6

Price Discrimination and Monopoly: Non-linear Pricing

If you buy the *New Yorker* magazine at the newsstand you will pay \$4.60 per issue, or \$216.20 if you buy all 47 issues. If instead you purchase an annual subscription you will pay \$47 for 47 issues—a saving of nearly 80 percent over the newsstand price. Similarly, if you are a baseball fan you will find out that the price per ticket on a season pass is much less than the price per ticket on a game-by-game basis. Likewise, when you go grocery shopping you will discover that a 24-pack of Coca-Cola costs less on a price per can basis than a six-pack or than a single can. These are all examples of price discrimination that reflect quantity discounts—the more you buy the cheaper it is on a per unit basis.

Quantity discounting is really a way of saying that the pricing is non-linear. The price per unit is not constant but varies with some feature of the buying arrangement depending, perhaps, on the consumer's income, value of time, the quantity bought or other characteristics. Such a pricing strategy is different from the linear price discrimination methods discussed in Chapter 5. Yet, similar to linear price discrimination, the goal of nonlinear pricing techniques is again to capture as much of the individual consumer's willingness to pay in the seller's revenues and profits. We shall see that such techniques are generally more profitable than the third-degree price discrimination or linear pricing, precisely because they permit the seller to set a price closer to willingness to pay *of each consumer*. As a result, nonlinear pricing can help the monopolist earn more profit.

The design and implementation of nonlinear pricing strategies are the focus of this chapter. We shall explore how a firm with monopoly power can implement such pricing schemes, to a greater or lesser extent, along with the welfare implications of such pricing. Traditionally, non-linear pricing is divided into two general categories called first-degree price discrimination and second-degree price discrimination or, as Shapiro and Varian (1999) categorize them, personalized pricing and menu pricing.

6.1 FIRST-DEGREE PRICE DISCRIMINATION OR PERSONALIZED PRICING

First degree or perfect price discrimination is practiced when the monopolist is able to charge the maximum price each consumer is willing to pay for *each* unit of the product sold. Suppose that you just inherited five antique cars, each a classic Ford Model T, and you want to sell them to finance your college education. They are of no other value to you. Your own market research tells you that there are several collectors interested in buying a model T. When you rank these collectors in terms of their willingness to pay for a car, you estimate that the keenest collector is willing to pay up to $10,000, the second up to $8,000, the third up to $6,000, the fourth $4,000 and the fifth $2,000. First-degree price discrimination means that you are able sell the first car for $10,000, the second for $8,000, the third for $6,000, the fourth for $4,000 and the fifth car for $2,000. The revenue from such a discriminatory pricing policy will be $30,000. Not surprisingly this strategy is also called personalized pricing.

What if, on the other hand, you chose to sell all your cars at the same, uniform price? It is easy to calculate that the best you can do is to set a price of $6,000 at which you will sell three cars for a total revenue (and profit) of $18,000. Any higher or lower price generates lower revenues. In short, under uniform pricing your highest possible revenue is $18,000 while successful first-degree price discrimination yields much higher revenue of $30,000. Simply put, first-degree price discrimination enables you to extract the entire surplus that selling your car generates. No consumer surplus remains if you can successfully discriminate to this extent whereas with a uniform price the keenest buyer has consumer surplus of $4,000 and the second keenest buyer has consumer surplus of $2,000.

Since first-degree price discrimination or personalized pricing redirects surplus from consumers to the firm it should be expected to raise the incentive for the monopolist to produce. In fact, under first-degree price discrimination, the monopolist chooses the same socially efficient amount that would be achieved under perfect competition. In our Model T example no mutually beneficial trades are left unmade: all five cars are sold. By contrast, with uniform pricing only three cars are sold leaving two of the cars in the "wrong" hands.

