i$. Assume for contradiction that there exist types $t \neq t'$ such that $\sum_i t_i = \sum_i t'_i$, and consider any other type t'' such that $\sum_i t_i \neq \sum_i t''_i$ (t'' exists since the first condition of the proposition does not hold). By applying our above analysis to pairs t, t'' and t', t'', we have $t_i/t_j = t''_i/t''_j = t'_i/t'_j$ for all i, j. Therefore the two types t and t' must be identical, which is a contradiction. We conclude that every pair of types $t \neq t'$ must satisfy $\sum_i t_i \neq \sum_i t'_i$ and therefore $t_i/t_j = t'_i/t'_j$ for all products i and j. Letting $\delta_i = t'_i/t'_1$, we have $t_i = t_1\delta_i$ for all types t and products i. □

Note that the above proposition concerns the optimality of pure bundling *for all* distributions over a set of types T. It is still possible that pure bundling is optimal for certain distributions over an additive set of types that satisfy neither of the conditions of Proposition 5. Indeed, Pavlov (2011), Menicucci et al. (2015), and Daskalakis et al. (2017) provide examples that fall in this category. Nonetheless, we take the above analysis as an indication that a general principle for the optimality of pure bundling is unlikely to exist with additive values.

4.3 Extension: Uniform Markup Pricing

In this section we relax two of the main assumptions of our model, namely that costs are zero and that a grand bundle exists. We generalize Theorem 1 and identify a condition for optimality of selling all bundles at a uniform markup above cost, defined formally below.[8]

Generalized Model and Uniform Markup Pricing. Let $c(b) \geq 0$ denote the cost of producing a bundle b, and assume no longer that a grand bundle necessarily exists. Extend the definition of an environment from Section 1 to include costs, (B, T, v, c). A mechanism (b, p) is optimal if it is IC and IR and maximizes the expected profit $\mathbf{E}[p(t) - c(b(t))]$ among all IC and IR mechanisms.

A mechanism (b, p) is a uniform markup pricing mechanism with markup $p \in \mathbb{R}$ if it offers each bundle $b \in B - \{\emptyset\}$ at price $p + c(b)$. If costs are zero, then a uniform markup pricing mechanism offers all bundles at a uniform price p. If, further, a grand bundle exists, then all types either choose the grand bundle or the outside option (type t may choose another bundle b if $v(b, t) = v(gb, t)$, but such a mechanism is equivalent in profit to a pure bundling mechanism). Thus, a uniform markup pricing mechanism is equivalent to a pure bundling mechanism.

[8]Armstrong (1999) considers a large number of products and identifies conditions under which uniform markup pricing is approximately optimal.

Optimality of Uniform Markup Pricing. The theorem below identifies a condition for the optimality of uniform markup pricing, and includes Theorem 1 as a special case. Define $\tilde{v}(b,t) = \max(v(b,t) - c(b), 0)$, and $\tilde{r}(b,t) = \tilde{v}(b,t)/\max_{b'} \tilde{v}(b',t)$. Uniform markup pricing is optimal if $\tilde{r}(\cdot,t)$ is stochastically non-decreasing in $\tilde{v}(b,t)$.

Theorem 2. *Uniform markup pricing is optimal if* $\mathbf{Pr}[\tilde{r}(\cdot,t) \in U \mid \max_b \tilde{v}(b,t) = \hat{v}]$ *is non-decreasing in* $\hat{v} > 0$ *for all upper sets* $U \subseteq \mathbb{R}^B$*.*

Proof. The proof consists of three steps, each transforming the environment to one in which an extra assumption can be made. First, values can be assumed to be no less than costs. Second, costs can be assumed to be zero. Third, a grand bundle can be assumed to exist. Then Theorem 1 applies to prove the result. The transformations affect neither the set of types T nor the distribution μ, so we fix the distribution μ satisfying the condition of the theorem, and drop it from the discussion.

First, values can be assumed to be no less than costs. Formally, uniform pricing is optimal in environment $E = (B, T, v, c)$ if it is optimal in environment $E^1 = (B, T, v^1, c)$, where $v^1(b,t) = \max(v(b,t), c(b))$. Values in environment E^1 are indeed no less than costs, $v^1(b,t) \geq c(b)$. To establish the equivalence, notice that profit weakly increases if the price of every bundle b that is offered at price below cost is increased to $c(b)$. Therefore, without loss of generality, an optimal mechanism offers each bundle b at price at least $c(b)$ or not at all. In such a mechanism, the choice made by a type t with values $v(\cdot,t)$ is identical to the choice of a type t with values $v^1(\cdot,t)$. Therefore, the optimal profit in environments E and E^1 are equal. Since a mechanism is a uniform pricing mechanism in environment E if and only if it is a uniform pricing mechanism in environment E^1, uniform pricing is optimal in environment E if it is optimal in environment E^1.

