When Is Pure Bundling Optimal?*

Nima Haghpanah† Jason Hartline‡

September 24, 2018

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 value for the grand bundle have lower relative disutility for consuming smaller bundles. Conversely, pure bundling is not optimal if consumers with higher value for the grand bundle have higher relative disutility for consuming smaller bundles. We prove the results using a decomposition approach that relies on identifying binding incentive constraints with multi-dimensional heterogeneity.

*We thank Nageeb Ali, 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, and seminar participants at Arizona State University, Northwestern University, University of Chicago and Becker Friedman Institute, University of Cologne, University of Pennsylvania, University of Rochester, Penn State University, Simons Institute for Theoretical Computer Science, as well as the participants at NBER Market Design 2014, Stony Brook Game Theory Conference 2015, ACM EC 2015, and the North American Econometric Society Meeting 2016. We thank 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). The monopolist may choose to offer different bundles in order to screen types (McAfee et al., 1989), or to sell products separately to obtain robust profit guarantees (Carroll, 2017). Normatively, the selling strategy directly affects producer and consumer value.[1] For instance, screening types may involve distorting allocations and thus may cause efficiency loss. We characterize when pure bundling, i.e., offering only the grand bundle of all products, is optimal. Our characterization is easy to state and has a straightforward intuition.

Consider a single-product seller and a buyer who needs at most two units of the product.[2] Assume that costs are zero.[3] The buyer's type, consisting of its value for one unit v_1 and its value for a bundle (two units) $v_2 \geq v_1$, is drawn privately from a distribution. To maximize profit, should the seller use pure bundling and only offer the bundle? Or should it use mixed bundling and offer one unit at a discounted price, and the bundle at a full price?

We show that optimality of pure bundling depends on the correlation between the buyer's value for the bundle v_2 and the buyer's *value-ratio* v_1/v_2. Pure bundling is optimal if the distribution of value-ratio is stochastically non-decreasing in the value for the bundle; i.e., types with higher value for the bundle are more likely to have higher value-ratios. A type with a high value-ratio has a low relative disutility for consuming only a unit instead of the bundle (i.e., $1 - v_1/v_2$). Thus, pure bundling is optimal if types with higher value for the bundle are more likely to have lower relative disutility for consuming smaller bundles. Conversely, pure bundling is not optimal if the distribution of value-ratio is stochastically decreasing in the value for the bundle. The exact same statements hold for selling multiple products, once value-ratio is appropriated defined: pure bundling is optimal if value-ratio is stochastically non-decreasing, and is not optimal if value-ratio is stochastically decreasing, in the value for the grand bundle.[4] The exact definition of value-ratio for multiple products

[1]For a recent empirical literature estimating the welfare effects of bundling, see Byzalov (2008); Crawford and Yurukoglu (2012) (cable TV), Ho et al. (2012) (video rentals), and Shiller and Waldfogel (2011) (music).

[2]A different interpretation is that the seller is facing a population of consumers, each needing at most two units, and can not discriminate based on consumers' identities.

[3]The monopolist's optimal selling strategy depends on the distribution of consumers' tastes as well costs of producing different bundles (e.g., existence of economies of scope). In order to focus only on how consumers' tastes affects optimal screening menus, we mainly assume that costs are zero. This assumption models markets for information goods (cable TV, software, movies, and music). We do consider extensions to allow for arbitrary costs, and provide conditions for optimality of mechanism that offers all bundles at a uniform markup above their costs.

[4]The stochastic monotonicity condition required for optimality of pure bundling is the standard first order stochastic dominance. On the other hand, the condition for non-optimality of pure bundling is a new

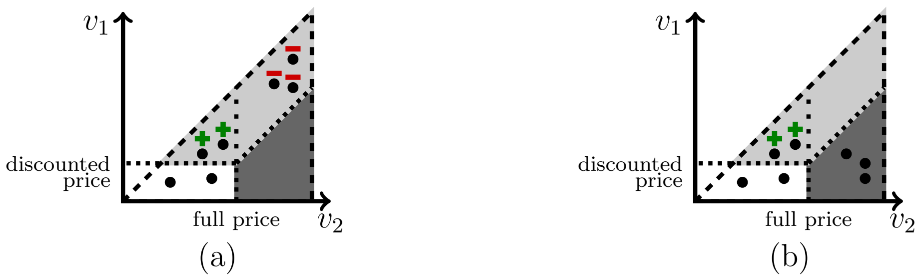

Figure 1: Given the option to buy two units at full price or one unit at a discounted price, types in the area shaded dark buy two units, in the area shaded light buy one unit, and otherwise buy nothing. (a) Types with higher v_2 have higher v_1/v_2. (b) Types with higher v_2 have lower v_1/v_2.

is given later.

The characterization has a straightforward economic intuition. Assume that types with higher value for the bundle have higher value-ratio, as in Figure 1a. Offering one unit at a discounted price in addition to the bundle has a gain and a loss, compared to offering no such discount. The gain is from types who are unwilling to pay the full price for the bundle, but are willing to buy one unit (types marked with a "+"). 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 (types marked with a "−"). These types have high value for the bundle and low disutility for consuming only a unit. If types with higher value for the bundle have lower disutility, the loss is relatively large, and mixed bundling is not profitable (pure bundling is optimal). Conversely, if types with higher value for the bundle have higher disutility, as in Figure 1b, then the loss is relatively small and mixed bundling is profitable (pure bundling is not optimal).[5]

A conventional view in the bundling literature is that bundling is profitable (compared to selling the products separately) if the values for individual products are negatively correlated (see, for example, Church and Ware, 2000). This view is mostly based on the assumption that preferences are additive (i.e., the value for a bundle is the sum of the values of its constituting products), and was first suggested through examples by Stigler (1963) and Adams and Yellen (1976).[6] On the other hand, the literature shows that pure bundling

notion of strict first order stochastic dominance. The statement regarding non-optimality of pure bundling also requires continuity and interiority of the distribution.

[5]This intuitive explanation does not consider the possibility of offering multiple bundles for selling multiple products. The proof of optimality of pure bundling is more involved and is discussed in detail later. On the other hand, to prove that pure bundling is *not* optimal, we directly analyze the gain and loss of offering a small discount and show that the net gain is positive (Appendix A.7).

[6]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. However,

is generally not optimal. Adams and Yellen (1976) and McAfee et al. (1989) show that pure bundling is generally strictly dominated by mixed bundling, i.e., offering prices for all bundles, even with negatively correlated values. Furthermore, mixed bundling is itself dominated by offering randomized bundles (Thanassoulis, 2004; Daskalakis et al., 2017).[7] Indeed, Vincent and Manelli (2007) and Hart and Nisan (2013) show that the optimal menu may offer infinitely many randomized bundles. Thus the literature does not explain the prevalence of pure bundling mechanisms (e.g., cable TV or software components such as Microsoft Office).

Our results show that pure bundling is optimal under general conditions, once the assumption of additivity is relaxed. Furthermore, these conditions are closely related to the negative correlation of product values. The appropriate notion of negative correlation for optimality of pure bundling is between the value for the grand bundle and the relative disutility for consuming smaller bundles.

Our condition for optimality of pure bundling is not knife edge. If the distribution of value-ratio is strictly stochastically increasing in the value for the bundle, then the same will hold for a small perturbation of the distribution. Thus pure bundling remains optimal for perturbations of the distribution. The previously known conditions for optimality of pure bundling do not satisfy this robustness property. Stokey (1979) and Riley and Zeckhauser (1983) (and Myerson, 1981, in the special case of a single buyer) show that when v_1/v_2 is constant, for example when $v_1 = v_2/2$, then pure bundling is optimal.[8] The analysis of Armstrong (1996), applied to our setting, shows that the same result holds when value for the bundle and value-ratio are independently distributed. In either case, the distribution of value-ratio is constant in the value for the 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. Robustness and generality of our condition for optimality of pure bundling suggest that prevalence of such mechanisms may not be solely due to their simplicity. Our result complements the numerical findings of Chu et al. (2011), that offering surprisingly few bundles is often nearly optimal, and rationalizes the widespread use of simple bundling mechanisms.

Schmalensee (1984) also shows that pure bundling may be profitable even with independent or positively correlated values. Fang and Norman (2006) obtain similar results for any number of products, but exclude mixed bundling (i.e., offering prices for all bundles) and only compare pure bundling with separate sales.

[7] Our setting also allows for randomized bundles.

[8] These results and our setting allow for a continuum of quantities.

Multiple Products. Assume now that the seller can produce bundles of possibly heterogeneous products. The buyer's type is its private information and specifies a value for each bundle. For each type, construct a profile of *value-ratios*, consisting of the ratio of value of each bundle to the value of the grand bundle. Pure bundling is optimal if the profile of value-ratios is stochastically non-decreasing in the value of the grand bundle. That is, types with higher value for the grand bundle are more likely to have higher value-ratio. Pure bundling is not optimal if there exists a bundle whose value-ratio stochastically decreases in the value of the grand bundle. Note that our conditions are stochastic ones and do not require structures such as additivity on values.

The intuition is identical to before. Types with high value-ratio have low relative disutility of consuming smaller bundles. If types with higher value for the grand bundle have lower relative disutility for consuming smaller bundles, the spillover of offering smaller bundles is large and mixed bundling is not profitable. The converse is similar.

An interpretation of the results is that pure bundling is optimal if consumers with higher value for the grand bundle consider the products to be less complementary. As an example, one way to express complementarities in preferences is to assume that the buyer's value for a bundle that includes a fraction $a_i \in [0, 1]$ of each product $i \in \{1, 2\}$ is

$$x \cdot \Big(y_1 a_1 + y_2 a_2 + (1 - y_1 - y_2) a_1 a_2\Big),$$

where the type of the buyer is (x, y_1, y_2).[9] Note that types with higher $y_1 + y_2$ consider the products to be less complementary. For instance, if $y_1 + y_2 < 1$, then the products are complements (value for a bundle is more than the values of its constituting products), and if $y_1 + y_2 > 1$, then the products are substitutes (value for a bundle is less than the values of its constituting products). Note also that the value for the grand bundle $(a_1, a_2) = (1, 1)$ is x. Pure bundling is optimal if (y_1, y_2) is stochastically non-decreasing in x, that is, if consumers with higher value for the grand bundle consider the products to be less complementary. Thus it is indeed possible for pure bundling to be optimal even if all consumers consider the products to be substitutes ($y_1 + y_2 > 1$), as long as the degree of complementarity is monotone. Conversely, pure bundling is not optimal if for some $i \in \{1, 2\}$, y_i is stochastically decreasing in x.

Our work sheds light on effective product differentiation strategies. Consider a monopolist who can create various low quality configurations of a product it sells. Which configuration

[9] This representation is without loss of generality, as we argue later.

should it offer in order to profit from price discrimination?[10] The low quality configuration should be one for which higher value consumers have higher relative disutility. For instance, imagine a hotel with two types of customers, leisure and business, where business travelers have higher values 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 have higher relative disutility, 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.

Welfare Comparisons. We explore two welfare consequences of our results. First, what does optimality of pure bundling imply about the overall efficiency? That is, does the fact that pure bundling is optimal have any implications about the fraction of the consumers that are excluded? The answer is negative. The correlation condition that implies optimality of pure bundling is orthogonal to how the value for the grand bundle is distributed. As a result, given any price, one can construct examples where it is robustly optimal (in the sense discussed before) to offer only the grand bundle at that price. In particular, if the price is equal to the lowest value in the support of the value for the grand bundle, then the optimal mechanism has efficient allocation (sells the grand bundle to all types), and if the price is equal to the highest value, then the optimal mechanism gives the buyer zero surplus.[11]

Second, how does preventing the seller from offering inferior products, e.g., one unit instead of a bundle in our starting example, affect efficiency? Keeping the price of the bundle fixed, offering one unit at a discounted price increases surplus.[12] However, the price of the bundle in an optimal mechanism may be lower or higher than it is in the mechanism that is optimal among pure bundling mechanisms. As a result, the surplus in the optimal menu may be lower or higher than the surplus in the optimal pure bundling mechanism.

[10] Our model discussed above is equivalent to a monopolist who sells different configurations of a product or service. For instance, in our starting example one can replace the bundle with a "premium" product, and one unit with a "standard" product. Our characterization gives conditions for offering only the premium product. Deneckere and McAfee (1996) provide several examples of firms creating damaged products in order to price discriminate. If offering only the premium product is optimal if costs are zero, it is also optimal if damaged products are more costly to produce.

[11] Armstrong (1999), Jehiel et al. (1999), and Severinov and Deneckere (2006) identify conditions under which the exclusion set is empty, although they do not address the efficiency of the allocation. Barelli et al. (2014) provide conditions under which the exclusion set is non-empty.

[12] Consumer surplus increases since prices decrease. Thus if it is optimal for the seller to offer one unit at a discounted price while keeping the price of the bundle fixed, total surplus increases as well.

Our Methodology. We contribute to the literature on multi-dimensional screening, roughly defined to be settings where single-crossing does not hold. Single-crossing is a condition that allows types to be ordered such that only local incentive constraints bind. Without single-crossing, such an ordering is not possible since the binding incentive constraints are endogenous. General methodologies for solving such problems are not known, with a few exceptions including Rochet and Chone (1998), Daskalakis et al. (2017), Carroll (2017), and the decomposition approaches discussed below.

Our approach consists of two components. The first component is to prove the results assuming that a type's value for the grand bundle uniquely determines its values for all bundles. In this case, pure bundling is optimal if value-ratio is (non-stochastically) 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 incentive constraints are not needed). Given this construction, we argue that pure bundling maximizes the virtual surplus and is therefore optimal, without relying on single-crossing.

