

Bundling Information Goods: Pricing, Profits and Efficiency

Yannis Bakos^{*} and Erik Brynjolfsson^{}**

First Draft: December, 1996

Current Draft: February 1999

Forthcoming in *Management Science*

Copyright © 1996, 1997, 1998 by Yannis Bakos and Erik Brynjolfsson

^{*} Stern School of Business, New York University; email: bakos@stern.nyu.edu

^{**} Sloan School of Management, Massachusetts Institute of Technology; email: erikb@mit.edu

Bundling Information Goods: Pricing, Profits and Efficiency

ABSTRACT

We study the strategy of bundling a large number of information goods, such as those increasingly available on the Internet, and selling them for a fixed price. We analyze the optimal bundling strategies for a multiproduct monopolist, and we find that bundling very large numbers of unrelated information goods can be surprisingly profitable. The reason is that the law of large numbers makes it much easier to predict consumers' valuations for a bundle of goods than their valuations for the individual goods when sold separately. As a result, this "predictive value of bundling" makes it possible to achieve greater sales, greater economic efficiency and greater profits per good from a bundle of information goods than can be attained when the same goods are sold separately. Our main results do not extend to most physical goods, as the marginal costs of production for goods not used by the buyer typically negate any benefits from the predictive value of large-scale bundling.

While determining optimal bundling strategies for more than two goods is a notoriously difficult problem, we use statistical techniques to provide strong asymptotic results and bounds on profits for bundles of any arbitrary size. We show how our model can be used to analyze the bundling of complements and substitutes, bundling in the presence of budget constraints and bundling of goods with various types of correlations. We find that when different market segments of consumers differ systematically in their valuations for goods, simple bundling will no longer be optimal. However, by offering a *menu* of different bundles aimed at each market segment, bundling makes traditional price discrimination strategies more powerful by reducing the role of unpredictable idiosyncratic components of valuations. The predictions of our analysis appear to be consistent with empirical observations of the markets for Internet and on-line content, cable television programming, and copyrighted music.

We thank Timothy Bresnahan, Hung-Ken Chien, Frank Fisher, Michael Harrison, Paul Kleindorfer, Thomas Malone, Robert Pindyck, Nancy Rose, Richard Schmalensee, John Tsitsiklis, Hal Varian, Albert Wenger, Birger Wernerfelt, four anonymous reviewers and seminar participants at the University of California at Berkeley, MIT, New York University, Stanford University, University of Rochester, the Wharton School, the 1995 Workshop on Information Systems and Economics and the 1998 Workshop on Marketing Science and the Internet for many helpful suggestions..

1. Introduction

1.1 Overview

Millions of digital information goods can be distributed almost costlessly via data networks such as the Internet. The technology continues to advance with breathtaking speed, yet existing theory and practice fail to provide clear guidance on how digital information goods should be packaged, priced and sold. At one end of the spectrum technologies such as micropayments increasingly enable the sale and delivery of small units of information, but in this article we draw attention to the opposite end. We analyze the strategy of bundling a large number of information goods and selling the bundle for a fixed price. We find that in a variety of circumstances, a multiproduct monopolist will extract substantially higher profits by offering one or more bundles of information goods than by offering the same goods separately. In addition, we provide criteria for the optimal design and pricing of bundles.

The key intuition behind the power of bundling is that consumer's valuation for a collection of goods typically has a probability distribution with a lower variance per good compared to the valuations for the individual goods. The larger the number of goods bundled, the greater the typical reduction in the variance. Because uncertainty about consumer valuations is the enemy of effective pricing and efficient transactions (Myerson and Satterthwaite, 1983), this "predictive value of bundling" can be very valuable. For instance, consumer valuations for an on-line sports scoreboard, a news service or a daily horoscope will vary. A monopolist selling these goods separately will typically maximize profits by charging a price for each good that excludes some consumers with low valuations for that good and forgoes significant revenues from some consumers with high valuations. Alternatively, the seller could offer all the information goods as a bundle. Under a very general set of conditions, the law of large numbers guarantees that the distribution of valuations for the bundle has proportionately fewer extreme values. As Schmalensee (1984) has argued, such a reduction in "buyer diversity" typically helps sellers extract higher profits from all consumers.

Historically, very large bundles of goods have typically been unprofitable and hence uncommon in practice. They have also been hopelessly complex to model; the number of possible interactions is simply too large to derive general results (Hanson and Martin, 1990; McAdams, 1997). As a result, large bundles have received little attention. However, as we show below, the advent of digital information goods with very low marginal costs can make bundling hundreds or even thousands of unrelated goods a profitable strategy. Furthermore, the modeling framework that we introduce provides strong results regarding the profitability of bundling even under relatively weak assumptions. Unlike earlier work, our model does not become more complex as the number of goods increases. Instead, the precision of our analysis increases with the number of goods considered, making our framework suitable for understanding the economics of large bundles. Our model can explain the prevalence of large bundles of information goods and provides guidelines for the use of more complex strategies such as mixed bundling, which involves simultaneously selling a large bundle and one or more subsets of the bundle.

1.2 The bundling literature

Bundling has many potential benefits, including cost savings in production and transaction costs, complementarities among the bundle components, and sorting consumers according to their valuations (Eppen, Hanson, et al. 1991). We focus on this last benefit of bundling, which was first discussed by Stigler (1963) in a paper showing how bundling could increase sellers' profits when consumer valuations for two goods were negatively correlated. Adams and Yellen (1976) introduced a two-dimensional graphical framework for analyzing bundling as a device for price discrimination. By introducing a setting with a multiproduct monopolist, two goods, no reselling, independent and additive consumer valuations, and linear "unit demands" (i.e., consumers buy either zero or one unit) for these two goods, they compare unbundled sales to pure bundling (offering only the complete bundle) and mixed bundling (offering both the complete bundle and subsets of the bundle).¹ Using stylized examples, they illustrate that various types of bundling can be more or less profitable than unbundling.

The more formal analyses by Schmalensee (1984), McAfee, McMillan and Whinston (1989) and Salinger (1995) also focused on bundles of two goods. Schmalensee assumed a bivariate Gaussian distribution of reservation prices, and through a combination of analytic derivation and numerical techniques showed that pure bundling typically reduces the diversity of the population of consumers, thereby enabling sellers to extract more consumers' surplus. He found that bundling can increase profits if the valuations of the two goods are negatively correlated (as suggested by Stigler and Adams and Yellen), but can also be true if the valuations are independent, or even positively but not perfectly correlated.

McAfee, McMillan and Whinston analyzed a setting with a multiproduct monopolist and a continuum of consumer valuations. They derived a set of conditions under which mixed bundling of two goods dominates unbundled sales. Salinger develops a graphical framework to analyze the profitability and welfare implications of bundling two goods, primarily in the context of independent linear demand functions. He finds that bundling two goods tends to be profitable when consumer valuations are negatively correlated and high relative to marginal costs.

More recently, Armstrong (1996) shows that for a special class of cases, the optimal tariff in the multiproduct case can be determined using the techniques typically used in the single-product case. He finds that, in his setting, the optimal bundle price will almost always inefficiently exclude some low-demand consumers. However, he does not explore the implications of increasing the number of goods.

There are few general results for bundles of more than two goods. McAdams (1997) found that the existing analytical machinery for analyzing mixed bundling could not be easily generalized to even three goods, because of the interactions among sub-bundles. In general, price-setting for mixed bundling of many goods is

¹ For the remainder of this article, the unmodified term "bundling" refers to *pure* bundling, unless otherwise specified.

an NP-complete problem, requiring the seller to determine a number of prices and quantities that grows exponentially as the size of the bundle increases (Hanson and Martin, 1990).

1.3 Approach in this article

Unlike the above articles, our approach is most applicable to large bundles of goods, such as the thousands of information goods available via a typical online service. We are able to bound the profits derived from any bundle of n goods with finite variance and to explore how the optimal bundle price changes under various conditions. In particular, we draw on well-established statistical theorems to characterize the probabilistic valuations of large collections of goods. We find that some of the results in the literature for bundles of two goods do not generalize to larger bundles. For instance, Salinger (1995) shows when consumers have independent linear demands, bundling two goods increases consumers' surplus when bundling was profitable. It turns out that this is not typical: bundles of more than two goods will always *reduce* consumers' surplus when the goods have independent linear demands. Other results from the bundling literature are strengthened in our setting, sometimes dramatically so: bundling is profitable for a broader set of conditions and may even be able to extract nearly all the value from a collection of goods.

Section 2 presents the basic modeling framework and key results regarding the predictive value of bundling for the general case of information goods with independent valuations. These include the asymptotic optimality of bundling when marginal costs are zero, the suboptimality of bundling when marginal costs exceed a critical threshold, and a sufficient condition for finite bundles to be more profitable than unbundled sales. Section 3 applies the general model to several specific cases, including bundles of goods that have i.i.d. (independent and identically distributed) valuations, bundles of complements or substitutes, and bundling in the presence of budget constraints. In each case, we derive an inequality for bundle profits as a function of bundle size. We also consider several types of correlation in the valuations of the information goods. We present discriminating mechanisms that significantly increase the benefits of bundling for goods with other types of correlated demands, provided the source of the underlying correlation can be identified, either directly, or indirectly through consumers' behavior. We also show that mixed bundling can be more profitable than pure bundling when consumer valuations are not drawn from the same distribution, as it induces consumers to self-select. Section 4 compares the implications of our analysis with some empirical evidence and provides some concluding remarks.

