A CHARACTERIZATION FOR OPTIMAL BUNDLING OF PRODUCTS WITH INTER-DEPENDENT VALUES

By

Soheil Ghili

February 2021

COWLES FOUNDATION DISCUSSION PAPER NO. 2273

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

A Characterization for Optimal Bundling of Products with Inter-dependent Values

Soheil Ghili*
Yale University

February 26, 2021

Abstract

This paper studies optimal bundling of products with inter-dependent values. I show that, under some conditions, a firm optimally chooses to sell only the full bundle of a given set of products if and only if the optimal sales volume of the full bundle is larger than the optimal sales volume for any smaller bundle. I then provide an interpretation of this characterization based on (i) the magnitude of the variation across consumers in how complementary they find different products, and (ii) how this variation correlates with price sensitivity.

1 Introduction

This paper has two objectives. First, I characterize necessary and sufficient conditions for when pure bundling of a finite number of products with inter-dependent values is optimal for a monopolist. Second I provide an interpretation of this characterization and examine the relevance and usefulness of that interpretation using simulation.

The literature on theoretical analysis of optimal bundling decisions is large and growing. To my knowledge, however, there has not been a full characterization of when it is optimal to bundle. This paper provides such characterization for "pure" bundling (i.e., the act of selling only the package of all available products together as one bundle). Under some assumptions, most notably weak complementarity across products and a form of vertical differentiation

*E-mail: soheil.ghili@yale.edu. Click here for the most current version.

among potential consumers, I prove that optimal bundling admits a simple characterization: Pure bundling is optimal if the optimal sales volume for the full bundle (if sold alone) is strictly larger than that for any other bundle. Conversely, if there is at least one bundle whose optimal sales volume (if sold alone) is strictly larger than that of the full bundle, then pure bundling is sub-optimal.[1] In simpler terms, bundling is optimal if it helps sell more.

The simplicity of this characterization and its tight conditions provide an opportunity to arrive at useful interpretations. I argue that this result has two implications. First, the variation across consumers in the complementarity levels among products is important for bundling decisions. The more such variation, the more likely it is for unbunbling to be optimal. Second, I argue that it is not just the magnitude of the variation in complementarity that matters. It also matters how correlated this variation across consumers is with the variation in price sensitivity. If more price sensitive consumers see more complementarity across products, then the optimality of bundling becomes more likely. Variants of this latter interpretation were previously mentioned in the literature. As such, in discussing it, I do not rely only on the characterization in this paper; I also draw on results and interpretations by Long (1984); Armstrong (2013) on the comparison between the price elasticity demand for the full bundle and those of individual products, as well as by Haghpanah and Hartline (2019) on the monotonicity of the ratio between the valuations of sub-bundles and that of the full bundle in the valuation of the full bundle itself. I show the relevance of these two interpretations for applied work using a simulation of random-coefficient discrete choice demand models a la Berry et al. (1995).

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 sets up the model, presents the characterization of optimal bundling decisions, and discusses its relation to other results in the literature. Section 4 discusses the interpretation of the main result. Section 5 showcases the relevance of the interpretation for applied work using a simulation. Section 6 concludes.

[1] Note that this is slightly short of a full characterization given it does not determine the optimal bundling strategy when the optimal sales volume for the full bundle is larger than those for some smaller bundles but exactly equal to some others. It can be shown that a strengthening of my model assumptions will imply full bundling is optimal in this case, completing the if-and-only-if characterization. Nevertheless, I see the benefit from being able to speak to this special scenario as too marginal to justify the stronger assumptions it requires. As such, I maintain the smaller set of assumptions that do not speak to this special case. See section 3 for more details.

2 Related Literature

The study of bundling has a large literature which dates at least as far back as Stigler (1963). The majority of the papers in this literature focus on the case of "independent values," meaning the valuation by each consumer of any given product i is not impacted by whether she also possesses product $i' \neq i$. Pioneering in this area was Adams and Yellen (1976), pointing out that bundling can be more profitable than unbundling when there is negative correlation among consumers in how they value individual products. Other studies such as McAfee et al. (1989); Menicucci et al. (2015); Pavlov (2011); Schmalensee (1984); Fang and Norman (2006) further develop results on optimal bundling under independent values. Most of these studies concentrate on sufficient conditions for bundling and many focus on a setting with two products only. Although most of this literature examines a monopolist seller (which is also the focus of this paper), some studies have analyzed multiple sellers (McAfee et al. (1989); Zhou (2017, 2019)).

The literature allowing for dependence in product valuations, to which this paper belongs, is considerably smaller. Part of this literature focuses directly on bundling (e.g., Haghpanah and Hartline (2019); Armstrong (2013, 2016); Long (1984)) whereas some study price discrimination settings which have implications for bundling (e.g., Anderson and Dana Jr (2009); Deneckere and Preston McAfee (1996)). This paper complements this literature in that it imposes a different set of assumptions (stronger only than those imposed by Haghpanah and Hartline (2019)) and delivers simple but necessary and sufficient conditions for bundling based on optimal quantities sold. I also contribute to this sub-literature in two other ways. First, by connecting the interpretation of my results to those based on price elasticities (such as Long (1984); Armstrong (2013)) and those based on ratio monotonicity (such as Haghpanah and Hartline (2019); Anderson and Dana Jr (2009); Deneckere and Preston McAfee (1996); Salant (1989)), I also illuminate the relationship between the intuitions from these two sets of results themselves, a connection not made before. Second, to my knowledge, this paper makes the first attempt to take insights from the analysis of optimal bundling under inter-dependent values to applied models of demand.

3 Model

3.1 Setup

There are n products indexed 1 through n. Possible bundles of these products are denoted $b \subseteq \{1, ..., n\}$. Set $\mathcal{B} = \{b | b \subseteq \{1, ..., n\}\}$ represents the set of all possible bundles.[2] By $\bar{b}$ denote the full bundle $\{1, ..., n\}$. There is a unit mass of customers whose types are represented by t with probability distribution $f(\cdot) > 0$. The willingness to pay by type t for bundle b is denoted $v(b, t)$. Without loss, assume $v(\emptyset, t) = 0$.

The problem the firm solves has two components. First, the firm makes a bundling decision. It chooses the optimal set B^* of bundles b among subsets B of $\mathcal{B}$ that satisfy $\emptyset \notin B$. Note that there are as many as $2^{2^n - 1}$ possible bundling strategies. Thus, characterizing the conditions under which the firm can simply choose $B^* = \{\bar{b}\}$ should indeed be of value.

The second decision by the firm is choosing prices $p(\cdot) : B \to \mathbb{R}$ for the bundles offered.[3] Denote by $\mathcal{P}_B$ the set of all possible such pricing functions.

Once the firm has decided on set B and prices $p(\cdot)$, customers decide which bundles to purchase. Each customer t's decision $\beta(t|B, p) \subseteq B$ is determined by:

$$\beta(t|B, p) = \arg\max_{\hat{\beta} \subseteq B} v(\cup_{b \in \hat{\beta}} b, t) - \Sigma_{b \in \hat{\beta}} p(b) \tag{1}$$

Throughout the paper, I assume customers break ties in favor of more expensive bundles and randomize evenly if similarly priced bundles tie for first. Also, note that equation 1 implies that customers want at most one unit of each product i and find additional units redundant.

Demand for bundle b is given by the measure of customers t who would choose to purchase bundle b:

$$D(b|B, p) = \int_t \mathbb{1}_{b \in \beta(t|B,p)} f(t) dt \tag{2}$$

Firm profit under bundling strategy B and pricing strategy p is given by:

$$\pi(B, p) = \Sigma_{b \in B} D(b|B, p) \left(p(b) - \Sigma_{i \in b} c_i \right) \tag{3}$$

[2]My notation, in part, follows Haghpanah and Hartline (2019).

[3]Note that, in principle, one could model the bundling decision through pricing; because not offering a product would be equivalent to pricing it so high that no customer would purchase it. As such, separating the bundling and pricing decisions in the model is, in some sense, redundant. Nevertheless, I decided to carry out this separation because it provides a more streamlined notation for the problem.

where c_i is the (constant) marginal cost of producing a unit of good i. Costs c_i can be negative, allowsing for "damaged-good" settings a la Deneckere and Preston McAfee (1996).