The second component of our approach is to extend the result to general type spaces. The idea is to decompose the type space into partitions, and solve a problem where all incentive constraints between partitions are relaxed. If all binding incentive constraints are within partitions (and are thus not relaxed), the solution to the relaxed problem is indeed feasible. Thus this approach relies on identifying binding incentive constraints. Wilson (1993) and Armstrong (1996) first use this approach where, translated to our setting, each partition is a ray from the origin in the type space. Eső and Szentes (2007) and Pavan et al. (2014) significantly advance this idea by allowing the partitioning to depend on the distribution of types. In particular, they decompose the type space into a base parameter and independent "shocks". We use a classical characterization from the statistics literature (Strassen, 1965; Kamae et al., 1977), which allows us to construct a partitioning such that the value-ratio is (non-stochastically) non-decreasing within each partition. We then use the result from the first component of our approach to show that pure bundling is optimal for the relaxed problem.

Our approach allows us to transform the screening problem with multi-dimensional heterogeneity to one where a single parameter matters (i.e., the value for the grand bundle). Thus, once such a transformation is made, the theory of Myerson (1981) can be applied to provide extensions to multi-agent auctions.[13] Nevertheless, in order to show that such a transformation is possible we need to show that pure bundling is optimal among *all* mechanisms and thus need to overcome the challenges with multi-dimensional heterogeneity and endogenous binding constraints.

Related Work. Salant (1989), Johnson and Myatt (2003), and Anderson and Dana Jr (2009) consider price discrimination based on quality, which is equivalent to the problem of selling multiple units of a single product.[14] Phrased in terms of bundling, they show that pure bundling is optimal if there are increasing returns to quantity, which is similar in spirit to our condition. However, these results allow for only a single dimension of product heterogeneity (number of units or quality levels) and assume single-crossing, and thus do not apply to selling multiple products. In addition, these results require a non-stochastic monotonicity condition, which is related to our monotonicity condition when a type's value for the grand bundle is unique and is more restrictive than our general stochastic monotonicity condition.

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 independent 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 conditions for pure bundling when selling two products with independently distributed values and additive preferences. These conditions require virtual valuation of each product to be positive on the entire support of values, and generally satisfy neither of our stochastic monotonicity conditions.[15] McAfee and McMillan (1988) and Manelli and Vincent (2006) find sufficient conditions for optimality of mixed bundling. Vincent and Manelli (2007) show that essentially every mechanism is optimal for *some* distribution. Jehiel et al. (2007) show that pure bundling auctions are not optimal with multiple buyers.

[13] We do not directly explore auctions in this paper.

[14] Johnson and Myatt (2003) further study price discrimination in a duopoly.

[15] Since a distribution may be neither stochastically non-decreasing nor stochastically decreasing, our conditions do not cover all distributions.

1 The Setting

Consider a bundling problem with a single seller and a single buyer. There is a compact set of bundles (assignments) $\mathcal{A}$ that the seller can produce.[16] There is a compact set of buyer types Θ. The utility of a type $\theta \in \Theta$ for a bundle a and expected payment p to the seller is $V(a, \theta) - p$.[17] Assume that V is non-negative and continuous.[18] It will be notationally convenient to represent a type θ directly by its induced profile of values $v \in \mathbb{R}_+^{\mathcal{A}}$, by setting $v_a = V(a, \theta)$ for all a. We will henceforth refer to v as a type, and denote the set of types by $\mathcal{V}$. The utility of a type $v \in \mathcal{V}$ for a bundle a and expected payment p to the seller is $v_a - p$. The buyer's type is its private information. The seller has a prior distribution $\mu \in \Delta(\mathcal{V})$ over the types of the buyer. Let $\tilde{a} \in \mathcal{A}$ denote the grand bundle and assume that $v_{\tilde{a}} \geq v_a$ for all v and a.[19] Let $a_0 \in \mathcal{A}$ denote the outside option, and normalize $v_{a_0} = 0$ for all v.

We invoke the revelation principle and focus on direct mechanisms. A mechanism is a pair of functions, an assignment rule $\alpha : \mathcal{V} \to \mathcal{A}$ and a payment rule $\rho : \mathcal{V} \to \mathbb{R}$.[20] The mechanism (α, ρ) is incentive compatible (IC) if no type increases its utility by misreporting,

$$v_{\alpha(v)} - \rho(v) \geq v_{\alpha(\hat{v})} - \rho(\hat{v}), \qquad \forall v, \hat{v} \in \mathcal{V}.$$

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

$$v_{\alpha(v)} - \rho(v) \geq 0, \qquad \forall v \in \mathcal{V}.$$

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

$$\mathbf{E}\left[\rho(v)\right]$$

over all IC and IR mechanisms.

A mechanism is a pure bundling mechanism if for some p, it offers only the grand bundle

[16] Even though we interpret $\mathcal{A}$ as a set of bundles, we do not impose any structure on it. Thus $\mathcal{A}$ can be more generally any mutually exclusive set of alternatives that the seller can assign to the buyer, for instance different configurations or quality levels of products.

[17] The set of bundles $\mathcal{A}$ may include randomized bundles. For instance, $a \in \mathcal{A} = [0,1]^n$ may specify the probability a_i of assigning product $i \in \{1, \ldots, n\}$ to the buyer, and $\theta \in \mathbb{R}_+^n$ may specify the buyer's value θ_i for product i. In this case we have $V(a, \theta) = \sum_i a_i \theta_i$. A special case is when $n = 1$, as in Stokey (1979); Riley and Zeckhauser (1983); Myerson (1981).

[18] Note that if $\mathcal{A}$ and Θ are finite, then any function V is continuous.

[19] This is a weaker assumption than assuming free disposal.

[20] By quasilinearity there is no loss in assuming that the payment rule is deterministic. As noted before, the set $\mathcal{A}$ already allows for randomized bundles.

at price p. In such a mechanism, $\alpha(v) = \tilde{a}$ and $\rho(v) = p$ if $v_{\tilde{a}} \geq p$, and $\alpha(v) = a_0$ and $\rho(v) = 0$ otherwise. 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 (among *all* mechanisms).

2 Paths: Value for the Grand Bundle Identifies Type

We start with a special case where the value for the grand bundle uniquely identifies the type, for two reasons. First, we use this result as a building block for our general theorem. Second, the cost-benefit analysis of discrimination is the most transparent for this case. The results of this section concern optimality of pure bundling for *all* distributions over a given support. In contrast, our main result in the next section concerns optimality of pure bundling for a given distribution μ.

Assume that $v_{\tilde{a}}$ uniquely identifies the values for all bundles. That is, there exists a function $C : \mathbb{R}_+ \to \mathbb{R}_+^{\mathcal{A}}$ such that

$$\mathcal{V} \subseteq \mathcal{V}(C) := \{v \in \mathbb{R}_+^{\mathcal{A}} \mid v = C(v_{\tilde{a}})\}, \tag{1}$$

where $C_{\tilde{a}}$ is the identity mapping, $C_{\tilde{a}}(v_{\tilde{a}}) = v_{\tilde{a}}$, and $C_{a_0}(v_{\tilde{a}}) = 0$ to ensure $v_{a_0} = 0$. In this case, we say that types are on a "path" C. To simplify notation we denote $t = v_{\tilde{a}}$. Since values for all bundles are identified given t, the only uncertainty is how t is distributed.

2.1 Optimality of Pure Bundling on Paths

Proposition 1. *Pure bundling is optimal for all distributions with support $\mathcal{V} \subseteq \mathcal{V}(C)$ if and only if $C_a(t)/t$ is monotone non-decreasing in $t > 0$ for all a.*

If $C_a(t)/t$ is monotone non-decreasing in $t > 0$ for all a, we say that the path C is *ratio-monotone*. Geometrically, ratio-monotonicity requires that the slope of a ray from the origin to a type is non-decreasing along the support (Figure 2a). Alternatively, a ray from the origin intersects the support from above and continues below. For the special case where $\mathcal{A} = [0, 1]$ and $v_a = t \times a$, pure bundling is known to be optimal from Stokey (1979) and Riley and Zeckhauser (1983). For this special case, optimality of pure bundling follows also from Proposition 1 since $C_a(t)/t = a$ is constant in t (Figure 2b). Notice that a perturbation of such an instance may violate ratio-monotonicity of C. Proposition 1 holds even though

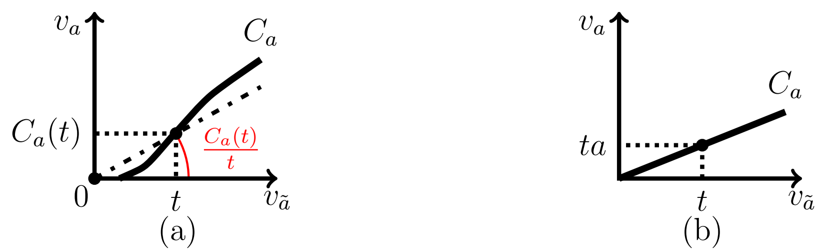

Figure 2: (a) The ratio $C_a(t)/t$ is non-decreasing along the support, and pure bundling is optimal. C_a need not be convex. (b) $C_a(t) = t \times a$, as in Stokey (1979) and Riley and Zeckhauser (1983).

single-crossing may be violated (we defer a formal definition of single-crossing and an example showing its violation to Appendix A.1).[21]

We start by proving the "only if" statement. Since $t = v_{\tilde{a}}$ uniquely identifies the type, we refer to t as a type.

Proof of Proposition 1, "only if" statement. Assume that there exists a bundle a such that $C_a(t)/t$ is not monotone non-decreasing in $t > 0$.[22] Therefore, there exist t, t' such that $0 < t < t'$ and $C_a(t)/t > C_a(t')/t'$. 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 from only offering the grand bundle is t, which is obtained by price t or t'.

Consider a *mixed bundling* mechanism that assigns bundle a to the low type at price $C_a(t)$, and bundle $\tilde{a}$ to the high type at price $t' - C_a(t)(t'/t - 1) + \epsilon$, for $\epsilon > 0$ to be identified shortly. See Figure 3. 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

$$\begin{aligned} & t - \Big(t' - C_a(t)(\tfrac{t'}{t} - 1) + \epsilon\Big) \\ & = (t - t')(1 - \tfrac{C_a(t)}{t}) - \epsilon \leq 0, \end{aligned}$$

where the inequality followed since $t - t' \leq 0$, $C_a(t) \leq t$, and $\epsilon \geq 0$. Therefore the IC and

[21]Single-crossing implies that types can be ordered in such a way that only local incentive constraints bind. Type spaces that satisfy single-crossing are often referred to as "single-dimensional". Even though types on a path can be parameterized by a single parameter t, such type spaces are not necessarily single-dimensional as the example in Appendix A.1 shows. We thus refrain from referring to such type spaces as single-dimensional to avoid confusion.

[22]Note that this is different that saying $C_a(t)/t$ is increasing in $t > 0$.

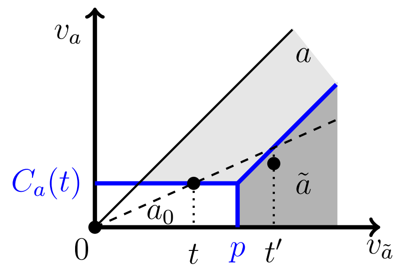

Figure 3: The mixed bundling mechanism of Proposition 1. If $\epsilon = 0$, the price of $\tilde{a}$, $p = t' - C_a(t)(\frac{t'}{t} - 1)$ is set such that the high type chooses $\tilde{a}$ if and only if $C_a(t')/t' \le C_a(t)/t$.

IR constraints for the low type are satisfied. To verify the incentive constraints for the high type, notice that the utility of the high type from truthtelling is

$$t' - \left(t' - C_a(t)(\tfrac{t'}{t} - 1) + \epsilon\right) = C_a(t)(\tfrac{t'}{t} - 1) - \epsilon. \tag{2}$$

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

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

The inequality is strict, and thus for $\epsilon > 0$ small enough, the utility of the high type calculated in equation (2) is at least $\max(0, C_a(t') - C_a(t))$, which is the utility the high type can receive from the outside option or reporting to be the low type. Thus the IC and IR constraints for the high type are satisfied, and the mechanism satisfies all constraints.

The revenue of the mixed bundling mechanism is

$$\begin{aligned}
&(1 - \tfrac{t}{t'})C_a(t) + \tfrac{t}{t'}\left(t' - C_a(t)(\tfrac{t'}{t} - 1) + \epsilon\right) \\
&= t + \tfrac{t}{t'}\epsilon,
\end{aligned}$$

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

Proof 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, assume that the marginal distribution of $v_{\tilde{a}}$ is supported over $[\underline{t}, \bar{t}]$ with strictly positive density. Second, assume that $C_a(t)$ is differentiable for each $a \in \mathcal{A}$. Third, assume that the marginal distribution of $v_{\tilde{a}}$ is regular, as defined next.

For each bundle a let $C'_a(t) := \frac{\mathrm{d}}{\mathrm{d}t} C_a(t)$ be the derivative of C_a with respect to t. Let $F_{\tilde{a}}$ be cumulative marginal distribution of $v_{\tilde{a}}$, and $f_{\tilde{a}}$ its density. For each $a \in \mathcal{A}$, define a virtual value function $\phi_a : [\underline{t}, \bar{t}] \to \mathbb{R}$ as follows.

$$\phi_a(t) = C_a(t) - C'_a(t) \times \frac{1 - F_{\tilde{a}}(t)}{f_{\tilde{a}}(t)}. \tag{3}$$

Recall that $C_{\tilde{a}}(t) = t$, which implies that $\phi_{\tilde{a}}(t) = t - \frac{1-F_{\tilde{a}}(t)}{f_{\tilde{a}}(t)}$.[23] Note that $\phi_{a_0}(t) = 0$. We say that the marginal distribution of $v_{\tilde{a}}$ is regular if $\phi_{\tilde{a}}(t)$ is monotone non-decreasing in t.