Second, costs can be assumed to be zero. Formally, uniform markup pricing is optimal in environment E^1 if it is optimal in environment $E^2 = (B, T, v^2, 0)$, where $v^2(b,t) = v^1(b,t) - c(b) = \max(v(b,t) - c(b), 0)$. Consider mechanisms (b, p^1) and (b, p^2) such that $p^2(t) = p^1(t) - c(b(t))$. Notice that the IC constraint for mechanism (b, p^1) in environment E^1,

$$v(b(t),t) - p^1(t) \geq v(b(t'),t) - p^1(t'),$$

is equivalent to the IC constraint for mechanism (b, p^2) in environment E^2

$$\Big(v(b(t),t) - c(b(t))\Big) - \Big(p^1(t) - c(b(t))\Big) \geq \Big(v(b(t'),t) - c(b(t'))\Big) - \Big(p^1(t') - c(b(t'))\Big).$$

Similarly the IR constraint for mechanism (b, p^1) in environment E^1 is equivalent to the IR constraint for mechanism (b, p^2) in environment E^2. Additionally, the profit of mechanism (b, p^1) in environment E^1, $\mathbf{E}[p^1(t) - c(b(t))]$, is equal to the profit of mechanism (b, p^2) in environment E^2. As a result, the optimal profit in environment E^1 is equal to the optimal profit in environment E^2. Finally, mechanism (b, p^1) is a uniform markup pricing mechanism in environment E^1 if and only if mechanism (b, p^2) is a uniform markup pricing mechanism in environment E^2. Therefore, uniform markup pricing is optimal in environment E^1 if it is optimal in environment E^2. Given the first step of the proof, we conclude that uniform markup pricing is optimal in environment E if it is optimal in environment E^2.

Third, a grand bundle can be assumed to exist. In particular, we can add a bundle to the set of bundles that represents each type's favorite bundle. Formally, for some $\tilde{b} \notin B$, define $B^3 = B \cup \{\tilde{b}\}$, and let $v^3(\tilde{b}, t) = \max_{b \in B} v^2(b, t)$, and $v^3(b, t) = v^2(b, t)$ for all other bundles $b \neq \tilde{b}$. Bundle $\tilde{b}$ is a grand bundle in environment $E^3 = (B^3, T, v^3, 0)$, that is, $v^3(\tilde{b}, t) \geq v^3(b, t)$ for all $b \in B^3$. Notice also that the optimal profit in environment E^2 is no higher than the optimal profit in environment E^3. Indeed, any IC and IR mechanism in environment E^2 is also IC and IR in environment E^3, and has identical profit in the two environments. Additionally, the profit of uniform markup pricing with markup p in environment E^2 is equal to the profit of pure bundling with price p in environment E^3. In each case, a type pays p if $\max_{b \in B} v^2(b, t) \geq p$, and pays zero otherwise. Therefore, uniform markup pricing is optimal in environment E^2 if pure bundling is optimal in environment E^3. Given the first two steps of the proof, we conclude that uniform markup pricing is optimal in environment E if pure bundling is optimal in environment E^3.

Optimality of pure bundling in environment E^3 follows directly from Theorem 1. Let $r^3(\cdot, t) = v^3(\cdot, t)/v^3(\tilde{b}, t)$, and notice that $r^3(b, t) = \tilde{r}(b, t)$. Theorem 1 states that pure bundling is optimal in environment E^3 if

$$\mathbf{Pr}\left[r^3(\cdot, t) \in U \mid v^3(\tilde{b}, t) = \hat{v}\right] = \mathbf{Pr}\left[\tilde{r}(\cdot, t) \in U \mid \max_b \tilde{v}(b, t) = \hat{v}\right]$$

is non-decreasing in $\hat{v} > 0$ for all upper sets $U \subseteq \mathbb{R}^B$, which holds by the assumption of the theorem. We conclude that uniform markup pricing is optimal for distribution μ in environment (B, T, v, c). □

As an application of Theorem 2, assume that there are two bundles $B = \{\emptyset, b_1, b_2\}$, $T \subseteq \mathbb{R}^3_+$, and the values for a type $t = (t_0, t_1, t_2)$ are $v(b_1, t) = t_0 t_1$ and $v(b_2, t) = t_0 t_2$. Assume further that $\max(t_1, t_2) = 1$ for all types. Different types may have different favorite bundles. A type t prefers bundle b_1 if $t_1 \geq t_2$, and prefers bundle b_2 if $t_1 \leq t_2$. Notice that $\max_b v(b,t) = t_0 \max(t_1, t_2) = t_0$, $r(b_1,t) = \frac{t_0 t_1}{t_0} = t_1$, and $r(b_2,t) = \frac{t_0 t_2}{t_0} = t_2$. Therefore, Theorem 2 states that uniform markup pricing, that is offering each bundle b_1, b_2 at a uniform price p, is optimal if (t_1, t_2) is stochastically non-decreasing in t_0. A special case is when the values for the two bundles $(v(b_1,t), v(b_2,t))$ are uniformly distributed on the uniform square $[0,1]^2$, where Pavlov (2011) show that uniform markup pricing is optimal.