We can now write out the firm's problem. The firm optimally chooses B^* and p^* maximize profit:

$$(B^*, p^*) = \arg \max_{B \in \mathcal{B}, p \in \mathcal{P}_B} \pi(B, p) \tag{4}$$

With the setup of the firm problem laid out, I now introduce a few more definitions and notations. For any disjoint bundles b and b', denote by $v(b, t|b')$ the valuation by type t for b conditional on possessing b'. Formally:

$$v(b, t|b') \equiv v(b \cup b', t) - v(b', t)$$

In a similar manner, denote $\beta(t|B, p, b') = \arg\max_{\hat{\beta} \subseteq (B \backslash \{b'\})} v(\cup_{b \in \hat{\beta}} b, t|b') - \Sigma_{b \in \hat{\beta}} p(b)$. Also denote $D(b|B, p, b') = \int_t 1_{b \in \beta(t|B,p,b')} f(t) dt$. Finally, denote

$$\pi(B, p|b') = \Sigma_{b \in B \backslash \{b'\}} D(b|B, p, b') \bigg(p(b) - \Sigma_{i \in b} c_i \bigg).$$

Finally, at times with some abuse of notation I will refer to $\cup_{b \in \beta(t|B,p,b')} b$ simply by $\beta(t|B, p, b')$. I next turn to the assumptions and the main result.

3.2 Main Result

The main results of the paper is about how optimal bundling decisions are informed by the comparison among optimal sales volumes for different bundles. I start with some necessary assumptions and definitions.

Assumption 1. ***Monotonicity:*** *Consider t and t' such that $v(\bar{b}, t) > v(\bar{b}, t')$. Then for all disjoint b and b', we have $v(b, t|b') > v(b, t'|b')$.*[4]

Assumption 2. ***Complementarity:*** *For any type t and disjoint bundles b and b', we have:*

$$v(b \cup b', t) \geq v(b, t) + v(b', t)$$

[4]The strict inequality in $v(b, t|b') > v(b, t'|b')$ is not necessary for the proof of the main result; the weak version $v(b, t|b') \geq v(b, t'|b')$ would suffice. Nevertheless, I decided to use the strict version as it simplifies the notations for the proof substantially.

Assumption 3. ***Quasi-concavity:*** *For any $b \in B$, profit function $\pi(B, p|b')$ is strictly quasiconcave in $p(b)$ for all values of $p(b)$ that yield strictly positive demand for b.*[5]

Assumption 4. ***Non-triviality:*** *for every product $i \in \{1, ..., n\}$, there is a non-measure-zero set of types t such that $v(\{i\}, t) > c_i$.*

Assumption 5. ***Continuity:*** *for any bundle b, value function $v(b, t)$ has at most finitely many points where it is discontinuous in t.*

Definition 1. *By $D^*(b)$ denote the "optimal quantity sold" of bundle b if no other bundle were offered by the firm. Formally, $D^*(b)$ is defined as $D(b|\{b\}, p_b^*)$ where p_b^* is the optimal price for bundle b when $B = \{b\}$.*

Definition 2. *A given firm strategy (B, p) involves pure bundling if for any customer of type t, we have:*

$$\cup_{b \in \beta(t|B,p)} b \in \{\emptyset, \bar{b}\}$$

This definition for pure bundling is "neat" in the sense that it avoids some technical issues. For instance, if B includes bundles other than $\bar{b}$ but at prices so high that no customer would purchase them, Definition 2 considers it pure bundling. In a similar vein, if under (B, p) multiple bundles are offered but the prices are such that each customer who buys anything combines them to construct $\bar{b}$, then this definition again detects pure bundling. One can verify that whenever the conditions in this definition hold, it is (not necessarily uniquely) optimal for the firm to offer only the full bundle $B = \{\bar{b}\}$. With this definition, we are now ready to state the main result.

Theorem 1. *Under assumptions 1 through 5, the optimal strategy (B^*, p^*) involves pure bundling if:*

$$D^*(\bar{b}) > \max_{b \in \mathcal{B} \setminus \{\bar{b}\}} D^*(b)$$

Conversely, the optimal strategy does not involve pure bundling if:

$$D^*(\bar{b}) < \max_{b \in \mathcal{B} \setminus \{\bar{b}\}} D^*(b)$$

[5]This simply means the profit peaks only once as we vary each price.

This theorem is proved in the appendix. Note that this result is slightly short of a full characterization because it does not specify whether pure bundling is optimal when $D^*(\bar{b}) = \max_{b\in\mathcal{B}\setminus\{\bar{b}\}} D^*(b)$. One can show that under this last possibility, pure bundling is optimal if instead of assuming profits are strictly quasi-concave in each price, we assume they are strictly concave and differntiable at peak. Even though this would yield a full characterization, I decided that the ability to speak to the "measure-zero" case of $D^*(\bar{b}) = \max_{b\in\mathcal{B}\setminus\{\bar{b}\}} D^*(b)$ is too small a return to justify such a restrictive assumption as strict concavity. As such, I maintain the quasi-concavity assumption.

In words, this result says that the firm would introduce smaller bundles to the market if some of those smaller bundles would "sell more" than the full bundle.

Theorem 1 has two features. First, it provides (an almost) *full characterization* of when pure bundling is optimal. Second, its characterization is simple. Both of these features, as I will discuss later, help with establishing an intuitive interpretation for the main result.

3.3 Interpretation of the assumptions: vertical differentiation among customers

It is worth further discussing the economic content of the underlying assumptions of the model. While assumptions 3 through 5 are technical ones, assumptions 1 and 2 have meaningful economic implications. Specifically, assumption 1 implies that there is vertical differentiation among the potential customers of the products the firm seeks to sell. The following lemma formalizes this idea:

Lemma 1. *There is a mapping τ from the set T of types t on to the interval $[0,1]$ such that:*

1. $\forall t, t' \in T : v(\bar{b}, t) > v(\bar{b}, t') \Leftrightarrow \tau(t) > \tau(t')$.
2. *τ is a sufficient statistic: Once $\tau(t)$ is known, one can fully pin down all $v(b, t|b')$ without having to know t.*

Proof of Lemma 1. Set $\tau(t) \triangleq \frac{v(\bar{b},t)-\min_{t'} v(\bar{b},t')}{\max_{t'} v(\bar{b},t')-\min_{t'} v(\bar{b},t')}$. By construction, it satisfies (1). It is straightforward to verify that, by monotonicity, it also satisfies (2). **Q.E.D.**

Note that, based on this lemma, it is without loss to think of t as $\tau(t)$ and, hence, the set of all possible t as $[0,1]$. Therefore, we can use expressions such as $t \geq t'$. The proofs in the appendix are all based on the assumption that $t \in [0,1]$. Also note that the monotonicity assumption *does not* rule out $v(b,t) > v(b',t)$ co-existing with $v(b,t') < v(b',t')$ (i.e., it does not force consumers to rank the products the same way). It, rather, rules out $v(b,t) > v(b,t')$ co-existing with $v(b',t) < v(b',t')$.

Lemma 1 shows that types t can be ordered based on their willingesses to pay for bundles, implying a vertical relationship among types. The next lemma shows that this vertical differentiation also manifests itself in the consumers' purchasing behaviors.

Lemma 2. *Consider bundling strategy B and pricing strategy p. It can be shown that*

$$\forall t,t' : \beta(t|B,p) \subsetneq \beta(t'|B,p) \Rightarrow t < t'$$

This lemma says that a higher type's purchase decision can never be a strict subset of a lower type's. The proof is in the appendix.

Obviously, the fact that quasi-concavity and continuity do not have as straightforward economic interpretations as those of monotonicity and complementarity does not mean that those assumptions are not restrictive. Specifically, one can prove a different version of Theorem 1 without quasi-concavity and continuity; but that version will be weaker and less straightforward to state and interpret.

3.4 Discussion

This subsection discusses the relationship between Theorem 1 and the literature. In particular, I focus on ratio-monotonicity results of which variants have been discussed in Anderson and Dana Jr (2009); Salant (1989); Deneckere and Preston McAfee (1996), and, most generally, in Haghpanah and Hartline (2019). These results (i) are related to mine, and (ii) help with the interpretation that I will put forth later in the paper.

