A Primer on Pricing
Benjamin E. Hermalin
University of California, Berkeley

Contents

1 Purpose .......................... 1

2 Buyers and Demand .................. 1
  2.1 Consumer Demand .................. 2
  2.1.1 Bibliographic Note ............... 4
  2.2 Firm Demand ....................... 4
  2.3 Demand Aggregation ................. 5

3 Simple Monopoly Pricing ............. 6
  3.1 Elasticity and the Lerner Markup Rule ... 7
  3.2 Welfare Analysis ................... 9
  3.3 An Example ........................ 11
  3.4 An Application ..................... 11

4 First-degree Price Discrimination .... 12
  4.1 Two-Part Tariffs ................... 13
  4.1.1 The Two-Instruments Principle .... 15
  4.1.2 Two-Part Tariffs without Apology ... 15
  4.2 Two-Part Tariffs with Heterogeneous Customers ... 17
  4.2.1 Bibliographic Note ............... 19

5 Third-degree Price Discrimination ..... 19
  5.1 Welfare Considerations ............. 20
  5.1.1 Bibliographic Note ............... 21
  5.2 Arbitrage .......................... 21
  5.3 Capacity Constraints ............... 22

6 Second-degree Price Discrimination ... 24
  6.1 Quality Distortions ................. 25
  6.2 Quantity Discounts ................ 27

7 Bundling .......................... 30
  7.0.1 Bibliographic Note ............... 32
1 Purpose

If one had to distill economics down to a single-sentence description, one probably couldn’t do better than describe economics as the study of how prices are and should be set. This primer is primarily focused on the normative half of that sentence, how prices should be set, although I hope it also offers some positive insights as well.

Because I’m less concerned with how prices are set, this primer doesn’t consider price setting by the Walrasian auctioneer or other competitive models. Nor is it concerned with pricing in oligopoly. Our attention will be exclusively on pricing by a single seller who is not constrained by competitive or strategic pressures (e.g., a monopolist).

Now, one common way to price is to set a price, \( p \), per unit of the good in question. So, for instance, I might charge $10 per coffee mug. You can buy as many or as few coffee mugs as you wish at that price. The \textit{revenue} I receive is $10 times the number of mugs you purchase. Or, more generally, at price \( p \) per unit, the revenue from selling \( x \) units is \( px \). Because \( px \) is the formula for a line through the origin with slope \( p \), such pricing is called \textit{linear pricing}.

If you think about it, you’ll recognize that linear pricing is not the only type of pricing you see. Generically, pricing in which revenue is not a linear function of the amount sold is called \textit{nonlinear pricing}.

Examples of nonlinear pricing would be if I gave a 10% discount if you purchased five or more mugs (e.g., revenue is $10x if \( x < 5 \) and $9x if \( x \geq 5 \)). Or if I had a “buy one mug, get one free” promotion (e.g., revenue is $10 if \( x = 1 \) or 2, $20 if \( x = 3 \) or 4, etc.). Or if I gave you a $3-dollar gift with each purchase (e.g., revenue is $10x−3). Alternatively, the price per mug could depend on some other factor (e.g., I offer a weekend discount or a senior-citizen discount). Or I could let you have mugs at $5 per mug, but only if you buy at least $50 worth of other merchandise from my store. Or I could pack 2 mugs in a box with a coffee maker and not allow you to buy mugs separately at all.

2 Buyers and Demand

A seller sets prices and buyers respond. To understand how they respond, we need to know what their objectives are. If they are consumers, the standard assumption is that they wish to maximize utility. If they are firms, the presumption is they wish to maximize profits.

\footnote{Remember in mathematics a function is linear if \( \alpha f(x_0) + \beta f(x_1) = f(\alpha x_0 + \beta x_1) \), where \( \alpha \) and \( \beta \) are scalars. Note, then, that a linear function from \( \mathbb{R} \) to \( \mathbb{R} \) is linear only if it has the form \( f(x) = Ax \).}
2.1 Consumer Demand

In the classic approach to deriving demand, we maximize an individual’s utility subject to a budget constraint; that is,

$$\max_x u(x) \text{ subject to } \mathbf{p} \cdot \mathbf{x} \leq I,$$

where $\mathbf{x}$ is an $N$-dimensional vector of goods, $\mathbf{p}$ is the $N$-dimensional price vector, and $I$ is income. Solving this problem yields the individual’s demand curve for each good $n$, $x^*_n(p_n, \mathbf{p}_{-n}, I)$ (where the subscript $-n$ indicates that it is the $N-1$-dimensional subvector of prices other than the price of the $n$th good). Unfortunately, while this analysis is fine for studying linear pricing, it is hard to utilize for nonlinear pricing because of the income effects that generally exist. In particular, much of the study of nonlinear pricing requires that the inverse of individual demand also represent the marginal benefit curve that the consumer derives from the marginal unit of the good. Unless there are no income effects, this isn’t a feature of demand curves.

For this reason, we will limit attention to quasi-linear utility. Assume that each individual $j$ purchases two goods. The amount of the one in which we’re interested (i.e., the one whose pricing we’re studying) is denoted $x$. The amount of the other good is denoted $y$. We can (and will) normalize the price of good $y$ to 1. If we like, we can consider $y$ to be the amount of consumption other than of good $x$. The utility function is assumed to have the form

$$\hat{u}(x, y) = v(x) + y.$$  

(2)

Because utility is defined only up to an affine transformation, there is no further lose of generality in redefining utility, expression (2), as

$$u(x, y) = v(x) + y - I - v(0),$$  

(3)

where, as above, $I$ is income. With two goods, we can maximize utility by first solving the constraint in (1) for $y$, yielding $y = I - px$ (recall $y$’s price is 1), and then substituting that into the utility function to get an unconstrained maximization problem:  

$$\max_x v(x) - px - v(0).$$  

(4)

Solving, we have the first-order condition

$$\hat{u}'(x) = p.$$  

(5)

Observe (5) also defines the inverse demand curve and, as desired, we have marginal benefit of $x$ equal to inverse demand. If we define $P(x)$ to be the

\footnote{As set forth, for instance, in Mas-Colell et al. (1995) or Varian (1992).}

\footnote{Well, actually, we need to be careful; there is an implicit constraint that $y \geq 0$. In what follows, we assume that this constraint doesn’t bind.}
inverse demand curve, then we have
\[ \int_0^x P(t)dt = \int_0^x v'(t)dt = v(x) - v(0). \]
Substituting this back into (4) we see that utility at the utility-maximizing quantity is equal to
\[ \int_0^x P(t)dt - xP(x). \]
In other words, utility equals the area below the inverse demand curve and above the price of \( x \). See Figure 1. You may also recall that (6) is the formula for consumer surplus (CS).

**Summary.** Given quasi-linear utility, the individual’s inverse demand curve for a good is his or her marginal benefit for that good. Moreover, his or her utility at the utility-maximizing quantity equals (to an affine transformation) his or her consumer surplus (i.e., the area below inverse demand and above the price).

Another way to think about this is to consider the first unit the individual purchases. It provides him or her (approximate) benefit \( v'(1) \) and costs him or her \( p \). His or her surplus or profit is, thus, \( v'(1) - p \). For the second unit the surplus is \( v'(2) - p \). And so forth. Total surplus from \( x \) units, where \( v'(x) = p \), is, therefore,
\[ \sum_{t=0}^x (v'(t) - p); \]
or, passing to the continuum (i.e., replacing the sum with an integral),
\[ \int_0^x (v'(t) - p) \, dt = \int_0^x v'(t) dt - px = \int_0^x P(t) dt - px. \]

Yet another way to think about this is to recognize that the consumer wishes to maximize his or her surplus (or profit), which is total benefit, \( v(x) \), minus his or her total expenditure (or cost), \( px \). As always, the solution is found by equating marginal benefit, \( v'(x) \), to marginal cost, \( p \).

2.1.1 Bibliographic Note

One of the best treatments of the issues involved in measuring consumer surplus can be found in Chapter 10 of Varian (1992). This is a good place to go to get full details on the impact that income effects have on measures of consumer welfare.

Quasi-linear utility allows us to be correct in using consumer surplus as a measure of consumer welfare. But even if utility is not quasi-linear, the error from using consumer surplus instead of the correct measures, compensating or equivalent variation (see Chapter 10 of Varian), is quite small under assumptions that are reasonable for most goods. See Willig (1976). Hence, as a general rule, we can use consumer surplus as a welfare measure even when there’s no reason to assume quasi-linear utility.

2.2 Firm Demand

Consider a firm that produces \( F(x) \) units of a good using inputs \( x \). Let the factor prices be \( p \) and let \( R(\cdot) \) be the revenue function. Then the firm maximizes
\[ R(F(x)) - p \cdot x. \] (7)

The first-order condition with respect to input \( x_n \) is
\[ R'(F(x)) \frac{\partial F}{\partial x_n} - p_n = 0. \] (8)

Let \( x^*[p] \) denote the set of factor demands, which is found by solving the set of equations (8). Define the profit function as
\[ \pi(p) = R(F(x^*[p])) - p \cdot x^*[p]. \]

Utilizing the envelope theorem, it follows that
\[ \frac{\partial \pi}{\partial p_n} = -x^*_n(p_n; p_{-n}). \] (9)

Consequently, integrating (9) with respect to the price of the \( n \)th factor, we have
\[ -\int_{p_n}^{\infty} \frac{\partial \pi(t; p_{-n})}{\partial p_n} \, dt = \int_{p_n}^{\infty} x^*_n(t; p_{-n}) \, dt. \] (10)
Buyers and Demand

The right-hand side of (10) is just the area to the left of the factor demand curve that’s above price $p_n$. Equivalently, it’s the area below the inverse factor demand curve and above price $p_n$. The left-hand side is $\pi(p_n; p_{-n}) - \pi(\infty; p_{-n})$. The term $\pi(\infty; p_{-n})$ is the firm’s profit if it doesn’t use the $n$th factor (which could be zero if production is impossible without the $n$th factor). Hence, the left-hand side is the increment in profits that comes from going from being unable to purchase the $n$th factor to being able to purchase it at price $p_n$. This establishes

**Proposition 1.** The area beneath the factor demand curve and above a given price for that factor is the total net benefit that a firm enjoys from being able to purchase the factor at that given price.