The same is true in more general cases. For a monopolist able to practice first-degree price discrimination, selling an additional unit never requires lowering the price on other units. Each additional unit sold generates revenue exactly equal to the price at which it is bought. Hence, with first-degree price discrimination marginal revenue is equal to price. Accordingly, for such a monopolist, the profit maximizing rule that marginal revenue equals marginal cost yields an output level at which price equals marginal cost as well. As we know, this is the output level that would be generated by a competitive industry.

Practice Problem 6.1

Suppose that a monopoly seller knows that her demand curve is linear, and knows that at a price of $40, she sells five units, while at a price of $25, she sells 10 units.

a. If each potential consumer buys only one unit, what is the reservation price of the consumer with the greatest willingness to pay?
b. Suppose that the monopolist discovers that the demand curve just worked out applies only to the first unit a consumer buys and that, in fact, each consumer will also buy a second unit at a price $8 below the price at which they purchase just one. How many units will be sold at a price of $33?

At first glance it might seem that first-degree discrimination is little more than a theoretical curiosity. How could a monopolist ever have sufficient information about potential buyers and the ability to prevent arbitrage so as to implement effectively a pricing scheme in which a different, personalized price is charged to each buyer and for each unit bought?

Reality Checkpoint

The More *You* Shop the More *They* Know

Internet shopping has undoubtedly brought with it considerable convenience in shopping for books, DVDs, wine, and gourmet foods. At the same time, however, it has provided e-commerce retailers such as Amazon.com and Wine.com the ability to track your purchases. The result is that companies such as these are able to tailor special offers to each individual consumer based on their predictions of the books, wines, condiments, and so on that the consumer is most likely to find attractive. In other words, the Internet has made possible a kind of personalized marketing that is not feasible through more traditional media.

Source: C. Shapiro and H. R. Varian, *Information Rules: A Strategic Guide to the Internet Economy*, Harvard Business School Press, Boston, 1999.

The problems of identification and arbitrage prevention seem insurmountable. However, in some cases the monopolist seller may have the ability to achieve the personalized pricing outcome. Think, for example, of the tax accountant who knows the financial situation of his or her clients. Another example, perhaps closer to home, can be found in the students who apply to any of the (expensive) private universities in the United States. When they apply for financial aid they are required to complete a detailed statement of financial means. The universities of course can use this information, as well as SAT scores and other data, to determine the aid that will be granted and so the net tuition that each prospective student will be required to pay. Look around you. If one breaks down the total tuition on a per class basis, chances are that many of your classmates are paying a different fee for this class than you are!

Of course, the accountant or university example may be somewhat special because often the fee is set *after* the customer has contracted to purchase the service. What we now want to consider is whether there are pricing strategies that will permit the seller to achieve the same effect even when she must announce her fees in advance. The answer, to a greater or lesser extent, is yes. One such strategy is a ***two-part*** pricing scheme and another is *block pricing*. We discuss each in turn.

6.1.1 Two-Part Pricing

A two-part pricing scheme is a pricing strategy that consists of: 1) a fee, such as a membership fee, that entitles the consumer to buy the good; and 2) a price or usage fee charged for each unit the consumer actually buys. Many clubs use such two-part pricing. They charge a flat annual fee for membership in the club (which is sometimes differentiated by age or some other member characteristic), and additional user fees to use particular facilities or buy particular goods or services. Country clubs, athletic clubs, and discount shopping clubs are all good examples of clubs that use this kind of pricing. A related example of two-part pricing is that used by theme parks under which a flat fee is charged to enter the park and additional fees (sometimes set to zero) are charged on a per ride or per amusement basis.[1]

[1] Versions of such a scheme are used, for example, at parks such as Disney World. See Oi (1971) for the seminal discussion. As Ekelund (1970) notes, much modern analysis was anticipated by the work of nineteenth-century French economist and engineer, Jules Dupuit. Varian (1989) provides a thorough survey of the price discrimination tactics discussed in this chapter.