Notice that if indeed a bundle gb that is an efficient bundle for all types exists, that is $\max(v(b,t) - c(b), 0) \leq \max(v(gb,t) - c(gb), 0)$ for all bundles b, then $\max_b \tilde{v}(b,t) = \tilde{v}(gb,t)$. The condition of Theorem 2 becomes that $\mathbf{Pr}[\frac{\tilde{v}(\cdot,t)}{\tilde{v}(gb,t)} \in U \mid \tilde{v}(gb,t) = \hat{v}]$ is non-decreasing in $\hat{v} > 0$ for all upper sets $U \subseteq \mathbb{R}^B$.

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

The literature on multi-product screening often rationalizes the use of simple bundling strategies. We contribute to this strand of literature by identifying when pure bundling is optimal. In practice, even if firms do not offer only 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.

Our work sheds light on effective product design strategies. Consider a monopolist who can create various low quality configurations of a product it sells. Which configuration should it offer in order to profit from price discrimination? The low quality configuration should be one for which higher value consumers have lower relative value. For instance, imagine a hotel with two types of customers, leisure and business, where business travelers have higher overall willingness to pay than leisure travelers. The hotel wants to price discriminate between the two types by creating two levels of service. To profitably do so, the hotel should offer a low quality service for which business travelers lower higher relative value, for example rooms with no Internet access (as opposed to, say, no access to recreational facilities). Thus the optimal menu with two levels of service offers rooms with and without Internet access.

Even though we focus on a multi-product monopolist, our analysis may have implications for oligopolistic competition and antitrust. Tying, the practice of conditioning the purchase of one product on the purchase of another product, has long been a focus of antitrust policy. On one hand, it is argued that tying can be an anticompetitive practice since it allows a firm to extend its market power from one market to another (see Rey and Tirole, 2007 for a survey). On the other hand, the literature has suggested alternative explanations for tying that are not necessarily anticompetitive, such as price discrimination, economies of joint sales, and risk sharing (Bowman Jr, 1957, Whinston, 1990). Our analysis lends support to the price discrimination motive by generalizing the known conditions under which tying can be optimal for a monopolist.[9]

[9]Tying and pure bundling are not exactly identical but are closely related. Tying occurs when product 1 is available only if the consumer purchases product 2, whereas product 2 may be available as a standalone product. One can view a pure bundling strategy as a tying strategy in which the price for the bundle is the same as the price for the individual product. Therefore, conditions for the optimality of tying are at least as general as the conditions for the 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.

Bowman Jr, W. S. (1957). Tying arrangements and the leverage problem. *Yale Lj*, 67:19.

Cai, Y., Devanur, N. R., and Weinberg, S. M. (2016). A duality based unified approach to bayesian mechanism design. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 926–939. ACM.

Carroll, G. (2017). Robustness and separation in multidimensional screening. *Econometrica*, 85(2):453–488.

Church, J. R. and Ware, R. (2000). *Industrial organization: a strategic approach*. Irwin McGraw Hill Boston.

Daskalakis, C., Deckelbaum, A., and Tzamos, C. (2017). Strong duality for a multiple-good monopolist. *Econometrica*, 85(3):735–767.

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.

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.

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).

Rey, P. and Tirole, J. (2007). A primer on foreclosure. *Handbook of industrial organization*, 3:2145–2220.

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.

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. *J. Economic Theory*, 117(2):217–245.

US Census Bureau (2006). American community survey. `https://web.archive.org/web/20060626001929/http://www.census.gov/hhes/www/income/medincsizeandstate.html`.

Whinston, M. D. (1990). Tying, foreclosure, and exclusion. *American Economic Review*, 80(4):837–59.

Wilson, R. B. (1993). *Nonlinear pricing*. Oxford University Press.

A Appendix

A.1 Proof of Corollary 1

Corollary 1. *Assume that* $B = \{\emptyset, sb, gb\}$. *Pure bundling is optimal if* $\mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$ *is non-decreasing in* $\hat{v} > 0$ *for all* $\hat{r} \in \mathbb{R}$.