In other words, as we could with quasi-linear utility, we can use the “consumer” surplus that the firm gets from purchasing a factor at a given price as the value the firm places on having access to that factor at the given price.

**Observation.** One might wonder why we have such a general result with factor demand, but we didn’t with consumer demand. The answer is that with factor demands there are no income effects. Income effects are what keep consumer surplus from capturing the consumer’s net benefit from access to a good at its prevailing price. Quasi-linear utility eliminates income effects, which allows us to treat consumer surplus as the right measure of value or welfare.

### 2.3 Demand Aggregation

Typically, a seller sells to more than one buyer. For some forms of pricing it is useful to know total demand as a function of price.

Consider two individuals. If, at a price of $3 per unit, individual one buys 4 units and individual two buys 7 units, then total or aggregate demand at $3 per unit is 11 units. More generally, if we have $J$ buyers indexed by $j$, each of whom has individual demand $x_j(p)$ as a function of price, $p$, then aggregate demand is $\sum_{j=1}^{J} x_j(p) \equiv X(p)$.

How does aggregate consumer surplus (i.e., the area beneath aggregate demand and above price) relate to individual consumer surplus? To answer this, observe that we get the same area under demand and above price whether we integrate with respect to quantity or price. That is, if $x(p)$ is a demand function and $p(x)$ is the corresponding inverse demand, then $\int_0^x (p(t) - p(x))dt = \int_p^\infty x(t)dt$. Consequently, if $CS(p)$ is aggregate consumer surplus and $cs_j(p)$ is
buyer $j$’s consumer surplus, then

$$CS(p) = \int_p^\infty X(t)dt$$

$$= \int_p^\infty \left( \sum_{j=1}^J x_j(t) \right) dt$$

$$= \sum_{j=1}^J \left( \int_p^\infty x_j(t)dt \right)$$

$$= \sum_{j=1}^J cs_j(p) ;$$

that is, we have

**Proposition 2.** Aggregate consumer surplus is the sum of individual consumer surplus.

### 3 Simple Monopoly Pricing

In this section, we consider a firm that sells all units at a constant price per unit. If $p$ is that price and it sells $x$ units, then its revenue is $px$. Such linear pricing is also called simple monopoly pricing.

Assume this firm incurs a cost of $C(x)$ to produce $x$ units. Suppose, too, that the aggregate demand for its product is $X(p)$ and let $P(x)$ be the corresponding inverse demand function. Hence, the maximum price at which it can sell $x$ units is $P(x)$, which generates revenue $xP(x)$. Let $R(x)$ denote the firm’s revenue from selling $x$ units; that is, $R(x) = xP(x)$. The firm’s profit is revenue minus cost, $R(x) - C(x)$. The profit-maximizing amount to sell maximizes this difference.

Assuming $R(x) - C(x)$ is a globally concave function and $R(x) > C(x)$ for some $x > 0$, the profit-maximizing quantity is some positive amount satisfying the first-order condition:

$$R'(x) - C'(x) = 0 ;$$

or, as it is sometimes written,

$$MR(x) = MC(x) ,$$

where $MR$ denotes marginal revenue and $MC$ denotes marginal cost.

Substituting $xP(x)$ for $R(x)$, we find that marginal revenue is

$$MR(x) = P(x) + xP'(x) .$$

Because demand curves slope down, $P'(x) < 0$; hence, $MR(x) < P(x)$ except at $x = 0$ where $MR(0) = P(0)$. See Figure 2. It might, at first, seem that the
Simple Monopoly Pricing

$\$/unit

Figure 2: Relation between inverse demand, $P(x)$, and marginal revenue, $MR$, under linear pricing; and the determination of the profit-maximizing quantity, $x^*_M$, and price, $P(x^*_M)$.

marginal revenue should equal the price received for the last unit sold. But such a naïve view ignores that to sell an additional item requires lowering the price (i.e., recall, $P(x+1) < P(x)$). So marginal revenue has two components: The price received on the marginal unit, $P(x)$, less the revenue lost on the infra-marginal units from having to lower the price, $|x P'(x)|$ (i.e., the firm gets $P'(x)$ less on each of the $x$ infra-marginal units).

Summary. Under simple monopoly pricing, the profit-maximizing quantity, $x^*_M$, solves

$$MR(x) = P(x) + x P'(x) = MC(x). \quad (11)$$

And the monopoly price, $p^*_M$, equals $P(x^*_M)$. Because $P'(x) < 0$, expression (11) reveals that $p^*_M > MC(x^*_M)$; that is, price is marked-up over marginal cost.

3.1 Elasticity and the Lerner Markup Rule

Recall that the elasticity of demand is the percentage change in demand per a one-percentage point change in price. That is,

$$\varepsilon = \left( \frac{\Delta x}{x} \times 100\% \right) \div \left( \frac{\Delta p}{p} \times 100\% \right)$$

$$= \frac{p \Delta x}{x \Delta p}$$

ELASTICITY OF DEMAND
or, passing to the continuum,
\[
\frac{d}{dp} \int p \, dx = \int p \frac{dX'(p)}{dp} \, dp = \int \frac{p}{xP'(x)} \, dx.
\]
(13)

Observe that (12) implies that \( \varepsilon = \frac{d \log x}{d \log p} \). Observe, too, that, because demand is downward sloping, \( \varepsilon < 0 \).

We know that the revenue from selling 0 units is 0. We also know that \( \lim_{x \to \infty} xP(x) = 0 \) because eventually price is driven down to zero. In between these extremes, revenue is positive. Hence, we know that revenue must increase over some range of output and decrease over another. Revenue is increasing if and only if
\[
P(x) + xP'(x) > 0 \quad \text{or} \quad xP'(x) > -P(x).
\]
Divide both sides by \( P(x) \) to get
\[
-1 < \frac{xP'(x)}{P(x)} = \frac{1}{\varepsilon},
\]
(14)

where the equality in (14) follows from (13). Multiplying both sides of (14) by \(-\varepsilon \) (a positive quantity) we have that revenue is increasing if and only if
\[
\varepsilon < -1.
\]
(15)

When \( \varepsilon \) satisfies (15), we say that demand is elastic. When demand is elastic, revenue is increasing with units sold. If \( \varepsilon > -1 \), we say that demand is inelastic. Reversing the various inequalities, it follows that, when demand is inelastic, revenue is decreasing with units sold. The case where \( \varepsilon = -1 \) is called unit elasticity.

Recall that a firm produces the number of units that equates MR to MC. The latter is positive, which means that a profit-maximizing firm engaged in linear pricing operates only on the elastic portion of its demand curve. This makes intuitive sense: If it was on the inelastic portion, then, if it produced less, it would both raise revenue and lower cost; that is, increase profits. Hence, it can’t maximize profits operating on the inelastic portion of demand.

**Summary.** *A profit-maximizing firm engaged in linear pricing operates on the elastic portion of its demand curve.*

Recall the first-order condition for profit maximization, equation (11). Rewrite it as
\[
P(x) - MC(x) = -xP'(x)
\]
and divide both sides by \( P(x) \) to obtain
\[
\frac{P(x) - MC(x)}{P(x)} = -x \frac{P'(x)}{P(x)} = \frac{1}{\varepsilon}.
\]
(16)
where the second equality follows from (13). Expression (16) is known as the Lerner markup rule. In English, it says that the price markup over marginal cost, \( P(x) - MC(x) \), as a proportion of the price is equal to \(-1/\varepsilon\). Hence, the less elastic is demand (i.e., as \( \varepsilon \) increases towards -1), the greater the percentage of the price that is a markup over cost. Obviously, the portion of the price that is a markup over cost can’t be greater than the price itself, which again shows that the firm must operate on the elastic portion of demand.

3.2 Welfare Analysis

Assuming that consumer surplus is the right measure of consumer welfare (e.g., consumers have quasi-linear utility), then total welfare is the sum of firm profits and consumer surplus. Hence, total welfare is

\[
\begin{align*}
\text{profit} &= xP(x) - C(x) + \int_0^x (P(t) - P(x))dt = xP(x) - C(x) + \int_0^x P(t)dt - xP(x) \\
\text{CS} &= \int_0^x P(t)dt - C(x).
\end{align*}
\]

Observe, first, that neither the firm’s revenue, \( xP(x) \), nor the consumers’ expenditure, \( xP(x) \), appear in (17). This is the usual rule that monetary transfers made among agents are irrelevant to the amount of total welfare. Welfare is determined by the allocation of the real good; that is, the benefit, \( \int P(t)dt \), that consumers obtain and the cost, \( C(x) \), that the producer incurs.

Next observe that the derivative of (17) is \( P(x) - MC(x) \). From (11) on page 7, recall that \( P(x^*_M) > MC(x^*_M) \), where \( x^*_M \) is the profit-maximizing quantity produced under linear pricing. This means that linear pricing leads to too little output from the perspective of maximizing welfare — if the firm produced more, welfare would increase.

**Proposition 3.** Under linear pricing, the monopolist produces too little output from the perspective of total welfare.

If we assume, as is typically reasonable given that demand slopes down, that demand crosses marginal cost once from above, then the welfare-maximizing quantity satisfies

\[
P(x) - MC(x) = 0.
\]

Let \( x^*_W \) be the solution to (18). From Proposition 3, \( x^*_W > x^*_M \).

