A Characterization for Optimal Bundling of Products with Interdependent Values

Soheil Ghili*
Yale University
June 3, 2021

Abstract

This paper studies optimal bundling of products with interdependent values. I show that, under some conditions, a monopolist optimally chooses to sell only the grand bundle of a given set of products if and only if the optimal sales volume for the grand 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. I also discuss several implications of the main result. In particular, I detail how it relates to “ratio monotonicity” results on bundling, and describe how it can provide a full characterization of the optimal tariff in nonlinear pricing problems.

1 Introduction

This paper theoretically studies optimal bundling decisions by a monopolist, and carries out an analysis with two main features. First, unlike most studies in the literature which assume independent values, I allow for interdependence: a consumer’s valuation for a given product can depend on whether s/he has purchased other products as well. Second, I seek to obtain a full characterization of when “pure” bundling (i.e., the act of selling only the package of

*E-mail: soheil.ghili@yale.edu. Click here for the most current version. I thank Nima Haghpanah, Johannes Horner, Peter Klibanoff, Barry Nalebuff, Larry Samuelson, Kai Hao Yang, Jidong Zhou, and seminar participants for helpful comments. All errors are mine.
all available products together as one bundle) is optimal. Under some assumptions, most notably weak complementarity across products and a form of vertical differentiation among consumers, I prove that optimal bundling admits a simple characterization: Pure bundling is optimal if the optimal sales volume for the grand 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 grand bundle, then pure bundling is sub-optimal. In simpler terms, bundling is optimal if it helps sell more.

I then proceed with multiple discussions regarding the implications of this characterization. First, I discuss the relation to “ratio-monotonicity” results from the bundling literature (see Haghpanah and Hartline (2019); Anderson and Dana Jr (2009) among others). Ratio monotonicity roughly states that pure bundling is optimal if the ratio between the value of any bundle and that of the grand bundle is everywhere increasing in the value of the grand bundle. Conversely, pure bundling is strictly sub-optimal if this ratio is everywhere decreasing for at least one bundle. These global requirements imposed by ratio monotonicity results are the reason why they do not provide a full characterization of when pure bundling is optimal. I show that my sales volumes result can be reinterpreted as requiring ratio monotonicity to hold only locally.

Next, I turn to the economic interpretation of the main characterization. I argue that this result has two intuitive takeaways. 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 unbundling 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. See Long (1984); Armstrong (2013) and especially Haghpanah and Hartline (2019)).

I also explore the implications of this bundling result for the traditional screening problem.

---

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 grand 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 pure 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 Of course this is made possible by some assumptions about which the paper is clear.
In a setting akin to the nonlinear pricing environment in Maskin and Riley (1984), I show that the optimal price schedule can be characterized by taking the following steps: Find the batch size $q$ such that if the monopolist were to sell only $q$-batches at a single (optimally chosen) price, it would generate the highest demand volume compared to other batch sizes. Then the optimal tariff will involve bundling of the first $q$ units of the product, by charging a flat fee for purchasing any quantity weakly smaller than $q$. The second batch size can be found in a similar way (i.e., by maximizing the sales volume) with a simple modification: demand functions should now be calculated conditional on each consumer already owning $q$ units of the product. Continuing this recursive procedure will fully characterize the optimal tariff. In this analysis, I provide two further discussions. First, I explain why in the nonlinear pricing setting we get a full characterization of the entire tariff whereas in the bundling setting we can only characterize when pure bundling is optimal. Second, I show that our economic intuitions for when bundling is optimal are also helpful for understanding when “bunching” is optimal in non-linear tariff design.

I finish the analysis in this paper by posing an open question. I argue that theoretical characterizations of optimal bundling strategies in the context of widely used empirical models would have substantial value. I demonstrate using simulation that the economic interpretations proposed in the literature and confirmed and expanded in this paper (i.e., variation in complementarity and co-variation between complementarity and price sensitivity) are indeed useful for deciding the right empirical specification for applied work on bundling. I observe, however, that neither the results from this paper nor other results from the literature are directly applicable to random-coefficient discrete-choice models of demand (Berry et al., 1995) which are perhaps the most widely used in applied work. I argue, as a result, that obtaining theoretical characterizations in such contexts would be a fruitful path for future theoretical research on bundling.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 sets up the model and formally presents the assumptions and the main result. Section 4 discusses the assumptions and delves into the implications and interpretations of the main result. Section 5 overviews the relevance of the interpretation for applied work and poses an open question. Section 6 concludes.
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); Manelli and Vincent (2007); Palfrey (1983) 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.\(^3\) I also contribute to this sub-literature in some 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, my work highlights some ties between bundling of products with interdependent values on the one hand and the nonlinear pricing problem on the other. Finally, 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.

\(^3\)Note that necessary and sufficient conditions for pure bundling of products with independent values do exist (see Daskalakis et al. (2017) for instance.) But this is not the case for interdependent values.
3 Main Result

3.1 Setup and Notations

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.\(^4\) By \( \bar{b} \) denote the grand bundle \( \{1, ..., n\} \). Also for any bundle \( b \), denote \( b^C = \bar{b} \setminus b \). There is a unit mass of customers whose types are represented by \( t \) with probability distribution \( f(\cdot) > 0 \).\(^5\) The willingness to pay by type \( t \) for bundle \( b \) is denoted \( v(b, t) \). Without loss, assume \( v(\emptyset, t) = 0 \). Also, for all \( b \), suppose that \( v(b, t) \) is continuous in \( t \) except, possibly, for finitely many points.

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 \( 2^{2n-1} \) 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 \( p(\cdot): B \rightarrow \mathbb{R} \) for the bundles offered.\(^6\) 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:

\[
\beta(t|B, p) = \arg \max_{\beta \subseteq B} v(\bigcup_{b \in \beta} b, t) - \sum_{b \in \beta} p(b)
\]

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 \( b \):

\[
D(b|B, p) = \int t f(t) dt
\]

\(^4\)My notation, in part, follows Haghpanah and Hartline (2019).
\(^5\)The main result should hold without strictly positive \( f \) as well. But I expect the proof to be less clean.
\(^6\)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.
Firm profit under bundling strategy $B$ and pricing strategy $p$ is given by:

\[
\pi(B, p) = \int_t \sum_{b \in \beta(t \mid B, p)} \left( p(b) - \sum_{i \in b} c_i \right) f(t) dt
\] (3)

where $c_i$ is the (constant) marginal cost of producing a unit of good $i$. Costs $c_i$ can be negative, allowing for “damaged-good” settings a la [Deneckere and Preston McAfee (1996)].

In order to avoid a situation in which the problem is trivial, assume that for each $i$ there is a non-measure-zero set of types $t$ with $v(\{i\}, t) \geq c_i$, with the inequality being strict for at least one product $i$.