Proof. Theorem 1 states that pure bundling is optimal if $\mathbf{Pr}[r(\cdot,t) \in U \mid v(gb,t) = \hat{v}]$ is non-decreasing in $\hat{v} > 0$ for all upper sets $U \in \mathbb{R}^B$. Since $r(\emptyset, t) = 0$ and $r(gb,t) = 1$ for all types t, we can restrict attention to sets U such that $x(\emptyset) = 0$ and $x(gb) = 1$ for all $x \in U$. Such a set is an upper set if and only if $U = \{x \mid x(\emptyset) = 0, x(sb) \geq \hat{r}, x(gb) = 1\}$ for some $\hat{r} \in \mathbb{R}$. Thus, by Theorem 1 pure bundling is optimal if $\mathbf{Pr}[r(\cdot,t) \in U \mid v(gb,t) = \hat{v}] = \mathbf{Pr}[r(sb,t) \geq \hat{r} \mid v(gb,t) = \hat{v}]$ is non-decreasing in $\hat{v} > 0$ for all $\hat{r} \in \mathbb{R}$. $\square$

A.2 Proof of Lemma 1

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

$$\mathbf{E}\left[p(t)\right] = \mathbf{E}\left[\phi(b(t), t)\right] - \Big(v(b(\underline{t}), \underline{t}) - p(\underline{t})\Big).$$

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

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = \partial_2 v(b(t), t), \tag{8}$$

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

$$\begin{aligned}
\mathbf{E}\left[u(t)\right] &= \int_{\underline{t}}^{\bar{t}} u(t) f_{gb}(t)\, \mathrm{d}t \\
&= \int_{\underline{t}}^{\bar{t}} \frac{\mathrm{d}}{\mathrm{d}t} u(t)(1 - F_{gb}(t))\, \mathrm{d}t + u(\underline{t}) \\
&= \int_{\underline{t}}^{\bar{t}} \partial_2 v(b(t), t)(1 - F_{gb}(t))\, \mathrm{d}t + u(\underline{t}) \\
&= \mathbf{E}\left[\partial_2 v(b(t), t) \frac{1 - F_{gb}(t)}{f_{gb}(t)}\right] + u(\underline{t})
\end{aligned}$$

where the second to last equality followed from substituting (8). Finally, the expected revenue can be written as the difference of surplus and consumer rents,

$$\begin{aligned}
\mathbf{E}\left[p(t)\right] &= \mathbf{E}\left[v(b(t), t) - u(t)\right] \\
&= \mathbf{E}\left[v(b(t), t)) - \partial_2 v(b(t), t) \frac{1 - F_{gb}(t)}{f_{gb}(t)}\right] - u(\underline{t}) \\
&= \mathbf{E}\left[\phi(b(t), t)\right] - \Big(v(b(\underline{t}), \underline{t}) - p(\underline{t})\Big).
\end{aligned}$$

□

A.3 Proof of Lemma 2

Lemma 2. *If $v(\cdot, t)/t$ is monotone non-decreasing in t, then $\phi(b,t) \leq \max(0, \phi(gb,t))$ for all b and t.*

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

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

As a result,

$$\frac{v(b,t)}{t} \leq \partial_2 v(b,t). \tag{9}$$

Directly from the definition of virtual values,

$$\begin{aligned}
\phi(b,t) &= v(b,t) - \partial_2 v(b,t) \times \frac{1 - F_{gb}(t)}{f_{gb}(t)} \\
&\leq v(b,t) - \frac{v(b,t)}{t} \times \frac{1 - F_{gb}(t)}{f_{gb}(t)} \\
&= \frac{v(b,t)}{t}[t - \frac{1 - F_{gb}(t)}{f_{gb}(t)}]
\end{aligned}$$

where the inequality followed from (9). Notice that the right hand side is $\frac{v(b,t)}{t}$ multiplied by the virtual value for the grand bundle $\phi(gb,t)$. Therefore,

$$\phi(b,t) \leq \frac{v(b,t)}{t}\phi(gb,t). \tag{10}$$

We now complete the proof by considering two cases and applying Inequality (10). If $0 \leq \phi(gb,t)$ we have

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

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

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

Given the above two inequalities, we have $\phi(b,t) \leq \max(0, \phi(gb,t))$ for all b and t. □

A.4 Proof of Proposition 1, the "if" Statement

The proof of the first statement of Proposition 1 requires a setup, with which we start.

A.4.1 Setting

Assume that set of types $T \subseteq \mathbb{R}_+$ is finite. That is, types are $t_0 (= \underline{t}) < t_1 < \ldots < t_I$. To avoid dividing by zero, assume without loss of generality that $t_0 > 0$. Let $f_{gb}(t_i)$ 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.4.2 Generalized Virtual Values

We now discuss a general construction of virtual values based on Lagrangian duality, identical to Cai et al. (2016); Carroll (2017). For any i and j from $\{0, \ldots, I\}$, let $\lambda(j, i) \geq 0$ be the Lagrangian multiplier of the IC constraint