The lemma below uses the envelope theorem to relate revenue with virtual surplus. Since $t = v_{\tilde{a}}$ uniquely identifies the type, we represent a mechanism by an assignment rule $\alpha : [\underline{t}, \bar{t}] \to \mathcal{A}$ and a payment rule $\rho : [\underline{t}, \bar{t}] \to \mathbb{R}$. The lemma below shows that the expected revenue of any IC mechanism (α, ρ) is equal to its expected virtual surplus $\mathbf{E}[\phi_{\alpha(t)}(t)]$, up to a constant which is the utility of type $\underline{t}$.

Lemma 1. *For any incentive compatible mechanism (α, ρ),*

$$\mathbf{E}\left[\rho(t)\right] = \mathbf{E}\left[\phi_{\alpha(t)}(t)\right] - \left(C_{\alpha(\underline{t})}(\underline{t}) - \rho(\underline{t})\right).$$

The lemma below follows directly from the definition of virtual values (Equation (3)), and lets us compare the virtual values of bundles a and $\tilde{a}$ given how $C_a(t)/t$ changes in t.

Lemma 2. *The following holds.*

- $\frac{C_a(t)}{t}\phi_{\tilde{a}}(t) \geq \phi_a(t)$ *for all $t > 0$ if $C_a(t)/t$ is monotone non-decreasing in $t > 0$.*
- $\frac{C_a(t)}{t}\phi_{\tilde{a}}(t) < \phi_a(t)$ *if $C_a(t)/t$ is decreasing at t and $t < \bar{t}$.*

To prove Proposition 1, we show that there exists a pure bundling mechanism with no less expected virtual surplus than any mechanism. Furthermore, utility of type $\underline{t}$ in the pure bundling mechanism is equal to zero, which by IR is lower than the utility of type $\underline{t}$ in any other mechanism. Thus Lemma 1 applies to imply that the pure bundling mechanism is optimal.

There exists a pure bundling mechanism that maximizes virtual surplus since any bundle a has lower virtual value than that of either $\tilde{a}$ or a_0. In particular, from Lemma 2, if C is

[23] This is identical to the virtual value of Myerson (1981). In fact, $\phi_a(t) = C_a(t) - \frac{1-F_a(t)}{f_a(t)}$ for any bundle a, where F_a is the cumulative marginal distribution of v_a, and f_a its density. That is, the virtual value of each bundle a is equal to the virtual value of the projected distribution of values of a. This follows from Myerson (1981), since the special case of our setting where the seller can only produce bundle a is equivalent to the setting of Myerson, and his analysis applies. We use equation (3) in our proof since it allows us to compare the virtual values of different bundles based on the curvature of C.

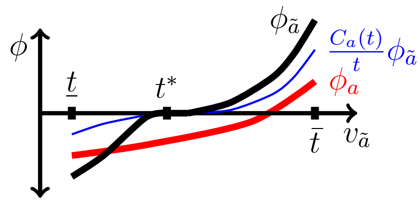

Figure 4: The "if" part of Proposition 1. If C is ratio-monotone, then $\frac{C_a(t)}{t}\phi_{\tilde{a}}(t)$ is an upper bound on $\phi_a(t)$ by Lemma 2. As a result, $\phi_{\tilde{a}} \geq \phi_a$ unless both virtual values are negative.

ratio-monotone, then $\phi_{\tilde{a}}(t) \geq \phi_a(t)$ unless both $\phi_{\tilde{a}}(t)$ and $\phi_a(t)$ are negative. In the later case, the virtual value of a_0, which is zero, is no less than the virtual value of a. As a result, assigning either $\tilde{a}$ or a_0, whichever has a higher virtual value, to each type optimizes the virtual surplus. See Figure 4. Details are below.

Since $\phi_{\tilde{a}}(t)$ is monotone non-decreasing in t, there exists a threshold t^* such that $\phi_{\tilde{a}}(t) \leq 0$ if $t \leq t^*$, and $\phi_{\tilde{a}}(t) \geq 0$ otherwise. Consider a pure bundling mechanism that only offers the grand bundle $\tilde{a}$ at price t^*. The mechanism assigns $\tilde{a}$ to a type $t \geq t^*$, and a_0 to a type $t < t^*$. We argue that the mechanism is optimal.

The assignment rule of the pure bundling mechanism with price t^* optimizes virtual surplus for any t, and therefore in expectation. First consider a type $t \geq t^*$. Such a type is assigned $\tilde{a}$, and the virtual surplus of the mechanism is $\phi_{\tilde{a}}(t) \geq 0$. Since $C_a(t) \leq t$ and $\phi_{\tilde{a}}(t) \geq 0$, the virtual value of $\tilde{a}$ is no less than the virtual value of a,

$$\phi_{\tilde{a}}(t) \geq \frac{C_a(t)}{t}\phi_{\tilde{a}}(t) \geq \phi_a(t), \qquad \forall a \in \mathcal{A},$$

where the last inequality is from Lemma 2. Similarly, consider $t < t^*$. Such a type is assigned a_0 and the virtual surplus of the mechanism is $\phi_{a_0}(t) = 0$. Since $\frac{C_a(t)}{t} \geq 0$ and $\phi_{\tilde{a}}(t) \leq 0$, the virtual value of a_0 is no less than the virtual value of a,

$$\phi_{a_0}(t) = 0 \geq \frac{C_a(t)}{t}\phi_{\tilde{a}}(t) \geq \phi_a(t), \qquad \forall a \in \mathcal{A},$$

where the last inequality is from Lemma 2. We conclude that the mechanism has a higher virtual surplus than any other mechanism for each t and, therefore, in expectation over t. Finally, the utility of $\underline{t}$ is zero, which is as low as possible given the IR constraint. Applying Lemma 1 implies that this pure bundling mechanism has higher revenue than any other mechanism.

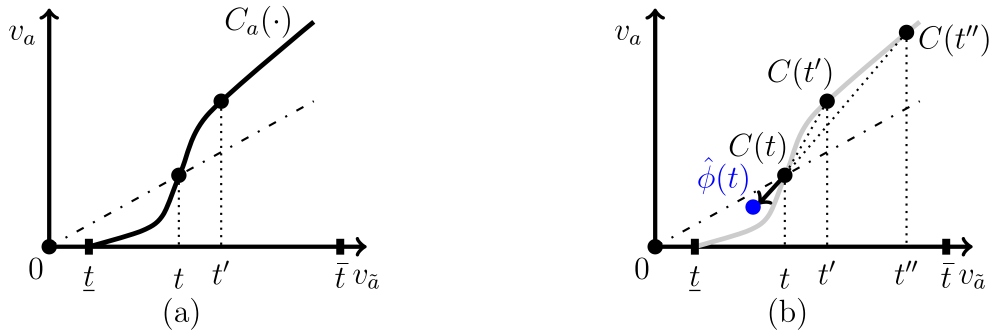

Figure 5: (a) Monotonicity of $C_a(t)/t$ implies that for any $t' > t$, the point $C(t')$ lies above the ray that connects the origin to $C(t)$. As a result, $C_a(t') \geq C_a(t)$. (b) For each t' such that the IC constraint from t' to t binds, $\hat{\phi}$ is pushed below the ray that connects the origin to the point $C(t)$.

Geometric Interpretation of the General Proof. The proof of Proposition 1 discussed above was facilitated by several simplifying assumptions. Let us provide a geometric interpretation of the general proof in Appendix A.4 that does not rely on these assumptions. The proof is based on two observations.

First, types can be ordered such that only "downward" IC constraints bind. In particular, order types by their value for the grand bundle $v_{\tilde{a}}$. Ratio-monotonicity of C implies that if two types v, v' satisfy $v_{\tilde{a}} \geq v'_{\tilde{a}}$, then v is larger in every dimension, $v_a \geq v'_a$ for all a. See Figure 5a. We show that the IC constraint that corresponds to a deviation of type v to higher type v' does not bind.

Second, each type can be assigned a virtual value based on binding IC constraints from higher types. In particular, each type t can be assigned a virtual value defined as

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

where $C(t) = (C_a(t))_a$ and $\lambda(t')$ is the Lagrangian multiplier corresponding to the possible deviation of t' to t. Viewed as vectors, the virtual value of t is defined by pushing $C(t)$ in the direction of the vector $C(t) - C(t')$, for each t' with binding IC constraints to t. Since any type t' with binding IC constraint is "above" $C(t)$ (above the ray from the origin to $C(t)$), the result is a virtual value vector $\hat{\phi}(t)$ that lies "below" $C(t)$, that is, $\frac{C_a(t)}{t}\hat{\phi}_{\tilde{a}}(t) \geq \hat{\phi}_a(t)$. See Figure 5b. Recall from the discussion before that $\frac{C_a(t)}{t}\hat{\phi}_{\tilde{a}}(t) \geq \hat{\phi}_a(t)$ implies that virtual surplus is never maximized by assigning bundle a. Furthermore, $\hat{\phi}_{\tilde{a}}(t)$ is monotone non-decreasing, and therefore there exists a pure bundling mechanism that optimizes virtual surplus.

2.2 Non-optimality of Pure Bundling With No Complementarities

An implication of Proposition 1, which we discuss in this section, is that pure bundling is generally not optimal with additive preferences (i.e., when the value for a bundle is the sum of the values for its constituting products). This is in accordance with the conclusions from the bundling literature (Adams and Yellen, 1976; McAfee et al., 1989). We nevertheless state such a result in our framework to allow for subsequent comparisons. The assumption of additivity removes complementarities from the buyer's preferences. After stating out main theorem, we will revisit complementarities and show that relaxing the assumption of no complementarities allows for pure bundling to be optimal under general conditions.

Consider selling n divisible products. That is, the set of bundles is $\mathcal{A} = [0,1]^n$, and a bundle $a \in \mathcal{A}$ includes a fraction a_i of each product i.[24] A set $\mathcal{V}$ of types is additive if each type v of the buyer is parameterized by a vector $(w_1, \ldots, w_n) \in \mathbb{R}^n_+$, and has value $v_a = \sum_i w_i a_i$ for a bundle a. That is, w_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 $\tilde{a} = (1, \ldots, 1)$.

Assume that pure bundling is optimal for all distributions over an additive set of types $\mathcal{V}$. Assume that for all types v, v', $\sum_i w_i \neq \sum_i w'_i$, that is, the value for the grand bundle is unique. Proposition 1 applies and states that for every bundle a, the ratio of the value for a to the value for the grand bundle $\frac{\sum_i w_i a_i}{\sum_i w_i}$ must be non-decreasing in $\sum_i w_i$. In particular, by letting a be a bundle that includes a unit of product j and no other product, we have that the ratio

$$\frac{w_j}{\sum_i w_i} \tag{4}$$

is non-decreasing in $\sum_i w_i$. However, the sum of the expression (4) over all j is 1. Thus, there must exist constants $\delta_1, \ldots, \delta_n$ such that $\frac{w_j}{\sum_i w_i} = \delta_j$ for all types.[25] This in turn implies that the ratio of the value for any bundle a to the value for the grand bundle is a constant $\sum_i \delta_i a_i$ across all types. In other words, the graph of the value for a bundle a as a function of the value for the grand bundle is a straight ray from the origin, as in Figure 2b. Thus the

[24] The conclusions of this section remain unaffected if the products are not divisible. We assume divisibility to make it easier to compare the conclusions with the subsequent analysis.

[25] Consider two types v, v' where $\sum_i w_i < \sum_i w'_i$. Note that $\sum_j \left(\frac{w_j}{\sum_{i=1}^n w_i}\right) = \sum_j \left(\frac{w'_j}{\sum_{i=1}^n w'_i}\right)$, since both expressions are equal to 1. If $\frac{w_j}{\sum_{i=1}^n w_i} < \frac{w'_j}{\sum_{i=1}^n w'_i}$ for some j, we must have $\frac{w_{\hat{j}}}{\sum_{i=1}^n w_i} > \frac{w'_{\hat{j}}}{\sum_{i=1}^n w'_i}$ for some $\hat{j}$, which violates monotonicity. Thus we must have $\frac{w_j}{\sum_{i=1}^n w_i} = \frac{w'_j}{\sum_{i=1}^n w'_i}$ for all j.

setting reduces to that of Stokey (1979) and Riley and Zeckhauser (1983). To summarize, if pure bundling is optimal for all distributions over an additive set of types, types must be "single-dimensional". Formally we have the following proposition (the "if" statement follows from Proposition 1).

Proposition 2. *Consider selling n divisible products to an additive set of types $\mathcal{V}$ where each type's value for the grand bundle $t = \sum_i w_i$ is unique. Pure bundling is optimal for all distributions over $\mathcal{V}$ if and only if there exist $\delta_1, \ldots, \delta_n$ such that the value of type t for any bundle a is $t \sum_{i \in a} \delta_i$.*

In contrast to Proposition 2, recall from Proposition 1 that it is indeed possible for pure bundling to be optimal for all distributions over a large class of (non-additive) type spaces, that is any $\mathcal{V} \subseteq \mathcal{V}(C)$ for any ratio-monotone path C.

Note that the above discussion required pure bundling to be optimal *for all* distributions over $\mathcal{V}$. It is still possible that pure bundling is optimal for certain distributions over additive but not single-dimensional types.[26] Nonetheless, we take the above analysis as an indication that a general principle for optimality of pure bundling is unlikely to exist with additive values. We will revisit bundling after proving our main theorem, and show that pure bundling is indeed optimal for a large class of preferences once the assumption of no complementarities is relaxed.