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

\[
(B^*, p^*) = \arg \max_{B \in \mathcal{B}, p \in \mathcal{P}} \pi(B, p)
\] (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 $v(b, t \mid b')$ the valuation by type $t$ for $b$ conditional on possessing $b'$. Formally:

\[ v(b, t \mid b') \equiv v(b \cup b', t) - v(b', t) \]

In a similar manner, denote $\beta(t \mid B, p, b') = \arg \max_{\beta \subseteq (B \setminus \{b'\})} v(\cup_{b \in \beta} b, t \mid b') - \sum_{b \in \beta} p(b)$. Also denote $D(b \mid B, p, b') = \int_t 1_{b \in \beta(t \mid B, p, b')} f(t) dt$. Moreover, denote

\[
\pi(B, p \mid b') = \int_t \sum_{b \in \beta(t \mid B, p, b')} \left( p(b) - \sum_{i \in \beta} c_i \right) f(t) dt.
\]

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

### 3.2 Assumptions and Characterization

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.

---

7Note that this is not really an assumption. If a product is so expensive to produce that it cannot profitably be sold to any consumer, then we can just remove that product from our analysis, starting with $n-1$ products instead of $n$. 

---

6
Assumption 1. **Monotonicity:** Consider $t$ and $t'$ such that $v(\bar{b}, t) \geq v(\bar{b}, t')$. Then for all $b \neq \emptyset$, we have $v(b, t) \geq v(b, t')$ and $v(b, t|b^C) \geq v(b, t'|b^C)$. If the first inequality is strict, then so are the latter two.

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

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

Assumption 3. **Quasi-concavity:** For any $b \in B$, profit functions $\pi(b, p)$ and $\pi(b, p|b^C)$ are strictly quasiconcave in $p(b)$ for all values of $p(b)$ that yield strictly positive demand for $b$.

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 $D(b\{\emptyset\}, p_b^*)$ 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 grand bundle $B = \{\bar{b}\}$. With this definition, we are now ready to state the main result.

**Theorem 1.** Under assumptions 1 through 3, the optimal strategy $(B^*, p^*)$ involves pure bundling if:

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

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

---

8This simply means the profit peaks only once as we vary each price.
\[ D^*(\bar{b}) < \max_{b \in B \setminus \{\bar{b}\}} D^*(b) \]

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 \( D^*(\bar{b}) = \max_{b \in B \setminus \{\bar{b}\}} D^*(b) \). 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 differentiable at peak. Even though this would yield a full characterization, I decided that the ability to speak to the “measure-zero” case of \( D^*(\bar{b}) = \max_{b \in B \setminus \{\bar{b}\}} D^*(b) \) 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 grand 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.

4 Discussion

4.1 Discussion of the assumptions

It is worth further discussing the economic content of the underlying assumptions of the model.

**Monotonicity:** 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 \( \tau \) 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. \( \tau \) is a sufficient statistic: Once \( \tau(t) \) is known, one can fully pin down all \( v(b, t) \) and \( v(b, t|b^C) \) 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 \( t \in [0,1] \).

Another implication of the model assumptions is that it is optimal for the firm to sell \( \bar{b} \) to at least some customers. Lemma \ref{lemma:2} formalizes this idea. It is proved in the appendix. Both monotonicity and complementarity used in the proof.

**Lemma 2.** Under assumptions \([\text{I} \text{ through } \text{I} \text{I}]\) and under firm optimal strategy \((B^*, p^*)\), there is a customer \( t \) such that \( \beta(t|B^*, p^*) = \bar{b} \).

Next, it is useful to observe that the monotonicity assumption imposes a vertical relationship not only on consumers’ preferences, but also on their purchase behaviors.

**Lemma 3.** Consider bundling strategy \( B \) and pricing strategy \( p \). Consider types \( t \) and \( t' \) such that \( \beta(t|B, p) \neq \beta(t'|B, p) \). Then the following statements hold:

1. If \( \beta(t'|B, p) = \emptyset \), we have \( t' < t \).
2. If \( \beta(t'|B, p) = \bar{b} \), we have \( t' > t \).

This lemma says that if there are types who buy \( \bar{b} \), they are the highest types. Also if there are types who buy nothing, they are the lowest types.

It is worth noting that monotonicity is quite common in the literature. A prevalent version of it in the screening literature is the single crossing condition imposed by Maskin and Riley (1984).\(^9\) Within the bundling literature most of the papers that focus on products with interdependent values focus on versions of the product-line pricing—which can be thought of as a special case of the bundling problem—problem and each impose a form of monotonicity (e.g., Anderson and Dana Jr (2009); Deneckere and Preston McAfee (1996); Long (1984)). This usually comes in the form of assumed increasing difference of values in the (unidimensional) type and the product quality level. Also the seminal paper by Mussa and Rosen (1978) on product line pricing assumes the valuation by each type of each quality level is proportional to both type and quality, which implies monotonicity.\(^10\) To my

---

\(^9\)Single crossing in Maskin and Riley (1984) is in fact stronger than my monotonicity condition. In order for monotonicity to be as strong as the single crossing condition in Maskin and Riley (1984), it has to be that \( v(\bar{b}, t') \geq (>) v(\bar{b}, t) \Rightarrow \forall b \cap b' = \emptyset : v(b, t'|b') \geq (>) v(b, t|b') \).

\(^10\)It is worth noting that unlike other papers on product line pricing such as Anderson and Dana Jr (2009), the Mussa and Rosen (1978) study does not focus on whether and when bundling is optimal. Mussa and Rosen (1978) make a series of assumptions that imply pure bundling (i.e., offering only the highest quality version) is always sub-optimal. One can indeed show that an appropriate translation of the Mussa and Rosen (1978) problem into the setting of this paper will satisfy the optimal sales volume conditions for sub-optimality of pure bundling. See end of section 4.2.4 for more details.
knowledge, the only study of bundling of products with interdependent values that does not impose a version of monotonicity is Haghpanah and Hartline (2019). That said, the ratio-monotonicity conditions in Haghpanah and Hartline (2019) –at least in one direction– have a similar implication to my monotonicity condition.

To be clear, there are many papers in the bundling literature that do not impose a version of monotonicity as a model assumption or as part of the conditions in their theorems. To the best of my knowledge, however, all such papers study environments with independent values, meaning they impose \( \forall b, t : v(b, t) = \sum_{i \in b} v(\{i\}, t) \).

Finally, note that the monotonicity assumption does not force consumers to rank the products the same way. That is, it does not rule out \( v(b, t) > v(b', t) \) co-existing with \( v(b, t') < v(b', t') \). It, rather, rules out \( v(b, t) > v(b', t') \) co-existing with \( v(b', t) < v(b', t') \).

**Complementarity**: The vast majority of the literature on bundling assumes that values are independent across products. Complementarity (i.e., super-additive values) is a relaxation of that. As mentioned before, the majority of the papers studying bundling with interdependent values focus on product line pricing. This setting usually consists of a “base product” and an “add-on” where the add-on is only valuable if consumed with the base product. This is a version of complementarity that is substantially stronger than what is imposed in this paper\(^{11}\). Again, to my knowledge, Haghpanah and Hartline (2019) is the only relevant paper that does not impose a version of complementarity.