2. The Basic Model with Independent Valuations

2.1 Asymptotic results for large bundles

We begin by considering a setting with a single seller providing n information goods to a set of consumers Ω . Each consumer can consume either 0 or 1 units of each information good, and resale is not permitted (or is

unprofitable for consumers).² For each consumer $\omega \in \Omega$, let $v_{ni}(\omega)$ denote the valuation of good i when a total of n goods are purchased. We allow v_{ni} to depend on n so that the distributions of valuations for individual goods can change as the number of goods purchased changes.³ Such a collection of random variables $v_{n1}(\omega), v_{n2}(\omega), \dots, v_{nn}(\omega)$ is sometimes referred to as a triangular array of random variables and can be denoted by $\mathbf{V}_n$.⁴

$$\mathbf{V}_n = \begin{bmatrix} v_{11} & & & \\ v_{21} & v_{22} & & \\ \vdots & & \ddots & \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix} \begin{array}{l} \text{(1-good bundle)} \\ \text{(2-good bundle)} \\ \vdots \\ \text{(n-good bundle)} \end{array}$$

Let $x_n = \frac{1}{n} \sum_{k=1}^n v_{nk}$ be the per-good valuation of the bundle of n information goods. Let p_n^* , q_n^* and π_n^* denote the profit-maximizing price per good for a bundle of n goods, the corresponding sales as a fraction of the population, and the seller's resulting profits per good. Assume the following conditions hold:

A1: The marginal cost for copies of all information goods is zero to the seller.⁵

A2: For all n , consumer valuations v_{ni} are independent and uniformly bounded,⁶ with continuous density functions, non-negative support, mean μ_{ni} and variance σ_{ni}^2 .

A3: Consumers have free disposal. In particular, for all $n > 1$, $\sum_{k=1}^n v_{nk} \geq \sum_{k=1}^{n-1} v_{(n-1)k}$.⁷

Under these conditions, we find that selling a bundle of all n information goods can be remarkably superior to selling the n goods separately. For the distributions of valuations underlying most common demand functions, bundling substantially reduces the average deadweight loss and leads to higher average profits for the seller. As n increases, the seller captures an increasing fraction of the total area under the demand curve,

² We assume that the producers of information goods can use technical, legal and social means to prevent unauthorized duplication and thus remain monopolists. However, Bakos, Brynjolfsson and Lichtman (1999) employ our framework to study a setting where users share the goods. Also, we do not consider strategic interactions with other parties, although our analysis is also applicable to monopolistic competition among providers of similar products. Bakos and Brynjolfsson (1998) extend our framework to study competition among bundlers.

³ For instance, the value of a weather report may be different when purchased alone from its value when purchased together with the morning news headlines, as they both compete for the consumer's limited time. Similarly, other factors such as goods that are complements or substitutes, diminishing returns and budget constraints may affect consumer valuations as additional goods are purchased.

⁴ To simplify the notation, we will omit the argument ω when possible.

⁵ In this article, we will typically use the phrase "information goods" as shorthand for "goods with zero or very low marginal costs of production".

⁶ I.e., $\sup_{n,i,\omega} (v_{ni}(\omega)) < \infty$, for all n, i ($i \leq n$), and $\omega \in \Omega$.

⁷ This assumption implies that adding a good to a bundle cannot reduce the total valuation of the bundle (although it may reduce the mean valuation).

correspondingly reducing both the deadweight loss and consumers' surplus relative to selling the goods separately. More formally:

Proposition 1: Asymptotic Profits, Consumers' Surplus and Efficiency for Bundling

Given assumptions A1, A2 and A3, as n increases, the deadweight loss per good and the consumers' surplus per good for a bundle of n information goods converges to zero, and the seller's profit per good is maximized.

Proof: All proofs are in the Appendix.

The intuition behind Proposition 1 is that as the number of information goods in the bundle increases, the law of large numbers assures that the distribution for the valuation of the bundle has an increasing fraction of consumers with “moderate” valuations near the mean of the underlying distribution. Since the demand curve is derived from the cumulative distribution function for consumer valuations, it becomes more elastic near the mean, and less elastic away from the mean (Figure 1).

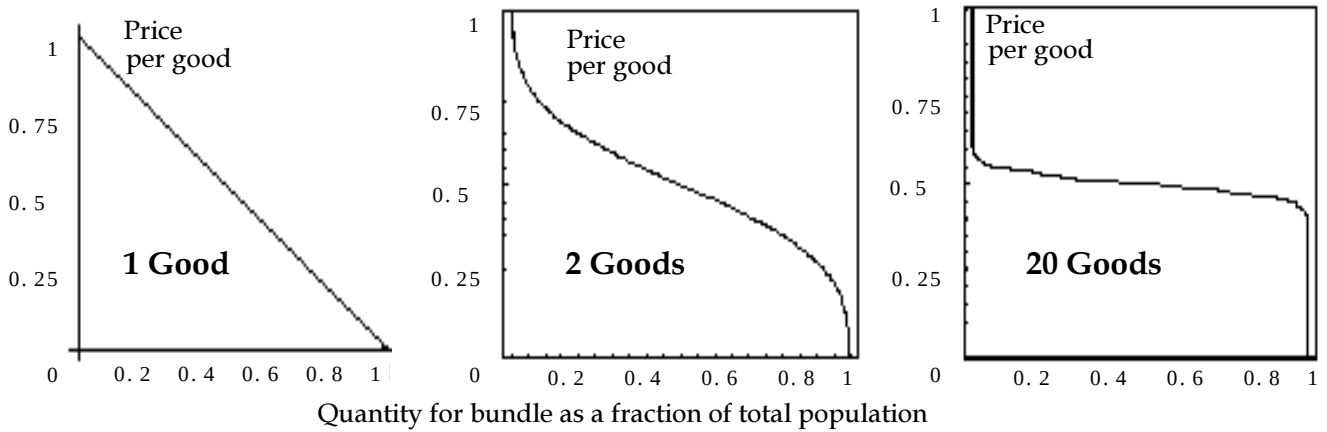


Figure 1: Demand for bundles of 1, 2, and 20 information goods with i.i.d. valuations uniformly distributed in $[0,1]$ (linear demand case).

Proposition 1 is fairly general. While it assumes independence of the valuations of the individual goods in a bundle of a given size, each valuation may be drawn from a different distribution.⁸ Furthermore, valuations may change as more goods are added to a bundle. As we show later, Proposition 1 can be invoked to study several specific settings, such as diminishing returns from the consumption of additional goods, or goods that are substitutes or complements.

⁸ The analysis in this section assumes independence to provide a useful baseline for isolating the effects of bundling. This assumption is likely to be most empirically relevant when consumers differ significantly in their tastes across goods, but not in their total expenditure on the entire set of goods. The case where consumer valuations are correlated is analyzed in section 3.

2.2 The Role of Marginal Costs

In the basic model, we assume that marginal costs are zero. While very large bundles will typically continue to be profitable even in the presence of non-zero (but small) marginal costs, bundling becomes unprofitable for goods with substantial marginal costs. Proposition 2 shows that, as expected, bundling goods with sufficiently high marginal costs is neither profitable nor socially efficient.⁹ Thus, our model predicts that since bits are dramatically cheaper to reproduce than atoms, the optimal bundling strategies differ substantially for information goods as compared to physical goods.

Proposition 2: Marginal Costs Can Make Bundling Unprofitable

Under assumptions A2 and A3, there is a marginal cost $c > 0$ for each information good that makes bundling result in lower profits and higher deadweight loss than selling the goods separately.

As pointed out by Schmalensee (1984), bundling can increase a seller's profits by reducing the dispersion of buyer valuations. However, if marginal costs are large, the seller will usually want to increase, rather than decrease, the dispersion of valuations. For example, if the marginal cost is greater than the mean valuation, bundling will decrease profits because it decreases the fraction of buyers with valuations in excess of the total marginal cost of the bundle. In general, the threshold at which bundling becomes less profitable than unbundled sales depends on the distribution of valuations for the individual goods.

Even with zero marginal costs, the benefits of bundling may be eliminated if the bundle includes goods that have negative value to some consumers (e.g. pornography or advertisements). In addition, while technology is rapidly reducing the marginal costs of reproduction and transmission, the time and energy a user must spend to identify an information good can present a barrier to the limiting result of Proposition 1. For example, Brynjolfsson and Kemerer (1996) find that the cognitive cost of learning the commands accounts for a substantial fraction of the price of spreadsheet software.

2.3 Results for Bundles of Finite Size

While Proposition 1 shows that for a sufficiently large n , selling goods as a bundle can be significantly more profitable than unbundled sales, pure bundling does not necessarily increase profits for small n .¹⁰ In particular, if the seller is able to extract a large fraction of the potential surplus even when the goods are sold

⁹ Salinger (1995) proves a similar proposition for bundling *two* goods. In general, the marginal cost threshold that makes bundling unprofitable at the limit as $n \rightarrow \infty$ is substantially higher for a large bundle than for a bundle of two goods. For example, if valuations are independent and uniformly distributed in $[0, v_{\max}]$, a marginal cost $c > .14v_{\max}$ suffices to make bundling two goods unprofitable, while the threshold becomes almost three times higher ($c > .41v_{\max}$) as $n \rightarrow \infty$. The threshold is much lower when the distribution of valuations has a long tail of low values: e.g. if 99% of the consumers have no value for each good, and the remaining consumers have uniformly distributed valuations in $[0, v_{\max}]$, then bundling is unprofitable for $c > .0041v_{\max}$.