2.3 Non-optimality of Pure Bundling on Paths

Proposition 1 states that if $C_a(t)/t$ is not monotone non-decreasing, then pure bundling is not optimal for *some* distribution. If $C_a(t)/t$ is not monotone non-decreasing, it is still possible for pure bundling to be optimal, for example if the non-monotonicity is at a set of types that has very low probability. However, if $C_a(t)/t$ is monotone *decreasing* for some bundle a, then pure bundling is not optimal for essentially all distributions.[27] To formalize such a claim, there are two subtleties to deal with. First, if the distribution puts all the mass on a single type (and vanishingly small mass on other types), then pure bundling is optimal regardless of the support. Indeed, in this case full surplus can be extracted by offering the grand bundle to the type with all the mass. To rule out such a possibility, we require the marginal distribution of $v_{\tilde{a}}$ to have continuous positive density $f_{\tilde{a}}$ over an interval $[\underline{t}, \bar{t}]$ such

[26] Indeed, Pavlov (2011), Menicucci et al. (2015), and Daskalakis et al. (2017) fall in this category. That is, values are additive, pure bundling is optimal, but types are not single-dimensional.

[27] An example with monotone decreasing $C_a(t)/t$ is $C_a(t) = t - t^2/2$ for $t \in [0, 1]$, which implies that $C_a(t)/t = 1 - t/2$.

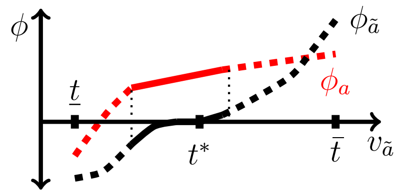

Figure 6: If $C_a(t)/t$ is decreasing at $t = t^*$ and $\phi_{\tilde a}(t^*) = 0$, then $\phi_a > \max(\phi_{\tilde a}, 0)$ around t^*.

that $\underline{t} < \bar{t}$. In this case we say that the distribution is continuous. Second, if the unique optimal pure bundling price is $\underline{t}$, then mixed bundling has no marginal gain. We exclude such a possibility by assuming that some $t > \underline{t}$ maximizes $t \times (1 - F_{\tilde a}(t))$, and say that the distribution is interior. Neither continuity nor interiority can be relaxed.[28]

Proposition 3. *Pure bundling is not optimal for all continuous and interior distributions with support $\mathcal{V} = \mathcal{V}(C)$ if $C_a(t)/t$ is monotone decreasing in t for some a.*

To see the idea behind Proposition 3, let $t^* > \underline{t}$ be an optimal pure bundling price, $t^* \in \arg\max_t t \times (1 - F_{\tilde a}(t))$. The first order condition of optimality is

$$(1 - F_{\tilde a}(t^*)) - t f_{\tilde a}(t^*) = 0.$$

Note that the first order condition implies that the virtual value (from Equation 3) of $\tilde a$ is zero, $\phi_{\tilde a}(t^*) = 0$. If $C_a(t)/t$ is monotone decreasing in t, Lemma 2 implies that the virtual value of a is strictly higher than the virtual value of a^* and zero,

$$\phi_{\tilde a}(t^*) = 0 = \frac{C_a(t^*)}{t^*}\phi_{\tilde a}(t^*) < \phi_a(t^*).$$

Continuity implies that there exists an interval around t^* wherein a has higher virtual value than a_0 and $\tilde a$. As a result, a mixed bundling mechanism that assigns a to the neighborhood of t^* obtains higher virtual surplus, and by Lemma 1 higher revenue, than pure bundling. See Figure 6. The full proof in the appendix shows that it is possible to assign a to a neighborhood of t^* in an incentive compatible manner.

[28]We already argued that continuity cannot be relaxed. To see the necessity of interiority, assume that there are only two non-trivial bundles, $\mathcal{A} = \{\tilde a, a, a_0\}$. Let $v_{\tilde a}$ be uniformly distributed on $[2, 3]$ and let $C_a(t) = \epsilon$ for some $\epsilon, 0 < \epsilon < 1$. Note that $C_a(t)/t = \epsilon/t$ strictly decreases in t. The virtual values are $\phi_{\tilde a} = 2t - 3 \geq 1$ and $\phi_a = \epsilon$. Thus, selling $\tilde a$ at price 2 is optimal.

3 Main Result: General Type Spaces

We now relax the assumption of Section 2 that types are on a path. Our main theorem relaxes the monotonicity conditions of Proposition 1 and Proposition 3 and replaces them with stochastic monotonicity. We start by defining the stochastic monotonicity conditions. The first one is the standard notion of (first order) stochastic dominance.[29] In what follows $\hat{x}$ and $\hat{x}'$ denote realizations of a random variable x (similarly for y etc).

Definition 1. *Consider jointly distributed random variables* $(x, y) \in \mathbb{R} \times \mathbb{R}^{\mathcal{S}}$ *for some set* $\mathcal{S}$*. The distribution of* y *is* stochastically non-decreasing in x *if*

$$\mathbf{Pr}\left[y \in \mathcal{Y} \mid x = \hat{x}\right] \leq \mathbf{Pr}\left[y \in \mathcal{Y} \mid x = \hat{x}'\right]$$

for all $\hat{x} < \hat{x}' \in \mathbb{R}$ *and all upper sets* $\mathcal{Y} \subseteq \mathbb{R}^{\mathcal{S}}$*.*[30]

For an upper set $\mathcal{Y}$, $y \in \mathcal{Y}$ can be interpreted as an event where y takes a high value, compared to $y \notin \mathcal{Y}$. Definition 1 states that conditioned on x taking higher values, it becomes more likely for y to take high values as well.

We now define a notion of strict stochastic dominance for scalar variables. For $(x, y) \in \mathbb{R} \times \mathbb{R}$, y is stochastically decreasing at the top in x if the upper bound of the support of y decreases in x. Formally, we have the following definition.

Definition 2. *Consider jointly distributed random variables* $(x, y) \in \mathbb{R} \times \mathbb{R}$*. The distribution of* y *is* stochastically decreasing at the top in x *if* $\bar{y}(\hat{x})$ *is monotone decreasing in* $\hat{x}$*, where* $\bar{y}(\hat{x})$ *is the upper bound of the support of* y *conditioned on* $x = \hat{x}$*.*[31]

Let us compare Definition 1 with Definition 2. Definition 1 implies that $\bar{y}_s$ is monotone non-decreasing for all $s \in \mathcal{S}$. To see this, note that Definition 1 implies that for all $s \in \mathcal{S}$, $\hat{y}_s$, and $\hat{x} < \hat{x}' \in \mathbb{R}$,

$$\mathbf{Pr}\left[y_s \geq \hat{y}_s \mid x = \hat{x}\right] \leq \mathbf{Pr}\left[y_s \geq \hat{y}_s \mid x = \hat{x}'\right], \tag{5}$$

which is the standard notion of (first order) stochastic dominance for scalar variables $(x, y_s) \in \mathbb{R} \times \mathbb{R}$.[32] Inequality (5) in turn implies that $\bar{y}_s$ is monotone non-decreasing. Assume for

[29] See for example Shaked and Shanthikumar (2007).

[30] A set $\mathcal{Y} \subseteq \mathbb{R}^{\mathcal{S}}$ is an upper set if $\hat{y} \in \mathcal{Y}$ implies that all values higher than $\hat{y}$ are in $\mathcal{Y}$ as well. That is, for all $\hat{y}, \hat{y}'$ such that $\hat{y} \in \mathcal{Y}$ and $\hat{y}_s \leq \hat{y}'_s$ for all $s \in \mathcal{S}$, we have $\hat{y}' \in \mathcal{Y}$.

[31] That is, $\bar{y}(\hat{x}) = \sup\{\hat{y} \mid \hat{y}$ is in the support of y conditioned on $x = \hat{x}\}$.

[32] Define $\mathcal{Y} = \{\hat{y}' \in \mathbb{R}^{\mathcal{S}} \mid \hat{y}'_s \geq \hat{y}_s\}$. Note that $\mathcal{Y}$ is an upper set, and that $y \in \mathcal{Y}$ is equivalent to $y_s \geq \hat{y}_s$.

contradiction that $\bar{y}_s(\hat{x})$ is not non-decreasing. Then there exists $\hat{x} < \hat{x}'$ and $\hat{y}_s$ such that $\bar{y}_s(\hat{x}) > \hat{y}_s > \bar{y}_s(\hat{x}')$, and thus $\mathbf{Pr}[y_s \geq \hat{y}_s \mid x = \hat{x}] > 0 = \mathbf{Pr}[y_s \geq \hat{y}_s \mid x = \hat{x}']$, violating (5). We discuss Definition 1 and Definition 2 in more detail after stating the main theorem.

The main theorem considers the stochastic monotonicity of value-ratios in the value for the grand bundle $v_{\tilde{a}}$. For each type v where $v_{\tilde{a}} > 0$, define the value-ratio profile $r \in [0,1]^{\mathcal{A}}$ by setting $r_a = v_a / v_{\tilde{a}}$ for each bundle a. Note that a type v is uniquely identified by the value for the grand bundle $v_{\tilde{a}}$ and r since $v = v_{\tilde{a}} r$. We call a distribution of types continuous if it has a continuous density function over $\mathcal{V}$, and interior if some $t > \underline{t}$ maximizes $t \times (1 - F_{\tilde{a}}(t))$, where $F_{\tilde{a}}$ is the marginal distribution of $v_{\tilde{a}}$, and $\underline{t}$ the smallest value in its support.

Theorem 1. *Pure bundling is optimal if the distribution of value-ratio profile r is stochastically non-decreasing in $v_{\tilde{a}} > 0$. Pure bundling is not optimal if the distribution of types is continuous and interior, and r_a is stochastically decreasing at the top in $v_{\tilde{a}}$ for some a.*

To prove Theorem 1, a first observation is that if a mechanism is optimal for each distribution in a set, then by linearity of expectation the mechanism is also optimal for any mixture of those distributions. Formally, for a mechanism (α, ρ), let $\mathcal{D}(\alpha, \rho)$ denote the set of distributions $\hat{\mu} \in \Delta(\mathcal{V})$ for which the mechanism (α, ρ) is optimal. The following holds.

Lemma 3. *For any mechanism (α, ρ), the set of distributions $\mathcal{D}(\alpha, \rho)$ is convex.*

By Lemma 3, in order to prove that a mechanism (α, ρ) is optimal for a distribution μ, it is sufficient to decompose μ into a distribution over a set of distributions, and prove optimality of (α, ρ) for each distribution in the set. If each decomposed distribution is supported on a ratio-monotone path, the pure bundling is optimal by Proposition 1. What distributions are decomposable into a set of distributions, each supported on a ratio-monotone path? The lemma below allows us to provide a characterization. It relates the stochastic monotonicity condition of Definition 1 with the existence of a monotone decomposition.

To build up to the lemma, consider jointly distributed random variables $(x, y) \in \mathbb{R} \times \mathbb{R}$. Let $F(\hat{y} \mid \hat{x}) = \mathbf{Pr}[y \leq \hat{y} \mid x = \hat{x}]$ be the cumulative density function of y conditioned on $x = \hat{x}$. Define a random variable q by setting $q = F(y \mid x)$. That is, q is the quantile of y conditioned on x. Let F^{-1} be the inverse of F, that is $F^{-1}(F(\hat{y} \mid \hat{x}) \mid \hat{x}) = \hat{y}$. Thus $F^{-1}(\hat{q} \mid \cdot)$ is a "contour curve" that maps any $\hat{x}$ to a $\hat{y}$ with quantile equal to $\hat{q}$. Note two properties of q and F^{-1} for future reference. First, q is independent from x. Indeed, the quantile q is distributed uniformly on $[0, 1]$, independently from x. And second, conditioned on $q = \hat{q}$, $y = F^{-1}(\hat{q} \mid x)$ with probability one. That is, conditioned on $q = \hat{q}$, x uniquely identifies y through F^{-1}.

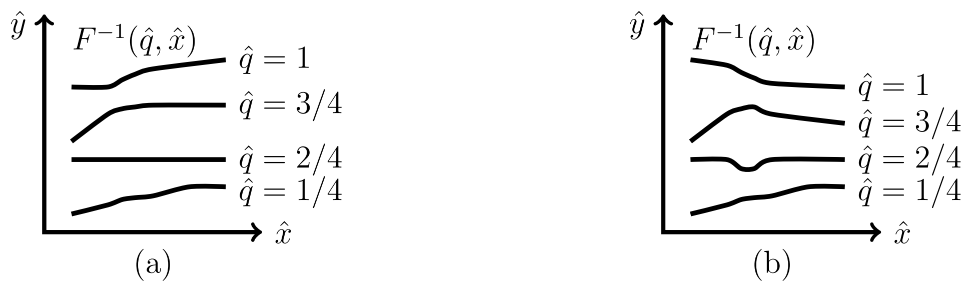

Figure 7: (a) y is stochastically increasing in x, and the contour curves of the distribution of y conditioned on x are non decreasing. (b) y is stochastically decreasing at the top in x, and the top contour curve is decreasing.