**Quasi-concavity**: Quasi-concavity simply requires that each relevant profit function be single-peaked. This assumption has been made in the literature before (e.g., see assumption 5 in Maskin and Riley (1984)). Without this assumption, one can still prove a version of Theorem 1 but that version would be weaker and less straightforward to state. Finally, it is worth specifying what this assumption would look like if expressed based on the model primitives rather than profit functions. This assumption would require that \( \frac{\partial \log(v(b, t) - \sum_{i \in b} c_i)}{\partial t} - \frac{f(t)}{1 - F(t)} \) cross zero only once from below for all \( b \neq \emptyset \).

**Other notes**: Most papers on optimal bundling of products with interdependent values assume there are only two products whereas this paper (along with Haghpanah and Hartline (2019)) examines an environment with arbitrarily many products. Additionally, it is worth noting that the second part of Theorem 1 (i.e., sufficient conditions for sub-optimality of pure bundling) holds even under non-linear production cost functions. Moreover, the first part of the theorem will also hold under non-linear cost functions if we make some additional

---

\(^{11}\)One can show that a product line with more than two quality levels (e.g., Mussa and Rosen (1978)) is also a special case of complementarity.
assumptions (either setting $n = 2$ or an appropriate strengthening of monotonicity).\footnote{Proofs will be available upon request. Also note that by non-liner cost functions, I mean economies or dis-economies of scale, meaning the marginal cost of producing product $i$ depends on how many units of $i$ are sold. I do not mean economies or dis-economies of scope. That is, the marginal cost of producing product $i$ cannot depend on how many units of other products $j \neq i$ are sold.}

In sum, this paper’s assumptions do not deviate substantially from the literature on bundling of products with interdependent values. In fact in many respects, these assumptions are weaker than those in the literature. The only exception I am aware of is \cite{Haghpanah and Hartline (2019)} who impose weaker restrictions than this paper does. On the other hand, this paper provides a stronger result by delivering full characterization of optimal bundling. In this sense, I see this paper as complementary to the literature.

\section*{4.2 Further Discussion of the Main Result}

This subsection delves deeper in the analysis of the main result. In particular, this subsection (i) explains how the main result relates to the model primitives, (ii) examines how it relates to ratio monotonicity results from the literature on optimal bundling, (iii) provides economic interpretations, and finally (iv) discusses the implications for screening and non-linear pricing.

\subsection*{4.2.1 Relation to Model Primitives}

Theorem 4 characterizes optimal bundling decisions in terms of the “endogenous” notion of optimal sales volume. Though I did not express the result based on primitives—i.e., values and costs—in the statement of the theorem (because I find the current statement more intuitive,) one can certainly do so. Here, I carry out this task focusing on the case of zero marginal costs to make the expressions slightly simpler.\footnote{That said, the case with non-zero marginal costs will be a simple extension.}

\begin{proposition} \label{prop:zero_marginal}
Suppose that assumptions 1 through 3 hold and that production costs are zero. Also suppose each $\pi(b, p)$ is differentiable in $p$ and that the derivative is only zero at the peak. Then for any two bundles $b$ and $b'$ we have $D^*(b') > D^*(b)$ if and only if for the unique $\tilde{t}$ with \(\frac{\partial \log(v(b', \tilde{t}))}{\partial t} = \frac{f(\tilde{t})}{1-F(\tilde{t})}\), we have: \(\frac{\partial \log(v(b, \tilde{t}))}{\partial t} > \frac{f(\tilde{t})}{1-F(\tilde{t})}\).
\end{proposition}

The proof follows directly from the assumptions. Combined with Theorem 4, this result characterizes when pure bundling is optimal based on model primitives.
4.2.2 Relation to “Ratio Monotonicity”

This section 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 \( v(b,t) > v(\bar{b},t) \) 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 \( v(b,t) < v(\bar{b},t) \) 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 2 clarifies how Theorem 1 relates to ratio monotonicity results.

**Proposition 2.** 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 2 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 $t \in (0,0.3)$, 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, $\frac{v(b,t)}{v(\bar{b},t)}$ 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 above also illustrate an important point that was implicitly made by Proposition 1. Though it may not be clear in the first glance, this proposition simply says that, under assumptions 1 through 3, ratio-monotonicity conditions for bundling are required only locally (as opposed to the global requirements in the literature.) To see this, note that according to Proposition 1, for $D^*(\bar{b}) > D^*(b)$, we need the $\tilde{t}$ that satisfies $\frac{\partial \log(v(\bar{b}, \tilde{t}))}{\partial t} = \frac{f(\tilde{t})}{1 - F(\tilde{t})}$ to also satisfy $\frac{\partial \log(v(b, \tilde{t}))}{\partial t} > \frac{f(\tilde{t})}{1 - F(\tilde{t})}$. In other words: $\frac{\partial \log(v(\bar{b}, \tilde{t}))}{\partial t} < \frac{\partial \log(v(b, \tilde{t}))}{\partial t}$. This means that for $t$ slightly larger than $\tilde{t}$, we get: $\frac{v(b, t)}{v(b, \tilde{t})} < \frac{v(\bar{b}, t)}{v(\bar{b}, \tilde{t})}$. Rearranging, we get $\frac{v(b, \tilde{t})}{v(b, t)} < \frac{v(\bar{b}, \tilde{t})}{v(\bar{b}, t)}$. This is exactly ratio monotonicity. As a result, assumptions 1 through 3 may be thought of as assumptions under which ratio monotonicity requirements reduce to local from global, allowing one to obtain a substantially tighter characterization. In light of this, the above examples may be better understood: changes to value functions that preserve (i) the model assumptions and (ii) the local ratio monotonicity condition will not lead to a change in the optimal bundling decision, even if they violate global ratio monotonicity.

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.

To sum up, this subsection shows how Theorem 1 closes the gap in the ratio monotonicity results. Before moving to the next subsection, there is another point worth making regarding the application of Theorem 1 and ratio monotonicity results. As mentioned before, the majority of the bundling literature assumes independent values (i.e., no complementarity or substitution). In such settings, ratio monotonicity will imply co-linearity of values across types: $\forall t, t' \exists a_1, a_2 \in \mathbb{R}: \forall b: v(b, t) = a_1 + a_2 b(b, t')$, effectively making the problem trivial. The assumptions required for the application of the sales volumes result do not impose as strong a restriction.\footnote{That said, Theorem 1 also imposes meaningful restrictions. One can show that under assumptions 1 through 3 plus independent values, mixed bundling is optimal with “measure zero” exceptions. This should not be surprising given that the combination of monotonicity and independent values bears a close resemblance with the “positive correlation” case in Adams and Yellen (1976).}
4.2.3 Interpretation of the result: complementarity, its variation, and its co-variation with price sensitivity

Although the characterization of optimal bundling in Theorem 1 based on optimal quantities 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 $D^*(\bar{b} | b) < D^*(\bar{b})$) 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\footnote{Though it is not necessary, to ease the interpretation assume all $c_i$ are zero.} (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); Deheuckere 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 $\frac{v(b, t)}{v(b, t)} \leq \frac{v(b, t')}{v(b, t)}$ and $\frac{v(b^C, t)}{v(b, t)} \leq \frac{v(b^C, t')}{v(b, t')}$. From these
inequalities, one can conclude
\[ \frac{v(\bar{b}, t)}{v(\bar{b}, t) + v(b^C, t)} \geq \frac{v(b, t')}{v(b, t') + v(b^C, t')} . \]