¹⁰ In contrast, McAfee, McMillan and Whinston (1989) find that *mixed* bundling of two goods always dominates unbundled sales when consumer valuations are independent. For a large number of goods and under the conditions for Proposition 1, pure bundling captures nearly the entire value created by the information goods, so mixed bundling cannot do substantially better. However, as shown in sections 2.2 and 3.5, the presence of marginal costs or correlated demands can make mixed bundling substantially more profitable than pure bundling.

separately, then there may be little or no benefit from small amounts of bundling. For example, if consumer valuations for individual goods are i.i.d., taking values either $v_H=10$ with probability $r=.9$, or $v_L=1$ with probability $1-r=.1$, the profit maximizing price $p_1^*=10$ will sell to all high valuation consumers and will extract most potential surplus, as $q_1^*=.9$. A bundle of two goods will have per-good valuations of v_H with probability $r^2=.81$, $\frac{1}{2}(v_H + v_L)=5.5$ with probability $2r(1-r)=.18$, and v_L with probability $(1-r)^2=.01$. The profit maximizing price for this bundle is $p_2^*=10$, and results in sales to a fraction $q_2^*=.81$ of consumers, yielding both lower profits and higher deadweight loss.¹¹

To further study under what conditions bundling will dominate separate sales, even for small n , we assume the following condition that implies a “single crossing” property for the per-good demands $p_n = p_n(q_n)$:

A4: Single-Crossing of Cumulative Distributions Condition (SCDC): The distribution of consumer valuations is such that $\text{Prob}[|x_n - \mu_n| < \varepsilon] \leq \text{Prob}[|x_{n+1} - \mu_{n+1}| < \varepsilon]$ for all n and ε .¹²

In practice, the SCDC is not very restrictive. It holds for most common demand functions, including linear, semi-log and log-log demand, as well as any demand function based on a Gaussian distribution of valuations. Given the SCDC, if it is more profitable to bundle a certain number of goods, say $\hat{n}$, than to sell them separately, and if the optimal price per good for the bundle is less than the mean valuation $\mu_{\hat{n}}$, then bundling any number of goods greater than $\hat{n}$ will further increase profits, compared to the case when the additional goods (or all goods) are sold separately. More formally:

Proposition 3: Monotonic Bundling Profits

Given assumptions A1, A2, A3 and A4, if $\pi_{\hat{n}}^* > \pi_1^*$ and $p_{\hat{n}}^* < \mu_{\hat{n}}$, then bundling any number of goods $n \geq \hat{n}$ will monotonically increase the seller’s profits, compared to selling them separately.

Since the uniform distribution of valuations underlying linear demand satisfies Assumption A4 when the valuations are independent, and since bundling two goods with independent linear demands and zero marginal cost is profit maximizing for the seller (Salinger 1995), the following corollary follows from Proposition 3:

Corollary 3a: With independent linear demands for the individual goods, bundling any number of goods with zero marginal cost increases the seller’s profits.¹³

¹¹ For simplicity, this example uses a discrete distribution of consumer valuations. The results would not be materially effected if a similar continuous distribution were used instead, e.g. a bimodal continuous distribution with a very small probability of any valuation between 0 and 10 and peaks with areas of just under 0.1 at 1 and 0.9 at 10.

¹² The weak law of large numbers requires this probability to decrease as $O(\sigma^2/\varepsilon^2 n)$ for all distributions with finite mean and variance, but it does not guarantee monotonicity.

It is interesting to contrast the bundling approach we analyze here with conventional price discrimination. Suppose there are m consumers in the set Ω . If, as in our setting, each of the consumers potentially has a different value for each of the n goods, then mn prices will be required to capture the complete surplus when the goods are sold separately. Furthermore, price discrimination requires that the seller can accurately identify consumer valuations and prevent consumers from buying goods at prices meant for others. Thus, the conventional approach to price discrimination operates by increasing the number of prices charged to accommodate the diversity of consumer valuations. In contrast, the bundling approach might be called "Procrustean price discrimination" since it operates on a "one-size-fits-all" principle.¹⁴ Bundling reduces the diversity of consumer valuations so that, in the limit, sellers need charge only one price, do not need to identify different types of consumers, and do not need to enforce any restrictions on which prices consumers pay.

As the number of goods in the bundle increases, total profit and profit per good increase. The profit-maximizing price per good for the bundle steadily increases, gradually approaching the per-good expected value of the bundle to the consumers. Figure 2 shows this for the case of linear demand. The number of goods necessary to make bundling desirable, and the speed at which deadweight loss and profit converge to their limiting values, depend on the actual distribution of consumer valuations.

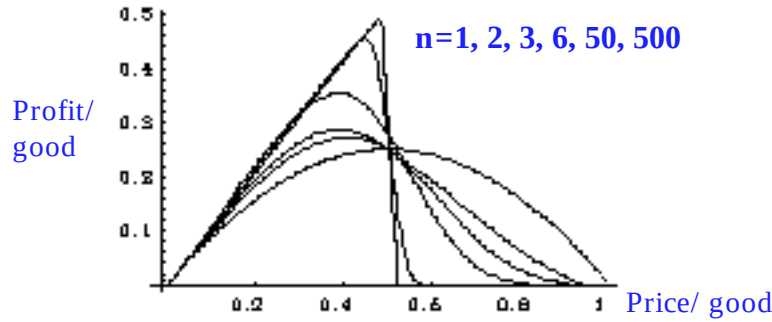


Figure 2: Profit as a function of price per good for bundles of varying number of goods n with identical linear demands and valuations in $[0, 1]$ (steeper curves reflect larger n). The profit-maximizing price is the point at the maximum of each curve. In the limit, the price per good approaches the mean valuation 0.5.

The efficiency and profit gains that bundling offers in our setting contrast with the more limited benefits identified in previous work, principally as a result of our focus on bundling large numbers of goods and on information goods with zero marginal costs. An important implication of our analysis is that the benefits of bundling grow as the number of goods in the bundle increases. This implies a form of superadditivity: bigger bundles will be more profitable than smaller bundles, even when the goods involved are identical.

¹³ It is straightforward to derive analogous corollaries for other common demand functions, such as semilog or cumulative (truncated) normal, that satisfy Assumption A4.

Corollary 3b: If bundles of n_1 goods and n_2 goods are profitable (per Proposition 3), then selling a bundle of $n_1 + n_2$ goods is more profitable than selling two separate bundles of n_1 and n_2 goods respectively.

When n_1 and n_2 are sufficiently large, the central limit theorem guarantees that A4 will hold for any initial demand function for the individual goods, provided that the corresponding distribution of valuations has finite mean and variance; this makes Corollary 3b fairly general.¹⁵

Proposition 3 and Corollary 3b have several implications for marketing strategy and competition. Bundling can create significant economies of scope even in the absence of technological economies in production, distribution, or consumption. In theory, profits under the bundling strategy can be an arbitrarily large multiple of the maximum profits obtainable when the same information goods are sold separately. To see this, assume that demand for the individual goods is approximated by a log-log (constant elasticity) function.¹⁶ If such goods are sold separately, total profits become an arbitrarily small fraction of the area under the demand curve as elasticity increases. In contrast, for a sufficiently large number of goods, Proposition 1 shows that bundling can convert a large fraction of the area under the demand curve into profits. An important empirical implication is that a monopolist selling a low-quality good as part of a bundle may enjoy higher profits and a greater market share than could be obtained by selling a higher-quality good outside the bundle. Bundling low marginal cost goods may therefore lead to "winner-take-all" outcomes similar to those for goods with network externalities or economies of scale in production.

3. Applications of the Basic Model

The basic model introduced in section 2 applies, *inter alia*, to consumers with budget constraints, goods that are complements or substitutes, goods with diminishing or increasing returns, and goods that are drawn from different distributions. In this section, we study each of these cases. To simplify the analysis, we assume that consumer valuations are i.i.d., *conditional on the number of goods in the bundle*. In this case, we can replace Assumption A2 with A2':

A2': For any given n , consumer valuations v_{ni} are independent and identically distributed (i.i.d), with continuous density functions, non-negative support, and finite mean μ_n and variance σ_n^2 .

¹⁴ Procrustes, the mythological Greek king, boasted that his bed would fit perfectly any guest, no matter how large or small. If an unfortunate visitor were too tall, Procrustes would chop off his legs to fit; a guest who was too short would be stretched as needed.

¹⁵ Thus, for large n , Corollary 3b can be seen as an application of Schmalensee's (1984) finding that it is often profitable to bundle two goods with Gaussian demand.

¹⁶ Information goods are often highly valued only by a relatively small set of consumers, which is consistent with such a demand function.

3.1 Minimum Bundle Profits as a Function of Bundle Size

The i.i.d assumption makes it possible to derive some stronger results regarding the "size" of the bundle, which can now be easily indexed by the number of goods, n . While Proposition 1 presents an asymptotic result, moderate-sized bundles suffice for economically-significant effects. We derive an upper bound for the number of goods in the bundle that are needed to enable the seller to capture any given fraction of the total area under the demand curve. Specifically, Corollary 1a provides a useful inequality that follows from the weak law of large numbers as used in the proof of Proposition 1. If the distribution of valuations is symmetric around the mean, a stronger inequality applies, as shown in the proof of the Corollary.

Corollary 1: Bundle Profits Inequality for i.i.d. Valuations

Given assumptions A1, A2' and A3, if the profits per good that can be extracted from a bundle of n goods are denoted by $\pi(n)$, then the following inequality holds:

$$\pi(n) \geq \mu_n \left(1 - 2 \left(\frac{\sigma_n^2 / \mu_n^2}{n} \right)^{\frac{1}{3}} + \left(\frac{\sigma_n^2 / \mu_n^2}{n} \right)^{\frac{2}{3}} \right).$$

In agreement with Proposition 1, as n approaches infinity, the per-good profits approach μ_n , which is the maximum possible value. Minimum profits for smaller bundles are also easy to compute. Figure 3 depicts the minimum profits for a bundle of information goods that have i.i.d. valuations with $\sigma^2 / \mu^2 = \frac{1}{3}$. Three cases are shown: the lower bound for any distribution implied by Corollary 1a, the lower bound for a distribution that is symmetric around the mean, and the actual profits for a normal distribution of valuations. For example, if consumer valuations are i.i.d. and symmetric around the mean with $\sigma^2 / \mu^2 = \frac{1}{3}$, then the seller can realize profits of at least 79% of the total area under the demand curve with a bundle of no more than 100 goods.