Reality Checkpoint
Call Options

Nonlinear pricing is an increasingly common feature of everyday life. Consider the packages available for cell phone service offered by the four major providers in the U.S., AT&T/Cingular, Verizon, Sprint/Nexus, and T-Mobile. Virtually all of these involve some variant of two-part pricing and quantity discount. Family plans, for example, offer two lines for a fixed monthly fee. After that, each minute of calling is free up to a specified maximum. Low-level plans offer something like 700 minutes for free at a monthly fee of, say, $70, while higher-use plans offer roughly twice as many free monthly minutes for a fee of about $90. There is also a 3,000-minute plan that usually sells for about $150. Additional phones can be added to a family plan at a fee of $10 per month. There are also single line plans and even pay-as-you-go plans. The latter are essentially calling-card plans that sell say, 30 minutes or 90 minutes of phone time for $15 or $25, respectively. They are clearly for those who cannot be induced to make more than a few calls even with a hefty discount.

Sources: L. Magid, "BASICS: Plain Cellphones Can Overachiever, With A Little Help," *New York Times*, January 25, 2007, p. c14; and "The Bottom Line on Calling Plans," *Consumer Reports, February*, 2004, pp. 11–18.

To see how a two-part pricing can work to achieve first-degree price discrimination, let us consider a jazz club where people meet for drinks and music. Assume that the club's customers are of two types: old and young (but both above the legal drinking age), and that there are just as many old members as young ones. A typical old consumer's inverse demand curve for the club's services is:

$$P = V_O - Q_O \tag{6.1}$$

While each young customer has the inverse demand curve:

$$P = V_Y - Q_Y \tag{6.2}$$

where Q_i is the number of drinks consumed in an evening by a customer of type i (O or Y), P is the price per drink and V_i is the maximum amount a consumer of type i will pay for just one drink. We shall assume that old customers are willing to pay more for a given number of drinks than young customers i.e. $V_O > V_Y$. We further assume that the jazz club owner incurs a cost of c dollars per drink served plus a fixed cost F of operating the club each night. That is, the cost function for the club is:

$$C(Q) = F + cQ \tag{6.3}$$

This example is illustrated in Figure 6.1. The demand curve for a typical consumer starts at Vi and declines with slope −1 until it hits the quantity axis. The constant marginal cost curve is a horizontal line through the value c.

Suppose that the jazz club owner is a "traditional" monopolist who employs simple linear pricing. Entry to the club is free, the club owner sets a price per drink and customers decide how many drinks to buy at that price. The jazz club owner would like to employ

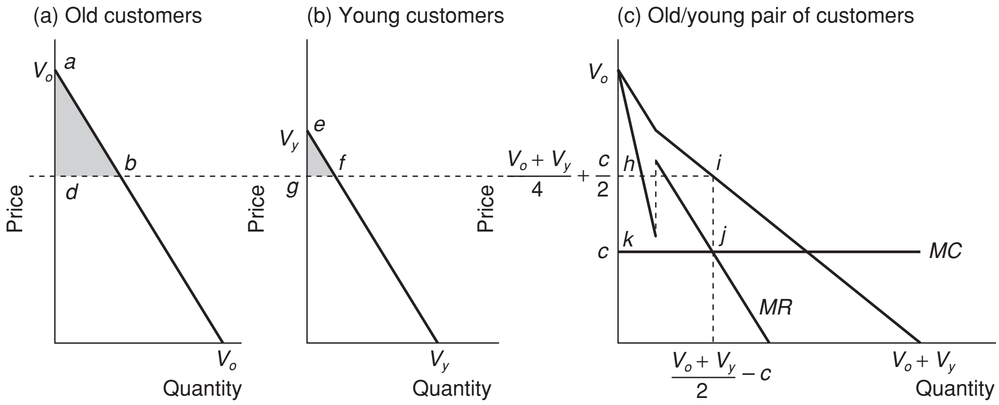

Figure 6.1 No price discrimination
Not price discriminating leaves both types of consumer with consumer surplus that the monopolist would like to convert to profit.