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

To sum up, this subsection does two things. First, it shows that the interpretation based on co-variation between complementarity and values, which in Haghpanah and Hartline (2019) is proposed based on ratio-monotonicity results, remains a valid and useful intuition in a setting where one can provide full characterization of optimal pure bundling (based on sales volumes). I find this reassuring given it suggests that the gap between the necessary and sufficient conditions in ratio monotonocity analysis does not flip the economic intuition behind the result. The discussion here also highlights the connection between this interpretation and the elasticity results in papers such as Long (1984); Armstrong (2013). The second thing this subsection highlights is that not only the how complementarity co-varies with valuation, but also how much it varies in the first place is also important for optimal bundling. Later in the paper, I argue that both of these dimensions may be useful for answering applied questions on optimal bundling.

4.2.4 Implications for non-linear pricing a la Maskin and Riley (1984)

The main focus of this paper is the problem of optimal bundling. Nevertheless, the main result also has implications for the non-linear pricing problem studied, among others, by Maskin and Riley (1984). Translated to the context of non-linear pricing, the analysis in this paper will take the following form: A monopolist selling a single product faces the problem of determining the optimal price schedule \( T^*(q) \) for different quantities \( q \in \{0, 1, \ldots, n\} \). The monopolist’s objective is to maximize total profit \( \pi(T^*) \). Valuations are denoted \( v(q, t) \) with \( v(0, t) = 0 \) for all \( t \). Conditional on a prior endowment of \( q' \), we denote the valuations by \( v(q, t|q') \). Similarly, we use notation \( \pi_q(p) \) to denote the firm’s profit when it sells only \( q \)-size batches of the product, pricing each batch at \( p \in \mathbb{R} \). Also \( \pi_q(p|q') \) denotes the same thing under the condition that all consumers have been pre-endowed with \( q' \leq n - q \) unites of the product. \( D^*(q) \), and \( D^*(q|q') \) are defined in the expected way.

Proposition 3 shows that in the above setting, one can fully characterize the optimal nonlinear tariff.
Proposition 3. Consider the non-linear pricing setup described above. Assume the following: (i) \( v(q,t) \) is increasing in both arguments (and strictly so whenever positive), and \( \forall t' > t, q > 0 \) we have \( v(q,t') - v(q-1,t') > v(q,t) - v(q-1,t) \). (ii) For all \( q + q' < n \) the profit function \( \pi_q(p|q') \) is strictly quasi-concave in \( p \) over the range of \( p \) that generates strictly positive demand. (iii) \( v(q,t) \) is continuous in \( t \) except possibly for finitely many points. (iv) For any \( q' < n \), the set \( \arg \max_{q \in \{1,...,n-q'\}} D^*(q|q') \) is a singleton denoted \( q^*(q') \) and then the optimal price schedule will involve \( m \leq n \) distinct quantities \((q_1^*,...,q_m^*)\) such that \( q_1^* = q^*(0) \) and \( \forall i \in \{2,...,m\} : q_i^* = q_{i-1}^* + q^*(q_{i-1}^*) \). Precisely:

\[
\forall q \in \{q_{i-1}^* + 1,...,q_i^*\} : T^*(q) = p^*(q_i^* - q_{i-1}|q_{i-1}^*)
\]

This proposition fully characterizes the optimal tariff \( T^* \). It starts by stating that the smallest quantity that any consumer can buy is \( q_1^* = q^*(0) \equiv \arg \max_{q \in \{1,...,n\}} D^*(q) \), i.e., the batch size that would sell the most if sold alone and priced optimally. In other words, at least \( q_1^* \) units of the products are always bundled together. The price of this \( q_1^* \)-bundle is the optimal price that the monopolist would set if it were to only sell this bundle. From this point on, the proposition takes a recursive structure and states that the second distinct quantity sold of the product is \( q_2^* = q_1^* + \arg \max_{q \in \{1,...,n-q_1^*\}} D^*(q|q_1^*) \) and so on. Though Proposition 3 in some ways resembles the demand profile approach of Wilson (1993), it has fundamental differences. The proof of this proposition is relegated to the appendix. I close this discussion by making several points that are noteworthy.

First, as mentioned shortly before, Proposition 3 takes a step beyond the main result in Theorem 1: it fully characterizes what the optimal tariff looks like as opposed to only characterizing the conditions under which the optimal tariff will involve a flat fee for selling all \( n \) units of the product. This property of Proposition 3 should naturally raise the question that whether Theorem 1 can also be strengthened to fully characterize the optimal bundling strategy as opposed to only characterizing when pure bundling is optimal. Unfortunately the answer turns out to be no; a bundling strategy constructed in a similar way to how the optimal tariff in Proposition 3 is constructed may or may not be optimal.

The second point is regarding complementarity. In Theorem 1, complementarity is one of the main assumptions; but it is not used in Proposition 3. Though this difference may,

---

\( ^{16} \) Similar to the main result, if instead of quasi-concavity we have concavity, one need not impose the assumption that the \( \arg \max \) is unique.

\( ^{17} \) That condition would be \( n = \arg \max_{q \in \{1,...,n\}} D^*(q) \), mirroring the condition in Theorem 1.

\( ^{18} \) Counterexamples are available upon request. Also, it is worth noting that the Maskin-Riley type monotonicity condition used in Proposition 3 is stronger than that used in Theorem 1. Nevertheless, this is not the reason why the optimal bundling strategy cannot be fully characterized.
in the first glance, seem at odds with the claim that there is a close tie between the two results, it is in fact not. When translating the setting of Proposition 3 to that of Theorem 1, complementarity naturally arises. In that translation, the optimal tariff design problem is cast as a bundling problem with $n$ products. Product $i$ in this bundling problem will be equivalent to the $i$-th unit of the single product in the non-linear pricing problem. That is, its value to type $t$ will be equal to the marginal value $v(i, t) - v(i - 1, t)$ if product $i$ is offered to $t$ as part of a bundle that includes all $j < i$, and the value is zero otherwise. This, by construction, imposes a special case of the complementarity condition on the bundling problem. This special case is the main reason why the result in Proposition 3 is stronger than that in Theorem 1.

Third, there is a difference between Proposition 3 and the usual form in which the nonlinear pricing problem is studied in the literature. The origin of Proposition 3 being in bundling makes the domain of quantities $q$ by construction discrete and bounded, whereas the literature (e.g., Maskin and Riley (1984)) examines a continuous and possibly unbounded environment. That said, I do not see this feature of Proposition 3 as too restrictive, given that one can always examine the limit case as $n \to \infty$.

Fourth, $q$ in this proposition need not be interpreted as quantity. One can also think of $q$ as quality in a similar fashion to Mussa and Rosen (1978). In that case, the result holds if the cost function is non-linear in quality, as long as it is still linear in the quantity of the consumers that purchase each quality level.