Under some conditions, Haghpanah and Hartline (2019) show that if $\frac{v(b,t)}{v(\bar{b},t)}$ is first-order stochastically increasing in $v(\bar{b},t)$ for all b, then pure bundling is optimal. Similarly, they show, that under some conditions if $\frac{v(b,t)}{v(\bar{b},t)}$ is strictly first-order stochastically decreasing in $v(\bar{b},t)$ for some b, then mixed bundling is optimal. Other variants of these two results are also given in Salant (1989); Anderson and Dana Jr (2009); Deneckere and Preston McAfee (1996). Proposition 1 clarifies how Theorem 1 relates to ratio monotonicity results.

Proposition 1. *If $\frac{v(b,t)}{v(\bar{b},t)}$ is increasing (decreasing) in $v(\bar{b},t)$ for bundle b, then $D^*(b) \leq (\geq)D^*(\bar{b})$.*

Proposition 1 is proved in the appendix. This result shows that Theorem 1 tightens the ratio monotonicity results in the literature to the extent that almost an if-and-only-if condition obtains (note that under assumption 1, the stochastic ratio monotonicity conditions in Haghpanah and Hartline (2019) boil down to deterministic ratio monotonicity). The following examples illustrate the gap in the ratio monotonicity conditions that Theorem 1 closes.

Example 1. Suppose $n = 2$, and t is uniformly distributed between 0 and 1. Assume the firm can produce these products at no cost. By b denote the bundle $\{1\}$. For simplicity, assume the complementary bundle $b^c = \{2\}$ is not valued by any type: $\forall t : v(\{2\}, t) = 0$. A common example of this is when $\{2\}$ is an "add on," which is not of value by itself but can add value once the "base product" is present (e.g., additional memory for a smart phone). Suppose $\forall t : v(b, t) = t$. For $v(\bar{b}, t)$, consider two cases.

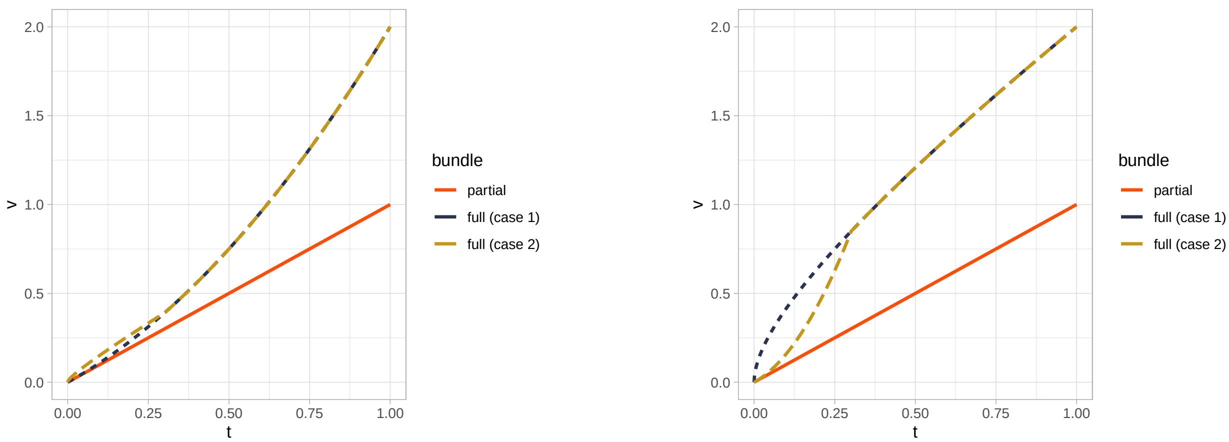
Case 1: suppose $\forall t : v(\bar{b}, t) = t + t^2$. That is, each type t's valuation of the add on on top of the original product is t^2. It is straightforward to verify that $\frac{v(b,t)}{v(\bar{b},t)}$ is strictly decreasing in $v(\bar{b}, t)$. It is also straightforward to verify that $D^*(b) = 0.5$ and $D^*(\bar{b}) = 0.42 < 0.5$. Finally, one can show that the optimal strategy for the firm would be to offer b at the price of 0.5 alongside $\bar{b}$ at the price of 0.94. In sum, the optimal strategy is in line with both what Theorem 1 predicted and what ratio monotonicity would predict.

Case 2: suppose $\forall t \geq 0.3 : v(\bar{b}, t) = t + t^2$ and $\forall t \leq 0.3 : v(\bar{b}, t) = t + \sqrt{t} \times \frac{0.3^2}{\sqrt{0.3}}$. That is, the "add on value" is initially concave in t and then becomes convex like Case 1 (see Figure 1a for a comparison between valuations in Case 1 v.s. that in Case 2). In this case, $\frac{v(b,t)}{v(\bar{b},t)}$ becomes strictly increasing in $v(\bar{b}, t)$ over $t \in (0, .3)$, which does not satisfy the necessary conditions in Haghpanah and Hartline (2019). Nevertheless, we still have $D^*(b) = 0.5$ and $D^*(\bar{b}) = 0.42 < 0.5$. One can also show that the optimal strategy by the firm is the same unbundling strategy that we arrived at in Case 1.

Example 2. With the exception of $v(\bar{b}, t)$ assume the exact same setup as in Example 1. For $v(\bar{b}, t)$ consider two cases:

Case 1. suppose $\forall t : v(\bar{b}, t) = t + \sqrt{t}$. That is, each type t's valuation of the add on on top of the original product is $\sqrt{t}$. It is straightforward to verify that $\frac{v(b,t)}{v(\bar{b},t)}$ is strictly increasing in $v(\bar{b}, t)$. It is also straightforward to verify that $D^*(b) = 0.5$ and $D^*(\bar{b}) = 0.59 > 0.5$. Finally, one can show that the optimal strategy for the firm would be to only offer $\bar{b}$ at the price of 1.05. In sum, the optimal strategy is in line with both what Theorem 1 predicted and with ratio monotonicity.

Case 2: suppose $\forall t \geq 0.3 : v(\bar{b}, t) = t + \sqrt{t}$ and $\forall t \leq 0.3 : v(\bar{b}, t) = t + t^2 \times \frac{\sqrt{0.3}}{0.3^2}$. That is, the "add on value" is initially convex in t and then becomes concave like Case 1 (see Figure 1b for a comparison between valuations in Case 1 v.s. that in Case 2). In this case,



(a) Example 1 (b) Example 2

Figure 1: Value functions in Examples 1 and 2

$\frac{v(b,t)}{v(\bar{b},t)}$ becomes strictly decreasing in $v(\bar{b},t)$ over the interval $(0, 0.3)$. This does not satisfy the necessary conditions in Haghpanah and Hartline (2019). Nevertheless, we still have $D^*(b) = 0.5$ and $D^*(\bar{b}) = 0.59 > 0.5$. One can also show that the optimal strategy by the firm is the same pure bundling strategy that we arrived at in Case 1.

The examples constructed here modified valuations of those who did not purchase the products. One can formulate other examples in which valuations of those who do purchase get modified but we arrive at similar conclusions.

In addition to relating the results of this paper to the bundling literature, the examples above should also highlight the simple implications these results have for price discrimination decisions based on quality: price discrimination is optimal if the low quality version, if priced optimally, sells more than the high quality version. A perhaps unsurprising consequence is that price discrimination is more likely optimal when the higher quality version is more costly to make (because higher marginal costs bring the optimal sales volume down.) This also applies to the case of "damaged goods" Deneckere and Preston McAfee (1996): damage the product if it helps sell more.

It is worth noting that in spite of providing (an almost) full characterization for pure bundling, the results in this paper are not a strengthening of the previous literature. This is because the background assumptions for Theorem 1 (i.e., assumptions 1 through 5) are different from those in the literature. In particular, my assumptions are stronger than those imposed in Haghpanah and Hartline (2019). As such, I view Theorem 1 as complementary to (rather than a substitute for) the related results in the literature.

4 Interpretation of the result: complementarity, its variation, and its co-variation with price sensitivity

Although the characterization of optimal bundling in Theorem 1 is intuitive, it is still worth further discussing how this can be interpreted in terms of the primitives of the model (i.e., valuation function v). This section interprets Theorem 1 based on (i) how much variation there is across consumers in the complementarity levels they see among products, and (ii) how this variation is correlated with variation in price sensitivity.