Monotonicity of the contour curves F^{-1} is related to the stochastic monotonicity conditions of Definition 1 and Definition 2. In particular, if y is stochastically non-decreasing in x (Definition 1), then $F^{-1}(\hat{q} \mid \hat{x})$ is monotone non-decreasing in $\hat{x}$ for all $\hat{q}$. See Figure 7a. To see this, note that since y is stochastically non-decreasing in x, then by inequality (5), $\hat{q} = F(\hat{y} \mid \hat{x}) \geq F(\hat{y} \mid \hat{x}')$ for all $\hat{x} < \hat{x}'$. Since F is non-decreasing in $\hat{y}$ and $\hat{q} \geq F(\hat{y} \mid \hat{x}')$, it must be that $F^{-1}(\hat{q} \mid \hat{x}') \geq \hat{y}$. Thus $\hat{y} = F^{-1}(\hat{q} \mid \hat{x}) \leq F^{-1}(\hat{q} \mid \hat{x}')$, that is, F^{-1} in monotone non-decreasing in $\hat{x}$. On the other hand, if y is stochastically decreasing at the top in x (Definition 2), then $F^{-1}(1 \mid \hat{x})$ is monotone decreasing in $\hat{x}$ (Figure 7b). This follows directly since $F^{-1}(1 \mid \hat{x})$ is the upper bound of the support of y conditioned on $x = \hat{x}$.[33]

The following lemma establishes a general connection between Definition 1 and the existence of a monotone decomposition.[34] It states that if y (not necessarily scalar) is stochastically non-decreasing in x, then there exists a random variable q such that q and x are independent and y is monotone non-decreasing in x conditioned on $q = \hat{q}$. In particular, conditioned on $q = \hat{q}$, x uniquely identifies y through a function that is monotone non-decreasing in x. As discussed above, when y is scalar, the quantile q is independent of x, and conditioned on $q = \hat{q}$, x uniquely identifies y through the function F^{-1} which is monotone non-decreasing in x if y is stochastically non-decreasing in x.[35]

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

[33] Note that this is weaker than requiring $F^{-1}(\hat{q} \mid \hat{x})$ to be monotone decreasing in $\hat{x}$ for all $\hat{q}$.

[34] See Shaked and Shanthikumar (2007) for a textbook exposition.

[35] The difference with the special case discussed above is that in general it need not be that $q \in [0, 1]$.

(I) q and x are independent.

(II) Conditioned on $q = \hat{q}$, $y = h(\hat{q}, x)$ with probability one.

(III) $h(\hat{q}, \hat{x})$ is monotone non-decreasing in $\hat{x}$ for all $\hat{q}$.[36]

We now use Lemma 4 to prove the first statement of Theorem 1. We defer the proof of the second statement to the appendix.

Proof of Theorem 1, the first statement. We prove the statement assuming $\mathcal{A}$ is finite and defer the general proof. Consider any maximizer t^* of $t \times (1 - F_{\tilde{a}}(t))$. We show that pure bundling with price t^* is optimal.

Assume that $r \in \mathbb{R}^{\mathcal{A}}$ is stochastically non-decreasing in $v_{\tilde{a}}$. By Lemma 4, there exists a random variable $q \in \mathcal{Q}$ and a function $h : \mathcal{Q} \times \mathbb{R} \to \mathbb{R}^{\mathcal{A}}$ such that

(I) q and $v_{\tilde{a}}$ are independent.

(II) Conditioned on $q = \hat{q}$, $r = h(\hat{q}, v_{\tilde{a}})$ with probability one.

(III) $h(\hat{q}, v_{\tilde{a}})$ is monotone non-decreasing in $v_{\tilde{a}}$ for all $\hat{q}$.

Let $\mu|\hat{q}$ be the distribution of types conditioned on $q = \hat{q}$. Note that μ is a distribution over the set of distributions $\{\mu|\hat{q}\}_{\hat{q}}$. Therefore, by Lemma 3, it is sufficient to prove optimality of pure bundling with price t^* for any $\mu|\hat{q}$.

Recall that $v = v_{\tilde{a}} r$. Define $C^{\hat{q}}(t) = t h(\hat{q}, t)$. By property (II), conditioned on $q = \hat{q}$, $v = v_{\tilde{a}} h(\hat{q}, v_{\tilde{a}}) = C^{\hat{q}}(v_{\tilde{a}})$ with probability one. That is, types in the support of $\mu|\hat{q}$ are on a path $C^{\hat{q}}$ as in Section 2. Furthermore, $C^{\hat{q}}$ is ratio-monotone by property (III), since $C^{\hat{q}}_a(t)/t = h_a(\hat{q}, t)$ is monotone non-decreasing in t. Therefore Proposition 1 applies to $C^{\hat{q}}$ and implies that pure bundling is optimal for $\mu|\hat{q}$. By property (I), the marginal distribution of $\mu|\hat{q}$ over $\tilde{a}$ is $F_{\tilde{a}}$. Therefore, to maximize profit for $\mu|\hat{q}$ it is optimal to set the price of the grand bundle $\tilde{a}$ equal to t^*. $\square$

3.1 A Closure Property

Preferences often have structures that simplify verifying the assumptions of Theorem 1. We first provide a standard closure property (see, for example, Shaked and Shanthikumar, 2007) and then provide its application to bundling.

[36] That is, $h_s(\hat{q}, \hat{x}) \leq h_s(\hat{q}, \hat{x}')$ for all $s, \hat{q}$ and $\hat{x} \leq \hat{x}'$.

Lemma 5. *Consider jointly distributed random variables* $(x, y) \in \mathbb{R} \times \mathbb{R}^{\mathcal{S}}$ *and a non-decreasing function* $g : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{S}'}$ *for some sets* $\mathcal{S}, \mathcal{S}'$.[37] *The distribution of* $g(y) \in \mathbb{R}^{\mathcal{S}'}$ *is stochastically non-decreasing in* x *if the distribution of* y *is stochastically non-decreasing in* x.

3.2 Optimality of Pure Bundling with Complementarities

We now present a reformulation of Theorem 1, and argue that pure bundling is optimal under general conditions, contrasting the results of Section 2.2. The reformulation allows us to relate optimality conditions of pure bundling with complementarities between products.

Consider again selling n divisible products. That is, the set of bundles is $\mathcal{A} = [0, 1]^n$, and a bundle $a \in \mathcal{A}$ includes a fraction a_i of each product i. Assume that a type v is parameterized by $(x, y) \in \mathbb{R} \times \mathbb{R}^{\mathcal{S}}$ for some set $\mathcal{S}$, and has value v_a for a bundle a,

$$v_a = x \cdot g_a(y), \tag{6}$$

for a function $g : \mathbb{R}^{\mathcal{S}} \to [0, 1]^{\mathcal{A}}$ that is continuous, non-decreasing, and $g_{(1,\ldots,1)}(y) = 1$ for all y.[38] A special case is where $\mathcal{S} = \{1, \ldots, n\}$ (thus $y \in \mathbb{R}^n$) and $g_a(y) = \prod_{i=1}^n h_i(a_i, y_i)$, where for each i, $h_i(a_i, y_i) \le 1$ is continuous, non-decreasing in y_i, and $h_i(1, y_i) = 1$ for all y_i. In this case we have,

$$v_a = x \cdot \prod_{i=1}^{n} h_i(a_i, y_i). \tag{7}$$

Proposition 4. *Assume that values are specified by equation* (6). *Pure bundling is optimal if the distribution of* y *is stochastically non-decreasing in* $x > 0$. *Conversely, assume that values are specified by equation* (7). *Pure bundling is not optimal if the distribution of types is continuous and interior, and* y_i *is stochastically decreasing at the top in* x *for some* i.[39]

As an example, with Cobb-Douglas preferences we have $v_a = x \cdot \prod_{i=1}^n a_i^{1/y_i}$. Proposition 4 states that with Cobb-Douglas preferences, pure bundling is optimal if y is stochastically non-decreasing in $x > 0$, and is not optimal if y_i is stochastically decreasing at the top in x for some i.

[37] That is, if $\hat{y} \le \hat{y}' \in \mathbb{R}^{\mathcal{S}}$, then $g(\hat{y}) \le g(\hat{y}') \in \mathbb{R}^{\mathcal{S}'}$. Comparisons are coordinate-wise, i.e., $\hat{y} \le \hat{y}'$ if $\hat{y}_s \le \hat{y}'_s$ for all $s \in \mathcal{S}$, and similarly for g.

[38] This representation is without loss of generality, by setting $x = v_{\tilde{a}}$, and $g_a(y) = v_a / v_{\tilde{a}}$.

[39] Proposition 4 holds even if the seller can randomize over bundles. We defer a formal treatment to the appendix.

An interpretation of Proposition 4 is that pure bundling is optimal if consumers with higher value for the grand bundle consider the products to be less complementary. To see this, assume that values for two products ($n = 2$) can be written as follows

$$v_a = x \cdot \Big(y_1 a_1 + y_2 a_2 + (1 - y_1 - y_2) a_1 a_2\Big),$$

where $x \geq 0$ and $y_1, y_2 \in [0,1]$.[40] Types with higher $y_1 + y_2$ consider the products to be less complementary. In particular, if $y_1 + y_2 < 1$, then the products are complements, $v_{(a_1,0)} + v_{(0,a_2)} < v_{(a_1,a_2)}$; if $y_1 + y_2 = 1$, then there is no complementarity, $v_{(a_1,0)} + v_{(0,a_2)} = v_{(a_1,a_2)}$; and if $y_1 + y_2 > 1$, then the products are substitutes, $v_{(a_1,0)} + v_{(0,a_2)} > v_{(a_1,a_2)}$. Proposition 4 applies and implies that pure bundling is optimal if (y_1, y_2) is stochastically non-decreasing in x.[41] That is, consumers with higher value for the grand bundle x consider the products to be less complementary.

Recall our discussion from Section 2.2 that pure bundling is rarely optimal with additive preferences (that is, if $y_1 + y_2 = 1$ for all types). Proposition 4 contrasts the results of Section 2.2 by showing that pure bundling is optimal under general conditions once the assumption of no complementaries is relaxed. Notice from the discussion above that it is indeed possible for pure bundling to be optimal even if all consumers consider the products substitutes ($y_1 + y_2 > 1$ for all types). In other words, for pure bundling to be optimal, it is not necessary for the degree of complementarity to be *positive*, but that it is *non-increasing*.

4 Welfare Comparisons

We now study welfare implications of our results. First, consider a setting in which pure bundling is optimal. Does the optimality of pure bundling imply anything about the price that the seller would set for the grand bundle? From Theorem 1, it is immediately clear that the answer is negative. That is, the condition that r is stochastically non-decreasing in $v_{\tilde{a}}$ is orthogonal to how $v_{\tilde{a}}$ is distributed, and therefore is orthogonal to the optimal monopoly price for $\tilde{a}$. In particular, we have the following proposition. It shows that it is possible for pure bundling to be optimal and efficient, and also it is also possible for pure bundling to be optimal and give zero utility to the buyer.

[40] This representation is without loss of generality if a_1 and a_2 are interpreted as probabilities of allocations and the buyer is risk neutral. The assumption that $y_1, y_2 \leq 1$ is required to ensure that the grand bundle maximizes value. Otherwise, if for instance $y_1 = 2$ and $y_2 = 0$, then $v_{(1,0)} = 2 > 1 = v_{(1,1)}$.

[41] Define $g_a(y) = y_1 a_1 + y_2 a_2 + (1 - y_1 - y_2) a_1 a_2$ and note that it is continuous, non-decreasing, and $g_{(1,1)}(y) = 1$.

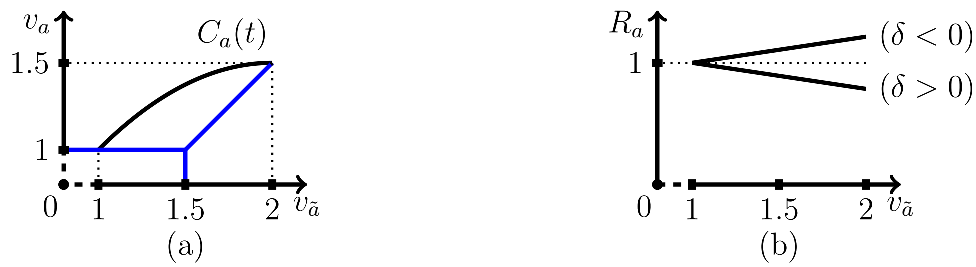

Figure 8: (a) The function C from Example 1. The optimal mechanism offers bundle a at price 1, and the grand bundle $\tilde{a}$ at price 2. (b) The revenue curve of the distribution in Example 1.

Proposition 5. *For any t in the support of $v_{\tilde{a}}$, there exists a distribution where pure bundling at price t is optimal.*

Now consider a setting in which pure bundling is not optimal. Does the welfare increase if the seller is restricted to only offer the grand bundle? If the price of the grand bundle is fixed, it is clear that mixed bundling improves the buyer's utility since it offers more choices. Thus, holding the price of the grand bundle fixed, if mixed bundling is optimal then it is indeed Pareto improving compared to pure bundling. But once the seller offers other bundles, it adjusts the price of the grand bundle. As the following example shows, the effect on price and surplus are ambiguous.

Example 1. *Consider a setting with two non-trivial bundles $\mathcal{A} = \{a, \tilde{a}, a_0\}$. The set of types is $\mathcal{V}(C)$ (as in Section 2), where*

$$C_a(t) = t - \frac{(t-1)^2}{2},$$

for $t \in [1, 2]$. See Figure 8. Let

$$F_{\tilde{a}}(t) = 1 - \frac{1}{t} + (t-1)\delta$$

be the marginal distribution of $v_{\tilde{a}}$. The profit of only selling $\tilde{a}$ at price t is $t \times (1 - F_{\tilde{a}}(t)) = 1 + (x-1)\delta$. When $\delta > 0$, the unique optimal price for $\tilde{a}$ is 2, and when $\delta < 0$, the unique optimal price for $\tilde{a}$ is 1. The consumer surplus is 0.69 when the price of $\tilde{a}$ is 1 and 0 when the price is 2.[42]

[42]It is also possible to add a small parametric function to $F_{\tilde{a}}(t) = 1 - \frac{1}{t}$ such that the price of $\tilde{a}$ varies smoothly between 1 and 2.