Finally, Proposition 3 highlights a tie between the notions of bunching (i.e., when the optimal tariff induces distinct types $t$ and $t'$ to purchase the same quantity $q$ and pay the same total amount) and bundling. Of course there is some bunching built into the setting of Proposition 3 due simply to the discrete nature of quantities $q$. But there is more than that. For instance if $q_1^* > 1$, then the seller will optimally not use any of the possible quantities $1, ..., q_1^* - 1$ in the tariff, bunching different types that buy a positive amount at $q_1^*$.

In light of this tie between bunching and bundling, the intuitive interpretation offered for optimality of bundling can also be useful to understand optimality of bunching: bunching in tariff design may be optimal if the marginal value of additional units diminishes, in relative

---

19 More precisely, for any $i$ and $b$ with $i \notin b$: $u(b \cup \{i\}, t) = u(b, t) + \mathbf{1}_{\{1, ..., i-1\} \subseteq b} \times (v(i, t) - v(i - 1, t))$ where $v$ represents the value function in the nonlinear pricing problem and, in order to avoid confusion, $u$ represents the value function in the equivalent bundling problem.

20 To be more precise, there is also a form of bunching that would not go away even in the limit case of $n \to \infty$. 

18
terms, more slowly for lower types compared to higher ones. To illustrate this, note that one can recast Example 2 from Section 4.2.2 as a non-linear pricing problem by thinking of the “base product” as the first unit and the “add on” as the second unit.\footnote{As can be seen there, the optimal tariff will involve bunching of the two units and selling only a two-package. This is also quite in line with the well-known fact (see Rochet and Stole (2003)) that bunching may arise when the “Myersonian Virtual Value” is not quasi-concave for some types in quantity purchased. One can verify that in Example 2, for types \( t \) just below 1, the virtual value is zero for \( q = 0 \), negative for \( q = 1 \), and positive for \( q = 2 \). Note that the above may seem at odds with Mussa and Rosen (1978) which is one of the first studies analyzing bunching. Mussa and Rosen (1978) argue that the optimal tariff should not “skip” some quantity (or in their setting, quality) levels because with such tariffs “the monopolist would not be making full use of his power to discriminate among different types of buyers.” That said, one can show this result in Mussa and Rosen (1978) arises from their assumption that valuations are linear in both type and quality/quantity \( q \), which would rule out Example 2 from section 4.2.2. They mention in a footnote that one can replace \( q \) with an increasing and concave function of \( q \) in the formula for valuations. Nevertheless, that modification, too, although allowing for Example 1 above, still rules out Example 2 which is the example under which pure bundling (i.e, bunching with skipped values) is optimal.}

5 Open Question for Future Research

This section poses a question for future research which, in my view, (i) is difficult to answer but (ii) answering it would provide substantial value for applied work. In practice, bundling/product line decisions must be made for a specific market. Thus, determining whether or not the necessary or sufficient conditions for optimal bundling decisions hold becomes ultimately an empirical question, requiring at least an estimated model of demand for the products being considered. This is where the difficult question arises: can we characterize necessary and/or sufficient conditions for the optimality of different bundling strategies in the context of commonly used empirical models?

I discuss this issue in the context of a simple class of random-coefficient discrete-choice demand models a la Berry et al. (1995) (BLP for short) which are the workhorse demand estimation models in the empirical literature. I examine a simple class of these BLP models.\footnote{It would also be fairly straightforward to make the quantities continuous by doing a piece-wise linear interpolation of values from \( \{0,1,2\} \) to \( [0,2] \) for each type \( t \). In this continuous setting, too, one can show that a flat fee is optimal.}
I argue that no theoretical model that we have to date (including the present paper) provides a characterization for how to make bundling decisions in this setting. However, I also argue, using simulations, that the economic interpretations from our theory results may be useful both as informal inputs for the design of empirical studies and, perhaps, as potential directions for future theory work.

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:

$$u_{it} = \alpha_0 + \alpha_{1,t}p_i + \alpha_{2,t}I_{i=2} + \varepsilon_{it}$$

where $\alpha_0$ is a constant, $\alpha_{1,t}$ is the price coefficient, $\alpha_{2,t}$ is the valuation of the add-on, and $\varepsilon_{it}$ 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:

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

where $\mu_1$ and $\mu_2$ are the means, $\sigma_1$ and $\sigma_2$ are the standard deviations, and $\rho$ 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 4.2.3 may be useful.

According to the interpretation provided in Section 4.2.3, 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 $\sigma_2$ and $\rho$ respectively.

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

- $\sigma_2$ is small, which means that the valuations for the add-on are so homogeneous that
Figure 2: Optimal bundling decision as a function of parameters $\rho$ and $\sigma_2$ 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 $\sigma_2$ decreases (i.e., when the population values the add on rather homogeneously) or when $\rho$ increases (i.e., when more price sensitive customers have higher relative valuations for the add on).

it makes sense to give the add on to all purchasing customers.

- $\rho$ 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 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, $\rho$ and $\sigma_2$, 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 $\sigma_2$ decreases (i.e., when the population values the add on rather homogeneously) or when $\rho$ increases (i.e., when more price sensitive customers have higher relative valuations for the add on).

Discussion. It should be straightforward to verify that theory results on bundling in the literature cannot be used to comment on optimal bundling strategy in the above BLP setting. The literature with independent values clearly does not apply. As for the interdependent values literature, it is easy to check that this setting never satisfies the ratio monotonicity condition in either direction. It also does not satisfy the monotonicity assumption in my
framework. Thus, Theorem 1 does not directly apply.

Nevertheless, as the simulation analysis suggests, the economic intuitions that came out of the previous literature and confirmed and expanded by the analysis in this paper are indeed useful. The simulation analysis has two practical implications: 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 for correlation between the random effects for add-on value and price sensitivity.

This analysis suggests that having formal results on optimal bundling in the context of empirical models such as BLP would be very useful for applied work on bundling. 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 the literature. This is because, in order to directly evaluate the sales volume conditions or ratio monotonicity, the econometrician 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.

6 Conclusion

This paper provided (an almost) full characterization for the optimality of pure bundling for products with interdependent values. Under some assumptions, 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 grand bundle.

Additionally, I discussed multiple implications of the main result, including its relation to “ratio monotonicity” results in the literature, its implications for “bunching” of types in the non-linear pricing problem a la [Maskin and Riley (1984)], and the economic interpretation

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.
of optimal bundling based on variation across types in perceived complementarity among products.

Finally, I argued that an important area for future theoretical research on bundling would be characterizing conditions for optimal bundling strategies in the context of widely used empirical models of demand (most notably random-coefficient discrete choice models a la Berry et al. (1995)).