Variation in complementarity levels: The condition $\forall b : D^*(b) \leq D^*(\bar{b})$ for optimal pure bundling means that for all b we have $D^*(b) \leq D^*(b^C|b)$ where $b^C = \bar{b} \setminus b$ (this latter inequality is implied by the quasi-concavity assumption. See proof of Theorem 1 in the appendix.) That is, how many units b^C would sell conditional on everyone having b plays a crucial role. One determinant of $D^*(b^C|b)$ would be the variation among customers in how much they value b^C conditional on having b. If $v(b^C, t|b)$ is fairly homogeneous across t (and if it is above $\Sigma_{i \in b^C} c_i$), then the firm would optimally sell b^C to the majority (or all) of customers, likely surpassing $D^*(b)$. If there is a large variation in $v(b^C, t|b)$, however, then the chance of $D^*(b^C|b) < D^*(b)$ (and hence that of $D^*(\bar{b}|b) < D^*(b)$) increases. As a result, the analysis in this paper suggests that the variation across customers in the complementarity level among products would be an important factor for optimal bundling decisions.

Correlation between complementarity and price sensitivity: The condition $\forall b : D^*(b) < D^*(\bar{b})$ for optimal pure bundling means that the demand level at which the price elasticity for $\bar{b}$ hits -1 is higher than the that for other bundles.[6] (note that this intuition bears some resemblance to elasticity-based results from Long (1984); Armstrong (2013)). In particular, for any bundle b, the aforementioned comparison holds both between b and $\bar{b}$ and between $b^C = \bar{b} \setminus b$ and $\bar{b}$. That is, if we go through types in a descending way based on $v(\bar{b}, t)$, then the willingness to pay for $\bar{b}$ dwindles less rapidly than does that for b or b^C. In other words, *more price sensitive types must see a higher degree of complementarity between b and b^C than do less price sensitive types.*

The aforementioned interpretation is also in line with the ratio monotonicity conditions from Haghpanah and Hartline (2019); Anderson and Dana Jr (2009); Salant (1989); Deneckere and Preston McAfee (1996) and has been mentioned by Haghpanah and Hartline (2019). Suppose $v(\bar{b}, t) \leq v(\bar{b}, t')$ for some t and t'. The sufficient ratio monotonicity condition for optimality of pure bundling says $\frac{v(b,t)}{v(\bar{b},t)} \leq \frac{v(b,t')}{v(\bar{b},t')}$ and $\frac{v(b^C,t)}{v(\bar{b},t)} \leq \frac{v(b^C,t')}{v(\bar{b},t')}$. From these

[6] Though it is not necessary, to ease the interpretation assume all c_i are zero.

inequalities, one can conclude

$$\frac{v(\bar{b},t)}{v(b,t)+v(b^C,t)} \geq \frac{v(\bar{b},t')}{v(b,t')+v(b^C,t')}.$$

This inequality, roughly, shows that the synergy between b and b^C is from the perspective of type t is higher than that for t'.

Based on this discussion, I propose that two key factors in optimal bundling decisions are (i) how much variation there is across customers in the complementarity they see among products, and (ii) how much this variation correlates with customers' price sensitivity levels. Next section verifies this interpretation using an empirical model of demand a la BLP.

5 Relevance of the interpretation of Theorem 1

This section tests the relevance and usefulness of the interpretation proposed in section 4 for applied work. Neither Theorem 1 in this paper nor, to my knowledge, any other theoretical result in the literature provides a characterization that could be directly applied to common econometric models of demand. Nevertheless, the informal interpretations that I obtain in section 4 may potentially be useful. In this section, I examine the usefulness of those interpretations for choosing the right model specification. I do this using a random coefficient discrete choice model a la Berry et al. (1995).

Setup. To keep things simple, I again focus on a setting with two products where one is the base product and the other an add on. Given that the add-on in and of itself is not valuable by customers, I use the notation $i \in \{1,2\}$ for the basic and premium versions of the product; $i=1$ represents the basic version and $i=2$ represents the version with the add on.

Each customer t has a utility u_{it} for product i. This utility is given by:

$$u_{it} = \alpha_0 + \alpha_{1,t}p_i + \alpha_{2,t}\mathbb{1}_{i=2} + \varepsilon_{it} \tag{5}$$

where α_0 is a constant, $\alpha_{1,t}$ is the price coefficient, $\alpha_{2,t}$ is the valuation of the add-on, and ε_{it} is the error term which has an Extreme Type I distribution. Note that both $\alpha_{1,t}$ and $\alpha_{2,t}$ are heterogeneous across customers t. I assume that for each customer t, the pair $(\alpha_{1,t}, \alpha_{2,t})$ is an independent draw from a bi-variate normal distribution:

$$(\alpha_{1,t}, \alpha_{2,t}) \sim \left((\mu_1, \mu_2), \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} \right) \tag{6}$$

where μ_1 and μ_2 are the means, σ_1 and σ_2 are the standard deviations, and ρ is the correlation between the two coefficients.

Results. The question I study in this section is: when is it optimal for the firm to offer both products $i = 1$ and $i = 2$ to the market, and when is it optimal to offer only $i = 2$? Unfortunately, Theorem 1 does not directly apply to a market where the customers' preferences are given by u_{it}. Nevertheless, the interpretation proposed in section 4 may be useful.

According to the interpretation provided in Section 4, there are two important objects when it comes to bundling decisions: (i) the variation across customers t the complementarity between products, and (ii) the correlation across customers between the complementarity level and the price sensitivity. In our BLP setting, these two concepts translates to parameters σ_2 and ρ respectively.

Based on our theoretical results, we expect pure bundling to be optimal when:

- σ_2 is small, which means that the valuations for the add-on are so homogeneous that it makes sense to give the add on to all purchasing customers.

- ρ is small (negative), which means more price sensitive customers (i.e., those with smaller, more negative $\alpha_{1,t}$) will consider the add on to deliver a higher relative value on top of the basic product.

To verify whether the above interpretation derived from Theorem 1 is indeed of relevance to the empirical context of this section, I simulate the optimal bundling decisions for the market described above. I use the following parameterization: $\alpha_0 = 2.5$, $\mu_1 = -2$, $\sigma_1 = 1$, $\mu_2 = 1.5$. The remaining two parameters, ρ and σ_2, are left flexible; and the model is simulated for a range of these parameters. Figure 2 shows the results.

As this figure depicts, and as the proposed intuition would suggest, bundling becomes optimal when σ_2 decreases (i.e., when the population values the add on rather homogeneously) or when ρ increases (i.e., when more price sensitive customers have higher relative valuations for the add on). Both of these are suggestive that, in order to properly capture the economic forces involved in the optimal bundling decision, one would need to (i) allow for random effects in the valuation of the add on across potential customers, and (ii) allow

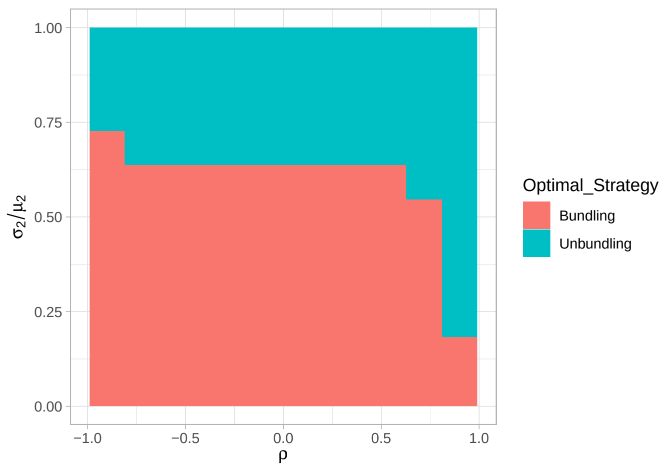


Figure 2: Optimal bundling decision as a function of parameters ρ and σ_2 when other parameters are fixed at $\alpha_0 = 2.5$, $\mu_1 = -2$, $\sigma_1 = 1$, $\mu_2 = 1.5$. As expected, bundling becomes optimal when σ_2 decreases (i.e., when the population values the add on rather homogeneously) or when ρ increases (i.e., when more price sensitive customers have higher relative valuations for the add on).

for correlation between the random effects for add-on value and price sensitivity. I close this section by making two concluding points.