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

Define the Lagrangian

$$\begin{aligned}
\mathcal{L}(\lambda, b, p) &= \Big(\sum_i p(t_i) f_{gb}(t_i)\Big) + \Big(\sum_{i,j} \lambda(j,i)\Big(v(b(t_j), t_j) - p(t_j) - (v(b(t_i), t_j) - p(t_i))\Big)\Big) \\
&= \sum_i \Big(p(t_i)\Big(f_{gb}(t_i) - \sum_j \lambda(i,j) + \sum_j \lambda(j,i)\Big)\Big) \\
&\quad + \sum_i \Big(\Big(\sum_j \lambda(i,j) v(b(t_i), t_i)\Big) - \Big(\sum_j \lambda(j,i) v(b(t_i), t_j)\Big)\Big).
\end{aligned}$$

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

$$f_{gb}(t_i) - \sum_j \lambda(i,j) + \sum_j \lambda(j,i) = 0. \tag{11}$$

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 (11) feasible. We interpret (11) 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 cancels and the Lagrangian becomes

$$\begin{aligned}
&\mathcal{L}(\lambda, b, p) \\
&= \sum_i \Big(\big(\sum_j \lambda(i, j) v(b(t_i), t_i) \big) - \big(\sum_j \lambda(j, i) v(b(t_i), t_j) \big) \Big) \\
&= \sum_i \Big(v(b(t_i), t_i) f_{gb}(t_i) - \sum_j \lambda(j, i)(v(b(t_i), t_j) - v(b(t_i), t_i)) \Big), \qquad (12)
\end{aligned}$$

where the equality followed from substituting (11). Define the *induced virtual value of* λ

$$\phi(b, t_i) := v(b, t_i) - \frac{1}{f_{gb}(t_i)} \sum_j \lambda(j, i)(v(b, t_j) - v(b, t_i)), \qquad \forall t_i \in T, b \in B. \qquad (13)$$

Substituting 13 into 12, the Lagrangian is the expected virtual surplus of b

$$\mathcal{L}(\lambda, b, p) = \sum_{t_i} \phi(b(t_i), t_i) \times f_{gb}(t_i). \qquad (14)$$

We summarize the analysis above in the following lemma.

Lemma 6 (Cai et al., 2016; Carroll, 2017)**.** *A mechanism (b, p) is optimal if and only if there exists feasible λ such that b maximizes the virtual surplus defined in Equations 13 and 14, and (b, p) and λ satisfy the complimentary slackness condition,*

$$\lambda(j, i)\Big(v(b(t_j), t_j) - p(t_j) - (v(b(t_i), t_j) - p(t_i))\Big) = 0. \qquad (15)$$

A.4.3 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 7. *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* (13)*, satisfy*

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

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

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

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

$$\begin{aligned}\frac{v(b,t_i)}{t_i}\phi(gb,t_i) &= v(b,t_i) - \frac{1}{f_{gb}(t_i)}\sum_{j:j>i} \lambda(j,i)(\frac{v(b,t_i)}{t_i}v(gb,t_j) - v(b,t_i))\\ &\geq v(b,t_i) - \frac{1}{f_{gb}(t_i)}\sum_{j:j>i} \lambda(j,i)(v(b,t_j) - v(b,t_i))\\ &= v(b,t_i) - \frac{1}{f_{gb}(t_i)}\sum_{j} \lambda(j,i)(v(b,t_j) - v(b,t_i))\\ &= \phi(b,t_i),\end{aligned}$$

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

To interpret Lemma 7 geometrically, note from definition (13) that viewed as a vector, $\phi(\cdot,t_i)$ is equal to the vector $v(\cdot,t_i)$, shifted proportional 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 8.

Given Lemma 7, 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 7 applies, (3) the induced virtual value $\bar{\lambda}$ of definition (13) for the favorite bundle $\bar{\phi}(gb,t)$ is monotone non-decreasing in t, and (4) the assignment rule that only assigns the grand bundle gb to types t with positive $\bar{\phi}(gb,t)$ satisfies the complementary slackness condition with $\bar{\lambda}$. The lemma below shows that such dual variables exist (we verify property (4) later).

Lemma 8. *There exist dual variables $\bar{\lambda}$ with induced virtual value $\bar{\phi}$ such that*

(I) $\bar{\lambda}$ is feasible, that is, it satisfies (11)*,*

(II) If $\bar{\lambda}(j,i) > 0$ *then* $i < j$*,*

(III) $\bar{\phi}(gb,t)$ *is monotone non-decreasing in* t*,*

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

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

$$R(F_i) = t_i F_i.$$

Define the ironed revenue function $\bar{R}$, defined over support $\{F_i\}_{i \in I}$, 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 gb satisfies

$$\bar{\phi}(gb,t_i) = \frac{\bar{R}(F_i) - \bar{R}(F_{i+1})}{f_{gb}(t_i)}.$$

That is, $\bar{\phi}(gb,t_i)$ is the slope of the ironed revenue curve at t_i. By concavity of $\bar{R}$, $\bar{\phi}(gb,t_i)$ 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,