The work in this paper can be extended in multiple directions. First, it would be valuable to investigate alternative assumptions to 1 through 3. 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. It would be valuable to find the extent to which the insights in this paper extend to multi-dimensional types. Of course it is well-known that multi-dimensional types substantially complicate the problem (Rochet and Stole, 2003). Therefore, if-and-only-if characterizations in those environments should be harder to obtain. Nevertheless, there may be specialized multi-dimensional settings with relevant applications which yield if-and-only-if conditions. Finally, another direction for future research, as mentioned in the previous section, would be to characterize optimal bundling decisions in the context of commonly used empirical models.

References


### 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 $t^*(b)$. 

25
Lemma 4. 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 4. 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 4 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 5. $\pi_b(t|b')$ is strictly quasi-concave in $t$.

Proof of Lemma 5. 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 $D^*(b) \geq D^*(\bar{b})$ 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 6. $D^*(b) > D^*(\bar{b})$ implies $D^*(b) > D^*(b^C|b)$.
Proof of Lemma 6. 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 7. 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 grand bundle alone.

Proof of Lemma 7. 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 $D^*(b,b^C|b)$ units of $b^C$ and $D^*(b)$ units of $b$ under optimal pricing.23

Also recall that under $\tilde{v}(\tilde{b},t) = v(\tilde{b},t)$ for all $t$. Therefore, under the modified problem and under $B^2$, the optimal price $p^{2,*}$ will be one that leads to exactly $D^*(\tilde{b})$ 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 $D^*(\tilde{b})$ unit of $b$ and $D^*(\tilde{b})$ units of $b^C$. This will yield exactly the profit of $\tilde{\pi}(B^2, p^{2,*})$. But we know at least one of these quantities sold is sub-optimal. This is because, by lemma 6 we have $D^*(b^C|b) < D^*(b)$. Therefore, by selling $D^*(\tilde{b})$ 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 $\hat{D}(b) \geq \hat{D}(b)$, the firm will get the exact same demand volumes for the two bundles under the original problem as it would 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 $\hat{D}(b) < \hat{D}(b)$, then by the construction of $\tilde{v}$ from $v$ and by the complementarity property of $v$, we have $D(b) \leq \hat{D}(b^C)$. This, in turn, due to complementarity, will lead to $D(b) \leq \hat{D}(b)$. Under these conditions, any pricing strategy that makes non-negative profit (i.e., any strategy that does not price below average cost,) 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

23Note 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.
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 \( v(\tilde{b}, t) = \tilde{v}(\tilde{b}, t) \) for all \( t \). Therefore, optimizing the price of \( \tilde{b} \) 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.

Notice that, as mentioned in the main text, the proof of the first side of the theorem did not make use of the assumption that the marginal costs are constant. Next, I turn to the proof of the sufficiency conditions (i.e., that \( D^*(\tilde{b}) > \max_{b \in B \setminus \{\tilde{b}\}} D^*(b) \) implies that pure bundling is optimal).

**Proof of sufficiency.** I start with some lemmas.

**Proof of Lemma 2 from the main text.** Assume, on the contrary that no type \( t \) purchases \( \tilde{b} \) under the optimal firm behavior. I show we can reach a contradiction. In particular, I show that type \( t = 1 \) not purchasing \( \tilde{b} \) leads to a contradiction.

Assume \( \cup_{t \in [0,1]} \beta(t|B^*,p^*) \neq \tilde{b} \). 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:

\[
v(b^C, 1) > \sum_{i \in b^C} c_i \tag{7}
\]

To see why (7) is true, note that by complementarity:

\[
v(b^C, 1) \geq \sum_{i \in b^C} v(\{i\}, 1) \tag{8}
\]

Also, by the fact that for each \( i \) there is some \( t \) with \( v(\{i\}, t) \geq c_i \) with the inequality being strict for at least one \( i \), and by monotonicity, we have

\[
\sum_{i \in b^C} v(\{i\}, 1) > \sum_{i \in b^C} c_i \tag{9}
\]

Together, inequalities (8) and (9) imply inequality (7)

Given (7), 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.

**Proof of Lemma 3 from the main text.** I prove the second statement in the lemma. The first statement would be proved in a similar way. Suppose that \( \beta(t|B,p) \neq \beta(t'|B,p) = \bar{b} \). Suppose, contrary to the statement of the lemma, that \( t' \leq t \). Given \( \beta(t|B,p) \neq \beta(t'|B,p) \), we know \( t \neq t' \) which implies \( t' < t \). Define:

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

In other words, \( \tilde{\beta} = (\beta(t|B,p))^C \). 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 \( \tilde{\beta} \) and form \( \bar{b} \). Formally, I show:

\[
v(\tilde{\beta}, t|\beta(t|B,p)) \geq \sum_{i \in \tilde{\beta}} p(i) \tag{10}
\]

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 \( \bar{b} \). Formally:

\[
v\left(\tilde{\beta}, t'|\beta(t|B,p)\right) \geq \sum_{i \in \tilde{\beta}} p(i) \tag{11}
\]

By monotonicity and \( t' < t \), we get:

\[
v\left(\tilde{\beta}, t|\beta(t|B,p)\right) \geq v\left(\tilde{\beta}, t'|\beta(t|B,p)\right) \tag{12}
\]

Together, inequalities 11 and 12 imply inequality 10 completing the proof. Q.E.D.

In light of lemma 2, the following two corollaries of lemma 3 are useful.

**Corollary 1.** Under \((B^*, p^*)\), 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 \((B^*, p^*)\), 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 strictly larger than \( t_2 \) in corollary 2.

Then, note that by corollary 1 and the piece-wise continuity of values in \( t \), 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 by piece-wise continuity of values in \( t \), 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.** \( t^*(b^C_1|b_1) = t_1 \).

In words, claim 4 says that the set of customers who purchase the grand bundle \( \beta(t|B^*, p^*) = \bar{b} \) under the firm optimal strategy \((B^*, p^*)\) is the same as those who purchase \( b^C_1 \) and construct the grand bundle if (i) everyone is endowed with \( b_1 \) and (ii) the firm offers only \( b^C_1 \), 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 3 implies \( t^*(\bar{b}) \leq t_2 \). By \( t_1 > t_2 \), we get \( t^*(\bar{b}) < t_1 = t^*(b^C_1|b_1) \). Also note that:

\[
\forall t : \pi_{\bar{b}}(t) = \pi_{b^C_1}(t|b_1) + \pi_{b_1}(t)
\]
As such, by strict quasi-concavity of profits, by \( t^*(\bar{b}) < t^*(b^C_1|b_1) \), and by remark 1 the peak of \( \pi_b(t) \) should happen in between those of \( \pi_{b^C}(t|b_1) \) and \( \pi_{b_1}(t) \). Therefore, we should have: \( t^*(b_1) \leq t^*(\bar{b}) \leq t^*(b^C_1|b_1) \). 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 \( t^*(b^C_1|b_1) \neq t_1 \). In that case, it can be shown that the firm can strictly improve its profit by slightly adjusting the price of \( b^C_1 \). That is, there is a pricing strategy \( p \) with \( p(b) = p^*(b) \) for all \( b \neq b^C_1 \) such that \( \pi(B^*, p) > \pi(B^*, p^*) \).