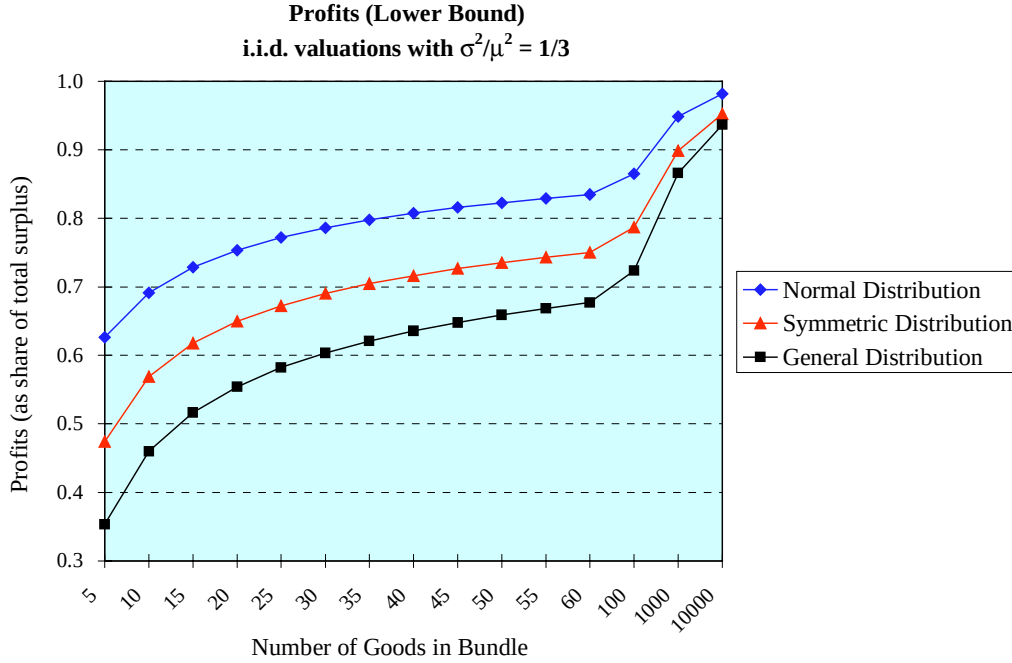


Figure 3: The lower bound on the profits per good from a bundle of i.i.d. information goods increases monotonically with n .

For slightly stronger assumptions about the distribution of consumer valuations, the theory of large deviations (e.g., Chernoff's theorem or Lyapounov's theorem for bounded sequences) provides better estimates of the number of goods needed for a seller to extract as profits a given fraction of the area under the demand curve. If the initial distribution of consumer valuations (and their correlation structure) is known, then our approach allows to explicitly compute the optimal price and bundling strategy, as we did for the case of i.i.d. linear demand in Figure 2. It can be seen from the bundle profits inequality or from Figure 3 that bundling may be unprofitable for a small number of goods, even if it is profitable when a large number of goods are bundled. For example, if the seller were able to extract as profits 80% of the total value through separate sales, then it might require over one hundred such goods to make an equally profitable bundle.

3.2 Bundles of Complements or Substitutes

Many goods are either complements or substitutes, in the sense that a consumer purchasing one good may experience increased utility from the consumption of complementary goods and decreased utility from the consumption of substitute goods. For example, reading successive news stories reviewing yesterday's baseball games is likely to decrease the reader's interest in more stories on the same subject. In such cases, the value of a bundle of goods does not simply equal to the sum of their separate values as assumed above. We now show how our basic model can include complementary and substitute goods. In particular, complementarities and substitution can be modeled by introducing Assumption A3':

A3': For all n, i ($i \leq n$), $v_{ni} = n^\alpha v_{11}$.

In this setting, a bundle of n goods has expected valuation per good $E[x_n] = E\left[\frac{1}{n}\left(\sum_{i=1}^n n^\alpha x_{i1}\right)\right] = n^\alpha \mu_1$. A value of $\alpha < 0$ indicates that the goods are substitutes. For instance, quadrupling the number of songs on a CD might only double its value for the average listener. Similarly, $\alpha > 0$ indicates complementary goods.¹⁷ The following corollary follows:

Corollary 1b (Bundle Profits Inequality for Complementary or Substitute Goods):

Given assumptions A1, A2' and A3', bundling n goods results in profits of π_B^* per good for the seller,

$$\text{where } \pi_B^* \geq n^\alpha \mu_1 \left[1 - 2 \left(\frac{(\sigma_1/\mu_1)^2}{n} \right)^{\frac{1}{3}} + \left(\frac{(\sigma_1/\mu_1)^2}{n} \right)^{\frac{2}{3}} \right].$$

Goods with network externalities exhibit a particularly interesting type of complementarity. For instance, it might be reasonable to treat a copy of Internet videoconferencing software on Alice's computer as a different good from a copy on Bob's computer. In this case, the total value of a set of such goods to the organization that employs Alice, Bob and other workers might be roughly proportional to the potential number of distinct two-way video links enabled as additional copies are purchased, $n(n-1)/2$, or order of n^2 . This property, sometimes referred to as "Metcalfe's Law", can be modeled by setting $\alpha = 2$ in the above setting.

Complementarities can obviously create additional incentives for bundling, and thus can lead to the bundling of goods for reasons that have nothing to do with the reshaping of demand that is modeled in this article (Eppen, Hanson, et al., 1991). In addition to the goods being complements or substitutes, but there may also be costs and benefits associated with producing, distributing or consuming the bundle as a whole, such as economies of scale in creating a distribution channel, administering prices, and making consumers aware of each product's existence. Such economies underlie most large "bundles" of physical goods. For example, technological complementarities affect the collective valuation of the millions of parts flying in close formation that constitute a Boeing 777. Similarly, it is cheaper to physically distribute newspaper or journal articles in "bundles" rather than individually.

One of the effects of the emerging information infrastructure is to dramatically decrease distribution costs for goods that can be delivered over networks. As noted by Metcalfe (1996), Chuang and Sirbu (1997) and others, this may be enough to make it profitable to unbundle certain goods, such as magazine and journal articles, packaged software and songs, to the extent they were formerly bundled simply to reduce distribution costs. The opposing effects on bundling of lower distribution costs due to networking and lower marginal costs due to digitization were first noted by Ward Hanson and are applied to the analysis of bundling, site licensing and subscriptions by Bakos and Brynjolfsson (1997).

¹⁷ There are, of course, many other ways to model complements and substitutes.

3.3 Budget Constraints

When there are explicit or implicit budget constraints, the average variance of valuations for the bundle is likely to decline more rapidly as new goods are added to the bundle. As a result, it may be easier for the seller to predict demand for the bundle, thereby increasing profits and reducing deadweight loss more rapidly. A related implication of monetary and time budget constraints is that the price of a bundle will be bounded even if the seller were to offer access to a practically infinite set of goods.

For instance, assume that the willingness to pay for purchasing a single good is uniformly distributed in $[0, 2\gamma B]$, where B is the total budget and γ is an appropriate scaling constant. For instance, the budget B might reflect the number of hours a consumer is willing and able to spend watching various football games on a Sunday afternoon. In this case, the expected valuation of the j -th good in a bundle can be expressed as $\gamma(1-\gamma)^{j-1}B$. In other words, its value is rescaled in proportion to the available budget.

The *total* valuation of the bundle converges to $\lim_{n \rightarrow \infty} \sum_{j=1}^n \gamma(1-\gamma)^{j-1}B = B$, because of the budget constraint.

While each new good available adds a positive increment to the consumer's utility, the average valuation per good clearly converges to zero as n gets large.

The combination of budget constraints and non-zero marginal costs creates a natural upper bound on the optimal bundle size. Because the expected contribution of each good converges monotonically toward zero as more goods are added to bundle, it will eventually become less than the marginal cost of the good.¹⁸

3.4 Asymmetric Bundling

In practice, information goods will have different means or variances. Even the same information good may have different valuations at different times: a movie or a news story is likely to command higher valuations when first released than a year later. Although Proposition 1 implies that bundling generally increases seller's profits for large numbers of goods with zero marginal cost, it is not always optimal to add an additional information good to a bundle. For instance, if potential surplus can be effectively extracted as profits when a good is sold separately, there is little to be gained by adding it to a bundle, as is the case for goods with only two possible valuations, 0 and v_H (see section 2.3 above). Furthermore, even when adding a good to a bundle does not affect the good's own profitability, it may adversely affect the seller's ability to earn profits on the other goods in the bundle.¹⁹

¹⁸ Other factors reducing the contribution of successive goods, such as substitutability, can limit bundle size as well.

¹⁹ For example, if a good with high variance is added, this may decrease the profitability of the bundle. Adding a new information good i to an existing bundle B will decrease the expected coefficient of variation, if and only if

This may explain why a typical cable TV bundle from providers like HBO or Cinemax offers access to hundreds of movies, but prize fights and other "special events" are typically offered on a "pay-per-view" basis. The cable companies may have established that valuations for the prize fight are concentrated among a small fraction of consumers willing to pay very high prices to watch the fight. Thus, the potential surplus of these consumers can be effectively extracted by selling the prize fight outside the bundle, while including the fight in the regular bundle might increase the bundle's coefficient of variation, σ_i/μ_i .²⁰

3.5 Correlated Demands and Price Discrimination

While Proposition 1 assumes that valuations of the information goods are independent, in practice these valuations may be positively correlated. We now explore how such correlation affects the profit-maximizing strategy of a monopolist who bundles information goods.