What is the welfare loss from linear pricing? It is the amount of welfare...
The deadweight loss from linear pricing is the shaded triangle.

forgone because only \( x^*_M \) units are traded rather than \( x^*_W \) units:

\[
\left( \int_{x^*_M}^{x^*_W} P(t) \, dt - C(x^*_W) \right) - \left( \int_{x^*_M}^{x^*_W} P(t) \, dt - C(x^*_M) \right)
= \int_{x^*_M}^{x^*_W} P(t) \, dt - \left( C(x^*_W) - C(x^*_M) \right)
= \int_{x^*_M}^{x^*_W} P(t) \, dt - \int_{x^*_M}^{x^*_W} MC(t) \, dt
= \int_{x^*_M}^{x^*_W} (P(t) - MC(t)) \, dt .
\]

The area in (19) is called the deadweight loss associated with linear pricing. It is the area beneath the demand curve and above the marginal cost curve between \( x^*_M \) and \( x^*_W \). Because \( P(x) \) and \( MC(x) \) meet at \( x^*_W \), this area is triangular (see Figure 3) and, thus, the area is often called the deadweight-loss triangle.

The existence of a deadweight-loss triangle is one reason why governments and antitrust authorities typically seek to discourage monopolization of industries and, instead, seek to encourage competition. Competition tends to drive price toward marginal cost, which causes output to approach the welfare-maximizing quantity.

We can consider the welfare loss associated with linear pricing as a motive to change the industry structure (i.e., encourage competition). We — or the

\[\text{DEADWEIGHT LOSS}\]

\[\text{DEADWEIGHT-LOSS TRIANGLE}\]

\[\text{A full welfare comparison of competition versus monopoly is beyond the scope of this primer. See, for instance, Chapters 13 and 14 of Varian (1992) for a more complete treatment.}\]
firm — can also consider it as encouragement to change the method of pricing. The deadweight loss is, in a sense, money left on the table. As we will see, in some circumstances, clever pricing by the firm will allow it to pick some, if not all, of this money up off the table.

3.3 An Example

To help make all this more concrete, consider the following example. A monopoly has cost function \( C(x) = 2x \); that is, \( MC = 2 \). It faces inverse demand \( P(x) = 100 - x \).

Marginal revenue under linear pricing is \( P(x) + xP'(x) \), which equals \( 100 - x + x \times (-1) = 100 - 2x \).\(^5\) Equating \( MR \) with \( MC \) yields 100 - 2x = 2; hence, \( x^*_M = 49 \). The profit-maximizing price is \( 100 - 49 = 51 \).

Profit is \( revenue - cost \); that is, \( 51 \times 49 - 2 \times 49 = 2401 \).

Consumer surplus is \( \int_4^{98} (100 - P(x) - 51) dt \). Total welfare, however, is maximized by equating price and marginal cost: \( P(x) = 100 - x = 2 = MC \). So \( x^*_W = 98 \). Deadweight loss is, thus, \( \int_{49}^{98} \left( \frac{100 - t - 2}{MC} \right) dt = 1200.5 \).

As an exercise, derive the general condition for deadweight loss for affine demand and constant marginal cost (i.e., under the assumptions of footnote 6).

3.4 An Application

We often find simple monopoly pricing in situations that don’t immediately appear to be linear pricing situations. For example, suppose that a risk-neutral seller faces a single buyer. Let the seller have single item to sell (e.g., an artwork). Let the buyer’s value for this artwork be \( v \). The buyer knows \( v \), but the seller does not. All the seller knows is that \( v \) is distributed according to the differential distribution function \( F(\cdot) \). That is, the probability that \( v \leq \hat{v} \) is \( F(\hat{v}) \). Assume \( F'(\cdot) > 0 \) on the support of \( v \). Let the seller’s value for the good — her cost — be \( c \). Assume \( F(c) < 1 \).

Suppose that the seller wishes to maximize her expected profit. Suppose, too, that she makes a take-it-or-leave-it offer to the buyer; that is, the seller quotes a price, \( p \), at which the buyer can purchase the good if he wishes. If he doesn’t wish to purchase at that price, he walks away and there is no trade. Clearly, the buyer buys if and only if \( p \leq v \); hence, the probability of a sale, \( x \), is given by the

\(^5\)Prove that if inverse demand is an affine function, then marginal revenue is also affine with a slope that is twice as steep as inverse demand.

\(^6\)Prove that, if inverse demand is \( P(x) = a - bx \) and \( MC = c \), a constant, then \( x^*_M = \frac{a - c}{2b} \) and \( P(x^*_M) = \frac{a - c}{2} \).

\(^7\)Prove that profit under linear pricing is \( \frac{1}{b} \left( \frac{a - c}{2b} \right)^2 \) under the assumptions of footnote 6.

\(^8\)Prove that consumer surplus under linear pricing is \( \frac{(a - c)^2}{2b} \) under the assumptions of footnote 6.
formula $x = 1 - F(p)$. The use of “$x$” is intentional — we can think of $x$ as the (expected) quantity sold at price $p$. Note, too, that because the formula $x = 1 - F(p)$ relates quantity sold to price charged, it is a demand curve. Moreover, because the probability that the buyer’s value is less than $p$ is increasing in $p$, this demand curve slopes down. Writing $F(p) = 1 - x$ and inverting $F$ (which we can do because it’s monotonic), we have $p = F^{-1}(1 - x) \equiv P(x)$. The seller’s (expected) cost is $cx$, so marginal cost is $c$. The seller’s (expected) revenue is $xP(x)$. As is clear, we have a standard linear-pricing problem. Marginal revenue is

$$P(x) + xP'(x) = F^{-1}(1 - x) + x \left( \frac{-1}{F'(F^{-1}(1 - x))} \right).$$

For example, if $c = 1/2$ and $v$ is distributed uniformly on $[0, 1]$, then $F(v) = v$, $F'(v) = 1$, and $F^{-1}(y) = y$. So $MR(x)$ is $1 - 2x$. Hence, $x^*_M = 1/4$ and, thus, the price the seller should ask to maximize her expected profit is $3/4$. Note that there is a deadweight loss: Efficiency requires that the good change hands whenever $v > c$; that is, in this example, when $v > 1/2$. But given linear pricing, the good only changes hands when $v > 3/4$ — in other words, half the time the good should change hands it doesn’t.

4 First-degree Price Discrimination

We saw in Section 3.2 that linear pricing “leaves money on the table,” in the sense that there are gains to trade — the deadweight loss — that are not realized. There is money to be made if the number of units traded can be increased from $x^*_M$ to $x^*_W$.

Why has this money been left on the table? The answer is that trade benefits both buyer and seller. The seller profits to the extent that the revenue received exceeds cost and the buyer profits to the extent that the benefit enjoyed exceeds the cost. The seller, however, does not consider the positive externality she creates for the buyer (buyers) by selling him (them) goods. The fact that his (their) marginal benefit schedule (i.e., inverse demand) lies above his (their) marginal cost (i.e., the price the seller charges) is irrelevant to the seller insofar as she doesn’t capture any of this gain enjoyed by the buyer (buyers). Consequently, she underprovides the good. This is the usual problem with positive externalities: The decision maker doesn’t internalize the benefits others derive from her action, so she does too little of it from a social perspective. In contrast, were the action decided by a social planner seeking to maximize social welfare, then more of the action would be taken because the social planner does consider the externalities created. The cure to the positive externalities problem is to change the decision maker’s incentives so she effectively faces a decision problem that replicates the social planner’s problem.

One way to make the seller internalize the externality is to give her the social benefit of each unit sold. Recall the marginal benefit of the $x$th unit

[^9]: An alternative approach, which is somewhat more straightforward in this context, is to solve $\max_p (p - c)(1 - F(p))$. 

First-degree Price Discrimination

is \( P(x) \). So let the seller get \( P(1) \) if she sells one unit, \( P(1) + P(2) \) if she sells two, \( P(1) + P(2) + P(3) \) if she sells three, and so forth. Given that her revenue from \( x \) units is \( \int_0^x P(t) dt \), her marginal revenue schedule is \( P(x) \). Equating marginal revenue to marginal cost, she produces \( x^*_W \), the welfare-maximizing quantity.

In general, allowing the seller to vary price unit by unit, so as to march down the demand curve, is impractical. But, as we will see, there are ways for the seller to effectively duplicate marching down the demand curve. When the seller can march down the demand curve or otherwise capture all the surplus, she’s said to be engaging in first-degree price discrimination. One sometimes sees this described as perfect price discrimination.

4.1 Two-Part Tariffs

Consider a seller who faces a single buyer with inverse demand \( p(x) \). Let the seller offer a two-part tariff: The buyer pays as follows:

\[
T(x) = \begin{cases} 
0, & \text{if } x = 0 \\
px + f, & \text{if } x > 0 
\end{cases}
\]  

(20)

where \( p \) is price per unit and \( f \) is the entry fee, the amount the buyer must pay to have access to any units. The scheme in (20) is called a two-part tariff because there are two parts to what the buyer pays (the tariff), the unit price and the entry fee.

The buyer will buy only if \( f \) is not set so high that he loses all his consumer surplus. That is, he buys provided

\[
f \leq \int_0^x (p(t) - p(x)) dt = \int_0^x p(t) dt - xp(x). 
\]  

(21)

Constraints like (21) are known as participation constraints or individual rationality (IR) constraints. These constraints often arise in pricing schemes or other mechanism design. They reflect that, because participation in the scheme or mechanism is voluntary, it must be induced.

The seller’s problem is to choose \( x \) (effectively, \( p \)) and \( f \) to maximize profit subject to (21); that is, maximize

\[
f + xp(x) - C(x) 
\]  

subject to (21). Observe that (21) must bind: If it didn’t, then the seller could raise \( f \) slightly, keeping \( x \) fixed, thereby increasing her profits without violating the constraint. Note this means that the entry fee is set equal to the consumer surplus that the consumer receives. Because (21) is binding, we can substitute it into (22) to obtain the unconstrained problem:

\[
\max_x \int_0^x p(t) dt - xp(x) + xp(x) - C(x).
\]
The first-order condition is \( p(x) = MC(x) \); that is, the profit-maximizing quantity is the welfare-maximizing quantity. The unit price is \( p(x^*_W) \) and the entry fee is \( \int_0^{x^*_W} p(t)dt - x^*_W p(x^*_W) \).