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

\[
B' = \{b_1, b^C_1\} \quad (15)
\]

Also construct pricing strategy \( p' \) by fixing \( p'(b_1) = \min_t v(b_1) \) but keeping \( p'(b^C_1) \) adjustable.

Now note that as long as \( \rho \in [p^*(b^C_1) - \epsilon, p^*(b^C_1) + \epsilon] \) for a small enough \( \epsilon \), then \( \pi(B^*, p) \) and \( \pi(B', p') \) move in parallel if we set \( p(b^C_1) = p'(b^C_1) = \rho \) and move \( \rho \). The range parameter \( \epsilon \) should be chosen so that for any pricing strategy \( p \) constructed with a \( \rho \) in this interval we have: \( D(b^C_1|B^*, p) < 1 - F(t'_1) \) where \( t'_1 \) was constructed in equation 13. In other words, \( \epsilon \) 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 \( b^C_1 \) 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 \( b^C_1 \) will exactly be given by \( v(b^C_1, t|b_1) - p(b^C_1) \) which is exactly how these types would value it under \( (B', p') \). Also, when some of these types drop \( b^C_1 \) in response to a change in \( \rho \), 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 \( \rho \) as part of pricing strategies \( p \) and \( p' \) will lead to the exact same reaction by customers. This, in turn, means the exact same change in total revenue. It also means the exact same change in the total cost given that the marginal costs
are assumed constant.\footnote{One can show we do not need the constant marginal cost assumption if \( n = 2 \). or if the monotonicity condition is strengthened.} Thus, we have the exact same change in the total profit. Therefore, the optimal value of \( \rho \) in this interval is the same under \((B^*, p)\) as it is under \((B', p')\) (by strict quasi-concavity, we know that this optimal \( \rho \) is unique in both cases). This common optimal value for \( \rho \) leads to the exact same demand for \( b_{C_1} \) under \((B^*, p)\) as it does under \((B', p')\). The optimal demand under \((B^*, p)\) is achieved by choosing \( \rho \) to equate \( p \) with \( p^* \), which by construction leads to all \( t \) with \( t \geq t_1 \) buying. The optimal \( \rho \) under \((B', p')\), by definition, should lead to all \( t \geq t^*(b_{C_1}|b_1) \) buying. Therefore, if \( t_1 \neq t^*(b_{C_1}|b_1) \), then one can modify \((B^*, p^*)\) by slightly changing \( p^*(b_{C_1}|b_1) \) 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 \( \rho \) 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 \( \rho \). 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 Proof of Proposition 2

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:

\[
\pi^*(\bar{b}) = v(\bar{b}, t^*(\bar{b})) \times D^*(\bar{b})
\]  

(16)

Using a similar notation for \( b \), we get:

\[
\pi^*(b) = v(b, t^*(b)) \times D^*(b)
\]  

(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 $D^*(\bar{b})$ 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:

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

and:

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

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

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

and:

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

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

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

C Proof of Proposition 3

First, let us introduce some notations. In a similar spirit to the consumer choice notation $\beta(t|B,p)$ in the bundling problem, define the following consumer choice function:

$$\beta(t|T) \equiv \arg \max_q v(t, q) - T(q)$$ (23)

I use the same notation $\beta$ as I did in the formulation of the bundling problem in order for the parallels between the two settings to be clearer. However, there are some differences. Most notably, the output of the $\beta$ function here is just an integer number, not a bundle.
Suppose consumers break ties in favor of higher quantities. I now proceed to state the following lemma.

**Lemma 8.** For any price schedule $T$ and any two types $t, t'$ with $t' > t$ we have: $\beta(t'|T) \geq \beta(t|T)$.

This lemma, whose proof is rather straightforward and is left to the reader, simply says that the consumption quantity is weakly monotonic in type.

Also, in this proof, I will assume that the optimal schedule $T^*(q)$ is weakly monotonic in $q$. This assumption is without loss, given that one can show that any non-monotonic $T$ can be modified in a way that (i) makes it weakly monotonic and (ii) delivers the same amount of profit to the monopolist. Given this weak monotonicity, $T^*$ has to take the following form for a strictly increasing sequence $q_0 = 0, q_1, ..., q_m = n$ and a weakly increasing sequence $T_1, ..., T_m$:

$$\forall q \in \{q_{i-1} + 1, ..., q_i\} : T^*(q) = T_i$$

(24)

The following lemma will be useful for the proof:

**Lemma 9.** $T_1 = p^*(q_1|0)$. That is, the lowest type $t$ purchasing $q_1$ under the optimal contract will satisfy $1 - F(t) = D^*(q_1)$.

**Proof of Lemma 9.** Suppose this lemma’s claim is not true: $T_1 \neq p^*(q_1|0)$. Then consider a scenario in which only $q_1$ is being sold by the seller at the price of $T_1$. Given that $T_1$ is not the optimal price, there is a small but nonzero $\epsilon$ such that if the seller prices $q_1$ at $T_1 + \epsilon$ instead of $T_1$, the seller will strictly improve its profit when selling only $q_1$.

Next, I move from the scenario of selling only a batch of $q_1$ units back to the full tariff design problem. I use the above deviation to construct a similar deviation from the full schedule $T^*$ and show that the seller can strictly improve its profit. Construct price schedule $T \equiv T^* + \epsilon$ for any positive $q$. I know claim that $T$ is strictly more profitable to the seller than is $T^*$.

To see why this claim is true, note that for any $q \in \{q_2, ..., q_m\}$ and any type $t$ such that $\beta(t|T^*) = q$, we have $\beta(t|T) = q$. This is because (i) for those types it is the IC constraint (and not the IR) that is binding; and (ii) the construction of $T$ from $T^*$ preserves all the IC constraints. Therefore a move from $T^*$ to $T$ will lead to the exact same revenue change that a move from $T_1$ to $T_1 + \epsilon$ does in the scenario of selling only $q_1$: the exact same new types are added to (or removed from) the set that purchases $q_1$, and the same change
(i.e., $\epsilon$) has been made to the amount made off of each type that buys. In addition to the change in the revenue, a move from $T^*$ to $T$ also leads to the exact same change in total costs as would a change from $T_1$ to $T_1 + \epsilon$. This is because in both cases, the only change made in the production is the number of $q_1$-size batches (or, in the Mussa and Rosen (1978) interpretation, the number of $q_1$-quality products). Given that we have assumed the cost function to be linear in this change, the changes in total cost is the same between the two scenarios.

As a result, a move from $T^*$ to $T$ will lead to the exact same change in the total profit as would a move from $T_1$ to $T_1 + \epsilon$. Therefore, a change from $T^*$ to the new schedule $T$ strictly profitable, finishing the proof of the lemma. Q.E.D.

Next, I introduce a lemma which will be the building block of the proof of this proposition.

**Lemma 10.** In the presentation of $T^*$ in equation 24, it has to be that $q_1 = q_1^*$ where $q_1^* = q^*(0)$ as defined in the statement of Proposition 3.