third-degree price discrimination, charging old customers more per drink than young customers. While the identification problem is easily resolved, by carding the customers, the arbitrage problem is not. Each old customer could ask (or bribe) a young customer to buy his drinks. So most likely best that the traditional linear pricing monopolist can do is to set a uniform price for drinks to customers of both types. Inverting (6.1) and (6.2) and adding gives the aggregate demand for each pair of customers consisting of one old and one young customer[2]

$$Q = Q_o + Q_y = (V_o + V_y) - 2P \tag{6.4}$$

Solving this demand curve for the price variable P to get the aggregate inverse demand for each old/young pair, assuming that both types of customer are allowed into the club:

$$P = (V_o + V_y)/2 - Q/2 \tag{6.5}$$

The jazz club monopolist maximizes profit by identifying the quantity, in this case the number of drinks, at which marginal cost equals marginal revenue and then identifying the price at which this quantity can be sold. Given the straight-line demand curve of equation 6.5 it is clear that the marginal revenue curve for each old/young pair is:

$$MR = (V_o + V_y)/2 - Q \tag{6.6}$$

Setting marginal revenue equal to marginal cost c requires that $(V_o + V_y)/2 - Q = c$, which gives the profit-maximizing output—number of drinks sold to each old/young pair:

$$Q_U = (V_o + V_y)/2 - c \tag{6.7}$$

[2] We can do this because we have assumed that there are equal numbers of each type of customer. With different numbers of each type we need a slightly different approach. See the end-of-chapter problems.

where the subscript U denotes uniform pricing. Substituting this into the demand function gives the profit-maximizing price per drink:

$$P_U = (V_o + V_y)/4 + c/2 \tag{6.8}$$

Each old customer buys $Q_o = V_o - P_U = (3V_o - V_y)/4 - c/2$ drinks and each young customer buys $Q_y = V_y - P_U = (3V_y - V_o)/4 - c/2$ drinks. The monopolist earns a surplus π_U from each pair of old and young customers of:

$$\pi_U = (P_U - c)Q_U = \frac{1}{8}(V_o + V_y - 2c)^2 \tag{6.9}$$

which is the area *hijk* in Figure 6.1(c). If there are n customers of each type per evening, the jazz-club owner's profit, Π_U is:

$$\Pi_U = n\pi_U - F = \frac{n}{8}(V_o + V_y - 2c)^2 - F \tag{6.10}$$

For example, if V_o is \$16, V_y is \$12 and c is \$4, then the optimal uniform price is \$9 per drink. Old customers each buy 7 drinks and young customers each buy 3 drinks. Under this strategy, the club owner earns a profit of (\$9 – \$4)*10 = \$50 for serving an old and a young customer. [Note that this is what we obtain when substituting the values for V_o, V_y and c, respectively, in equation (6.10).] If there were 100 old and 100 young customers per evening, the jazz club owner would earn a profit of \$5,000 each night less any fixed costs F that are incurred. To see that the jazz-club owner can improve on this outcome, first note that at the uniform price P_U = \$9 every customer of the jazz club enjoys some consumer surplus. Each old customer has consumer surplus given by the shaded triangle *abd* in Figure 6.1(a) and each young customer has consumer surplus given by area *efg* in Figure 6.1(b). These areas are, by standard geometric techniques, $CS_o^U = \frac{1}{2}(V_o - P_U)\cdot Q_o = \frac{1}{2}(Q_o)^2 = \frac{1}{2}\left(\frac{3V_o - V_y}{4} - \frac{c}{2}\right)^2$ for each old customer and $CS_y^U = \frac{1}{2}(V_y - P_U)\cdot Q_y = \frac{1}{2}(Q_y)^2 = \frac{1}{2}\left(\frac{3V_y - V_o}{4} - \frac{c}{2}\right)^2$ for each young customer. In our numerical example, each old customer has consumer surplus of \$24.50 and each young customer has consumer surplus of \$4.50. This is a measure of the surplus that the club owner has failed to extract. He will clearly prefer any pricing scheme that appropriates at least some, or even better, the entire surplus.