In the first case, valuations for the information goods are positively correlated, but not to the same underlying variables. For example, a trader's valuations for a sequence of stock quotations may be serially correlated over time or across industries. If these correlations become lower the more "distant" one gets from the initial topic or item, eventually converging to zero, then the law of large numbers and the central limit theorem apply, as do the limiting results obtained in earlier sections. As a result, the following more general proposition follows from the proof of Proposition 1 and the law of large numbers for stationary (in the wide sense) sequences:²¹

Proposition 1A

The results of Proposition 1 hold if Assumptions A1 and A3 are satisfied, and the sequences of consumer valuations $v_{n1}, v_{n2}, \dots, v_{nn}$ are uniformly bounded, not perfectly correlated, and stationary in the wide sense for all n , with continuous density functions, non-negative support, and finite mean μ_n and variance σ_n .

Thus, bundling of information goods can significantly increase profits even when the valuations of individual goods are highly correlated, but not to the same underlying variables. However, the number of goods required to achieve a given level of profits and efficiency gains generally increases.

$\frac{\sigma_i^2}{\sigma_B^2} + \frac{2 \text{cov}(v_i, v_B)}{\sigma_B^2} < \left(\frac{\mu_2}{\mu_1} \right)^2 + 2 \left(\frac{\mu_2}{\mu_1} \right)$. For example, if the valuations of i and B are uncorrelated and $\mu_i = \mu_B$, the coefficient of variation will decrease if $\sigma_i^2 < 3\sigma_B^2$.

²⁰ Schmalensee (1984) shows that for a Gaussian distribution of consumer valuations, bundling will be profitable when it decreases the coefficient of variation σ_i/μ_i , as long as mean valuations are sufficiently high. While this criterion is also predictive of bundling's profitability for several other distributions, it is not always a sufficient criterion.

²¹ A sequence $\{v_i\}$ is called stationary in the wide sense if $E\{v_i^2\} < \infty$ for all i , and the covariance $\text{cov}(v_{s+i}, v_s)$ does not depend on s . This condition is satisfied, for example, if all v_i are identically distributed with finite mean and variance, and $\rho_{i,j} = \rho^{|j-i|}$ for some ρ in $(0,1)$ and for all i and j .

In the second type of positive correlation, the valuations for all goods are correlated to one or more underlying variables, which can be thought of as characterizing different market segments. For instance, if business users have higher valuations than home users for both a stock quotation and a financial news story, they will also have a higher valuation for a bundle of both these goods. In this case, the distribution of consumer valuations for the bundle does not converge to a Gaussian distribution as more goods are added. Instead, the limiting distribution reflects the mean valuations of each market segment, in this example the probability that a consumer uses the computer for fun or for profit. In general, when valuations are correlated with underlying variables, bundling may not reduce deadweight loss even for very large bundles, and a simple bundling strategy may not be the profit-maximizing strategy for sellers of information goods.

The results of Proposition 1 can be restored if the market can be segmented according to consumer types. The strategy is to create submarkets defined by different values of the underlying variable, so that consumers' demands are independent, conditional to a given value of the underlying variable. Then, the seller offers discounts to consumers in market segments with lower mean valuations. For instance, while home and business users may have different valuations for a bundle, their valuations for individual information goods may be approximately i.i.d. *within* each category of user. By identifying a given consumer's market segment *ex ante*, a seller can maximize profits by offering an appropriately priced bundle for each type of consumer – third degree price discrimination.²²

In principle, demand might be segmented into an arbitrary number of subcategories, with separate demand curves and prices for each subcategory as illustrated in Figure 4. If consumer valuations for individual goods are correlated to a common underlying variable such as consumer type, but are i.i.d. conditional on this variable, then bundling increases profits, reduces deadweight loss, and reduces consumers' surplus if the seller can segment the market through third-degree price discrimination.

²² Such price discrimination is common among software and information vendors (Varian, 1995). For example, Network Associates, Inc. has separate price schedules for home and business users for identical bundles of anti-virus software.

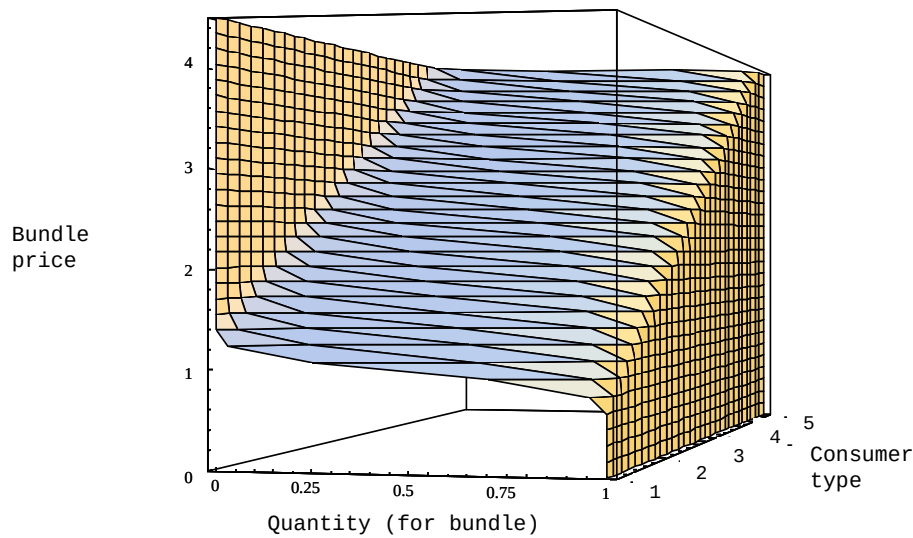


Figure 4: Bundling with third-degree price discrimination

The third-degree price discrimination strategy can be generalized to multiple underlying variables. If a seller segments consumers using one variable, and then finds that consumer valuations remain correlated to a different common variable, the process can be repeated to remove this residual correlation.²³ For instance, it might be possible to segment consumers by business vs. home use, zip code, educational background, age, sex, credit rating, etc, although legal and ethical issues may limit the use of some of this data for price discrimination. Third-degree price discrimination strategies will be facilitated by widespread computer networking and public key encryption and authentication technologies that enable the cost-effective delivery of non-transferable rebate coupons to individual consumers. The rebate amount can be a function of the underlying variables that are correlated with the targeted consumers' expected valuation for the bundle.

In order to execute the above strategy, the seller must be able to charge different prices based on observable characteristics of various market segments. In some cases, this is infeasible. However, in many cases, consumers can be induced to reveal information about their valuations through their choices by offering them a menu of bundles at different prices. For instance, consumers with low valuations may be willing to incur a delay before getting stock market data in exchange for a price discount.²⁴ Thus, consumer behavior can be used to segment the market. In a related strategy, the monopolist may profit by pursuing a mixed bundling strategy of offering several bundles, each including a subset of the available information goods. As with product features, such a menu of bundles can be used to screen consumers by market segment.

²³ This process is somewhat analogous to adding independent variables to a regression equation, or identifying the components of returns in a diversified portfolio of securities (Ross, 1976).

²⁴ See Varian (1996) or Clemons and Weber (1997) for models that use such delays to help segment markets.

Suppose the seller can remove or degrade a feature that disproportionately affects high-demand consumers. This could be achieved, for instance, by delaying stock quotations by 15 minutes. By offering appropriately-priced bundles with and without this feature, the seller may be able to infer consumers' expected valuations for the bundle based on their purchasing decisions. Conditioning on this information can thus eliminate undesired correlations and restore the strong results of proposition 1. Interestingly, while segmenting types is often infeasible for unbundled goods, bundling can create new opportunities for price discrimination by reducing the importance of idiosyncratic factors that add "noise" to the valuations for individual goods.²⁵

Another way to sort customers is to leave certain items out of some bundles. For example, the seller can offer an "economy" bundle that is a subset of the "premium" bundle. Such a mixed bundling strategy forces consumers to signal their valuations by their choice of bundles. While the smaller bundles need not be any less expensive to create or provide, offering them at a reduced price can increase profits by enabling the seller to service low-demand consumers without giving up a lot of potential revenue from high-demand consumers.²⁶

Whether using features or bundle size, the seller's strategy is similar to third-degree discrimination, except that the seller must provide incentives to prevent high-demand consumers from mimicking low-demand consumers when setting the price schedule, as consumers can strategically modify their behavior. This need to maintain incentive compatibility typically reduces the efficiency benefits of bundling as some consumers with low valuations are inefficiently excluded from some goods, and introduces some "rent spillover" as surplus is not completely extracted from some consumers with high valuations (Wilson 1993 ch.10).²⁷

In summary, sellers of information goods will often find it advantageous to segment their markets based on observable characteristics or revealed consumer behavior. This approach can reduce or eliminate the correlation of values by market segment and works synergistically with bundling to increase profits: the two types of price discrimination are more powerful in combination than separately. The optimal strategy will typically involve mixed bundling, the practice offering different bundles to different groups. Accordingly, even if marginal costs are zero, mixed bundling will dominate pure bundling when consumer valuations are correlated with an underlying variable.

4. Implications, Evidence and Conclusions

4.1 Implications for market structure

Our analysis shows that, because of the power of the predictive value of bundling, a multiproduct monopolist of information goods may achieve higher profits and greater efficiency by using a bundling strategy than by

²⁵ See Proposition 4 in the Appendix for a formal statement and proof of this result.