**Proposition 4.** A seller who sells to a single buyer with known demand does best to offer a two-part tariff with the unit price set to equate demand and marginal cost and the entry fee set equal to the buyer’s consumer surplus at that unit price.

Of course, a seller rarely faces a single buyer. If, however, the buyers all have the same demand, then a two-part tariff will also achieve efficiency and allow the seller to achieve the maximum possible profits. Let there be \( J \) buyers all of whom are assumed to have the same demand curve. As before, let \( P(\cdot) \) denote aggregate inverse demand. The seller’s problem in designing the optimal two-part tariff is

\[
\max_{f,x} J f + x P(x) - C(x)
\]

subject to consumer participation,

\[
f \leq cs_j(P(x)),
\]

where \( cs_j(p) \) denotes the \( j \)th buyer’s consumer surplus at price \( p \). Because the buyers are assumed to have identical demand, the subscript \( j \) is superfluous and constraint (24) is either satisfied for all buyers or it is satisfied for no buyer. As before, (24) must bind, otherwise the seller could profitably raise \( f \). Substituting the constraint into (23), we have

\[
\max_x J \times cs(P(x)) + x P(x) - C(x),
\]

which, because aggregate consumer surplus is the sum of the individual surpluses (recall Proposition 2 on page 6), can be rewritten as

\[
\max_x \int_0^{x^*_W} P(t)dt - x^*_W P(x^*_W)
\]

The solution is \( x^*_W \). Hence, the unit price is \( P(x^*_W) \) and the entry fee, \( f \), is

\[
\frac{1}{J} \left( \int_0^{x^*_W} P(t)dt - x^*_W P(x^*_W) \right).
\]

**Proposition 5.** A seller who sells to \( J \) identical buyers does best to offer a two-part tariff with the unit price set to equate demand and marginal cost and the entry fee set equal to \( 1/J \)th of aggregate consumer surplus at that unit price. This maximizes social welfare and allows the seller to capture all of social welfare.
In real life, we see many examples of two-part tariffs. A classic example is an amusement park that charges an entry fee and a per-ride price (the latter, sometimes, being set to zero). Another example is a price for a machine (e.g., a Polaroid instant camera or a punchcard sorting machine), which is a form of entry fee, and a price for an essential input (e.g., instant film or punchcards), which is a form of per-unit price. Because, in many instances, the per-unit price is set to zero, some two-part tariffs might not be immediately obvious (e.g., an annual service fee that allows unlimited “free” service calls, a telephone calling plan in which the user pays so much per month for unlimited “free” phone calls, or amusement park that allows unlimited rides with paid admission).

4.1.1 The Two-Instruments Principle

When the seller was limited to just one price parameter, \( p \) — that is, engaged in linear pricing — she did less well than when she controlled two parameters, \( p \) and \( f \). One way to explain this is that a two-part tariff allows the seller to face the social planner’s problem of maximizing welfare and, moreover, capture all welfare. Since society can do no better than maximize welfare and the seller can do no better than capture all of social welfare, she can’t do better than a two-part tariff in this context.

But this begs the question of why she couldn’t do as well with a single price parameter. Certainly, she could have maximized social welfare; all she needed to do was set \( P(x) = MC(x) \). But the problem with that solution is there is no way for her to capture all the surplus she generates. If she had an entry fee, then she could use this to capture the surplus; but with linear pricing we’ve forbidden her that instrument.

The problem with using just the unit price is that we’re asking one instrument to do two jobs. One is to determine allocation. The other is to capture surplus for the seller. Only the first has anything to do with efficiency, so the fact that the seller uses it for a second purpose is clearly going to lead to a distortion. If we give the seller a second instrument, the entry fee, then she has two instruments for the two jobs and she can “allocate” each job an instrument. This is a fairly general idea — efficiency is improved by giving the mechanism designer more instruments — call this the two-instruments principle.

4.1.2 Two-Part Tariffs without Apology

It might seem that the analysis of two-part tariffs is dependent on our assumption of quasi-linear utility. In fact, this is not the case. To see this, consider a single consumer with utility \( u(x, y) \). Normalize the price of \( y \) to 1. Assume the individual has income \( I \). Define \( Y(x) \) to be the indifference curve that passes through the bundle \((0, I)\); that is, the bundle in which the consumer purchases only the \( y \)-good. See Figure 4. Assume \( MC = c \).

Consider the seller of the \( x \) good. If she imposes a two-part tariff, then she transforms the consumer’s budget constraint to be the union of the vertical line segment \( \{(0, y) | I - f \leq y \leq I\} \) and the line \( y = (I - f) - px, x > 0 \). If we define
\( \bar{y} = I - f \), then this budget constraint is the thick dark curve shown in Figure 4. Given that the consumer can always opt to purchase none of the \( x \) good, the consumer can’t be put below the indifference curve through \((0, I)\); that is, below \( Y(x) \). For a given \( p \), the seller increases profit by raising \( f \), the entry fee. Hence, the seller’s goal is to set \( f \) so that this kinked budget constraint is just tangent to the indifference \( Y(x) \). This condition is illustrated in Figure 4, where the kinked budget constraint and \( Y(x) \) are tangent at \( x^* \). If the curves are tangent at \( x^* \), then

\[
-p = Y'(x^*).
\]  

(25)

At \( x^* \), the firm’s profit is

\[
(p - c)x^* + f
\]

(26)

(recall we’ve assumed \( MC = c \)). As illustrated, \( f = I - \bar{y} \). In turn, \( \bar{y} = Y(x^*) + px^* \). We can, thus, rewrite (26) as

\[
(p - c)x^* + I - Y(x^*) - px^* = -Y(x^*) - cx^* + I.
\]

(27)

Maximizing (27) with respect to \( x^* \), we find that \( c = -Y'(x^*) \). Substituting for \( Y'(x^*) \) using (25), we find that \( c = p \); that is, as before, the seller maximizes profits by setting the unit price equal to marginal cost. The entry fee is \( I - (Y(x^*) + cx^*) \), where \( x^* \) solves \( c = -Y'(x^*) \). Given that \( MC = c \), it is clear this generalizes for multiple consumers.

**Summary.** The conclusion that the optimal two-part tariff with one consumer or homogeneous consumers entails setting the unit price equal to marginal cost is not dependent on the assumption of quasi-linear utility.
As a “check” on this analysis, observe that $Y'(\cdot)$ is the marginal rate of substitution (MRS). With quasi-linear utility; that is, $u(x, y) = v(x) + y$, the MRS is $-v'(x)$. So $x^*$ satisfies $c = -(v'(x)) = P(x)$, where the last equality follows because, with quasi-linear utility, the consumer’s inverse demand curve is just his marginal benefit (utility) of the good in question. This, of course, corresponds to what we found above (recall Proposition 4).

4.2 Two-Part Tariffs with Heterogeneous Customers

The analysis to this point has assumed homogeneous customers. In real life, of course, different customers have different preferences and, thus, different demands. This section briefly considers the design of two-part tariffs when consumers are heterogeneous. One caveat, though: When consumers are heterogeneous, a two-part tariff is typically not the profit-maximizing pricing scheme from a theoretical perspective. Other means of price discrimination typically do better — and we’ll consider some later — however, if the transaction costs associated with these schemes get too large (e.g., because of difficulty monitoring individual consumption or preventing arbitrage), then a two-part tariff could be the best practical solution.

When considering heterogeneous consumers, it is necessary to switch from working with quantity to working with price. Let $x(p, \theta)$ be the individual demand of a buyer whose type is $\theta \in [\theta_0, \theta_1] \subset \mathbb{R}$. The word “type” is economics shorthand for characteristics, such as preferences, that help determine an agent’s actions or behavior. Recall that consumer surplus can be also be found by integrating the area to the left of the demand curve from price to infinity (alternatively, to the price at which demand goes to zero): 

$$cs(p, \theta) \equiv \int_p^\infty x(t, \theta)dt.$$  

(28)

We will impose an order assumption on $cs(p, \cdot)$, namely that it is non-increasing for all $p$. What this means is that, if $\theta > \theta'$, then $cs(p, \theta) \leq cs(p, \theta')$ regardless of $p$. For a fixed price, this would be an innocuous assumption — we could always define an index $\hat{\theta}$ to make this hold for a given price. But assuming it for all prices is a stronger assumption — it has implications for whether and how the demand curves of the different types can cross. As you will see, many pricing (and more general) mechanisms, rely on order assumptions.

Let $\Psi(\hat{\theta})$ denote the number of people in the population whose type is less than or equal to $\hat{\theta}$. Let $\psi(\cdot)$ denote the derivative of $\Psi(\cdot)$. Define $\theta(p, f)$ as follows

$$\bar{\theta}(p, f) = \begin{cases} 
\theta_0, & \text{if } cs(p, \theta_0) < f \\
\max\{\bar{\theta}(p, \theta) = f\}, & \text{if } \{\bar{\theta}(p, \theta) = f\} \neq \emptyset \\
\theta_1, & \text{if } cs(p, \theta_1) \geq f
\end{cases}.$$  

Clearly, the seller would never set $(p, f)$ such that $cs(p, \theta_1) > f$; she could increase profit without changing buying behavior simply by raising $f$. Given
that she’s out of business if \( cs(p, \theta_0) < f \), we see that the only relevant domain of \( \bar{\theta} \) is the middle part of the above expression. Note that means that \( f = cs(p, \theta) \) for some \( \theta \) and we may interpret it as the largest such \( \theta \) for which that equality holds.

Assume that \( MC = c \), a constant. Then the seller’s profit from a two-part tariff \((p, f)\), in which \( f = cs(p, \theta) \), is

\[
\Psi(\theta)cs(p, \theta) + \int_{\theta_0}^{\theta} x(p, t)(p - c)\psi(t)dt. \tag{29}
\]

As a “check,” note that if the population were homogeneous, so \( \theta_0 = \theta_1 \), expression (29) reduces to

\[
J \times cs(p) + Jx(p)(p - c),
\]

which, if we maximize with respect to \( p \), yields the first-order condition

\[
cs'(p) + x(p) + x'(p)(p - c) = 0.
\]

From (28), \( cs'(p) = -x(p) \), so the solution to that first-order condition is \( p = c \); the solution we found in the previous section.