One possibility is for the jazz-club owner to switch to a non-linear pricing scheme that has two parts—a cover charge just to enter the club and an additional charge for every drink consumed. This pricing design is often referred to as a two-part tariff. The old customer is charged a cover or entry fee $E_o = \frac{1}{2}\left(\frac{3V_o - V_y}{4} - \frac{c}{2}\right)^2$, whereas the young customer is charged a cover fee $E_y = \frac{1}{2}\left(\frac{3V_y - V_o}{4} - \frac{c}{2}\right)^2$. The cover charge is the first part of the tariff.

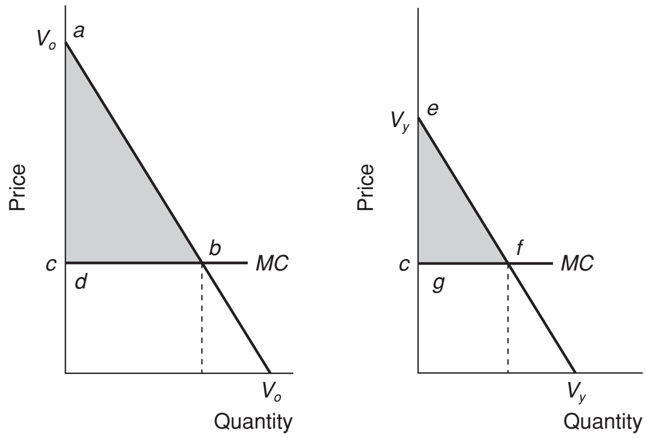

Figure 6.2 First-degree price discrimination with a two-part tariff
The monopolist sets a unit price to each type of consumer equal to marginal cost. It then charges each consumer an entry or membership fee equal to the resulting consumer surplus.

The second part is the charge of a price per drink of P_U. In our numerical example, each old customer is charged a cover of \$24.50 and each young customer is charged a cover \$4.50 for entry, while drinks are prices at \$9 each. Checking IDs at the door easily solves both arbitrage and identification problems.[3] Moreover, the customers will still be willing to patronize the club. Paying the entry fee reduces their surplus to zero but does not make it negative. The surplus is a measure of their willingness to pay. Finally, since the entry fee is independent of the amount the customer actually drinks, each customer will also continue to buy the same number of drinks as before. Because the entry fee is equal to the consumer surplus each customer previously enjoyed under the uniform pricing policy, the immediate effect of this two-part tariff is to extract the entire consumer surplus and to convert it into profit for the club owner. This implies a profit increase of E_o per old customer and E_y per young customer. Again, in our example, this yields a profit increase of \$24.50 per old customer and \$4.50 per young one.

However, the club owner can do better still. By *reducing* the price per drink the club-owner can increase the potential consumer surplus each consumer could have. In turn, this permits him to increase the entry fees, which enables him to extract that additional surplus and further increase his profit. The profit-maximizing two-part pricing scheme is illustrated in Figure 6.2. It has the following properties[4]:

[3] We assume that the expense of serious facelifts, hair coloring, and falsifying IDs is more than the surplus older consumers lose by paying the higher price.

[4] To see why these properties hold, denote the fixed portion of the two-part tariff for a particular type of consumer as T and the user charge as p. Express the demand curve for this type of consumer in inverse form, $p = D(q)$ and assume that the firm's total cost function is $C(q)$. The monopolist's problem is to choose the production level for this type of consumer, q^*, implying a price $p^* = D(q^*)$, that maximizes profits, $\Pi(q)$, where $\Pi(q)$ is given by: $\Pi(q) = \int_0^q D(x)dx - C(q)$. Standard calculus then reveals that maximizing this profit always requires setting a price or user charge equal to marginal cost, and a fixed charge T equal to the consumer surplus generated at that price.

1. Set the price per unit (drink) equal to marginal cost c.
2. Set the entry fee for each type of customer equal to that customer's consumer surplus.