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

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 (13), and the revenue curve is,

$$R^\kappa(F_i) = \sum_{j \geq i} \phi^\kappa(gb,t_j) f_{gb}(t_j), \tag{16}$$

which is equivalent to

$$\phi^\kappa(gb, t_i) = \frac{R^\kappa(F_i) - R^\kappa(F_{i+1})}{f_{gb}(t_i)}.$$

In each iteration κ, the variables satisfy four properties: (i) λ^κ is feasible; (ii) λ^κ is positive only for downward constraints; (iii) $R^\kappa(F_i) = \bar{R}(F_i)$ if $i \geq \kappa$ and $R^\kappa(F_i) = R(F_i)$ if $i < \kappa$; and, (iv) if $\lambda^\kappa(j, i) > 0$ and $j > i + 1$, then $R(F_{j'}) < \bar{R}(F_{j'})$ for all j' such that $i < j' < j$. Properties (i), (ii), and (iv) are trivially satisfied when $\kappa = n$. To see property (iii) when $\kappa = n$, write

$$\begin{aligned}
R^n(F_i) &= \sum_{j \geq i} \phi^n(gb, t_j) f_{gb}(t_j) \\
&= \sum_{j \geq i} \Big(v(gb, t_j) f_{gb}(t_j) - \sum_{j'} \lambda^n(j', j)(v(gb, t_{j'}) - v(gb, t_j)) \Big) \\
&= \sum_{j \geq i} \Big(t_j f_{gb}(t_j) - \sum_{j'} \lambda^n(j', j)(t_{j'} - t_j) \Big) \\
&= \sum_{j \geq i} \Big(t_j f_{gb}(t_j) - F_{j+1}(t_{j+1} - t_j) \Big) \\
&= \sum_{j \geq i} \Big(t_j F_j - F_{j+1} t_{j+1} \Big) \\
&= t_i F_i = R(F_i).
\end{aligned}$$

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

$$\lambda^\kappa(j, \kappa) = \gamma \lambda^{\kappa+1}(j, \kappa), \qquad \forall j > \kappa, \tag{17}$$

$$\lambda^\kappa(j, \kappa - 1) = \lambda^{\kappa+1}(j, \kappa - 1) + (1 - \gamma)\lambda^{\kappa+1}(j, \kappa), \qquad \forall j > \kappa, \tag{18}$$

$$\lambda^\kappa(\kappa, \kappa - 1) = \lambda^{\kappa+1}(\kappa, \kappa - 1) - (1 - \gamma) \sum_{j > \kappa} \lambda^{\kappa+1}(j, \kappa), \tag{19}$$

for a parameter γ to be identified shortly. See Figure 11. Define ϕ^κ via (13) and R^κ via (16).

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 (11) 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

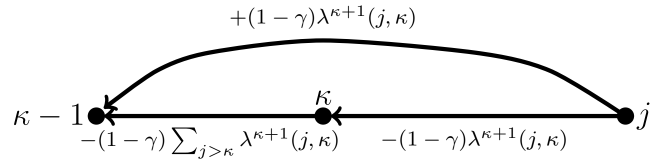

Figure 11: The change in λ^κ compared to $\lambda^{\kappa+1}$. For each $j > \kappa$, A fraction of the flow from j to κ and then to $\kappa-1$ is rerouted to go directly from j to $\kappa-1$.

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 $R^\kappa(F_{\kappa-1}) = R(F_{\kappa-1})$, notice that

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

and,

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

Summing up the above two equations, we get

$$\phi^\kappa(gb,\kappa)f_{gb}(t_\kappa) + \phi^\kappa(gb,\kappa-1)f_{gb}(t_{\kappa-1}) = \phi^{\kappa+1}(gb,\kappa)f_{gb}(t_\kappa) + \phi^{\kappa+1}(gb,\kappa-1)f_{gb}(t_{\kappa-1}).$$

From 16, we conclude that $R^\kappa(F_{\kappa-1}) = R^{\kappa+1}(F_{\kappa-1}) = R(F_{\kappa-1})$.

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

$$\bar{\phi}(gb,\kappa) = \frac{R(j) - R(j')}{F_j - F_{j'}}$$

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

$$\bar{\phi}(gb, \kappa) \leq \frac{t_j F_j - t_j F_{j'}}{F_j - F_{j'}} = t_j.$$

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

$$\begin{aligned} R^\kappa(F_\kappa) &= \phi^\kappa(gb, t_\kappa) f_{gb}(t_\kappa) + R^\kappa(F_{\kappa+1}) \\ &\geq \bar{\phi}(gb, \kappa) + \bar{R}(F_{\kappa+1}) \\ &= \bar{R}(F_{\kappa+1}). \end{aligned}$$