²⁶ See Proposition 5 in the Appendix for a formal statement and proof of this result.

²⁷ As Armstrong (1996) shows, inefficient exclusion of low-demand consumers is also common in multidimensional mechanism design, when consumers' private information (or "type") cannot be captured in a single scalar variable.

selling the goods separately. If it would be difficult (or illegal) for a collection of single-good monopolists to coordinate on a unified bundling strategy and price, our analysis suggests that they may benefit from selling their information goods to a single firm. Similarly, an information good that is unprofitable (net of development costs) if sold separately could become profitable when sold as part of a larger bundle. Thus, bundling confers size-based advantages which are distinct from technological economies of scope, scale or learning (e.g. Spence, 1981) or network externalities (e.g. Farrell and Saloner, 1985).

In addition to having a single firm develop and market a full collection of information goods, a variety of alternative market structures might also emerge. Bundling could be implemented by a broker that remarkets goods produced by information "content" producers. This is essentially the strategy of aggregators like America Online. Alternatively, a consortium or club of consumers could purchase access to a variety of information goods and make them available to all members for a fixed fee. Some user groups and certain site licensing arrangements for software resemble this approach. Finally, the government could fund the creation and distribution of information goods through taxes that do not depend on which individual goods are consumed, but only on access to the whole set. For instance, the United Kingdom funds public television programming via a use tax on television sets. Each of these institutional approaches is likely to involve different marketing strategies and trade-offs.

4.2 Empirical evidence

Our models for bundling information goods can help explain some empirical phenomena. For instance, an interesting contrast in pricing and bundling strategies is evident at commercial sites on the World Wide Web which sell different types of goods. At web sites that sell physical goods like computer accessories, each item is usually associated with a distinct price, while at sites that sell digital information, all of the items displayed are often available when the consumer pays a *single* price for access to the bundle.²⁸ Since both types of companies market their products over the Internet, it is reasonable to assume that they face similar transaction costs. However, the marginal costs of their goods differ markedly and so our theory of bundling as a pricing strategy for information goods provides a clear explanation for the difference in their pricing strategies. It is also interesting to contrast sellers of digital information with physical world newsstands or print publishers. A conventional newsstand or publisher may sell dozens of newspapers and magazines, but they do not typically pursue a bundling strategy the way their online counterparts often do.

Cable television firms also sell goods with nearly zero marginal costs of reproduction. In general, pay-per-view has been less common than bundling-oriented pricing schemes. Typically, a few standard bundles are

²⁸ Contrast for instance, the Internet Shopping Network (<http://internet.net>) and E-library (<http://www.elibrary.com>). Many aspects of these sites are similar, such as the colorful icons representing a variety of products for sale. However, as of 1997 and 1998, Internet Shopping Network charged separately for each connector, modem or computer accessory sold while E-library sold a bundle of 150 newspapers, 800 magazines, 2,000 works of literature,

offered, as predicted by our theory, in an attempt to achieve some degree of price discrimination. For example, these firms typically offer a “basic” bundle from which certain goods are excluded.²⁹ When similar video entertainment is packaged in the form of videocassettes, the marginal costs rise dramatically and bundling vanishes as a pricing strategy. How about the more recent emergence of direct satellite broadcast? Here the marginal cost is again close to zero and bundling again dominates.

Interestingly, Microsoft has often incorporated into its operating systems applications and functionality that were developed by other firms and previously sold separately; this may be consistent with our model. In 1992, Microsoft's Windows operating system incorporated most of the capabilities of Artisoft's Lantastic; in 1993, it incorporated memory management similar to Quarterdeck's QEMM product, disk compression like Stac's Double Space, and faxing like Delrina's Winfax product; and in 1995, email like Lotus's cc:mail (Markoff, 1996). Current versions of Windows include web-browsing software similar to browsers that were previously sold separately. Similarly, Wordperfect and Lotus have also sought to compete by bundling their products with applications that previously were sold separately.

Numerous technologists have predicted that the Internet would lead to the unbundling not only of application suites, but even of the applications themselves. For instance, Metcalfe (1997) writes: "Why should you pay for an unused spelling checker? Why not download a word processor for the evening, with or without fax, into your hotel room's network computer?" While Internet technology is certainly making it much cheaper to deliver and charge for small components of information goods, our analysis suggests that the ultimate equilibrium will include an important role for bundling-based strategies, including mixed bundling and menus of bundles targeted to different types of consumers.

4.3 Concluding remarks

A strategy of selling a bundle of many distinct information goods for a single price often yields higher profits and greater efficiency than selling the same goods separately. The bundling strategy takes advantage of the law of large numbers to “average out” unusually high and low valuations, and can therefore result in a demand curve that is more elastic near the mean valuation of the population and more inelastic away from the mean. As a result of this predictive value of bundling, profits and sales can be increased, even as inefficiency (deadweight loss) is reduced. While the profitability and efficiency benefits of bundling are easiest to quantify when the consumer valuations are identically distributed and not closely correlated for different products, a bundling strategy can be profitable in a variety of situations. For instance, we show how the framework we introduce in this article can be applied to the bundling of complements and substitutes and bundling in the

18,000 photos, and thousands of additional information goods for a fixed price of \$59.95 per year for individual users. They charged other categories of users, such as schools and libraries, different prices for this same bundle.

²⁹ The pay-per-view approach has been used mainly for unusual special events such as boxing matches; this can be explained as a strategy of excluding “big” goods from the bundle and charging for them separately if some aspects of the nature of consumers' demand for these goods is known *a priori*.

presence of budget constraints. Furthermore, when different market segments differ systematically in their average valuations of goods, we find that bundling can make price discrimination profitable even if it would have been unprofitable when selling the goods separately. In general, the predictive value of bundling can be a surprisingly powerful tool for digital goods, not only by itself, but also in leveraging other strategies.

Historically, it has been considered unprofitable and inefficient to bundle together large numbers of unrelated goods. However, our analysis suggests that the increasing availability of digital information goods should increase the significance of pricing strategies that leverage the predictive power of large-scale bundling. Further inquiry may challenge the direct applicability of other long-standing pricing, marketing and distribution principles to the case of digital information goods.

References

- Armstrong, M. "Multiproduct Nonlinear Pricing," *Econometrica*, Vol. 64, No. 1, January 1996, pp. 51-75.
- Adams, W.J. and Yellen, J.L. "Commodity bundling and the burden of monopoly," *Quarterly Journal of Economics* 90 (August): 1976. 475-98.
- Bakos, Y. and Brynjolfsson, E. "Aggregating and Disaggregating Information Goods: Implications for Bundling, Site Licensing and Subscriptions", in *Proceedings of Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property*, Kennedy School of Government, Harvard University, January 1997.
- Bakos, Y. and Brynjolfsson, E. " Bundling and Competition on the Internet: Aggregation Strategies in the Provision of Information Goods." Prepared for the INFORMS Mini-Conference on Marketing Science and the Internet, Cambridge, MA, March 1998.
- Bakos, Y. , Brynjolfsson, E. and Lichtman, D. "Shared Information Goods". *Journal of Law and Economics*, April, 1999.
- Brynjolfsson, E. and C. F. Kemerer, "Network Externalities in Microcomputer Software: An Econometric Analysis of the Spreadsheet Market," *Management Science*, 42(12), pp. 1627-1647, (1996).
- Chuang, J. C.-I., and Sirbu, M. A. (1997). "Network delivery of information goods: Optimal pricing of articles and subscriptions." In *Proceedings of Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property*, Kennedy School of Government, Harvard University, January 1997. Available at <http://www.epp.cmu.edu/~chuang>.
- Clemons, E. K. and B. W. Weber "Information Technology and Screen-Based Securities Trading: Pricing the Stock and Pricing the Trade", *Management Science*, 43(12), pp. 1693-1708, (1997).
- Eppen, G. D., W. A. Hanson, et al. "Bundling-New Products, New Markets, Low Risk." *Sloan Management Review* 32(4): pp. 7-14. (1991)
- Farrell, J. and G. Saloner, "Standardization, compatibility, and innovation," *Rand Journal of Economics*, 16 (1): 442-455, (1985).
- Hanson, W. and Martin, K., "Optimal Bundle Pricing," *Management Science*, 32(2) (February 1990).
- Harmon, S., "Media Comparables Uncovered: What's An Access Subscriber Worth Anyway?", *iWORLD*, (<http://netday.iworld.com/stocks/index.shtml>) (August 20, 1996).
- Markoff, A. "Tomorrow, the World Wide Web." *New York Times*, July 16, 1996, p. D1.
- McAdams, D. L. "Multiproduct Monopoly Bundling." Mimeo, Graduate School of Business, Stanford University (1997).
- McAfee, R.P., McMillan, J., and Whinston, M.D. "Multiproduct monopoly, commodity bundling, and correlation of values." *Quarterly Journal of Economics* 114 (May): 1989. 371-84.
- Metcalfe, R., "A penny for my thoughts is more than I could hope for on the next Internet", *Infoworld*, (January 22, 1996).
- Metcalfe, R., "Pollinate lets you rent the software you need for just the right amount of time", *Infoworld*, (June 9, 1997).

Myerson, R. B., and Satterthwaite, M. A. "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 28, pp. 265-281, (1983).

Ross, Stephen A. "The Arbitrage Theory of Capital Asset Pricing". *Journal of Economic Theory*, 13 (3) December, (1976) 22-531

Salinger, M. A., "A Graphical Analysis of Bundling", *Journal of Business*, 68 (1): 85-98, (1995).