First, as mentioned before, this simulation analysis does not *prove* that the observed direction in the relationship between the optimal bundling decision and ρ or σ_2. Though I have not been able to find parameters that would reverse the direction, its possibility is not ruled out.

The second point pertains to the relative importance of σ_2 versus ρ in optimal bundling decisions. Figure 2 suggests that there is a sense in which σ_2 is more important than ρ. To see this, note that there are values for σ_2 under which the optimal decision is bundling (or unbundling) regardless of what value ρ takes. The converse is not true, however; for any ρ, a large enough σ_2 implies unbundling is optimal and a small enough σ_2 will make bundling the optimal decision. The observation that a small (large) enough σ_2 can always make bundling (unbundling) optimal has been confirmed under all other parameterizations of the model that I have examined. This latter point should, in my view, be considered good news for the empirical analysis of bundling decisions; because it implies that if empirically identifying the co-variation between $\alpha_{2,t}$ and $\alpha_{i,t}$ is not possible, then only capturing and identifying the variation in $\alpha_{2,t}$ may provide a reasonable approximation of the optimal bundling decision.

6 Conclusion And Future Research

In this paper, I completed two tasks. First, under a set of assumptions, I provided (an almost) full characterization for the optimality of pure bundling.[7] I showed that pure bundling is optimal if the optimal quantity sold for the pure bundle (if sold alone) is strictly larger than that for any sub-bundle. Conversely, pure bundling is sub-optimal if there is at least one smaller bundle whose optimal sales volume (if sold alone) strictly surpasses that of the full bundle.

Second, I used the characterization to arrive at an informal interpretation for when bundling is optimal. I argued that to know whether to unbundle products, a firm would need to know (i) the variability –across potential customers– in the complementarity level among the products offered, and (ii) how much this variability correlates with variability in price sensitivity. Finally, I showed the relevance of these interpretations to applied work using a simulation of a random-coefficient discrete choice demand models.

The work in this paper can be extended in multiple directions. First, it would be valuable to investigate alternative assumptions to 1 through 5. In particular, it would be worth studying whether the main takeaways of this paper would hold under alternatives to monotonicity and complementarity. For instance, will bundling be so closely tied to sales volumes if products were substitutes instead of complements? If not, would there be any other notion that would fully characterize optimal bundling of substitute-able products in the same way that sales volumes do for complementary ones? Similar questions apply to the role of monotonicity: if instead of vertical differentiation, we had horizontal differentiation among consumers, would there be a criterion that, under some conditions, fully characterize optimal bundling in the same way that sales volumes do for vertically differentiated consumers?

Another useful direction for future research is one that–to the best of my knowledge–has received little or no attention in the literature in spite of its importance: characterizing necessary and sufficient conditions for bundling decisions in the context of widely used empirical models. Ideally, such work would shed light on how to choose the right *model specification* for studying bundling. To illustrate, the sales volume result in this paper, even if developed using an empirical model, would not be directly useful (same is true of ratio monotonicity results in Haghpanah and Hartline (2019); Anderson and Dana Jr (2009); Salant (1989); Deneckere and Preston McAfee (1996) and elasticity results in Long (1984); Armstrong (2013)). This is because, in order to directly evaluate the sales volume conditions, the econometrician

[7]As mentioned before, by replacing strict quasi-concavity of profits with strict concavity, we can bridge the small gap in the characterization and obtain if-and-only-if.

needs an already-estimated model. But in that case, s/he can directly use the estimated model to obtain the optimal bundling decision. Therefore, a useful theoretical result should guide the empirical research process before, rather than after, the econometrician chooses the empirical specification. I expect this task to be difficult. This is because commonly used econometric models are set up to facilitate estimation rather than facilitate derivation of theoretical results on matters such as optimal bundling. Thus, absent such direct results on empirical models, using the "next best approach" taken by this paper may prove fruitful: (i) obtain full characterization for pure/mixed bundling decisions using a different model; (ii) obtain an economic interpretation for the developed theoretical result(s); (iii) verify the relevance and insightfulness of the interpretation by simulating on the applied model of interest.

References

William James Adams and Janet L Yellen. 1976. Commodity bundling and the burden of monopoly. *The quarterly journal of economics* (1976), 475–498.

Eric T Anderson and James D Dana Jr. 2009. When is price discrimination profitable? *Management Science* 55, 6 (2009), 980–989.

Mark Armstrong. 2013. A more general theory of commodity bundling. *Journal of Economic Theory* 148, 2 (2013), 448–472.

Mark Armstrong. 2016. Nonlinear pricing. *Annual Review of Economics* 8 (2016), 583–614.

Steven Berry, James Levinsohn, and Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* (1995), 841–890.

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

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

Nima Haghpanah and Jason Hartline. 2019. When Is Pure Bundling Optimal? *Penn State University and Northwestern University* (2019).

John B Long. 1984. Comments on" Gaussian Demand and Commodity Bundling". *The Journal of Business* 57, 1 (1984), S235–S246.

R Preston McAfee, John McMillan, and Michael D Whinston. 1989. Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics* 104, 2 (1989), 371–383.

Domenico Menicucci, Sjaak Hurkens, and Doh-Shin Jeon. 2015. On the optimality of pure bundling for a monopolist. *Journal of Mathematical Economics* 60 (2015), 33–42.

Gregory Pavlov. 2011. Optimal mechanism for selling two goods. *The BE Journal of Theoretical Economics* 11, 1 (2011).

Stephen W Salant. 1989. When is inducing self-selection suboptimal for a monopolist? *The Quarterly Journal of Economics* 104, 2 (1989), 391–397.

Richard Schmalensee. 1984. Gaussian demand and commodity bundling. *Journal of business* (1984), S211–S230.

George J Stigler. 1963. United States v. Loew's Inc.: A note on block-booking. *The Supreme Court Review* 1963 (1963), 152–157.

Jidong Zhou. 2017. Competitive bundling. *Econometrica* 85, 1 (2017), 145–172.

Jidong Zhou. 2019. Mixed Bundling in Oligopoly Markets. (2019).

A Proof of Theorem 1

I start by some preliminary remarks, definitions, and lemmas.

Remark 1. *Suppose functions $f_1(x)$, $f_2(x)$ and $f_1(x) + f_2(x)$ are all strictly quasi-concave over the interval $[a, b]$. Then either (i) $\arg\max f_1 \leq \arg\max f_2$ or (ii) $\arg\max f_1 \leq \arg\max(f_1 + f_2)$ will imply:*

$$\arg\max f_1 \leq \arg\max(f_1 + f_2) \leq \arg\max f_2.$$

The proof of this remark is left to the reader.

Definition 3. *For disjoint bundles b and b', denote by $D^*(b|b')$ the "optimal quantity sold" of bundle b if all customers are already endowed with b' but no other bundle is offered by the firm. Formally, $D^*(b|b')$ is defined as $D(b|\{b\}, p^*_{b|b'}, b')$ where $p^*_{b|b'} : \{b\} \to \mathbb{R}$ is effectively one real number, and it is chosen among other possible p so that $\pi(\{b\}, p|b')$ is maximized.*

Next, I show that the problem of finding the optimal price for a bundle is equivalent to the problem of finding the right type t^* and sell to types $t \geq t^*$.

Definition 4. *Define by $t^*(b|b')$ the largest t such that $1 - F(t) \geq D^*(b|b')$. Also, for simplicity, denote $t^*(b|\emptyset)$ by $t^*(b)$.*

Lemma 3. *Consider disjoint bundles b and b'. Suppose that all types are endowed with bundle b', and that the firm is selling only bundle b, optimally choosing $p^*_{b|b'}$. The set of types who will buy the product at this price is the interval $[t^*(b|b'), 1]$.*

Proof of Lemma 3. Follows directly from monotonicity. Monotonicity implies that the optimal sales volume $D^*(b|b')$ would be purchased by the highest types t with t weakly above some cutoff $\hat{t}$. Definition 4 says that for the demand volume to equal $D^*(b|b')$, the cutoff $\hat{t}$ has to equal $t^*(b|b')$. **Q.E.D.**

Lemma 3 is important in that it shows the problem of choosing $p^*_{b|b'}$ can equivalently be thought of as the problem of choosing $t^*_{b|b'}$. This allows us to set up the firm's problem based on t. Next definition introduces a necessary notation for this purpose.