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

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 $R(F_\kappa) < \bar{R}(F_\kappa)$ (as otherwise $\gamma = 1$) and $\lambda^{\kappa+1}(j, \kappa) > 0$. In this case, by induction hypothesis, we must have $R(F_{j'}) < \bar{R}(F_{j'})$ 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 $R(F_{j'}) < \bar{R}(F_{j'})$ for all j' such that $i < j' < j$. Note from the definition of $\bar{\phi}$ and $\bar{R}$ that if $R(F_{j'}) < \bar{R}(F_{j'})$ then $\bar{\phi}(gb, t_{j'-1}) = \bar{\phi}(gb, t_{j'})$. As a result, $\bar{\phi}(gb, t_{j'}) = \bar{\phi}(gb, t_{j''})$ for all j', j'' such that $i \leq j', j'' < j$. □

A.4.4 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 $T_1, \ldots, T_n$ such that $v(b, t) - v(b, t') \leq \delta$ for all $t, t' \in T_i$, and $\mathbf{Pr}_\mu[t \in T_i] = \mathbf{Pr}_{\mu'}[t \in T_i]$.

Lemma 9 (Madarász and Prat, 2017; Lemma 4.3 in Carroll, 2017)**.** *For every $\epsilon > 0$, there exists $\delta > 0$ such that for every mechanism (b, p), there exists a mechanism (b', 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.4.5 The Proof

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

Proposition 1. *Assume that $T \subseteq \mathbb{R}^+$ and $v(gb, t) = t$. 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. The "only if" Statement is proved in the main body. We here prove the "if" Statement.

We start by assuming that T is finite. We later extend the proof to arbitrary T using Lemma 9.

Consider $\bar{\lambda}$ and $\bar{\phi}$ from Lemma 8. Since $\bar{\phi}(gb, t)$ is monotone non-decreasing in t, there exists a threshold t^* such that $\bar{\phi}(gb, t) < 0$ if $t \leq t^*$, and $\bar{\phi}(gb, t) \geq 0$ otherwise. Consider a pure bundling mechanism (b, p) that only offers the grand bundle gb for price t^*. We show that (b, p) is optimal by arguing that the mechanism and $\bar{\lambda}$ satisfy conditions of Lemma 6.

Feasibility of $\bar{\lambda}$ follows directly from property (1) of Lemma 8. By property (2) of Lemma 8, and since $v(b, t_i)/t_i$ is monotone non-decreasing in t_i, $\bar{\phi}$ satisfies

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

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

It only remains to verify the complementary slackness condition 15. If a pair of types get the same assignment in (b, p) (either gb 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 8, $\bar{\lambda}(i, j) = 0$. 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 8 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 exists 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.

Assume 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 $\{t'_0, t'_1, \ldots\}$ of μ' inductively from T. Let $t'_0 = \inf(T)$. Set $\epsilon = \sigma/3$, and consider δ from Lemma 9. 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 9, 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. □

A.5 Proof of Theorem 1

In this section we extend the proof of Theorem 1 to infinite B.

Proof of Theorem 1, general B. Assume for contradiction that there exists a mechanism with revenue at least $\epsilon > 0$ larger than any pure bundling mechanism. By continuity of v, there exists a set of bundles $\bar{B}$ such that for any bundle $b \in B$, there exists a bundle $\bar{b} \in \bar{B}$ such that $|v(b,t) - v(gb,t)| < \epsilon/2$ for all $t \in T$. By Lemma 9, when the set of bundles is $\bar{B}$, there exists a mechanism with revenue at least $\epsilon/2$ larger than any pure bundling mechanism. This contradicts the proof of Theorem 1 for finite $\bar{B}$. □

A.6 Proof of Proposition 2

We now prove Section A.6 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 $\bar{r}(b, \hat{v})$ is monotone decreasing in $\hat{v}$ for some b, and the distribution of values is continuous and interior.*

Proof. Consider the optimal price p for only selling the grand bundle gb. We prove the optimal pure bundling revenue can be improved by offering another bundle, contradicting

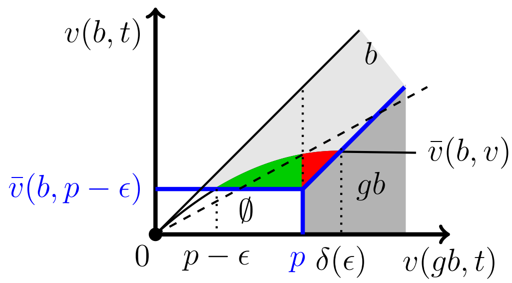

Figure 12: As a result of adding an offer with price $V(p-\epsilon)$ for bundle a to the existing offer of price p for bundle gb, the types in darker shaded part of curve V will change decisions and contribute to a change in revenue.

the optimality of pure bundling. Let $f = f_{gb,b}$ denote the density function of the joint distribution of the value for the grand bundle and value for bundle b.