Take \( cs(p, \cdot) \) to be strictly decreasing and differentiable for relevant values of \( p \), so that we can take derivatives of (29) with respect to \( \theta \), as well as \( p \). To determine the appropriate prices, we can differentiate (29) with respect to \( \theta \) and \( p \), set the derivatives equal to zero, and solve the equations for the optimal \( \theta \) and \( p \). At this level of abstraction, we can’t achieve a closed-form solution, but we can gain some insights. Using (28), the derivative with respect to \( p \) is

\[
-\Psi(\theta)x(p, \theta) + \int_{\theta_0}^{\theta} \left( \frac{\partial x(p, t)}{\partial p} (p - c) + x(p, t) \right) \psi(t)dt = 0. \tag{30}
\]

Observe that if \( x(p, t) \) were invariant in \( t \) (i.e., we had homogeneous customers), then we would again get our familiar result that \( p = c \).

We are, however, concerned with the situation in which customers are heterogeneous. We previously assumed that \( \partial cs(p, t)/\partial t < 0 \). Given that \( cs(p, t) = \int_{p}^{\infty} x(q, t) dq \), a sufficient condition for that assumption is that \( x(p, t) \) be decreasing in \( t \). If we assume that condition, then observe we can rewrite (30) as follows

\[
\int_{\theta_0}^{\theta} \frac{\partial x(p, t)}{\partial p} (p - c)\psi(t)dt = \Psi(\theta)x(p, \theta) - \int_{\theta_0}^{\theta} x(p, t)\psi(t)dt < \Psi(\theta)x(p, \theta) - x(p, \theta) \int_{\theta_0}^{\theta} \psi(t)dt = 0. \tag{31}
\]

Because \( \partial x(p, t)/\partial p < 0 \) — demand curves slope down — (31) holds only if \( p > c \). In other words, if the types’ demand curves don’t cross, then the profit-maximizing two-part tariff sets the unit price above marginal cost. Intuitively,
suppose the firm set \( p = c \). Then all its profit comes from the entry fee, which is determined by the smallest consumer surplus of the types served. If customers were homogeneous and it raised \( p \), then the gain in profit per unit sold would just be offset by the reduction in the entry fee. But with heterogeneous customers, the infra-marginal types (i.e., \( t < \theta \)) buy more than the marginal type, so the gain from increasing the profit per unit more than offsets the reduction in the entry fee.

### 4.2.1 Bibliographic Note

For more on two-part and multi-part tariffs, see Wilson (1993). The result that, with heterogeneous consumers, \( p > c \) in a two-part tariff is not a general result (recall we made a number of special assumptions, including, at the end, ruling out crossing demand curves). In some contexts, it is possible to have \( p < c \) — the seller loses on each sale but is more than compensated by the entry fees she collects. For a more in depth discussion see, again, Wilson or Varian (1989).

### 5 Third-degree Price Discrimination

The introduction of heterogeneous customers raises the question of conditioning prices on the customers’ types. Clearly, in the last section, were we able to condition the prices directly on type, \( \theta \), then the optimal solution, assuming the seller knew which type was which, would have been to set \( p = c \) for everyone, but to vary the entry fee by type so that \( f(\theta) = cs(c, \theta) \).

How we condition prices on type depends on whether the seller can observe consumers’ types or not. If she can, then we’re in the world of third-degree price discrimination. If she can’t, then we’re in the world of second-degree price discrimination.

Actually, when economists refer to third-degree price discrimination, what they typically mean is engaging in linear pricing in distinct markets. That is, for example, when a seller sets distinct prices in two geographically distinct markets. Sometimes the markets aren’t geographically distinct; for instance, they might be distinguishable by some observable characteristic such as age, gender, student status, or temporally different markets.\(^{11}\)

Consider a seller who faces \( \Theta \) distinct markets. Assume that a good sold in the \( \theta \)th market cannot be resold in another market (i.e., there is no arbitrage across markets). Let \( P_\theta \) denote inverse demand in the \( \theta \)th market and let \( x_\theta \) be the quantity sold in that market. Then the seller’s problem is

\[
\max_{\{x_1, \ldots, x_\Theta\}} \sum_{\theta=1}^{\Theta} x_\theta P_\theta(x_\theta) - C \left( \sum_{\theta=1}^{\Theta} x_\theta \right).
\]

\(^{10}\)What happened to second-degree price discrimination? Despite the conventional ordering, it makes more sense to cover third-degree price discrimination before second-degree price discrimination.

\(^{11}\)Although pricing differently at different times could also be part of second-degree price discrimination scheme.
Assuming (32) is concave, the solution is given by

\[ P_\theta + x_\theta P'_\theta(x_\theta) - MC \left( \sum_{\theta=1}^{\Theta} x_\theta \right) = 0, \text{ for } \theta = 1, \ldots, \Theta. \]  

(33)

Some observations based on conditions (33):

- If marginal cost is a constant (i.e., \(MC = c\)), then third-degree price discrimination is nothing more than setting optimal linear prices independently in \(\Theta\) different markets.

- If marginal cost is not constant, then the markets cannot be treated independently; how much the seller wishes to sell in one market is dependent on how much she sells in other markets. In particular, if marginal cost is not constant and there is a shift in demand in one market, then the quantity sold in all markets can change.

- Marginal revenue across the \(\Theta\) markets is the same at the optimum; that is, if the seller found herself with one more unit of the good, it wouldn’t matter in which market she sold it.

### 5.1 Welfare Considerations

Does allowing a seller to engage in third-degree price discrimination raise or lower welfare. That is, if she were restricted to set a single price for all markets, would welfare increase or decrease?

We will answer this question for the case in which \(MC = c\) and there are two markets, \(\theta = 1, 2\). Let \(v_\theta(x) = \int_0^x p_\theta(t) dt\); that is, \(v_\theta(x)\) is the gross aggregate benefit enjoyed in market \(\theta\). Welfare is, therefore,

\[ W(x_1, x_2) = v_1(x_1) + v_2(x_2) - (x_1 + x_2)c. \]

In what follows, let \(x^*_\theta\) be the quantity traded in market \(\theta\) under third-degree price discrimination and let \(x^U_\theta\) be the quantity traded in market \(\theta\) if the seller must charge a uniform price across the two markets.\(^{12}\) Because demand curves slope down, \(v_\theta(\cdot)\) is a concave function, which means

\[ v_\theta(x^*_\theta) < v_\theta(x^U_\theta) + v'_\theta(x^U_\theta) \cdot (x^*_\theta - x^U_\theta) \]

\[ = v_\theta(x^U_\theta) + p_\theta(x^*_\theta) \cdot (x^*_\theta - x^U_\theta). \]  

(34)

Likewise,

\[ v_\theta(x^*_\theta) < v_\theta(x^U_\theta) + p_\theta(x^U_\theta) \cdot (x^*_\theta - x^U_\theta). \]  

(35)

\(^{12}\)To determine \(x^U_\theta\), define \(X_\theta(p)\) as demand in market \(\theta\), let \(X(p) = X_1(p) + X_2(p)\) be aggregate demand across the two markets, and let \(P(x) = X^{-1}(p)\) be aggregate inverse demand. Solve \(P(x) + xP'(x) = c\) for \(x\) (i.e., solve for optimal aggregate production assuming one price). Call that solution \(x^*_M\). Then \(x^U_\theta = X_\theta(P(x^*_M))\).
If we let $\Delta x_\theta = x^*_\theta - x^U_\theta$, $p^*_\theta = p_\theta(x^*_\theta)$, $p^U = p_\theta(x^U_\theta)$ (note, by assumption, this price is common across the markets), and $\Delta v_\theta = v_\theta(x^*_\theta) - v_\theta(x^U_\theta)$, then we can combine (34) and (35) as

$$p^U \Delta x_\theta > \Delta v_\theta > p^*_\theta \Delta x_\theta. \quad (36)$$

Going from a uniform price across markets to different prices (i.e., to 3rd-degree price discrimination) changes welfare by

$$\Delta W = \Delta v_1 + \Delta v_2 - (\Delta x_1 + \Delta x_2)c.$$  

Hence, using (36), the change in welfare is bounded by

$$(p^U - c)(\Delta x_1 + \Delta x_2) > \Delta W > (p^*_1 - c)\Delta x_1 + (p^*_2 - c)\Delta x_2. \quad (37)$$

Because $p^U - c > 0$, if $\Delta x_1 + \Delta x_2 \leq 0$, then switching from a single price to third-degree price discrimination must reduce welfare. In other words, if aggregate output falls (weakly), then welfare must be reduced. For example, suppose that $c = 0$ and $X_\theta(p) = a_\theta - b_\theta p$, then $x^*_\theta = a_\theta/2.13$ Aggregate demand across the two markets is $X(p) = (a_1 + a_2) - (b_1 + b_2)p$ and $x^U_1 + x^U_2 = (a_1 + a_2)/2$. This equals $x^*_1 + x^*_2$, so there is no increase in aggregate demand. From (37), we can conclude that third-degree price discrimination results in a loss of welfare relative to a uniform price in this case.

But third-degree price discrimination can also increase welfare. The quickest way to see this is to suppose that, at the common monopoly price, one of the two markets is shut out (e.g., market 1, say, has relatively little demand and no demand at the monopoly price that the seller would set if obligated to charge the same price in both markets). Then, if price discrimination is allowed, the already-served market faces the same price as before — so there’s no change in its consumption or welfare, but the unserved market can now be served, which increases welfare in that market from zero to something positive.

5.1.1 Bibliographic Note
This discussion of welfare under third-degree price discrimination draws heavily from Varian (1989).