Definition 5. *Consider disjoint bundles b and b'. Suppose that all potential customers have already been endowed with b', and that the firm is to sell only bundle b. By $\pi_b(t|b')$ denote the profit to the firm if it chose a price for bundle b such that all types $t' \geq t$ would purchase bundle b. More formally:*

$$\pi_b(t|b') = \pi(\{b\}, v(b, t|b')|b')$$

Lemma 4. *$\pi_b(t|b')$ is strictly quasi-concave in t.*

Proof of Lemma 4. Suppose $\pi_b(t|b')$ is not quasi-concave in t. This means there are $t_1 < t_2 < t_3$ such that $\pi_b(t_2|b') \leq \min(\pi_b(t_1|b'), \pi_b(t_3|b'))$. Then construct p_1, p_2 and p_3 from t_1, t_2 and t_3 according to the procedure in definition 5. That is, set $p_i = v(b, t|b')$ for each i. Monotonicity puts p_2 strictly between p_1 and p_3. Note that for these prices, we have:

$$\pi(\{b\}, p_2|b') \leq \min(\pi(\{b\}, p_1|b'), \pi(\{b\}, p_3|b'))$$

which violates the quasi-concavity assumption in p. **Q.E.D.**

With the above definitions and lemmas in hand, we are ready to prove the main theorem. I start by the necessity condition (i.e., the condition that $D^*(\bar{b}) \geq D^*(b)$ for all b is necessary for pure bundling to optimal).

Proof of necessity. We want to show that if there is some b such that $D^*(b) > D^*(\bar{b})$, then pure bundling is sub-optimal. Specifically, I show that offering bundles b and $b^C = \bar{b} \setminus b$ would be strictly more profitable to the firm compared to offering $\bar{b}$ alone. The argument follows.

Lemma 5. *$D^*(b) > D^*(\bar{b})$ implies $D^*(b) > D^*(b^C|b)$.*

Proof of Lemma 5. Suppose, on the contrary, that $D^*(b) \leq D^*(b^C|b)$. This means $t^*(b) \geq t^*(b^C|b)$. We know:

$$t^*(b) = \arg\max_t \pi_b(t)$$

and

$$t^*(b^C|b) = \arg\max_t \pi_{b^C}(t|b).$$

Also, given definition 4, it is straightforward to verify that:

$$\pi_{\bar{b}}(t) \equiv \pi_{b^C}(t|b) + \pi_b(t)$$

By strict quasi-concavity of all profits in t and by remark 1, it has to be that the argmax of $\pi_{\bar{b}}(t)$ falls in between the argmax values $t^*(b^C|b)$ and $t^*(b)$. Therefore, we get: $t^*(\bar{b}) \leq t^*(b)$, which implies $D^*(b) \leq D^*(\bar{b})$, contradicting a premise of the lemma. **Q.E.D.**

Lemma 6. *Selling $D^*(b^C|b)$ units of bundle b^C along with $D^*(b)$ units of bundle b would be strictly more profitable to the firm compared to selling $D^*(\bar{b})$ units of the full bundle alone.*

Proof of Lemma 6. In order to complete this proof, I first introduce a modified problem for the firm.

Modified Firm Problem: Suppose the firm is to choose the optimal set B^* of bundles and optimal prices p^* under the following conditions:

- The set B^* can only be constructed from members of $\{\emptyset, b, b^C, \bar{b}\}$.
- The valuation function is $\tilde{v}$ rather than v. The function $\tilde{v}$ over the set $\{\emptyset, b, b^C, \bar{b}\}$ is defined by:

$$\forall t : \begin{cases} \tilde{v}(\emptyset, t) = v(\emptyset, t) = 0, \\ \tilde{v}(b, t) = v(b, t) \\ \tilde{v}(b^C, t) = v(b^C, t|b) \\ \tilde{v}(\bar{b}, t) = v(\bar{b}, t) = \tilde{v}(b, t) + \tilde{v}(b^C, t) \end{cases}$$

The only bundle for which $\tilde{v}$ deviates from v is b^C. By this construction, there is no complementarity or substitution between b and b^C under $\tilde{v}$. Also note that $\tilde{v}$ is always greater than or equal to v. Finally note that $\tilde{v}$ inherits monotonicity and quasi-concavity.

Denote the profit and demand functions under the modified problem by $\tilde{\pi}(\cdot)$ and $\tilde{D}(\cdot)$ respectively.

The rest of the proof of the necessity conditions of Theorem 1 is organized as follows. I first make a series of claims (without proving them) about the optimal solution to the modified problem and its relationship with the optimal solution to the original problem. Then I use these claims to prove the necessity conditions of the theorem. Finally, I go back providing the proofs to these claims.

Claim 1. *Consider the modified problem. Denote $B^1 = \{b, b^C\}$. Also denote by $\tilde{p}^{1,*}$ the optimal pricing strategy given B^1 under the modified problem. Similarly, construct $B^2 = \{\bar{b}\}$ and $\tilde{p}^{2,*}$. Then, the following is true:*

$$\tilde{\pi}(B^1, \tilde{p}^{1,*}) > \tilde{\pi}(B^2, \tilde{p}^{2,*})$$

Claim 2. *Consider the original problem (i.e., under value function v). Construct the set $B^1 = \{b, b^C\}$, the same way as in the previous claim. Also denote by $p^{1,*}$ the optimal pricing strategy given B^1 under the original problem. Then one can show that $p^{1,*} = \tilde{p}^{1,*}$ and:*

$$\pi(B^1, p^{1,*}) = \tilde{\pi}(B^1, \tilde{p}^{1,*})$$

Claim 3. *Consider the original problem (i.e., under value function v). Set $B^2 = \{\bar{b}\}$, the same way as in claim 1. Also denote by $p^{2,*}$ the optimal pricing strategy given B^2 under the original problem. Then one can show that $p^{2,*} = \tilde{p}^{2,*}$ and:*

$$\pi(B^2, p^{2,*}) = \tilde{\pi}(B^2, \tilde{p}^{2,*})$$

Together, claims 1 through 3 yield:

$$\pi(B^2, p^{2,*}) < \pi(B^1, p^{1,*})$$

This completes the proof of the necessity conditions of the theorem, provided that claims 1 through 3 are correct. That is, it is optimal for the firm to offer $B^1, p^{*,1}$, which will lead types $t^*(b^C|b)$ and above to buy both of the bundles and form $\bar{b}$, and types in the interval $[t^*(b), t^*(b^C|b))$ to buy only b.

Next, I show claims 1 through 3 indeed hold.

Proof of Claim 1. Recall that under valuations $\tilde{v}$, the two products are independent of each other. Therefore, for bundling strategy B^1, the firm will choose the optimal prices for b and b^C separately. This will lead to selling $D^*(b^C|b)$ units of b^C and $D^*(b)$ units of b under optimal pricing.[8]

Also recall that under $\tilde{v}(\bar{b}, t) = v(\bar{b}, t)$ for all t. Therefore, under the modified problem and under B^2, the optimal price $p^{2,*}$ will be one that leads to exactly $D^*(\bar{b})$ units sold.

Note that given monotonicity and the independence feature of $\tilde{v}$, the firm could replicate using B^1 any strategy that it can implement with B^2. In particular, the firm could replicate the profit from $(B^2, p^{2,*})$ using B^1 by setting prices $p(b)$ and $p(b^C)$ such that each product sells exactly $D^*(\bar{b})$ unit of b and $D^*(\bar{b})$ units of b^C. This will yield exactly the profit of $\tilde{\pi}(B^2, \tilde{p}^{2,*})$. But we know at least one of these quantities sold is sub-optimal. This is because, by lemma 5, we have $D^*(b^C|b) < D^*(b)$. Therefore, by selling $D^*(\bar{b})$ units for b and b^C, at least one of the quantities will be strictly sub-optimal.. This finishes the proof of this claim. **Q.E.D.**

Proof of Claim 2. To see why this claim is true, consider bundling strategy B^1 under the original problem. Assume the firm sets $p^{1,*}$ to be equal to $\tilde{p}^{1,*}$. It is straightforward to verify that the demand volumes for b and b^C in these conditions will be exactly equal to those under the modified problem when the firm strategy is $(B^1, \tilde{p}^{1,*})$. Therefore, the firm can achieve $\tilde{\pi}(B^1, \tilde{p}^{1,*})$ under the original problem.