In particular, let $\bar{v}(b,t)$ be the highest value of $v(b,t)$ among all types t with $v(gb,t) = v$. By assumption, $\bar{v}(b,v)/v$ is decreasing in v. Consider the change in revenue as a result of offering bundle b for a price $\bar{v}(b,p-\epsilon)$ in addition to gb for price p. In this new mechanism, a type t chooses gb if $v(gb,t)-p \geq \max(0, v(b,t)-\bar{v}(b,p-\epsilon))$, chooses b if $v(b,t)-\bar{v}(b,p-\epsilon) \geq \max(0, v(gb,t)-p)$, and chooses $\emptyset$ otherwise. See Figure 12.

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 green in Figure 12). The gain in revenue $\nabla_+(\epsilon)$ is

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

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

$$\mu_+(\epsilon) := \mu(T_+(\epsilon)) = \int_{p-\epsilon}^{p} \int_{\bar{v}(b,p-\epsilon)}^{\bar{v}(b,t)} f(v,z)\, \mathrm{d}z\, \mathrm{d}v. \tag{20}$$

On the other hand, a set of types $T_-(\epsilon)$ with value slightly higher than p for gb will change their decision from selecting bundle gb to bundle b (shaded red in Figure 12). 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 $T_-(\epsilon)$,

$$\mu_-(\epsilon) := \mu(T_-(\epsilon)) = \int_p^{\delta(\epsilon)} \int_{t-p+\bar{v}(b,p-\epsilon)}^{\bar{v}(b,t)} f(v,z)\, \mathrm{d}z\, \mathrm{d}v.$$

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 follow

$$\begin{aligned}
\mu'_+(\epsilon) &= -\int_{\bar{v}(b,p-\epsilon)}^{\bar{v}(b,p-\epsilon)} f(p-\epsilon,z)\,\mathrm{d}z - \int_{p-\epsilon}^{p} \partial_2\bar{v}(b,p-\epsilon) f(v,\bar{v}(b,p-\epsilon))\,\mathrm{d}v\\
&= -\int_{p-\epsilon}^{p} \partial_2\bar{v}(b,p-\epsilon) f(v,\bar{v}(b,p-\epsilon))\,\mathrm{d}v, \text{ and} \qquad (21)\\
\mu''_+(\epsilon) &= \partial_2\bar{v}(b,p-\epsilon) f(p-\epsilon,\bar{v}(b,p-\epsilon)) - \int_{p-\epsilon}^{p} \frac{\mathrm{d}}{\mathrm{d}\epsilon}\partial_2\bar{v}(b,p-\epsilon) f(v,\bar{v}(b,p-\epsilon))\,\mathrm{d}v.
\end{aligned}$$

Now notice from 20 that $\mu_+(0) = 0$ and from 21 that $\mu'_+(0) = 0$. Therefore, $\nabla'_+(0) = 0$. Now we can calculate $\nabla''_+(0)$ as follows

$$\nabla''_+(0) = \bar{v}(b,p) \times \mu''_+(0) = \bar{v}(b,p) \times \partial_2\bar{v}(b,p) f(p,\bar{v}(b,p)). \qquad (22)$$

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_-,

$$\mu'_-(\epsilon) = \delta'(\epsilon) \int_{\delta(\epsilon)-p+\bar{v}(b,p-\epsilon)}^{\bar{v}(b,\delta(\epsilon))} f(\delta(\epsilon), z) \, \mathrm{d}z - \int_p^{\delta(\epsilon)} \partial_2 \bar{v}(b, p-\epsilon) f(v, v-p+\bar{v}(b,p-\epsilon)) \, \mathrm{d}v$$
$$= -\int_p^{\delta(\epsilon)} \partial_2 \bar{v}(b, p-\epsilon) f(v, v-p+\bar{v}(b,p-\epsilon)) \, \mathrm{d}v.$$
$$\mu''_-(\epsilon) = -\delta'(\epsilon)\partial_2 \bar{v}(b, p-\epsilon) f(\delta(\epsilon), \delta(\epsilon)-p+\bar{v}(b,p-\epsilon))$$
$$-\int_p^{\delta(\epsilon)} \frac{\mathrm{d}}{\mathrm{d}\epsilon} \partial_2 \bar{v}(b, p-\epsilon) f(v, v-p+\bar{v}(b,p-\epsilon)) \, \mathrm{d}v$$

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

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

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

$$\bar{v}(b,p) > (p - \bar{v}(b,p))\delta'(0). \tag{23}$$

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 gb 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 23, we need to show that

$$\bar{v}(b,p) > (p - \bar{v}(b,p)) \frac{\partial_2 \bar{v}(b,p)}{1 - \partial_2 \bar{v}(b,p)}. \tag{24}$$

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 24, completing the proof. □