Now consider the optimal mechanism. For small enough δ, the optimal mechanism offers a for price 1, and the grand bundle $\tilde{a}$ for price 1.5. See Figure 8. The consumer surplus is 0.38. Note that depending on δ, the price of $\tilde{a}$ in the optimal mechanism can be higher or lower than the price of $\tilde{a}$ in the optimal pure bundling mechanism. Similarly, the consumer surplus may be higher or lower in the pure bundling mechanism compared to the optimal mechanism.

5 Extension: Costs

In this section we discuss how to incorporate a cost c_a for each bundle a. Our result identifies conditions for optimality of offering certain bundles at a uniform markup above costs. In certain cases, it is sufficient to offer only the grand bundle. In particular, a mechanism *uniformly prices* $\bar{\mathcal{A}} \subseteq \mathcal{A}$ if it can be represented as a menu that offers each bundle $a \in \bar{\mathcal{A}} \backslash \{a_0\}$ at price $p + c_a$, for some $p \geq 0$.[43] We identify conditions on the distribution μ, generalizing Theorem 1, under which uniformly pricing $\bar{\mathcal{A}}$ is optimal for some $\bar{\mathcal{A}}$.

The modification compared to Theorem 1 is that for each bundle a, the distribution of value-ratios needs to be considered conditioned on a being an efficient bundle. For a type v and bundle a, let $\bar{v}_a = \max(v_a - c_a, 0)$. For a type v, we say that a is an efficient bundle if $\bar{v}_a \geq \bar{v}_{a'}, \forall a'$. For each a and v, let $\bar{r}(a) \in \mathbb{R}^{\mathcal{A}}$ be a profile of value-ratios where $\bar{r}_{a'}(a) = \bar{v}_{a'}/\bar{v}_a$. We say that the distribution of $\bar{r}(a)$ is stochastically non-decreasing in $\bar{v}_a$ if

$$\mathbf{Pr}\left[\bar{r}(a) \in \mathcal{R} \mid a \text{ is efficient}; \bar{v}_a = t\right] \leq \mathbf{Pr}\left[\bar{r}(a) \in \mathcal{R} \mid a \text{ is efficient}; \bar{v}_a = t'\right]$$

for all $t < t' \in \mathbb{R}$ and upper sets $\mathcal{R} \subseteq \mathbb{R}^{\mathcal{A}}$. Let $\bar{F}_a$ be the marginal distribution of $\bar{v}_a$ conditioned on a being an efficient bundle, and let $\bar{P}_a$ be the probability of the event that a is an efficient bundle.

Theorem 2. *Uniformly pricing the set of bundles $\bar{\mathcal{A}} = \{a | \bar{P}_a > 0\}$ is optimal if the following conditions hold.*

(I) For all $a \in \bar{\mathcal{A}}$, the distribution of $\bar{r}(a)$ is stochastically non-decreasing in $\bar{v}_a > 0$.

(II) There exists t^* such that $t^* \in \arg\max_t t \cdot (1 - \bar{F}_a(t))$ for all a where $\bar{P}_a > 0$.

[43] If the cost are additive (cost of a bundle is the sum of the cost of its constituting products) and $\mathcal{A} = \bar{\mathcal{A}}$, such a mechanism charges a subscription fee p for the option to buy each product at cost.

The second condition is a symmetry condition that holds if $\bar{F}_a = \bar{F}_{a'}$ for all a and a', which in turn holds if the distribution of $\bar{v}$ is symmetric. If there exists a bundle a such that $\bar{v}_a \geq \bar{v}_{a'}$ for all v and a', then $\bar{\mathcal{A}} = \{a\}$, the second condition is trivially satisfied, and thus offering only a (pure bundling) is optimal. However, the assumption that $v_{\tilde{a}} \geq v_{a'}$ for all v and a' does not necessarily implies that $\bar{v}_{\tilde{a}} \geq \bar{v}_{a'}$ for all $\bar{v}$ and a.[44]

6 Conclusions

Dating back to Spence (1977) and Mussa and Rosen (1978), the extensive literature on price discrimination studies, among other questions, *how* to optimally price discriminate (see Stole, 2007, for a survey). This paper studies *when* it is optimal to price discriminate. Offering smaller bundles in addition to the grand bundle has three effects on the profit of a multi-product monopolist: (a) the seller may be able to reduce costs if smaller bundles are cheaper to produce; (b) the seller may be able to attract consumers who are unwilling to pay a high price for the grand bundle; and (c) the seller may lose profit due to spillover. We abstract away from the first effect in order to focus on the second and third effects that involve incentives of the consumers.

Our approach is to decompose the type spaces into paths, and to prove optimality of pure bundling for each path. Even though types on a path can be parameterized by a single value (the value for the grand bundle), such type spaces are not single-dimensional in the sense that they do not satisfy single-crossing, even if the path is ratio-monotone. As such, the binding incentive constraints are endogenous. Nevertheless, we show that given ratio-monotonicity, only downward incentive constraints bind (possibly non-local ones) and use those constraints to construct appropriate virtual valuations. Our decomposition approach extends the orthogonal decomposition approaches that are common in dynamic mechanism design to static but multi-dimensional screening.

Our main motivation for studying pure bundling is to rationalize the widespread use of such mechanisms. The existing bundling literature argues that pure bundling mechanisms are generally sub-optimal and thus does not explain the prevalence of such mechanisms. In practice, even if sellers do not use pure bundling, they tend to offer only a few bundles. Thus, ideally, a theory of bundling would explain when this is the case and which small set of bundles would be offered by the sellers. Even though our results do not address how to

[44] For example, consider selling two products to two types, whose values for product one, product two, and the bundle are $(3, 1, 3)$ and $(3, 3, 6)$, respectively. If costs are $(2, 2, 4)$, then the efficient bundle for the first type is product one, and for the second type is the grand bundle.

offer only a few bundles, we hope that our work sheds light on possible answers.

Our results provide two contrasting insights regarding optimality of pure bundling. On one hand, pure bundling is generally not optimal if the degree of complementarity is constant among all consumers (e.g., additive preferences, unless the type space is single-dimensional). On the other hand, pure bundling is optimal if the degree of complementarity decreases with the value for the grand bundle, even if the products are substitutes for all consumers.

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). Optimal regulation with unknown demand and cost functions. *Journal of Economic Theory*, 84(2):196–215.

Bakos, Y. and Brynjolfsson, E. (1999). Bundling information goods: Pricing, profits, and efficiency. *Management science*, 45(12):1613–1630.

Barelli, P., Basov, S., Bugarin, M., and King, I. (2014). On the optimality of exclusion in multi-dimensional screening. *Journal of Mathematical Economics*, 54:74–83.

Börgers, T. (2015). *An introduction to the theory of mechanism design*. Oxford University Press, USA.

Byzalov, D. (2008). Unbundling cable television: An empirical investigation. Technical report.

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.

Chu, C. S., Leslie, P., and Sorensen, A. (2011). Bundle-size pricing as an approximation to mixed bundling. *The American Economic Review*, 101(1):263–303.

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

Crawford, G. S. and Yurukoglu, A. (2012). The welfare effects of bundling in multichannel television markets. *American Economic Review*, 102(2):643–85.

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

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

Eső, P. and Szentes, B. (2007). Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies*, 74(3):705–731.

Fang, H. and Norman, P. (2006). To bundle or not to bundle. *The RAND Journal of Economics*, 37(4):946–963.

Hart, S. and Nisan, N. (2013). The menu-size complexity of auctions. *arXiv preprint arXiv:1304.6116*.

Ho, K., Ho, J., and Mortimer, J. H. (2012). The use of full-line forcing contracts in the video rental industry. *American Economic Review*, 102(2):686–719.

Jehiel, P., Meyer-Ter-Vehn, M., and Moldovanu, B. (2007). Mixed bundling auctions. *Journal of Economic Theory*, 134(1):494–512.

Jehiel, P., Moldovanu, B., and Stacchetti, E. (1999). Multidimensional mechanism design for auctions with externalities. *Journal of economic theory*, 85(2):258–293.

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.

Madarász, K. and Prat, A. (2017). Sellers with misspecified models. *The Review of Economic Studies*, 84(2):790–815.

Manelli, A. M. and Vincent, D. R. (2006). Bundling as an optimal selling mechanism for a multiple-good monopolist. *Journal of Economic Theory*, 127(1):1–35.

McAfee, R. P. and McMillan, J. (1988). Multidimensional incentive compatibility and mechanism design. *Journal of Economic Theory*, 46(2):335–354.

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.

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic theory*, 18(2):301–317.

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). A property of solutions to linear monopoly problems. *The BE Journal of Theoretical Economics*, 11(1).

Riley, J. and Zeckhauser, R. (1983). Optimal selling strategies: When to haggle, when to hold firm. *The Quarterly Journal of Economics*, 98(2):267–289.

Rochet, J.-C. and Chone, P. (1998). Ironing, sweeping, and multidimensional screening. *Econometrica*, 66(4):783–826.

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.

Severinov, S. and Deneckere, R. (2006). Screening when some agents are nonstrategic: does a monopoly need to exclude? *The RAND Journal of Economics*, 37(4):816–840.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer Science & Business Media.

Shiller, B. and Waldfogel, J. (2011). Music for a song: an empirical look at uniform pricing and its alternatives. *The Journal of Industrial Economics*, 59(4):630–660.

Spence, M. (1977). Nonlinear prices and welfare. *Journal of public economics*, 8(1):1–18.

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.

Stole, L. A. (2007). Price discrimination and competition. *Handbook of industrial organization*, 3:2221–2299.

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.

Vincent, D. and Manelli, A. (2007). Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *Journal of Economic Theory*, 137(1):153–185.

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

A Appendix

A.1 Violation of Single-crossing

We first provide a definition of single-crossing (see, for example, Börgers, 2015).

Definition 3. *Given a complete and transitive order R over bundles, define its induced order $\succ_R$ over types $\mathcal{V}$ as follows. For two types $v, v' \in \mathcal{V}$, $v \succ_R v'$ if*

(I) $v_a - v_{a'} > v'_a - v'_{a'}$ for all a, a' such that aRa' but not $a'Ra$.

(II) $v_a - v_{a'} = v'_a - v'_{a'} = 0$ for all a, a' such that aRa' and $a'Ra$.

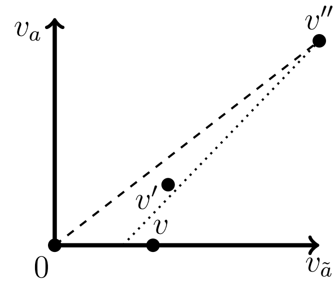

Figure 9: An instance that violates single-crossing where $C_a(t)/t$ is monotone non-decreasing. The dotted line which has slope 1 and goes through v'' demonstrates that $v'_a - v'_{\tilde{a}} > v''_a - v''_{\tilde{a}} > v_a - v_{\tilde{a}}$.

Definition 4 (Börgers, 2015)**.** *A set of types $\mathcal{V}$ satisfies single-crossing with respect to a complete and transitive order R over bundles if for every pair of types v, v', either $v \succ_R v'$ or $v' \succ_R v$.*

We now construct an example with two non-trivial bundles, $\mathcal{A} = \{\tilde{a}, a, a_0\}$, and three types $\mathcal{V} = \{v, v', v''\}$ such that each type's value for $\tilde{a}$ is unique and single-crossing is violated for any order R. In particular, assume $v'' > v' > v$, and $v'_a - v'_{\tilde{a}} > v''_a - v''_{\tilde{a}} > v_a - v_{\tilde{a}}$ (Figure 9).

Assume for contradiction that $\mathcal{V}$ satisfies single-crossing with respect to some R. First assume that $\tilde{a}Ra_0$ and $a_0R\tilde{a}$. By Definition 3 and Definition 4, we must have $v_{\tilde{a}} = v'_{\tilde{a}} = v''_{\tilde{a}} = 0$, which violates the requirement that the value for $\tilde{a}$ is unique. Now that $\tilde{a}Ra_0$ and not $a_0R\tilde{a}$. The definition of single crossing implies that $v'' \succ_R v' \succ_R v$. Since $v' \succ_R v$, we must have $aR\tilde{a}$ and not $\tilde{a}Ra$. Similarly, since $v'' \succ v'$, we must have $\tilde{a}Ra$, which is a contradiction. The case where $a_0R\tilde{a}$ and not $\tilde{a}Ra_0$ is similar. In this case we have $v \succ_R v' \succ_R v''$. Since $v' \succ v''$, we have $\tilde{a}Ra$ and not $aR\tilde{a}$. But $v \succ_R v'$ implies that $aR\tilde{a}$.

Given the above analysis, one can construct an example where single-crossing is violated and even though $C_a(t)/t$ is monotone non-decreasing, for instance $(v_{\tilde{a}}, v_a) = (1, 0), (v'_{\tilde{a}}, v_a) = (2, 1.1), (v''_{\tilde{a}}, v''_a) = (3, 1.9)$.