Schmalensee, R.L. "Gaussian demand and commodity bundling." *Journal of Business* 57 (January): 1984. S211-S230.

Spence, A. M., "The Learning Curve and Competition", *Bell Journal of Economics*, 12 49-70, (1981).

Stigler, G.J. "United States v. Loew's, Inc.: A note on block booking." *Supreme Court Review*, 1963, pp. 152-157.

Varian, H. "Pricing Information Goods." *Proceedings of Scholarship in the New Information Environment Symposium*. Harvard Law School. May 1995.

Varian, H. "Differential Pricing and Efficiency", SIMS working paper, Berkeley, June, 1996.

Wilson, R. *Nonlinear Pricing*. Oxford University Press. New York, 1993.

Appendix 1: Proofs of Propositions

Proposition 1

Consider a bundle of n goods with zero marginal cost and independent consumer valuations. Let μ_n and σ_n be the mean and standard deviation for the valuation of the bundle adjusted for n ; i.e.,

$$\mu_n = E[x_n] = E\left[\frac{1}{n} \sum_{k=1}^n v_{nk}\right] \text{ and } \sigma_n^2 = E\left[|x_n - \mu_n|^2\right]. \text{ Let } \lim_{n \rightarrow \infty} \mu_n = \mu \text{ and } \lim_{n \rightarrow \infty} \sigma_n = \sigma \text{ (these limits exist}$$

because the sequences v_{ni} are uniformly bounded). Denote by p_n^* , q_n^* the optimal mean price for the bundle (per good, i.e., adjusted for n) and the corresponding quantity ($0 \leq q_n^* \leq 1$), and let π_n^* be the resulting profits per good $\pi_n^* = p_n^* q_n^*$. Let $\lim_{n \rightarrow \infty} p_n^* = P$ and $\lim_{n \rightarrow \infty} q_n^* = Q$. We show that $P = \mu$ and $Q = 1$. (If these limits do not exist, the same reasoning can be applied to convergent subsequences of $\{p_n^*\}$ and $\{q_n^*\}$, as $\{q_n^*\}$ is bounded, and so is $\{p_n^*\}$ because of the finite variance assumption.)

If $P > \mu$, there exists some $\varepsilon > 0$ such that for all large enough n , $p_n^* > \mu + \varepsilon$. By the weak law of large numbers, $\text{Prob}[|x_n - \mu| < \varepsilon] \geq 1 - \delta$, where $n \geq \frac{\sigma^2}{\varepsilon^2 \delta}$ or $\delta \leq \frac{\sigma^2}{\varepsilon^2 n}$. Thus if $P > \mu$, $\{q_n^*\} \rightarrow 0$, and since $\{p_n^*\}$ is bounded, $\{\pi_n^*\} \rightarrow 0$, which contradicts the optimality of p_n^* and q_n^* .

If $P < \mu$, there exists some $\varepsilon > 0$ such that for all large enough n , $p_n^* < \mu - \varepsilon$. Let $\hat{p}_n = P + \frac{\varepsilon}{2}$, and $\hat{q}_n$ the corresponding quantity. The weak law of large numbers implies that $\lim_{n \rightarrow \infty} q_n^* = \lim_{n \rightarrow \infty} \hat{q}_n = 1$, and $\lim_{n \rightarrow \infty} (q_n^* - \hat{q}_n) = 0$. Since for large enough n , $\hat{p}_n - p_n^* \geq \frac{\varepsilon}{2}$, it follows that $\hat{p}_n \hat{q}_n > p_n^* q_n^*$, which again contradicts the optimality of p_n^* and q_n^* . Thus $\lim_{n \rightarrow \infty} p_n^* = \mu$.

If $Q < 1$, let $Q' = \frac{1}{3}(1 + 2Q)$ and $Q'' = \frac{1}{3}(2 + Q)$, so that $Q < Q' < Q'' < 1$. Since q_n^* converges to Q and $Q < Q'$, there exists some n' such that $q_n^* < Q'$ for all $n > n'$. Choose $\varepsilon > 0$ such that

$$(\mu + \varepsilon)Q' < (\mu - \varepsilon)Q'', \text{ which is satisfied for } 0 < \varepsilon < \frac{1 - Q}{3(1 + Q)}\mu, \text{ and let } q_n^{\mu - \varepsilon} \text{ be the quantity sold at price}$$

$\mu - \varepsilon$. By the weak law of large numbers, $q_n^{\mu - \varepsilon} \geq \text{Prob}[|x_n - \mu| < \varepsilon] \geq 1 - \frac{\sigma^2}{\varepsilon^2 n}$, and thus there exists some n''

such that $q_n^{\mu - \varepsilon} > Q''$ for all $n > n''$. Finally, since p_n^* converges to μ as shown above, there exists some n''' such that $p_n^* < \mu + \varepsilon$ for $n > n'''$. Let $\hat{n} = \max(n', n'', n''')$. Then for $n > \hat{n}$, setting a price $\hat{p} = \mu - \varepsilon$ yields corresponding sales $\hat{q}_n$ and revenues $\hat{p}\hat{q}_n > (\mu - \varepsilon)Q''$. Since ε was chosen so that $(\mu - \varepsilon)Q'' > (\mu + \varepsilon)Q'$, we get $\hat{p}\hat{q}_n > (\mu - \varepsilon)Q'' > (\mu + \varepsilon)Q' > p_n^* Q' > p_n^* q_n^*$, contradicting the optimality of p_n^* and q_n^* .

Corollary 1a

Using the same notation as in Proposition 1, the weak law of large numbers implies that

$q_n^{\mu-\varepsilon} \geq \text{Prob}[|x_n - \mu_n| < \varepsilon] \geq 1 - \delta$ where $\delta \leq \frac{\sigma_n^2}{\varepsilon^2 n}$. Thus $q_n^{\mu-\varepsilon} \geq \text{Prob}[|x_n - \mu_n| < \varepsilon] \geq 1 - \frac{\sigma_n^2}{\varepsilon^2 n}$. Pricing a bundle of n goods at $p_n = \mu_n - \varepsilon$ per good will result in bundle sales $q_n^{\mu-\varepsilon}$. Thus $\pi_n^* \geq (\mu_n - \varepsilon)q_n^{\mu-\varepsilon}$, and it follows that $\pi_n^* \geq (\mu_n - \varepsilon) \left(1 - \frac{\sigma_n^2}{\varepsilon^2 n}\right)$. By choosing $\varepsilon^3 = \frac{\sigma_n^2 \mu_n}{n}$, i.e., $\varepsilon = \sigma_n^{\frac{2}{3}} \mu_n^{\frac{1}{3}} n^{-\frac{1}{3}}$ we get

$$\pi_n^* \geq (\mu_n - \varepsilon) \left(1 - \frac{\varepsilon}{\mu_n}\right) = \mu_n \left(1 - \frac{\varepsilon}{\mu_n}\right)^2, \text{ which implies that } \pi_n^* \geq \mu_n \left[1 - 2 \left(\frac{(\sigma_n/\mu_n)^2}{n}\right)^{\frac{1}{3}} + \left(\frac{(\sigma_n/\mu_n)^2}{n}\right)^{\frac{2}{3}}\right].$$

Note that for distributions of valuations that are symmetric around the mean, half of the probability that

$|x_n - \mu| \geq \varepsilon$ corresponds to values of x_n above μ , which enables us to write $q_n^{\mu-\varepsilon} \geq 1 - \frac{1}{2} \frac{\sigma_n^2}{\varepsilon^2 n}$. In this case,

choosing $\varepsilon^3 = \frac{\sigma_n^2 \mu_n}{n}$ results in $\pi_n^* \geq \mu_n \left(1 - \frac{3\varepsilon}{2\mu_n} + \frac{\varepsilon^2}{2\mu_n^2}\right)$, which can be written as

$$\pi_n^* \geq \mu_n \left[1 - \frac{3}{2} \left(\frac{(\sigma_n/\mu_n)^2}{n}\right)^{\frac{1}{3}} + \frac{1}{2} \left(\frac{(\sigma_n/\mu_n)^2}{n}\right)^{\frac{2}{3}}\right].$$

Proposition 2

If the marginal cost is close enough to the maximum valuation (which is finite as the valuations are uniformly bounded), it is easy to see that bundling even two goods will result in virtually (or exactly) zero sales and profits, as the total marginal costs must be recovered, and only an infinitesimal fraction of consumers will value the bundle above the sum of marginal costs. Separate sale if the goods will still be profitable, however. Since bundling in this case reduces sales to virtually (or exactly) zero, it will increase deadweight loss in addition to reducing profits. Asymptotically, if the marginal cost is higher than the mean valuation, it is easily seen that bundling is unprofitable at the limit as $n \rightarrow \infty$. Separate sales are still profitable as long as *some* consumers' valuations are higher than the marginal cost.

Proposition 3

Using the same notation as in Proposition 1 we assume that, according to Assumption A4, for all integer $n > 0$ and all $\varepsilon > 0$, $\text{Prob}[|x_n - \mu_n| < \varepsilon] \leq \text{Prob}[|x_{n+1} - \mu_{n+1}| < \varepsilon]$. This assumption implies that the quantity of the

bundle of $\hat{n} + 1$ goods sold at price $p_{\hat{n}}^*$ per good will increase compared to the bundle of $\hat{n}$ goods, i.e.,

$q_{\hat{n}+1}(p_{\hat{n}}^*) > q_{\hat{n}}^*$. This guarantees that $\pi_{\hat{n}+1}^* > \pi_{\hat{n}}^*$. Adding the $(\hat{n} + 1)$ th good to the bundle is desirable for the seller, because a bundle of $\hat{n} + 1$ goods is more profitable than a bundle of $\hat{n}$ goods plus a single good sold separately, since $(\hat{n} + 1)\pi_{\hat{n}+1}^* \geq \hat{n}\pi_{\hat{n}}^* + \pi_{\hat{n}}^* \geq \hat{n}\pi_{\hat{n}}^* + \pi_1^*$.