In our jazz club case the price per drink is set at c. The area of the triangles *abd* and *efg* describe the consumer surplus at this price for old customers and for young customers respectively. These areas are $CS_o = \frac{1}{2}(V_o - c)^2$ and $CS_y = \frac{1}{2}(V_y - c)^2$. As a result, the jazz-club owner can now increase the entry fee to CS_0 for old customers and CS_y for young customers.

Under this optimal pricing scheme, the profit per drink from each consumer is zero, since drinks are sold at cost. This pricing strategy has the advantage of encouraging consumers to purchase many drinks, thereby yielding more consumer surplus. In turn, the jazz club owner can appropriate that surplus for himself by imposing the optimal cover charge. Since the funds claimed by the entry fees *are* profit, total profit has, therefore, been increased to:

$$\Pi_f = \frac{n}{2}((V_0 - c)^2 + (V_y - c)^2) - F \tag{6.11}$$

In our example, profit per old customer is now \$72 while it is \$32 per young customer instead of the \$24.50 and \$4.50 earned from each when the cover charge was associated with a \$9 drink price. This is a hefty profit increase.[5]

While the increase in profit is sizable and important, the two-part tariff has had another result that is equally significant. Note that each customer is now buying the quantity of drinks, $V_o - c$ for the typical old customer and $V_y - c$ for the typical young one, that each would have bought if the drinks had been priced competitively. The ability to practice first-degree price discrimination leads the monopolist to expand output to the competitive level. That is, the market outcome is now efficient. The total surplus is maximized—and that total surplus is claimed entirely by the monopolist.

Practice Problem 6.2

Consider an amusement park operating as a monopoly. Figure 6.3 shows the demand curve of a typical consumer at the park. There are no fixed costs. The marginal cost associated with each ride is constant. It is comprised of two parts, each also a constant. There is the cost per ride of labor and equipment, k, and there is the cost per ride of printing and collecting tickets, c. A management consultant has suggested two alternative pricing policies for the park. Policy A: charge a fixed admission fee, T, and a fee per ride of r. Policy B: simply charge a fixed admission fee, say T', and a zero fee per ride.

a. For pricing policy A, show on the graph the admission fee, T, and the per ride price, p, that will maximize profits.
b. For pricing policy B, show on the graph the single admission fee, T', that will maximize profits.
c. Compare the two policies. What are the relative advantages of each policy? What determines which policy leads to higher profits?

[5] It is also easy to show that the jazz club owner's profit would be smaller than that achieved by the two-part tariff if he could somehow engage in third degree price discrimination by somehow charging a different drink price for each group. We invite you to work this out for yourselves.

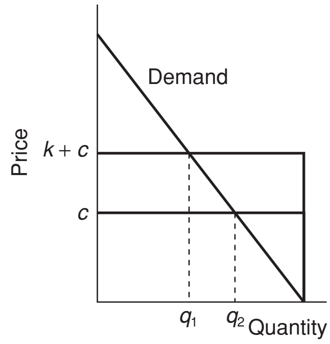

Figure 6.3 Diagram for the amusement park problem

6.1.2 Block Pricing

There is a second non-linear pricing scheme by which the jazz-club owner can achieve the same level of profit. This scheme is often called block-pricing. Using this type of pricing a seller *bundles* the quantity that he is willing to sell with the total charge that he wishes to set for that quantity. In our jazz club example, the owner sets a pricing policy of the form "Entry plus X drinks for Y dollars." In order to earn maximum profit and appropriate all potential consumer surplus, two simple rules determine the optimal block-pricing strategy:

1. Set the quantity offered to each consumer type equal to the amount that type of consumer would buy at competitive pricing, i.e., the quantity bought at a price equal to marginal cost;
2. Set a fixed charge for each consumer type at the total willingness to pay for the quantity identified above.