5.2 Arbitrage
We have assumed, so far, in our investigation of price discrimination that arbitrage is impossible. That is, for instance, a single buyer can’t pay the entry fee, then resell his purchases to other buyers, who, thus, escape the entry fee. Similarly, a good purchased in a lower-price market cannot be resold in a higher-price market.

13One can quickly verify this by maximizing profits with respect to price. Alternatively, observe that inverse demand is

$$P(x) = \frac{a}{b} - \frac{x}{b}.$$  

Hence, $x^* = a/2$ (see footnote 6 on page 11).
In real life, however, arbitrage can occur. This can make utilizing nonlinear pricing difficult; moreover, the possibility of arbitrage helps to explain why we see nonlinear pricing in some contexts, but not others. For instance, it is difficult to arbitrage amusement park rides to those who haven’t paid the entry fee. But it is easy to resell supermarket products. Hence, we see two-part tariffs at amusement parks, but we typically don’t see them at supermarkets. Similarly, senior-citizen discounts to a show are either handled at the door (i.e., at time of admission), or through the use of color-coded tickets, or through some other means to discourage seniors from reselling their tickets to juniors.

If the seller cannot prevent arbitrage, then the separate markets collapse into one and there is a single uniform price across the markets. The welfare consequences of this are, as shown in the previous section, ambiguous. Aggregate welfare may either be increased or decreased depending on the circumstances. The seller, of course, is made worse off by arbitrage — given that she could, but didn’t, choose a uniform price indicates that a uniform price yields lower profits than third-degree price discrimination.

5.3 Capacity Constraints

Third-degree price discrimination often comes up in the context of discounts for certain groups to some form of entertainment (e.g., a play, movie, or sporting event). Typically, the venue for the event has limited capacity and it’s worth considering the implication that has for third-degree price discrimination.

Consider an event for which there are two audiences (e.g., students and non-students). Assume the (physical) marginal cost of a seat is essentially 0. The number of seats sold if unconstrained would be \( x^*_1 \) and \( x^*_2 \), where \( x^*_\theta \) solves \( P_\theta(x) + xP'_\theta(x) = MC = 0 \).

If the capacity of the venue, \( K \), is greater than \( x^*_1 + x^*_2 \), then there is no problem. As a convention, assume that \( P_2(x^*_2) > P_1(x^*_1) \) (e.g., group 1 are students and group 2 are non-students).

Suppose, however, that \( K < x^*_1 + x^*_2 \). Then a different solution is called for. It might seem, given a binding capacity constraint, that the seller would abandon discounts (e.g., eliminate student tickets), particularly if \( x^*_2 \geq K \) (i.e., the seller could sell out charging just the high-paying group its monopoly price). This view, however, is naïve, as we will see.

The seller’s problem can be written as

\[
\max_{x_1, x_2} x_1P_1(x_1) + x_2P_2(x_2)
\]

(recall we’re assuming no physical costs that vary with tickets sold) subject to

\[
x_1 + x_2 \leq K.
\]

Given that we know the unconstrained problem violates the constraint, the constraint must bind. Let \( \lambda \) be the Lagrange multiplier on the constraint. The
first-order conditions are, thus,

\[ P_1(x_1) + x_1P'_1(x_1) - \lambda = 0 \quad \text{and} \quad P_2(x_2) + x_2P'_2(x_2) - \lambda = 0. \]

Observe that the marginal revenue from each group is set equal to \( \lambda \), the shadow price of the constraint. Note, too, that the two marginal revenues are equal. This makes intuitive sense: What is the marginal cost of selling a ticket to a group-1 customer? It’s the opportunity cost of that ticket, which is the forgone revenue of selling it to a group-2 customer; that is, the marginal revenue of selling to a group-2 customer.

Now we can see why the seller might not want to sell only to the high-paying group. Suppose, by coincidence, that \( x^*_2 = K \); that is, the seller could sell out the event at price \( P_2(x^*_2) \). She wouldn’t, however, do so because

\[ P_1(0) > P_2(x^*_2) + x^*_2P'_2(x^*_2) = 0; \]

(the equality follows from the definition of \( x^*_2 \) given that physical marginal cost is 0). The marginal revenue of the \( K \)th seat, if sold to a group-2 customer, is clearly less than its marginal (opportunity) cost.

As an example, suppose that \( P_1(x) = 40 - x \) and \( P_2(x) = 100 - x \). Suppose \( K = 50 \). You should be able to readily verify that \( x^*_1 = 20 \) and \( x^*_2 = 50 \); that is, the seller could just sell out if she set a price of $50, which would yield sales only to group-2 customers (no group-1 customer would pay $50 for a seat). Her (accounting) profit would be $2500. This, however, is not optimal. Equating the marginal revenues, we have

\[ 40 - 2x_1 = 100 - 2x_2. \]

Substituting the constraint, \( x_1 = 50 - x_2 \), into (38) yields

\[ 40 - 2(50 - x_2) = 100 - 2x_2; \] or

\[ 4x_2 = 160. \]

So, optimally, \( x_2 = 40 \) and, thus, \( x_1 = 10 \). The seller’s profit is \( 40 \times (100 - 40) + 10 \times (40 - 10) = 2700 \) dollars; which, as claimed, exceeds her take from naively pricing only to the group-2 customers.

Although the seller’s profit is greater engaging in third-degree price discrimination (i.e., charging $30 for student tickets and $60 for regular tickets) than it is under uniform pricing (i.e., $50 per ticket), welfare has been reduced. We know this, of course, from the discussion in Section 5.1 — output hasn’t changed (it’s constrained to be 50) — so switching from uniform pricing to price discrimination must lower welfare. We can also see this by considering the last 10 tickets sold. Under uniform pricing, they go to group-2 consumers, whose value for them ranges from $60 to $50 and whose aggregate gross benefit is \( \int_{40}^{50} (100 - t) dt = 550 \) dollars. Under price discrimination, they are reserved for group-1 consumers (students), whose value for them ranges from $40 to $30.
and whose aggregate gross benefit is just \( \int_0^{10} (40 - t) \, dt = 350 \) dollars. In other words, to capture more of the total surplus, the seller distorts the allocation from those who value the tickets more to those who value them less.

6 Second-degree Price Discrimination

In many contexts, a seller knows that different types or groups of consumers have different demand, but she can’t readily identify from which group any given buyer comes. For example, it is known that business travelers are willing to pay more for most flights than are tourists. But it is impossible to know whether a given flier is a business traveler or a tourist.

A well-known solution used by airlines is to offer different kinds of tickets. For instance, because business travelers don’t wish to stay over the weekend or often can’t book much in advance, the airlines charge more for round-trip tickets that don’t involve a Saturday-night stayover or that are purchased within a few days of the flight (i.e., in the latter situation, there is a discount for advance purchase). Observe an airline still can’t observe which type of traveler is which, but by offering different kinds of service it hopes to induce revelation of which type is which. When a firm induces different types to reveal their types for the purpose of differential pricing, we say the firm is engaged in second-degree price discrimination.

Restricted tickets are one example of price discrimination. They are an example of second-degree price discrimination via quality distortions. Other quality distortions examples include:

- **Different classes of service (e.g., first and second-class carriages on trains).** The classic example here is the French railroads in the 19th century, which removed the roofs from second-class carriages to create third-class carriages.

- **Hobbling a product.** This is popular in high-tech, where, for instance, Intel produced two versions of a chip by “brain-damaging” the state-of-the-art chip. Another example is software, where “regular” and “pro” versions (or “home” and “office” versions) of the same product are often sold.

- **Restrictions.** Saturday-night stayovers and advance-ticketing requirements are a classic example. Another example is limited versus full memberships at health clubs.

The other common form of second-degree price discrimination is via quantity discounts. This is why, for instance, the liter bottle of soda is typically less than twice as expensive as the half-liter bottle. Quantity discounts can often be operationalized through multi-part tariffs, so many multi-part tariffs are examples of price discrimination via quantity discounts (e.g., choices in calling plans between say a low monthly fee, few “free” minutes, and a high per-minute charge thereafter versus a high monthly fee, more “free” minutes, and a lower per-minute charge thereafter).
6.1 Quality Distortions

Consider an airline facing $N_b$ business travelers and $N_\tau$ tourists on a given round-trip route. Suppose, for convenience, that the airline’s cost of flying are essentially all fixed costs (e.g., the fuel, aircraft depreciation, wages of a fixed-size crew, etc.) and that the marginal costs per flier are effectively 0. Let there be two possible kinds of round-trip tickets,\textsuperscript{14} restricted (e.g., requiring a Saturday-night stayover) and unrestricted (e.g., no stayover requirements); let superscripts $r$ and $u$ refer to these two kinds of tickets, respectively. Let $\kappa$ denote an arbitrary kind of ticket (i.e., $\kappa \in \{r, u\}$).

A type $\theta$-flier has a valuation (gross benefit or utility) of $v_\kappa^\theta$ for a $\kappa$ ticket. Assume, consistent with experience, that

- $v_u^\theta \geq v_r^\theta$ for both $\theta$ and strictly greater for business travelers. That is, fliers prefer unrestricted tickets \textit{ceteris paribus} and business travelers strictly prefer them.
- $v_b^\kappa > v_\tau^\kappa$ for both $\kappa$. That is, business travelers value travel more.

Clearly, if the airline offered only one kind of ticket, it would offer unrestricted tickets given that they cost no more to provide and they can command greater prices. There are, then, two possibilities with one kind of ticket (linear pricing): (i) sell to both types, which means $p = v_u^\tau$; or (ii) sell to business travelers only, which means $p = v_b^r$. Option (i) is as good or better than option (ii) if and only if

$$N_b v_b^u \leq (N_b + N_\tau) v_u^\tau \quad \text{or, equivalently,} \quad N_b \leq \frac{v_r^\tau}{v_b^u - v_r^\tau} N_\tau. \quad (39)$$

Alternatively, the airline could offer both kinds of tickets (price discriminate via quality distortion). In this case, the restricted ticket will be intended for one type of flier and the unrestricted ticket will be intended for the other. At some level, which kind of ticket goes to which type of flier needs to be determined, but experience tells us, in this context, that we should target the unrestricted ticket to the business traveler and the restricted ticket to the tourist. Let $p^\kappa$ be the price of a $\kappa$ ticket.