Next, I show that the firm cannot achieve more than $\tilde{\pi}(B^1, \tilde{p}^{1,*})$ under the original problem by choosing other values for $p(b)$ and $p(b^C)$. To see this, consider two cases regarding the firm's pricing strategy:

Case 1. If the firm sets $p(b)$ and $p(b^C)$ in a way that $\tilde{D}(b) \geq \tilde{D}(b^C)$, the firm will get the exact same demand volumes for the two bundles under the original problem as it would

[8] Note that independent values of b and b^C under $\tilde{v}$ would not generally imply that the firm would optimally unbundle and set the optimal prices for b and b^C separately. An clear example of when the firm strictly prefers pure bundling over selling b and b^C separately is Adams and Yellen (1976). Nevertheless, under monotonicity, one can indeed show that independence of values leads to optimality of unbundling and independence of optimal prices from each other. The proof of this is straightforward and left to the reader.

under the modified one. Hence, the firm will also get the exact same profits under the two problems with such prices: $\pi(B^1, p) = \tilde{\pi}(B^1, p)$.

Case 2. If, however, the firm sets $p(b)$ and $p(b^C)$ such that $\tilde{D}(b) < \tilde{D}(b^C)$, then by the construction of $\tilde{v}$ from v and by the complementarity property of v, we have $D(b^C) \leq \tilde{D}(b^C)$. This, in turn, due to complementarity, will lead to $D(b) \leq \tilde{D}(b)$. Under these conditions, any pricing strategy such that $p(b) \geq \Sigma_{i \in b} c_i$ and $p(b^C) \geq \Sigma_{i \in b^C} c_i$ will lead to $\pi(B^1, p) \leq \tilde{\pi}(B^1, p)$.

Therefore, for all pricing strategies with non-negative profit (which include all the candidates for $p^{1,*}$) we have $\pi(B^1, p) \leq \tilde{\pi}(B^1, p)$. This, combined with the fact that $\tilde{p}^{1,*}$ delivers the exact same profit under v as it does under $\tilde{v}$ (where it is the unique optimum,) implies that $\tilde{p}^{1,*}$ also uniquely maximizes $\pi(B^1, p)$ over different possible p. **Q.E.D.**

Proof of Claim 3. Note that $v(\bar{b}, t) = \tilde{v}(\bar{b}, t)$ for all t. Therefore, optimizing the price of $\bar{b}$ under v and $\tilde{v}$ is identical, implying this claim. **Q.E.D.**

Given the proofs of the claims, the proof of the if side of the theorem is now complete. **Q.E.D.**

Next, I turn to the proof of the sufficiency conditions (i.e., that $D^*(\bar{b}) > \max_{b \in \mathcal{B} \setminus \{\bar{b}\}} D^*(b)$ implies that pure bundling is optimal).

Proof of sufficiency. I start with some lemmas.

Proof of Lemma 2 from the main text. Suppose, on the contrary, that $\beta(t|B, p) \subsetneq \beta(t'|B, p)$. for some $t' < t$. That is:

$$\tilde{\beta} \triangleq \beta(t|B, p) \setminus \beta(t'|B, p) \neq \emptyset$$

I demonstrate that we can arrive at a contradiction by showing that type t', when endowed with $\beta(t', B, p)$, would have the incentive to buy $\tilde{\beta}$. Formally, I show:

$$v(\tilde{\beta}, t'|\beta(t'|B, p)) \geq \Sigma_{i \in \tilde{\beta}} p(i) \tag{7}$$

To see this, first note that by construction, type t, conditional on being endowed with $\beta(t'|B, p)$, would find it optimal to purchase $\tilde{\beta}$ in order to obtain $\beta(t|B, p)$. Formally:

$$v\Big(\tilde{\beta}, t|\beta(t'|B, p)\Big) \geq \Sigma_{i \in \tilde{\beta}} p(i) \tag{8}$$

By monotonicity and $t' > t$, we get:

$$v\Big(\tilde{\beta}, t'|\beta(t'|B, p)\Big) \geq v\Big(\tilde{\beta}, t|\beta(t'|B, p)\Big) \tag{9}$$

Together, inequalities 8 and 9 imply inequality 7, completing the proof of the lemma. **Q.E.D.**

Lemma 7. *Under assumptions 1 through 4, and under firm optimal strategy* (B^*, p^*)*, there is a customer* t *such that* $\beta(t|B^*, p^*) = \bar{b}$*.*

Proof of Lemma 7. Assume, on the contrary that no type t purchases $\bar{b}$ under the optimal firm behavior. I show we can reach a contradiction. In paricular, I show that type $t = 1$ not purchasing $\bar{b}$ leads to a contradiction.

Assume $\cup_{t\in[0,1]}\beta(t|B^*, p^*) \neq \bar{b}$. That is, if we denote $\cup_{t\in[0,1]}\beta(t|B^*, p^*)$ by b, then $b^C \neq \emptyset$. One can show that:

$$v(b^c, 1) > \Sigma_{i\in b^C} c_i \tag{10}$$

To see why (10) is true, note that by complementarity:

$$v(b^c, 1) \geq \Sigma_{i\in b^C} v(\{i\}, 1) \tag{11}$$

Also, by the fact that for each i there is some t with $v(\{i\}, t) > c_i$, and by monotonicity, we have

$$\Sigma_{i\in b^C} v(\{i\}, 1) > \Sigma_{i\in b^C} c_i \tag{12}$$

Together, inequalities 11 and 12 imply inequality 10.

Given 10, and given that we are assuming no customer is buying any product within b^C, the firm can (i) drop from B^* any bundle that includes any element of b^C, and (ii) then introduce b^C at the price of $v(b^C, 1)$. This move will lead at least type 1 to purchase the bundle, which is profitable to the firm. Also, this move will not hurt the profit of the firm by leading customers to not purchase bundles they bought before the introduction of b^C. This is because, for any type t, there are two cases. Case 1- type t will not buy newly introduced b^C: in this case her preferences over other bundles, and hence her purchase decisions on other bundles remain unchanged. Case 2- type t does buy b^C: in this case, by complementarity, the valuations by t of all of the other product t has bought increases, which means t will still buy those other products.

Therefore, we showed that if no type purchases $\bar{b}$ under (B^*, p^*), there will be a contradiction. This completes the proof of the lemma. **Q.E.D.**

In light of lemma 7, the following two corollaries of lemma 2 are useful.

Corollary 1. *Under any (B, p), the set of types to for which $\beta(t|B, p) = \bar{b}$ takes the form of $[t_1, 1]$ for some $t_1 < 1$.*

Corollary 2. *Under any (B, p), the set of types to for which $\beta(t|B, p) = \emptyset$ takes the form of $[0, t_2)$ for some $t_2 < 1$.*

With these lemmas in hand, I next turn to the proof of the sufficiency conditions. The strategy is, again, contrapositive.

Assume on the contrary that we have, at the same time: (i) $\forall b : D^*(b) \leq D^*(\bar{b})$ and (ii) the firm's optimal strategy does not involve pure bundling. This latter statement implies that the set of all distinct bundles chosen by customers under (B^*, t^*) includes members other than $\emptyset$ or $\bar{b}$. Formally, if we denote

$$\beta^* = \{b | \exists t : \beta(t|B^*, p^*) = b\}$$

then $\beta^* \setminus \{\emptyset, \bar{b}\} \neq \emptyset$. In other words, our contrapositive assumption implies that t_1 in corollary 1 is stictly larger than t_2 in corollary 2.

Then, note that by corollary 1 and the continuity assumption, there is some bundle $b_1 \in \beta^* \setminus \{\emptyset, \bar{b}\}$ such that for t'_1 close enough to but smaller than t_1, we have:

$$\forall t \in [t'_1, t_1) : \beta(t|B^*, p^*) = b_1 \tag{13}$$

Also, by corollary 2 and the continuity assumption, there is some bundle $b_2 \in \beta^* \setminus \{\emptyset, \bar{b}\}$ such that for t'_2 close enough to but larger than t_2, we have:

$$\forall t \in [t_2, t'_2] : \beta(t|B^*, p^*) = b_2 \tag{14}$$

The rest of the proof of the sufficiency conditions of the theorem is organized as follows. I first make a series of claims (without proving them). Next I use the claims to prove the sufficiency conditions of the theorem. Finally, I will return to the proofs of the claims.