A.2 Proof of Lemma 1

Proof of Lemma 1. Fix an IC mechanism (α, ρ). Let $\tilde{C}(\hat{t}, t)$ be the value of bundle $\alpha(\hat{t})$ to type t, that is, $\tilde{C}(\hat{t}, t) = C_{\alpha(\hat{t})}(t)$. Notice that

$$\partial_2 \tilde{C}(\hat{t}, t) = \frac{\mathrm{d}}{\mathrm{d}t} C_{\alpha(\hat{t})}(t) = C'_{\alpha(\hat{t})}(t) \tag{8}$$

Incentive compatibility requires that,

$$t \in \arg\max_{\hat{t}} \tilde{C}(\hat{t}, t) - \rho(\hat{t}).$$

The first order condition of optimality is

$$\partial_1 \tilde{C}(t,t) - \frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = 0.$$

Let u be the indirect utility function of mechanism (α, ρ), that is $u(t) := \tilde{C}(t,t) - \rho(t)$. Using the above equation, the derivative of the indirect utility function can be simplified,

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = \partial_1 \tilde{C}(t,t) + \partial_2 \tilde{C}(t,t) - \frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = \partial_2 \tilde{C}(t,t) = C'_{\alpha(t)}(t) \tag{9}$$

where the last equality followed from Equation 8.

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

$$\begin{aligned}
\mathbf{E}\left[u(t)\right] &= \int_{\underline{t}}^{\bar{t}} u(t) f_{\tilde{a}}(t)\ \mathrm{d}t \\
&= \int_{\underline{t}}^{\bar{t}} \frac{\mathrm{d}}{\mathrm{d}t}u(t)(1 - F_{\tilde{a}}(t))\ \mathrm{d}t + u(\underline{t}) \\
&= \int_{\underline{t}}^{\bar{t}} C'_{\alpha(t)}(t)(1 - F_{\tilde{a}}(t))\ \mathrm{d}t + u(\underline{t}) \\
&= \mathbf{E}\left[C'_{\alpha(t)}(t)\frac{1 - F_{\tilde{a}}(t)}{f_{\tilde{a}}(t)}\right] + u(\underline{t})
\end{aligned}$$

where the second to last equality followed from substituting (9). Finally, the expected revenue can be written as the difference of surplus and consumer rents,

$$\begin{aligned}
\mathbf{E}\left[\rho(t)\right] &= \mathbf{E}\left[C_{\alpha(t)}(t) - u(t)\right] \\
&= \mathbf{E}\left[C_{\alpha(t)}(t) - C'_{\alpha(t)}(t)\frac{1 - F_{\tilde{a}}(t)}{f_{\tilde{a}}(t)}\right] - u(\underline{t}) \\
&= \mathbf{E}\left[\phi_{\alpha(t)}(t)\right] - \left(C_{\alpha(\underline{t})} - \rho(\underline{t})\right).
\end{aligned}$$

□

A.3 Proof of Lemma 2

Proof of Lemma 2. Assume that $C_a(t)/t$ is monotone non-decreasing in t. Note that this is equivalent to

$$C'_a(t) \geq \frac{C_a(t)}{t}. \tag{10}$$

Directly from the definition of virtual values,

$$\begin{aligned}
\frac{C_a(t)}{t}\phi_{\tilde{a}}(t) &= \frac{C_a(t)}{t}[C_{\tilde{a}}(t) - \frac{1-F_{\tilde{a}}(t)}{f_{\tilde{a}}(t)}] \\
&= C_a(t) - \frac{C_a(t)}{t} \times \frac{1-F_{\tilde{a}}(t)}{f_{\tilde{a}}(t)} \\
&\geq C_a(t) - C'_a(t) \times \frac{1-F_{\tilde{a}}(t)}{f_{\tilde{a}}(t)} = \phi_a(t),
\end{aligned}$$

where the inequality followed from (10). The case where $C_a(t)/t$ is strictly decreasing at t is similar. In that case, $C'_a(t) < \frac{C_a(t)}{t}$. The only subtlety is to make sure that the inequality in the above chain is strict by requiring $1 - F_{\tilde{a}}(t) > 0$. □

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

The setup assumes that set of values $\mathcal{V}(C)$ is finite. That is, $v_{\tilde{a}}$ takes values from $t_0(= \underline{t}) < t_1 < \ldots < t_I$, where $t_0 > 0$.[45] Let $f_{\tilde{a}}(t_i)$ 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

[45] The assumption $t_0 > 0$ is to avoid dividing by zero when defining $C_a(t)/t$.

Lagrangian multiplier of the IC constraint

$$C_{\alpha(t_j)}(t_j) - \rho(t_j) \geq C_{\alpha(t_i)}(t_j) - \rho(t_i).$$

Define the Lagrangian

$$\begin{aligned}
\mathcal{L}(\lambda, \alpha, \rho) &= \Big(\sum_i \rho(t_i) f_{\tilde{a}}(t_i)\Big) + \Big(\sum_{i,j} \lambda(j,i)\Big(C_{\alpha(t_j)}(t_j) - \rho(t_j) - (C_{\alpha(t_i)}(t_j) - \rho(t_i))\Big)\Big) \\
&= \sum_i \Big(\rho(t_i)\Big(f_{\tilde{a}}(t_i) - \sum_j \lambda(i,j) + \sum_j \lambda(j,i)\Big)\Big) \\
&+ \sum_i \Big(\Big(\sum_j \lambda(i,j) C_{\alpha(t_i)}(t_i)\Big) - \Big(\sum_j \lambda(j,i) C_{\alpha(t_i)}(t_j)\Big)\Big).
\end{aligned}$$

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

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

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

$$\begin{aligned}
&\mathcal{L}(\lambda, \alpha, \rho) \\
&= \sum_i \Big(\Big(\sum_j \lambda(i,j) C_{\alpha(t_i)}(t_i)\Big) - \Big(\sum_j \lambda(j,i) C_{\alpha(t_i)}(t_j)\Big)\Big) \\
&= \sum_i \Big(C_{\alpha(t_i)}(t_i) f_{\tilde{a}}(t_i) - \sum_j \lambda(j,i)(C_{\alpha(t_i)}(t_j) - C_{\alpha(t_i)}(t_i))\Big), \qquad (12)
\end{aligned}$$

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

$$\phi_a(t_i) := C_a(t_i) - \frac{1}{f_{\tilde{a}}(t_i)} \sum_j \lambda(j,i)(C_a(t_j) - C_a(t_i)), \qquad \forall t_i, a \in \mathcal{A}. \tag{13}$$

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

$$\mathcal{L}(\lambda, \alpha, \rho) = \sum_{t_i} \phi_{\alpha(t)}(t_i) \times f_{\tilde{a}}(t_i). \tag{14}$$

We summarize the analysis above in the following lemma.

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

$$\lambda(j,i)\Big(C_{\alpha(t_j)}(t_j) - \rho(t_j) - (C_{\alpha(t_i)}(t_j) - \rho(t_i))\Big) = 0. \tag{15}$$

A.4.3 A Construction of Ironed Virtual Values

The lemma below is similar to Lemma 1 and provides a comparison of virtual values based on the change in $C_a(t)/t$, 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* $C_a(t_i)/t_i$ *is monotone non-decreasing in* t_i*, then the induced virtual values, defined via Equation* (13)*, satisfy*

$$\frac{C_a(t_i)}{t_i}\phi_{\tilde{a}}(t_i) \geq \phi_a(t_i), \qquad \forall t_i, a \in \mathcal{A}.$$

Proof. Recall that $C_{\tilde{a}}(t_i) = t_i$ for all i. Directly from the definition of virtual values,

$$\begin{aligned}\frac{C_a(t_i)}{t_i}\phi_{\tilde{a}}(t_i) &= \frac{C_a(t_i)}{t_i}\left(t_i - \frac{1}{f_{\tilde{a}}(t_i)} \sum_j \lambda(j,i)(C_{\tilde{a}}(t_j) - C_{\tilde{a}}(t_i))\right) \\ &= C_a(t_i) - \frac{1}{f_{\tilde{a}}(t_i)} \sum_j \lambda(j,i)(\frac{C_a(t_i)}{t_i} C_{\tilde{a}}(t_j) - C_a(t_i)).\end{aligned}$$

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

$$\begin{aligned}
\frac{C_a(t_i)}{t_i}\phi_{\tilde{a}}(t_i) &= C_a(t_i) - \frac{1}{f_{\tilde{a}}(t_i)} \sum_{j:j>i} \lambda(j,i)(\frac{C_a(t_i)}{t_i} C_{\tilde{a}}(t_j) - C_a(t_i)) \\
&\geq C_a(t_i) - \frac{1}{f_{\tilde{a}}(t_i)} \sum_{j:j>i} \lambda(j,i)(C_a(t_j) - C_a(t_i)) \\
&= C_a(t_i) - \frac{1}{f_{\tilde{a}}(t_i)} \sum_{j} \lambda(j,i)(C_a(t_j) - C_a(t_i)) \\
&= \phi_a(t_i),
\end{aligned}$$

where the inequality followed from monotonicity of $C_a(t_i)/t_i$. □

To interpret Lemma 7 geometrically, note from definition (13) that viewed as a vector, $\phi(t_i)$ is equal to the vector $C(t_i)$, shifted proportional to $C(t_j) - C(t_i)$ for all $t_j > t_i$ with strictly positive $\lambda(j,i)$. The resulting vector is "below" the ray that connects the origin to $C(t)$, as depicted in Figure 5.

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 $\bar{\phi}_{\tilde{a}}$ is monotone non-decreasing, and (4) the assignment rule that only assigns the grand bundle $\tilde{a}$ to types with positive $\bar{\phi}_{\tilde{a}}$ 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) $\bar{\phi}_{\tilde{a}}(t)$ is monotone non-decreasing in t,

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

Proof. For each i, let $F_i = \sum_{j \geq i} f_{\tilde{a}}(t_j)$, 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 $\tilde{a}$ satisfies

$$\bar{\phi}_{\tilde{a}}(t_i) = \frac{\bar{R}(F_i) - \bar{R}(F_{i-1})}{f_{\tilde{a}}(t_i)}.$$

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

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

which is equivalent to

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

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

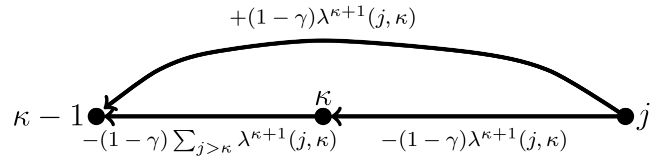

Figure 10: The change in λ^κ compared to $\lambda^{\kappa+1}$. For each $j>\kappa$, A fraction of the flow from j to κ and then to $\kappa-1$ is rerouted to go directly from j to $\kappa-1$.

(iii) when $\kappa = n$, write

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

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

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

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

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

for a parameter γ to be identified shortly. See Figure 10. 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 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

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

and,

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

Summing up the above two equations, we get

$$\phi^\kappa_{\tilde{a}}(\kappa) f_{\tilde{a}}(t_\kappa) + \phi^\kappa_{\tilde{a}}(\kappa-1) f_{\tilde{a}}(t_{\kappa-1}) = \phi^{\kappa+1}_{\tilde{a}}(\kappa) f_{\tilde{a}}(t_\kappa) + \phi^{\kappa+1}_{\tilde{a}}(\kappa-1) f_{\tilde{a}}(t_{\kappa-1}).$$

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

$$\bar{\phi}_{\tilde{a}}(\kappa) = \frac{R(j) - R(j')}{F_j - F_{j'}}$$

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

$$\bar{\phi}_{\tilde{a}}(\kappa) \leq \frac{t_j F_j - t_j F_{j'}}{F_j - F_{j'}} = t_j.$$

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

$$
\begin{aligned}
R^\kappa(F_\kappa) &= \phi^\kappa_{\tilde{a}}(t_\kappa) f_{\tilde{a}}(t_\kappa) + R^\kappa(F_{\kappa+1}) \\
&\geq \bar{\phi}_{\tilde{a}}(\kappa) + \bar{R}(F_{\kappa+1}) \\
&= \bar{R}(F_{\kappa+1}).
\end{aligned}
$$

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 $\bar{\phi}_{\tilde{a}}(t_{j'-1}) = \bar{\phi}_{\tilde{a}}(t_{j'})$. As a result, $\bar{\phi}_{\tilde{a}}(t_{j'}) = \bar{\phi}_{\tilde{a}}(t_{j''})$ 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 shows that a mechanism is optimal for a distribution if it is optimal for arbitrarily close approximations of that distribution.

Lemma 9 (Madarász and Prat, 2017)**.** *Consider distributions μ and μ' with supports $\mathcal{V}$ and $\mathcal{V}'$. Consider a mapping $M : \mathcal{V} \to \mathcal{V}'$ such that $|v_a - v'_a| \leq \epsilon$ for all a if $v' = M(v)$. Assume that for any $\mathcal{V}_1 \subseteq \mathcal{V}$, $\mathbf{Pr}_\mu[v \in \mathcal{V}_1] = \mathbf{Pr}_{\mu'}[v' \in M(\mathcal{V}_1)]$. Then the difference between the optimal revenue of μ and μ' is at least $-\epsilon$ and at most ϵ.*

A.4.5 The Proof

Proof of Proposition 1, the "if" Statement. We start by assuming that the support $\mathcal{V}(C)$ is finite. We later extend the proof to arbitrary support using Lemma 9.