The airline’s seeks to do the following:

$$\max_{\{p^r, p^u\}} N_b p^u + N_\tau p^\tau. \quad (40)$$

It, of course, faces constraints. First, fliers need to purchase. This means that buying the kind of ticket intended for you can’t leave you with negative consumer surplus, because, otherwise, you would do better not to buy. Hence, we have the participation or IR constraints:

$$v_b^u - p^u \geq 0 \quad (41)$$
$$v_\tau^r - p^\tau \geq 0. \quad (42)$$

\textsuperscript{14}Assume all travel is round-trip.
But we also have an additional set of constraints because the different types of flier need to be induced to reveal their types by purchasing the kind of ticket intended for them. Such revelation or incentive compatibility constraints are what distinguish second-degree price discrimination from other forms of price discrimination. Moreover, because such constraints are a feature of screening models, they indicate that second-degree price discrimination is a form of screening. The revelation constraints say that a flier type must do better (gain more consumer surplus) purchasing the ticket intended for him rather than purchasing the ticket intended for the other type:

\begin{align*}
v_b^u - p^u &\geq v_b^r - p^r \quad (43) \\
v_r^r - p^r &\geq v_r^u - p^u. \quad (44)
\end{align*}

To summarize, the airline’s problem is to solve (40) subject to (41)–(44). This is a linear programming problem. To solve it, begin by noting that the constraint set is empty if \( v_u^r - v_r^r > v_u^b - v_r^b \) because, then, no price pair could satisfy (43) and (44). Hence, let’s assume:

\begin{equation}
v_u^u - v_r^r \leq v_u^b - v_r^b. \quad (45)
\end{equation}

As you will see, most screening problems require such an order assumption about the marginal utilities. In many contexts, these are referred to as single-crossing conditions because they are or can be interpreted as assumptions about the relative steepness of the types’ indifference curves when they cross.

Visualize the constraints in \((p^r, p^u)\) space (i.e., the space with \(p^r\) on the horizontal axis and \(p^u\) on the vertical axis). There are two constraints on the maximum \(p^u\): It must lie below the horizontal line \(p^u = v_b^u\) and it must lie below the upward-sloping line \(p^u = p^r + v_b^u - v_r^u \equiv \ell(p^r)\). The former condition is just (41) and the latter is (43). From (42), we know we can restrict attention to \(p^r \leq v_r^r\). This means

\begin{equation*}
p^u \leq \ell(p^r) \leq \ell(v_r^r) \\
= v_b^u - (v_r^r - v_r^r) \\
< v_b^u,
\end{equation*}

where the last inequality follows because \(v_b^u > v_r^u\). From this, we see that (43) is a binding constraint, while (41) is slack. If (43) is binding, it’s clear that (44) is slack. Clearly, we can’t set \(p^r = \infty\), so we can conclude that (42) is binding.

**Summary.** The incentive compatibility constraint (43) for the business traveler is binding, but the participation constraint (41) is not. The incentive compatibility constraint (44) for the tourist is slack, but the participation constraint (42) is binding.

If you think about it intuitively, it is the business traveler who wishes to keep his type from the airline. His knowledge, that he is a business traveler, is valuable information because he’s the one who must be induced to reveal
his information; that is, he’s the one who would have incentive to pretend to
be the low-willingness-to-pay other type. Hence, it is not surprising that his
revelation constraint is binding. Along the same lines, the tourist has no in-
centive to keep his type from the airline — he would prefer the airline know he
has a low willingness to pay. Hence, his revelation constraint is not binding,
only his participation constraint is. These are general insights: The type who
wishes to conceal information (has valuable information) has a binding revela-
tion constraint and the type who has no need to conceal his information has
just a binding participation constraint.

As summarized above, we have just two binding constraints and two un-
known parameters, \( p^r \) and \( p^u \). We can, thus, solve the maximization problem
by solving the two binding constraints. This yields \( p^*_r = v^r \) and \( p^*_u = \ell(p^*_r) = v^u - (v^r - v^*_r) \). Note that the tourist gets no surplus, but the business traveler
enjoys \( v^u - v^*_r > 0 \) of surplus. This is a general result: The type with the valu-
able information enjoys some return from having it. This is known as his or her
information rent. A type whose information lacks value fails, not surprisingly;
to capture any return from it.

**Summary.** The business traveler enjoys an information rent. The tourist does
not.

Under price discrimination, the airline’s profit is

\[
N_b \left( v^u - (v^u - v^*_r) \right) + N_r v^*_r. \tag{46}
\]

It is clear that (46) is dominated by uniform pricing if either \( N_b \) or \( N_r \) gets suf-
ficiently small relative to the other. But provided that’s not the case, then (46)
— that is, second-degree price discrimination — can dominate. For instance, if
\( v^u_b = 500, v^u_r = 200, v^u_u = 200, \) and \( v^*_r = 100, \) then \( p^*_u = 100 \) and \( p^*_u = 400. \) If
\( N_b = 60 \) and \( N_r = 70, \) then the two possible uniform prices, \( v^u_b \) and \( v^u_r, \) yield
profits of \( $30,000 \) and \( $26,000, \) respectively; but price discrimination yields a
profit of \( $31,000. \)

Observe, too, that, in this example, going from a world of profit-maximizing
uniform pricing to second-degree price discrimination raises welfare — the tour-
ists would not get to fly under the profit-maximizing uniform price \( ($500), \) but
would with price discrimination. Given that tourists value even a restricted
ticket more than the marginal cost of flying them \( ($0), \) getting them on board
must increase welfare.

### 6.2 Quantity Discounts

Consider two consumer types, 1 and 2, indexed by \( \theta. \) Assume the two types
occur equally in the population. Assume that each consumer has quasi-linear
utility

\[
v(x, \theta) - T,
\]

where \( x \) is consumption of a good and \( T \) is the payment (transfer) from the
consumer to the seller of that good. Assume the following order condition on
marginal utility
\[ \frac{\partial}{\partial \theta} \left( \frac{\partial v(x, \theta)}{\partial x} \right) > 0. \]  (47)

Expression (47) is called a Spence-Mirrlees condition; it is a single-crossing condition. As noted in the previous section, we often impose such an order assumption on the steepness of the indifference curves across types. Another way to state (47) is that the marginal utility of consumption is increasing in type for all levels of consumption.

Although we could analyze the case in which \( v(0, 1) > v(0, 2) \), that case is somewhat messy, so we further assume:

\[ v(0, 1) \leq v(0, 2). \]  (48)

For convenience assume a constant marginal cost, \( c \). Given this, we can consider the seller’s optimal strategy against a representative customer, who is, as previously assumed, as likely to be type 1 as type 2.

In analyzing this problem, we can view the seller’s problem as one of designing two “packages.” One package will have \( x_1 \) units of the good and be sold for \( T_1 \) and the other will have \( x_2 \) units and be sold for \( T_2 \). Obviously, the \( x_\theta \)-unit package is intended for the type-\( \theta \) consumer. (One can think of these as being different size bottles of soda with \( x_\theta \) as the number of liters in the \( \theta \) bottle.) Hence, the seller’s problem is

\[ \max_{\{x_1, x_2, T_1, T_2\}} \frac{1}{2}(T_1 - cx_1) + \frac{1}{2}(T_2 - cx_2) \]  (49)

subject to participation (IR) constraints,

\[ v(x_1, 1) - T_1 \geq 0 \quad \text{and} \quad v(x_2, 2) - T_2 \geq 0, \]  (50)

and subject to revelation (IC) constraints,

\[ v(x_1, 1) - T_1 \geq v(x_2, 1) - T_2 \]  (52)

\[ v(x_2, 2) - T_2 \geq v(x_1, 2) - T_1. \]  (53)

As is often true of mechanism-design problems, it is often easier to work with net utility (in this case, consumer surplus) rather than payments. To that end, let

\[ U_\theta = v(x_\theta, \theta) - T_\theta. \]

Also define

\[ I(x) = v(x, 2) - v(x, 1) \]

\[ = \int_1^2 \frac{\partial v(x, t)}{\partial \theta} dt. \]
Observe then, that
\[ I'(x) = \int_1^2 \frac{\partial^2 v(x,t)}{\partial \theta \partial x} dt > 0, \]
where the inequality follows from the Spence-Mirrlees condition (47). Note, given condition (48), this also implies \( I(x) > 0 \) if \( x > 0 \). The use of the letter “I” for this function is not accidental; it is, as we will see, related to the information rent that the type-2 consumer enjoys.

We can rewrite the constraints (50)–(53) as
\[
\begin{align*}
U_1 &\geq 0 \quad (54) \\
U_2 &\geq 0 \quad (55) \\
U_1 &\geq U_2 - I(x_2) \quad (56) \\
U_2 &\geq U_1 + I(x_1). \quad (57)
\end{align*}
\]

We can also rewrite the seller’s problem (49) as
\[
\max_{\{x_1,x_2,U_1,U_2\}} \frac{1}{2}(v(x_1,1) - U_1 - cx_1) + \frac{1}{2}(v(x_2,2) - U_2 - cx_2). \quad (58)
\]

We could solve this problem by assigning four Lagrange multipliers to the four constraints and crank through the problem. This, however, would be way tedious and, moreover, not much help for developing intuition. So let’s use a little logic first.

• Unless \( x_1 = 0 \), one or both of (56) and (57) must bind. To see this, suppose neither was binding. Then, since the seller’s profits are decreasing in \( U_\theta \), she would make both \( U_1 \) and \( U_2 \) as small as possible, which is to say 0. But given \( I(x_1) > 0 \) if \( x_1 > 0 \), this would violate (57).