Assumption A5 also implies that $p_{\hat{n}+1}^* < \mu_{n+1}$ (otherwise $p_{\hat{n}}^*$ would not be optimal), allowing the reasoning above to be applied inductively, which proves the proposition for all $n \geq \hat{n}$.

Propositions 4 and 5

The key insights for bundling in the presence of correlated values can be obtained using a simplified setting with two consumer types and two levels of valuation for the information goods:³⁰

- A5: Consumers have either a high or a low valuation for each information good, respectively denoted by v_H and v_L with ($v_H > v_L$).
- A6: There are two types of consumers, denoted by τ_1 and τ_2 , that differ in their probabilities of having a high valuation for each good. We denote by α the fraction of consumers of type τ_1 ; then a fraction $1 - \alpha$ of consumers will be of type τ_2 . Consumers of type τ_1 (the “high demand” consumers) value each good at v_H with probability θ_1 , and at v_L with probability $1 - \theta_1$. Consumers of type τ_2 (the “low demand” consumers) value each good at v_H with probability θ_2 ($\theta_2 < \theta_1$), and at v_L with probability $1 - \theta_2$.

Proposition 4

Given Assumptions A1, A3, A5 and A6, suppose that the seller can “degrade” the information goods by a factor δ ($v_L/v_H < \delta < 1$), so that their value to high-valuation consumers is reduced to δv_H . If $\alpha\theta_1 > \theta_2$ and $\frac{2\alpha(\theta_1 - \theta_2)}{\alpha\theta_1 - \theta_2} < \frac{v_H}{v_L} - 1$, the seller will maximize profits (compared to either separate sales or pure bundling)

by also offering a bundle with $\delta = 1 - \frac{v_L}{v_H} \frac{2\alpha(\theta_1 - \theta_2)}{\alpha\theta_1 - \theta_2}$ targeted to consumers of type τ_2 .

Proof:

Let the seller offer a “full” bundle intended for consumers of type τ_1 at price p_1 , and a “degraded” bundle where the value of goods for the high-valuation consumer is reduced to δv_H , intended for consumers of type τ_2 at price p_2 . As $v_L/v_H < \delta < 1$, it follows that $v_L < \delta v_H < v_H$. The seller faces the following constraints: Participation constraints (i.e., consumers of both types stay in the market): $p_1 \leq \theta_1 v_H + (1 - \theta_1)v_L$ and $p_2 \leq \theta_2 \delta v_H + (1 - \theta_2)v_L$.

Self-selection constraints (i.e., consumers of a certain type must prefer consuming the bundle intended for this type to the bundle intended for the other type): $\theta_1 v_H + (1 - \theta_1)v_L - p_1 \geq \theta_1 \delta v_H + (1 - \theta_1)v_L - p_2$ and $\theta_2 \delta v_H + (1 - \theta_2)v_L - p_2 \geq \theta_2 v_H + (1 - \theta_2)v_L - p_1$.

Rearrange the above inequalities as $p_1 \leq \theta_1 v_H + (1 - \theta_1)v_L$, $p_1 \leq \theta_1 v_H + (1 - \theta_1)v_L - \theta_1 \delta v_H - (1 - \theta_1)v_L + p_2$, $p_2 \leq \theta_2 \delta v_H + (1 - \theta_2)v_L$ and $p_2 \leq \theta_2 \delta v_H + (1 - \theta_2)v_L - \theta_2 v_H - (1 - \theta_2)v_L + p_1$.

The seller would like to choose p_1 and p_2 to be as large as possible, and thus one of the first two inequalities will be binding, and one of the second two inequalities will be binding. Since $\theta_1 > \theta_2$ and $v_L < \delta v_H < v_H$, the binding constraints are $p_1 \leq \theta_1 v_H + (1 - \theta_1)v_L - \theta_1 \delta v_H - (1 - \theta_1)v_L + p_2$ and $p_2 \leq \theta_2 \delta v_H + (1 - \theta_2)v_L$. Since these constraints will be satisfied as equalities, we get

$p_1 = (\theta_1 - \delta\theta_1 + \delta\theta_2)v_H + (1 - \theta_2)v_L$ and $p_2 = \theta_2 \delta v_H + (1 - \theta_2)v_L$. The corresponding profit is $\pi_d^* = \alpha p_1 + (1 - \alpha)p_2 = (\alpha\theta_1 - \alpha\delta\theta_1 + \delta\theta_2)v_H + (1 - \theta_2)v_L$.

Price discrimination must be more profitable for the seller than setting a low price and selling to everyone, or setting a high price and selling only to the high demand consumers, whether the goods are sold separately or

in bundles. We write $\pi_{S1}^* = (\alpha\theta_1 + (1 - \alpha)\theta_2)v_H$, $\pi_{S2}^* = v_L$, $\pi_{B1}^* = \alpha(\theta_1 v_H + (1 - \theta_1)v_L)$,

$\pi_{B2}^* = \theta_2 v_H + (1 - \theta_2)v_L$, $\pi_S^* = \max(\pi_{S1}^*, \pi_{S2}^*)$, $\pi_B^* = \max(\pi_{B1}^*, \pi_{B2}^*)$.

³⁰ This setting can be extended to allow for multiple or continuous consumer types and consumer valuations, and our results qualitatively apply to these cases as well.

Since $\pi_{s2}^* \leq \pi_{B2}^*$, the seller can successfully price discriminate on the feature when $\pi_d^* \geq \max(\pi_{s1}^*, \pi_{B1}^*, \pi_{B2}^*)$.

This is the case when the following three conditions are met:

$\alpha\theta_1 > \theta_2$, $\delta \leq \frac{1-\theta_2}{\alpha\theta_1-\theta_2} \frac{v_L}{v_H} - \frac{(1-\alpha)\theta_2}{\alpha\theta_1-\theta_2}$, and $\delta \leq \left(1 + \frac{1-\alpha}{\alpha\theta_1-\theta_2}\right) \frac{v_L}{v_H}$. The optimal δ is the largest one that satisfies the last two of the above three conditions.

It can be shown that this is true when $\alpha\theta_1 > \theta_2$ and $\frac{2\alpha(\theta_1-\theta_2)}{\alpha\theta_1-\theta_2} < \frac{v_H}{v_L} - 1$; the corresponding optimal value of δ is given by $\delta^* = 1 - \frac{v_L}{v_H} \frac{2\alpha(\theta_1-\theta_2)}{\alpha\theta_1-\theta_2}$.

Proposition 4 implies that a seller can price a bundle contingent on the level of feature δ chosen by each consumer (and the corresponding implied type τ_1 or τ_2), thereby making the bundling strategy profitable relative to separate sales, even when consumers are not homogeneous. Alternatively, the seller may be able to offer a “complete” bundle intended for consumers of type τ_1 at price p_1 per good, and a smaller bundle that contains a fraction β of the goods in the “full” bundle ($0 < \beta < 1$), intended for consumers of type τ_2 at price p_2 per good. In this case Proposition 5 holds:

Proposition 5

Given Assumptions A1, A3, A5 and A6, suppose that $\alpha\theta_1 > \theta_2$, $\frac{1-\alpha}{\alpha\theta_1-\theta_2} - 1 \leq \frac{v_H}{v_L} \leq \frac{1-\alpha}{\alpha\theta_1-\theta_2} + 1$, and

$\frac{\alpha(\theta_1-\theta_2)v_H}{(\alpha\theta_1-\theta_2)(v_H-v_L)-(1-\alpha)v_L} < 1$. In this case, the seller will maximize profits (compared to either separate sales or pure bundling) by also offering a bundle that only includes a fraction β of the information goods targeted to consumers of type τ_2 , where $\beta = 1 - \frac{\alpha(\theta_1-\theta_2)v_H}{(\alpha\theta_1-\theta_2)(v_H-v_L)-(1-\alpha)v_L}$.

Proof:

Participation and self-selection constraints can be derived in a similar way as in the proof of Proposition 6, leading to equilibrium prices

$p_1 = (\theta_1 - \beta\theta_1 + \beta\theta_2)v_H + (1-\theta_1 + \beta\theta_1 - \beta\theta_2)v_L$ and $p_2 = \theta_2\beta v_H + (1-\theta_2)\beta v_L$. The corresponding profit is $\pi_d^* = \alpha p_1 + (1-\alpha)p_2 = (\alpha\theta_1 - \alpha\beta\theta_1 + \beta\theta_2)v_H + (\alpha + \beta - \alpha\beta - \alpha\theta_1 + \alpha\beta\theta_1 - \beta\theta_2)v_L$.

In order for price discrimination to be profitable, it must be preferred by the seller to either selling the goods separately, or selling a single bundle, i.e., $\pi_d^* \geq \max(\pi_{s1}^*, \pi_{B1}^*, \pi_{B2}^*)$.

This is the case when the following three conditions are satisfied:

$\alpha\theta_1 > \theta_2$, $\frac{1-\alpha}{\alpha\theta_1-\theta_2} - 1 \leq \frac{v_H}{v_L} \leq \frac{1-\alpha}{\alpha\theta_1-\theta_2} + 1$, and $\beta \leq 1 - \frac{\alpha(\theta_1-\theta_2)v_H}{(\alpha\theta_1-\theta_2)(v_H-v_L)-(1-\alpha)v_L}$