Consider $\bar{\lambda}$ and $\bar{\phi}$ from Lemma 8. Since $\bar{\phi}_{\tilde{a}}(t)$ is monotone non-decreasing in t, there exists a threshold t^* such that $\bar{\phi}_{\tilde{a}}(t) < 0$ if $t \leq t^*$, and $\bar{\phi}_{\tilde{a}}(t) \geq 0$ otherwise. Consider a pure bundling mechanism (α, ρ) that only offers the grand bundle $\tilde{a}$ for price t^*. We show that (α, ρ) 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 $C_a(t_i)/t_i$ is monotone non-decreasing in t_i, $\bar{\phi}$ satisfies

$$\frac{C_a(t_i)}{t_i}\phi_{\tilde{a}}(t_i) \geq \phi_a(t_i), \forall t_i, a \in \mathcal{A}.$$

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

It only remains to verify the complementary slackness condition 15. If a pair of types get the same assignment in (α, ρ) (either $\tilde{a}$ or a_0), 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 distributions with arbitrary support (not necessarily finite). Consider a distribution μ with arbitrary support $\mathcal{V}(C)$. 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 $\mathcal{V}(C)$ for which pure bundling is not optimal. If $\mathcal{V}_a(t)/t$ is monotone non-decreasing over $\mathcal{V}(C)$, then it is also monotone non-decreasing over any subset of $\mathcal{V}(C)$. 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 (a, p) whose revenue is $\epsilon > 0$ more than the optimal pure bundling revenue. We construct the support $\{t'_0, t'_1, \ldots\}$ of μ' inductively from $\mathcal{V}(C)$. Let $t'_0 = \inf(\mathcal{V}(C))$. Choose $\epsilon' < \epsilon$. Given t'_i, define t'_{i+1} as follows,

$$t'_{i+1} = \sup\{t \in \mathcal{V}(C) \mid C_a(t) \leq C_a(t'_i) + \epsilon', \forall a\}.$$

Let the probability of t'_i in μ' be the probability of $[t'_i, t'_{i+1})$ in μ. From Lemma 9, we conclude that the optimal revenue under μ' is strictly higher than the optimal pure bundling revenue under μ', which is a contradiction since $C_a(t)/t$ is non-decreasing over the support of μ'. □

A.5 Proof of Proposition 3

Proof of Proposition 3. Proposition 3 is a special case of Theorem 1. □

A.6 Proof of Lemma 3

Proof of Lemma 3. The proof is by linearity of expectation. Consider $0 \leq \gamma \leq 1$ and two distributions μ_1 and μ_2 in $\mathcal{D}(\alpha, \rho)$. Consider any mechanism $(\hat{\alpha}, \hat{\rho})$ defined over the support of $\gamma\mu_1 + (1-\gamma)\mu_2$. Since (α, ρ) is optimal for μ_1, $\mathbf{E}_{\mu_1}[\rho(v)] \geq \mathbf{E}_{\mu_1}[\hat{\rho}(v)]$. Similarly $\mathbf{E}_{\mu_2}[\rho(v)] \geq \mathbf{E}_{\mu_2}[\hat{\rho}(v)]$. Taking the combination of the two inequalities, and using linearity of expectation, yields the result, $\mathbf{E}_{\gamma\mu_1+(1-\gamma)\mu_2}[\rho(v)] \geq \mathbf{E}_{\gamma\mu_1+(1-\gamma)\mu_2}[\hat{\rho}(v)]$. □

A.7 Proof of Theorem 1

Proof of Theorem 1. We first complete the proof of the first statement by extending the proof to non-finite $\mathcal{A}$. 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 finite set of bundles $\bar{\mathcal{A}}$ such that for any bundle $a \in \mathcal{A}$, there exists a bundle $\bar{a} \in \bar{\mathcal{A}}$ such that $|v_a - v_{\bar{a}}| < \epsilon/2$ for all $v \in \mathcal{V}$. By Lemma 9, when the set of bundles is $\bar{\mathcal{A}}$, there exists a mechanism with revenue at least $\epsilon/2$ larger than any mechanism. This contradicts the proof of Theorem 1 for finite $\bar{\mathcal{A}}$.

We now prove the second statement. Consider the optimal price p for only selling the grand bundle $\tilde{a}$. We prove the optimal pure bundling revenue can be improved by offering another bundle, contradicting the optimality of pure bundling.

In particular, let $\bar{C}_a(t) = \max_z (t, z) \in \mathcal{V}$. By assumption, $\bar{C}_a(t)/t$ is decreasing in t. Consider the change in revenue as a result of offering a for a price $\bar{C}_a(p-\epsilon)$ in addition to $\tilde{a}$ for price p. In this new mechanism, a type v chooses $\tilde{a}$ if $v_{\tilde{a}} - p \geq \max(0, v_a - \bar{C}_a(p-\epsilon))$, chooses a if $v_a - \bar{C}_a(p-\epsilon) \geq \max(0, v_{\tilde{a}} - p)$, and chooses a_0 otherwise. See Figure 11.

Now consider the change in revenue as a result of the discounted offer. The results are twofold. On one hand, a set of types $\mathcal{V}_+(\epsilon)$ will pay $V_a(p-\epsilon)$ for a (shaded green in Figure 11). The gain in revenue $\nabla_+(\epsilon)$ is

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

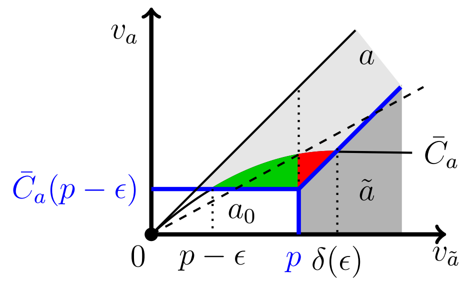

Figure 11: As a result of adding an offer with price $V(p-\epsilon)$ for bundle a to the existing offer of price p for bundle $\tilde{a}$, the types in darker shaded part of curve V will change decisions and contribute to a change in revenue.

where $\mu_+(\epsilon)$ is the probability of $\mathcal{V}_+(\epsilon)$,

$$\mu_+(\epsilon) := \mu(\mathcal{V}_+(\epsilon)) = \int_{p-\epsilon}^{p} \int_{\bar{C}_a(p-\epsilon)}^{\bar{C}_a(t)} f(t,z)\, \mathrm{d}z\, \mathrm{d}t. \tag{20}$$

On the other hand, a set of types $\mathcal{V}_-(\epsilon)$ with value slightly higher than p for $\tilde{a}$ will change their decision from selecting bundle $\tilde{a}$ to bundle a (shaded red in Figure 11). The loss of revenue $\nabla_-(\epsilon)$ is

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

where $\mu_-(\epsilon)$ is the probability of $\mathcal{V}_-(\epsilon)$,

$$\mu_-(\epsilon) := \mu(\mathcal{V}_-(\epsilon)) = \int_{p}^{\delta(\epsilon)} \int_{t-p+\bar{C}_a(p-\epsilon)}^{\bar{C}_a(t)} f(t,z)\, \mathrm{d}z\, \mathrm{d}t.$$

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

$$\nabla'_+(\epsilon) = -\bar{C}'_a(p-\epsilon) \times \mu_+(\epsilon) + \bar{C}_a(p-\epsilon) \times \mu'_+(\epsilon),$$
$$\nabla''_+(\epsilon) = \bar{C}''_a(p-\epsilon) \times \mu_+(\epsilon) + \bar{C}_a(p-\epsilon) \times \mu''_+(\epsilon).$$

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

$$\begin{aligned}
\mu'_+(\epsilon) &= -\int_{\bar{C}_a(p-\epsilon)}^{\bar{C}_a(p-\epsilon)} f(p-\epsilon, z)\,\mathrm{d}z - \int_{p-\epsilon}^{p} \bar{C}'_a(p-\epsilon) f(t, \bar{C}_a(p-\epsilon))\,\mathrm{d}t \\
&= -\int_{p-\epsilon}^{p} \bar{C}'_a(p-\epsilon) f(t, \bar{C}_a(p-\epsilon))\,\mathrm{d}t, \text{ and} \qquad (21)\\
\mu''_+(\epsilon) &= \bar{C}'_a(p-\epsilon) f(p-\epsilon, \bar{C}_a(p-\epsilon)) - \int_{p-\epsilon}^{p} \frac{\mathrm{d}}{\mathrm{d}\epsilon} \bar{C}'_a(p-\epsilon) f(t, \bar{C}_a(p-\epsilon))\,\mathrm{d}t.
\end{aligned}$$

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

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

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

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

We calculate the derivatives of μ_-,

$$\begin{aligned}
\mu'_-(\epsilon) &= \delta'(\epsilon) \int_{\delta(\epsilon)-p+\bar{C}_a(p-\epsilon)}^{\bar{C}_a(\delta(\epsilon))} f(\delta(\epsilon), z)\,\mathrm{d}z - \int_{p}^{\delta(\epsilon)} \bar{C}'_a(p-\epsilon) f(t, t-p+\bar{C}_a(p-\epsilon))\,\mathrm{d}t \\
&= -\int_{p}^{\delta(\epsilon)} \bar{C}'_a(p-\epsilon) f(t, t-p+\bar{C}_a(p-\epsilon))\,\mathrm{d}t. \\
\mu''_-(\epsilon) &= -\delta'(\epsilon)\bar{C}'_a(p-\epsilon) f(\delta(\epsilon), \delta(\epsilon)-p+\bar{C}_a(p-\epsilon)) \\
&\quad - \int_{p}^{\delta(\epsilon)} \frac{\mathrm{d}}{\mathrm{d}\epsilon} \bar{C}'_a(p-\epsilon) f(t, t-p+\bar{C}_a(p-\epsilon))\,\mathrm{d}t
\end{aligned}$$

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

$$\nabla''(0) = (p - \bar{C}_a(p)) \times \delta'(0)\bar{C}'_a(p) f(p, \bar{C}_a(p)).$$

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

$$\bar{C}_a(p) > (p - \bar{C}_a(p))\delta'(0). \qquad (23)$$

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 $\tilde{a}$ and a,

$$\delta(\epsilon) - p = \bar{C}_a(\delta(\epsilon)) - \bar{C}_a(p - \epsilon).$$

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

$$\delta'(0)(1 - \bar{C}'_a(p)) = \bar{C}'_a(p).$$

Substituting into 23, we need to show that

$$\bar{C}_a(p) > (p - \bar{C}_a(p))\frac{\bar{C}'_a(p)}{1 - \bar{C}'_a(p)}. \tag{24}$$

Since by assumption $\bar{C}'_a(p) < \bar{C}_a(p)/p$, we have

$$\frac{\bar{C}'_a(p)}{1 - \bar{C}'_a(p)} < \frac{\bar{C}_a(p)}{p - \bar{C}_a(p)},$$

which is identical to 24, completing the proof. □

A.8 Proof of Proposition 4

Proof of Proposition 4. Assume that y is stochastically non-decreasing in $x > 0$. There exists an grand bundle, namely the grand bundle. The value of type (x, y) for the grand bundle is x. The value-ratio of type (x, y) for a bundle a is $r_a = g_a(y)$. Lemma 5 implies that r is stochastically non-decreasing in the value for the bundle x. Theorem 1 then implies that pure bundling is optimal.

To prove the second statement, assume that y_i is stochastically decreasing at the top in x. By definition, $\bar{y}_i(x)$ is decreasing in x. Let a be a bundle that includes a_i unit of product i, and one unit of all other products. Thus $r_a = h_i(a_i, y_i)$. By monotonicity of h_i, $\bar{r}_a(y)$ is decreasing in x. Thus r_a is stochastically decreasing at the top in x, and Theorem 1 implies that pure bundling is not optimal. □

A.9 Randomized Alternatives

Assume that a bundle $a \in \mathcal{A}$ is a distribution over outcomes, $a \in \Delta(\mathcal{O})$, for some set of outcomes $\mathcal{O}$. Each type v is parameterized by $w \in \mathbb{R}^{\mathcal{O}}$, and have value $v_a = \mathbf{E}_{o\sim a}[w_o]$ for bundle a. Assume $\tilde{o}$ is the grand bundle, that is $w_{\tilde{o}} \geq w_o$ for all $o \in \mathcal{O}$. Define a profile $\theta \in \mathbb{R}^{\mathcal{O}}$ by setting $\theta_o = w_o/w_{\tilde{o}}$ for each o. The following proposition states that it is sufficient to verify stochastic order of the value-ratios only for outcomes, instead of the larger set of bundles.

Proposition 6. *Pure bundling is optimal if the distribution of θ is stochastically non-decreasing in $w_{\tilde{o}}$. Pure bundling is not optimal if the distribution of types is continuous and interior and for some o, θ_o is stochastically decreasing at the top in $w_{\tilde{o}}$.*

Proof. Note that $\mathbf{E}_{o\sim a}[w_o]$ is linear in each w_o. Thus by Lemma 5, if θ is stochastically non-decreasing in $w_{\tilde{o}}$, then so is the profile of value-ratios $r \in \mathbb{R}^{\mathcal{A}}$ where $r_a = v_a/w_{\tilde{o}}$. For the second statement, note that allowing for randomization enlarges the set of mechanisms. □

A.10 Proof of Theorem 2

Proof of Theorem 2. We show that given the conditions, the uniform-pricing mechanism that offers any bundle at price t^* is optimal. The proof uses the fact that the set of distributions over which a mechanism is optimal is convex, Lemma 3. For each a, consider the distribution of types $\mu|a$ conditioned on a being an grand bundle. The distribution μ is a distribution over the set of distributions $\{\mu|a\}_{a:P_a>0}$. Therefore, by Lemma 3 it is sufficient to show that the uniform-pricing mechanism with price t^* is optimal for any distribution $\mu|a$ where $P_a > 0$. As a result, Theorem 1 applies to imply offering only a at price t^* is optimal. The uniform-pricing mechanism that offers *all* bundles at price t^*, is also optimal for $\mu|a$. Thus the theorem follows. □

A.11 Proof of Proposition 5

Proof of Proposition 5. Consider any random variable r that is stochastically non-decreasing in $v_{\tilde{a}}$. Choose any marginal distribution $F_{\tilde{a}}$ such that t maximizes $t \times (1 - F_{\tilde{a}}(t))$. By Theorem 1, it is optimal to only offer $\tilde{a}$ at price t. □