• Observe that (56) and (57) can be combined so that \( I(x_2) \geq U_2 - U_1 \geq I(x_1) \). Ignoring the middle term for the moment, the fact that \( I(\cdot) \) is increasing means that \( x_2 \geq x_1 \). Moreover, if \( x_1 > 0 \), then \( U_2 - U_1 \geq I(x_1) > 0 \). Hence \( U_2 > 0 \), which means (55) is slack.

• Participation constraint (54), however, must bind at the seller’s optimum. If it didn’t, then there would exist an \( \varepsilon > 0 \) such that \( U_1 \) and \( U_2 \) could both be reduced by \( \varepsilon \) without violating (54) or (55). Since such a change wouldn’t change the difference in the \( U \)s, this change also wouldn’t lead to a violation of the ic constraints, (56) and (57). But from (58), lowering the \( U \)s by \( \varepsilon \) increases profit, so we’re not at an optimum if (54) isn’t binding.

• We’ve established that, if \( x_1 > 0 \), then (54) binds, (55) is slack, and at least one of (56) and (57) binds. Observe that we can rewrite the ic constraints as \( I(x_2) \geq U_2 \geq I(x_1) \). The seller’s profit is greater the smaller is \( U_2 \), so it is the lower bound in this last expression that is important. That is, (57) binds. Given that \( I(x_2) \geq I(x_1) \), as established above, we’re thus free to ignore (56).
So our reasoning tells us that, provided \( x_1 > 0 \), we need only pay attention to two constraints, (54) and (57). Using them to solve for \( U_1 \) and \( U_2 \), we can turn the seller’s problem into the following unconstrained problem:

\[
\max_{\{x_1, x_2\}} \frac{1}{2} (v(x_1, 1) - cx_1) + \frac{1}{2} (v(x_2, 2) - I(x_1) - cx_2).
\]

The first-order conditions are:

\[
\frac{\partial v(x_1^*, 1)}{\partial x} - I'(x_1^*) - c = 0
\]

\[
\frac{\partial v(x_2^*, 2)}{\partial x} - c = 0.
\]

Note that (61) is the condition for maximizing welfare were the seller selling only to type-2 customers; that is, we have efficiency in the type-2 “market.” Because, however, \( I'(\cdot) > 0 \), we don’t have the same efficiency vis-à-vis type-1 customers; in the type-1 “market,” we see too little output relative to welfare-maximizing amount. As we will see, when we study mechanism design more generally, this is a standard result — efficiency at the top and distortion at the bottom.

To make this more concrete, suppose \( v(x, \theta) = 5(\theta + 1) \ln(x + 1) \) and \( c = 1 \). Then \( x_2^* = 14 \) and \( x_1^* = 4 \). Consequently, \( T_1 \approx 16.1 \) and \( T_2 = v(x_2^*, 2) - I(x_1^*) \approx 32.6 \). Note the quantity discount: A type-2 consumer purchases more than three times as much, but pays only roughly twice as much as compared to a type-1 consumer.

### 7 Bundling

Often we see goods sold in packages. For instance, a CD often contains many different songs. A restaurant may offer a prix fixe menu that combines an appetizer, main course, and dessert. Theater companies, symphonies, and operas may sell season tickets for a variety of different shows. Such packages are called bundles and the practice of selling such packages is called bundling.

In some instances, the goods are available only in the bundle (e.g., it may be impossible to buy songs individually). Sometimes the goods are also available individually (e.g., the restaurant permits you to order à la carte). The former case is called pure bundling, the latter case is called mixed bundling.

Why bundle? One answer is it can be a useful competitive strategy; for instance, it is claimed that the advent of Microsoft Office, which bundled a wordprocessor, spreadsheet program, database program, presentation program, etc., helped Microsoft “kill off” strong competitor products that weren’t bundled (e.g., WordPerfect, Lotus 1-2-3, Harvard Graphics, etc.). See Nalebuff (2000) for details.

Another answer, and one relevant to this primer, is that it can help price discriminate. To see this, suppose a Shakespeare company will produce two plays, a comedy and a tragedy, during a season. Type-C consumers tend to
prefer comedies and, thus, value the comedy at $40 and the tragedy at $30. Hence, a type-C consumer will pay $70 for a season ticket (i.e., access to both shows). Type-D consumers tend to prefer dramas and, thus, value the comedy at $25 and the tragedy at $45. Hence, a type-D consumer will pay $70 for a season ticket. Assume no capacity constraint and a constant marginal cost, which, for convenience, we will normalize to 0. Let $N_{\theta}$ denote the number of type-$\theta$ theater goers. If the company sold the shows separately, then its profit is

$$\max\{25(N_C + N_D), 40N_C\} + \max\{30(N_C + N_D), 45N_D\} < 70(N_C + N_D).$$

But if the theater company sold season tickets, it would get $70 from both types and this, as indicated in (62), would yield greater profit. This is an example in which pure bundling does better than selling the goods separately.

For an example where mixed bundling is optimal, change the assumptions so that type-D consumers are now only willing to pay $20 for the comedy. If $5N_C > 20N_D$ (i.e., type-C consumers are more than 80% of the market), then the profit-maximizing solution is to sell season tickets for $70, but now make the tragedy available separately for $45.

Observe how the negative correlation between preferences for comedy versus tragedy helps the theater company price discriminate. Effectively, this negative correlation can be exploited by the company to induce the two types to reveal who they are for the purpose of price discrimination. It follows that bundling is related to the forms of second-degree price discrimination considered earlier, particularly quantity discounts.

As a more general treatment, assume that each customer has a value $v_1$ for good 1 and $v_2$ for good 2. Assume that these values are drawn independently for each consumer from a uniform distribution on [0, 1]. For convenience, again assume $MC = 0$. Consider the firm pricing against a single customer. If it sells the two goods separately, at $p_1$ and $p_2$, its expected profit is

$$p_1(1 - p_1) + p_2(1 - p_2).$$

It is readily shown that the profit-maximizing prices are $p_1^* = p_2^* = 1/2$, which yields an expected profit of 1/2.

If the firm bundles the two goods, at price $p_b$ for the bundle, then it sells the bundle if and only if $v_1 + v_2 \geq p_b$. If $p_b \leq 1$, then the probability that it fails to sell to the consumer is the probability that $(v_1, v_2)$ lies in the lower left-hand corner of the unit square below the isoquant $v_1 + v_2 = p_b$; that is, the probability that $(v_1, v_2)$ is in the set \{$(v_1, v_2)|0 \leq v_1, 0 \leq v_2, \text{ and } v_1 + v_2 \leq p_b$\}. Given the assumption of independent uniform distributions, this is just the area of the corner (triangle): $\frac{1}{2}p_b^2$. Thus the probability of selling the bundle if $p_b \leq 1$ is $1 - \frac{1}{2}p_b^2$. Expected profit is

$$p_b \left(1 - \frac{1}{2}p_b^2\right).$$

Calculations reveal this is maximized by \( p_b = \sqrt{\frac{2}{3}} \approx .82 \), which yield an expected profit of approximately .54. Note that such a \( p_b \) yields greater expected profits than selling the goods separately; that is, we’ve already established that pure bundling is superior to selling the goods separately.

We do, however, also need to consider the possibility of setting \( p_b > 1 \). In this case, the probability of a sale is the area of the upper right-hand corner of the unit square; that is, \( \frac{1}{4} (2-p_b)^2 \). Expected profit is \( p_b (2-p_b)^2 / 2 \). It’s derivative is \( (2-p_b)(1-\frac{3}{2}p_b) \leq 0 \) for \( 2 \geq p_b > 1 \) (note, obviously, \( p_b \leq 2 \)). Expected profit evaluated at \( p_b = 1 \) is \( 1/2 \). Hence, we can conclude that the profit-maximizing price for the bundle is \( \sqrt{\frac{2}{3}} \), which, as we’ve already seen, yields greater profits than selling the goods separately.

We’ve just seen that pure bundling beats selling the goods separately. What about mixed bundling? Suppose the firm sells both goods separately at \( p_1 \) and \( p_2 \) respectively and \( p_b \) for the bundle. Obviously, no one buys the bundle if \( p_b > p_1 + p_2 \), so we assume \( p_b \leq p_1 + p_2 \). No one buys a separate good if \( p_1 > p_b \) and \( p_2 > p_b \), so we assume that \( p_n \leq p_b \), where \( n = 1, 2 \). Note that the customer would prefer to buy just good \( n \) if \( v_n - p_n > v_n + v_m - p_b \); that is, if \( v_m < p_b - p_n \), where \( n = 1, 2, m = 1, 2, \) and \( m \neq n \). We can thus divide the unit square into four regions:

Region 1 = \{ \( (v_1, v_2) | v_1 \geq p_1 \) and \( v_2 \leq p_b - p_1 \) \}

Region 2 = \{ \( (v_1, v_2) | v_2 \geq p_2 \) and \( v_1 \leq p_b - p_2 \) \}

Region 3 = \{ \( (v_1, v_2) | v_1 \geq p_b - p_2, v_2 \geq p_b - p_1, \) and \( v_1 + v_2 \geq p_b \) \}

Region no sale = the unit square minus Regions 1, 2, and 3

A little algebra reveals that expected profit is

\[
(1 - p_1)(p_b - p_1)p_1 + (1 - p_2)(p_b - p_1)p_2 + p_b \left( \frac{1}{2}(p_1 + p_2 - p_b)^2 + (1 - p_1)(1 + p_1 - p_b) \right) + (1 - p_2)(1 + p_2 - p_b) - (1 - p_1)(1 - p_2).
\]

Tedious calculations then reveal \( p_1 = p_2 = \frac{2}{3} \) and \( p_b = \frac{1}{3}(4 - \sqrt{2}) \approx .86 \). Expected profits are approximately .55; slightly greater than under pure mixing. So the optimal pricing is mixed bundling in this case.

7.0.1 Bibliographic Note

A good, short, exposition on bundling can be found in Varian (1989) (although, be warned, the figure illustrating mixed bundling is misleading).
References