Claim 4. $t^*(b_1^C | b_1) = t_1$.

In words, claim 4 says that the set of customers who purchase the full bundle $\beta(t|B^*, p^*) = \bar{b}$ under the firm optimal strategy (B^*, p^*) is the same as those who purchase b_1^C and construct the full bundle if (i) everyone is endowed with b_1 and (ii) the firm offers only b_1^C, pricing it optimally.

Claim 5. $t^*(b_2) = t_2$.

Claim 5 says that the set of customers who purchase b_2 under the firm optimal strategy (B^*, p^*) is the same as those who purchase b_2 if the firm offers only b_2 and prices it optimally.

Next, note that the assumption $D^*(\bar{b}) > D^*(b_2)$, combined with monotonicity and claim 5, implies $t^*(\bar{b}) \leq t_2$. By $t_1 > t_2$, we get $t^*(\bar{b}) < t_1 = t^*(b_1^C|b_1)$. Also note that:

$$\forall t : \pi_{\bar{b}}(t) = \pi_{b_1^C}(t|b_1) + \pi_{b_1}(t)$$

As such, by strict quasi-concavity of profits, by $t^*(\bar{b}) < t^*(b_1^C|b_1)$, and by remark 1, the peak of $\pi_{\bar{b}}(t)$ should happen in between those of $\pi_{b_1^C}(t|b_1)$ and $\pi_{b_1}(t)$. Therefore, we should have: $t^*(b_1) \leq t^*(\bar{b}) \leq t^*(b_1^C|b_1)$. But $t^*(b_1) \leq t^*(\bar{b})$ implies:

$$D^*(b_1) \geq D^*(\bar{b})$$

which is a contradiction. Therefore, the sufficiency part of the theorem is true provided that claims 4 and 5 are true. I now turn to the proofs of these claims.

Proof of Claim 4. Suppose on the contrary that $t^*(b_1^C|b_1) \neq t_1$. In that case, it can be shown that the firm can strictly improve its profit by slightly adjusting the price of b_1^C. That is, there is a pricing strategy p with $p(b) = p^*(b)$ for all $b \neq b_1^C$ such that $\pi(B^*, p) > \pi(B^*, p^*)$.

To see why this is the case, construct bundling strategy B' in the following way:

$$B' = \{b_1, b_1^C\} \tag{15}$$

Also construct pricing strategy p' by fixing $p'(b_1) = \min_t v(b_1)$ but keeping $p'(b_1^C)$ adjustable.

Now note that as long as $\rho \in [p^*(b_1^C) - \epsilon, p^*(b_1^C) + \epsilon]$ for a small enough ϵ, then $\pi(B^*, p)$ and $\pi(B', p')$ move *in parallel* if we set $p(b_1^C) = p'(b_1^C) = \rho$ and move ρ. The range parameter ϵ should be chosen so that for any pricing strategy p constructed with a ρ in this interval we have: $D(b_1^C|B^*, p) < 1 - F(t_1')$ where t_1' was constructed in equation 13. In other words, ϵ should be small enough so that every type t in this interval purchases a (weak) super-set of b_1 under the optimal strategy.

Profits move in parallel because a small price change for bundle b_1^C only changes the purchase decisions of those types t who are sufficiently close to t_1. All such customers have decided to purchase b_1 under (B^*, p^*). Therefore, under our constructed (B^*, p), these types' valuations of b_1^C will exactly be given by $v(b_1^C, t|b_1) - p(b_1^C)$ which is exactly how these types would value it under (B', p'). Also, when some of these types drop b_1^C in response to a change in ρ, they will *not* drop any subset of b_1 alongside it. This is because, even though complementarities exist, these types t are all larger enough than t_1' so that by monotonicity they value all components of b_1 in B^* above the collective price charged by p^* for b_1. To sum up, a small enough change in ρ as part of pricing strategies p and p' will lead to the exact same reaction by customers and, hence, the exact same change in profits. Therefore, the optimal value of ρ in this interval is the same under (B^*, p) as it is under B', p' (by strict quasi-concavity, we know that this optimal ρ is unique in both cases). This common optimal value for ρ leads to the exact same demand for b_1^C under (B^*, p) as it does under (B', p'). The optimal demand under (B^*, p) is achieved by choosing ρ to equate p with p^*, which by construction leads to all t with $t \geq t_1$ buying. The optimal ρ under (B', p'), by definition, should lead to all $t \geq t^*(b_1^C|b_1)$ buying. Therefore, if $t_1 \neq t^*(b_1^C|b_1)$, then one can modify (B^*, p^*) by slightly changing $p^*(b_1^C)$ and improve the profit, a contradiction. **Q.E.D.**

Proof of Claim 5. The proof of this claim is fairly similar to that of the previous claim. We start by assuming, on the contrary, that $t^*(b_2) \neq t_2$ and reach a contradiction. Construct (B', p') by assuming $B' = \{b_2\}$, which makes p' just one number (for the price of b_2). Similar to the previous claim, one can show that for prices ρ for b_2 sufficiently close to $p^*(b_2)$ the two profit functions $\pi(B^*, p)$ and $\pi(B', p')$ move in parallel as we move ρ. Again, similarly to the previous claim, this implies that (B^*, p^*) can be improved upon if $t_2 \neq t^*(b_2)$. **Q.E.D.**

The completion of the proofs for claims 4 and 5 finishes the proof of the sufficiency side of the theorem, and hence the theorem itself. **Q.E.D.**

B Proofs of Other Results

Proof of Proposition 1. I prove the statement outside of parentheses: *If* $\frac{v(b,t)}{v(\bar{b},t)}$ *is decreasing in* $v(\bar{b}, t)$ *for bundle* b*, then* $D^*(b) \leq D^*(\bar{b})$. The version inside parentheses can be proven in a similar way.

Assume on the contrary that $\frac{v(b,t)}{v(\bar{b},t)}$ is increasing but $D^*(b) > D^*(\bar{b})$. Denote by $\pi^*(\bar{b})$ the amount of profit the firm obtains by selling only $\bar{b}$ and optimally pricing it, which yields the demand level $D^*(\bar{b})$. Given that production costs are assumed zero, we have:

$$\pi^*(\bar{b}) = v(\bar{b}, t^*(\bar{b})) \times D^*(\bar{b}) \tag{16}$$

Using a similar notation for b, we get:

$$\pi^*(b) = v(b, t^*(b)) \times D^*(b) \tag{17}$$

By our assumption that $D^*(b) > D^*(\bar{b})$, we get $t^*(b) < t^*(\bar{b})$. Also, by $\pi^*(\bar{b})$ and $\pi^*(b)$ being the profits from optimal decisions, and by quasi-concavity, we know that the firm's profit would be strictly lower than $\pi^*(\bar{b})$ if it were to sell $D^*(b)$ units of $\bar{b}$ instead of $D^*(\bar{b})$ units. Likewise, its profit would fall strictly below $\pi^*(b)$ if it were to sell $D^*(\bar{b})$ units of b instead of $D^*(b)$ units. Formally:

$$\pi^*(\bar{b}) > v(\bar{b}, t^*(b)) \times D^*(b) \tag{18}$$

and:

$$\pi^*(b) > v(b, t^*(\bar{b})) \times D^*(\bar{b}) \tag{19}$$

Replacing from 16 and 18, and also 17 and 19, we get:

$$v(\bar{b}, t^*(\bar{b})) \times D^*(\bar{b}) > v(\bar{b}, t^*(b)) \times D^*(b) \tag{20}$$

and:

$$v(b, t^*(b)) \times D^*(b) > v(b, t^*(\bar{b})) \times D^*(\bar{b}) \tag{21}$$

Multiplying the left-hand-side terms of inequalities 20 by each other and doing the same for the right-hand-side terms, then removing the terms that cancel out and rearranging, one can obtain:

$$\frac{v(b, t^*(b))}{v(\bar{b}, t^*(b))} > \frac{v(b, t^*(\bar{b}))}{v(\bar{b}, t^*(\bar{b}))} \tag{22}$$

But inequality 22, combined with $t^*(b) < t^*(\bar{b})$ and monotonicity, violates the premise of the proposition, a contradiction. **Q.E.D.**