As a reminder, $q^*(0) = \arg\max_{q=1,...,n} D^*(q)$. Thus, the lemma simply says that the smallest quantity that the optimal schedule $T^*$ offers to consumers is the quantity that, if sold alone, would sell the highest volume.

**Proof of Lemma 10** Suppose $q_1 \neq q_1^*$. Then one can construct a deviation from $T^*$ that would strictly improve the seller’s profit. This will suffice to finish the proof of the lemma.

First, for convenience, assume that even though $q_1 \neq q_1^*$, there is some $k > 1$ such that $q_k = q_k^*$. In other words, even though $q_1^*$ is not the smallest package on the schedule, it is nonetheless somewhere on the schedule. Later I will show this assumption is not necessary. But for now, it will make the steps of the proof more straightforward.

For each $i \in \{1, ..., m\}$ denote by $t_i$ the lowest type that buys $q_i$ units under $T^*$. That is: $t_i = \min\{t : \beta(t|T^*) = q_i\}$. From Lemma 8 we know larger types buy weakly more units of the product. That is: $t_1 < t_2 < ... < t_m$.

From Lemma 9 we know that $1 - F(t_1) = D^*(q_1)$. Given $t_{k-1} \geq t_1$, we get $1 - F(t_{k-1}) \leq D^*(q_1)$. But we also know, by assumption, that $D^*(q_1^*) > D^*(q_k^*)$. Therefore:

$$1 - F(t_{k-1}) < D^*(q_1^*) \equiv D^*(q_k) \quad (25)$$

In addition, again by the definition of $q_k = q_k^*$, we know that $D^*(q_{k-1}) < D^*(q_k)$. This inequality, along with a similar quasi-concavity argument to that used in the proof of the main result yields:
\[ D^*(q_k) \leq D^*(q_k - q_{k-1}|q_{k-1}) \] (26)

Together, inequalities 25 and 26 yield:

\[ 1 - F(t_{k-1}) < D^*(q_k - q_{k-1}|q_{k-1}) \] (27)

Obviously, by \( t_k > t_{k-1} \), we also know:

\[ 1 - F(t_k) < D^*(q_k - q_{k-1}|q_{k-1}) \] (28)

Note that equation 28 implies that if all consumers have already been endowed with \( q_{k-1} \) units of the product and the monopolist is selling only batches of size \( q_k - q_{k-1} \) and pricing them so that types \( t_k \) and above purchase, then the monopolist, by quasi-concavity, will strictly profit from a small price reduction \( \rho \).

Next, I move from the scenario of selling only \( q_k - q_{k-1} \) packages under a pre-endowment of \( q_{k-1} \) to the main scenario of designing the full schedule \( T^* \). I argue that the monopolist will enjoy the same profit increase as the one described in the previous paragraph if it modifies \( T^* \) by reducing all \( T_j \) for \( j \geq k \) by \( \rho \). The argument is similar to that in the proof of Lemma 9. This modification does not alter the behavior of any type \( t \) that purchases \( q_k + 1 \) or more units due to the fact that it preserves all of the binding IC constraints for those types. As a result, the impact of this change on the firm profit is (i) a cut in margin by \( \rho \) across all consumers, combined by the change in the behavior of those who used to purchase less than \( q_k \) units under \( T^* \) but will now switch to \( q_k \). Note that if the price reduction \( \rho \) is small enough, these types will only consist of those who under \( T^* \) purchase \( q_{k-1} \). From equation 27, we know there is a non-zero mass of such types. Therefore, the effect of this price change parallels that of the price change described in the previous paragraph, making it strictly profitable.

But the above argument contradicts the optimality of \( T^* \). Therefore, the contrapositive assumption must have been incorrect. That is: it has to be that \( q_1 = q_1^* \).

With the above argument, the proof is complete for the case where there is some \( k > 1 \) with \( q_k = q_1^* \). That is, when \( q_1^* \) is “on the price schedule.” Thus, it remains to show that the proof also works when \( \nexists k : q_k = q_1^* \). Suppose this is the case, and take \( k \) to be the smallest index with \( q_k \geq q_1^* \). Like before, denote by \( t_i \) the smallest type that purchases \( q_i \) under \( T^* \).

I now construct schedule \( T^{**} \) in the following way:

\[ \text{None of these types would switch to buying more than } q_k \text{ units, due to the single-crossing condition.} \]
• For any \( q \leq q_{k-1} \) or \( q > q_1^* \), set \( T^{**}(q) = T^*(q) \)

• For all \( q_{k-1} < q \leq q_1^* \), set \( T^{**}(q) = T^*(q_k) + v(q_1^*, t_k) - v(q_k, t_k) \)

Next, I take two steps. First, I show that \( T^{**} \) delivers the same profit to the monopolist as does \( T^* \). Then, I will construct a deviation from \( T^{**} \) that yields a strict profit improvement.

As for the first step, note that by construction, \( T^{**}(q_1^*) \) is designed to make the type \( t_k \) consumer indifferent between purchasing \( q_k \) units and \( q_1^* \) units. But we know, by construction, that this type is also indifferent between buying \( q_k \) units and buying \( q_k-1 \) units. This makes this type indifferent among all three quantities \( q_k-1 < q_1^* < q_k \) under the tariff \( T^{**} \). But by the monotonicity condition (i.e., single crossing,) any type \( t < t_k \) will strictly prefer \( q_k-1 \) to \( q_1^* \) and any type \( t > t_k \) will strictly prefer \( q_k \) over \( q_1^* \). In other words, this “addition of \( q_1^* \) to the schedule” will not change any consumer’s purchase behavior: \( \forall t : \beta(t|T^*) = \beta(t|T^{**}) \).

Thus the two tariffs deliver the same profit to the monopolist.

But now the structure of \( T^{**} \) allows us to construct the same profit enhancing modification that we applied to \( T^* \) when we assumed it did have \( q_1^* \) on the schedule. All of the steps are the same. This finishes the proof of the lemma. \textbf{Q.E.D.}

The rest of the proof is straightforward and involves recursive use of Lemma 10. First note that by quasi-concavity, and by \( \forall q \in \{1, ..., n-q_1^*\} : D^*(q + q_1^*) < D^*(q_1^*) \), we have:

\[
\forall q \in \{1, ..., n-q_1^*\} : D^*(q|q_1^*) < D^*(q_1^*)
\]

Thus, all of the optimal strategies will sell only to types \( t_1 \) and above. This means that in order to construct “the rest of the optimal schedule” conditional on having set \( T^*(q) \) for all \( q \leq q_1^* \) equal to \( p^*(q|0) \), one can just focus attention on designing the optimal schedule for selling \( n-q_1^* \) units when all consumers have been endowed with \( q_1^* \) units already. This means we are facing a version of the same problem. It is straightforward to check that all of the conditions of the main problem are satisfied for this “sub-problem.” Thus, one can apply Lemma 10 again and find that \( q_2 \) in the optimal schedule should be equal to \( q_1^* \) which was defined as \( q_1^* + q^*(q_1^*) \). Repeating this procedure will fully characterize the optimal tariff and the outcome matches what the statement of the proposition predicted. \textbf{Q.E.D.}