Let's examine how this would work in our jazz club example. Applying rule 1, we know that each old customer will buy $V_o - c$ drinks and each young customer will buy $V_y - c$ drinks if the drinks are priced at marginal cost. The total willingness to pay for these quantities by old and young customers respectively is the area under the relevant demand curve at these quantities. In the case of old and young customers, respectively, this is:

$$WTP_o = \frac{1}{2}(V_o - c)^2 + (V_o - c)c = \frac{1}{2}(V_o^2 - c^2) \tag{6.12a}$$

$$WTP_y = \frac{1}{2}(V_y - c)^2 + (V_y - c)c = \frac{1}{2}(V_y^2 - c^2) \tag{6.12b}$$

Applying rule 2 we then have the following pricing policy. Offer each old customer entry plus $V_o - c$ drinks for a total charge of $\frac{1}{2}(V_o^2 - c^2)$ dollars and each young customer entry plus $V_y - c$ drinks for a total charge of $\frac{1}{2}(V_y^2 - c^2)$ dollars.

How would we implement this strategy in our jazz club example? One way would be to card the customers at the door and then to give each customer the appropriate number of tokens that can be exchanged (at no additional charge) for drinks. Profit from a customer of type i is the charge WTP_i minus the cost of the drinks, $c(V_i - c)$, or $\frac{1}{2}(V_o - c)^2$ or \$72 from each old customer and $\frac{1}{2}(V_y - c)^2$ or \$32 from each young customer, exactly as in the two-part pricing system.

Before leaving this section, we wish to point out a further interesting feature of both types of first-degree price discrimination that we have discussed. Both the two-part tariff and the block pricing schemes result in the jazz club owner serving each old customer entry plus 12 drinks for a total charge of \$120 and each young customer entry plus 8 drinks for a total charge of \$64. Therefore, in each case, the average price paid by an old customer is $\frac{1}{2}(V_o^2 - c^2)/(V_o - c) = \frac{1}{2}(V_o + c)$ or \$10. Similarly, each young customer pays an average price per drink of $\frac{1}{2}(V_y + c) = \$8$. You can easily check that these are exactly the same prices per drink that would be levied if the club owner were able to apply third-degree price discrimination. Yet the profit outcome is different.

The reason that first-degree and third-degree price discrimination lead to very different profits, despite the fact that the average price is the same in each case, lies in the very different nature of the two pricing schemes. Recall that a demand function measures the marginal benefit that a consumer obtains from the last unit consumed. The quantity demanded equates marginal benefit with the marginal cost to the consumer of buying the last unit where, of course, marginal cost to the consumer is just the price for that last unit. With third-degree price discrimination or linear pricing, there is no difference between the price paid for the first unit and the price paid for the last unit. Hence, average price and marginal price are the same. By contrast, the non-linear scheme pricing permits the club owner effectively to charge an old person \$16 for the first drink, \$15 for the second and so on, while charging a young person \$12 for the first drink, then \$11, etc. With the linear pricing scheme, the old person would pay \$10 for each drink, from the first to the last, while a young person would pay \$8. The average price to each type of customer is the same under either first or third-degree price discrimination. Yet the non-linear pricing of first-degree price discrimination so lowers the price of the last unit purchased, that consumers are willing to buy (many) more units. Effectively this permits the owner to charge very high prices on the first few drinks purchased. As a result, the average price under first-degree discrimination is just as high as it is under third-degree discrimination. Yet since first-degree discrimination so greatly increases sales at that average price, and since that average price is greater than the firm's marginal cost, such non-linear pricing generates considerably more profit.

6.2 SECOND-DEGREE PRICE DISCRIMINATION OR MENU PRICING

First-degree price discrimination, or personalized pricing, is possible for the jazz club owner for two reasons. First, the club's different types of customers could be distinguishable by means of a simple, observable characteristic. Second, the club has the ability to deny access to those not paying the entry charge that was designed for them. Not all services can be marketed in this way. For example, if instead of a jazz club the monopoly seller is a refreshment stand located in a campus center then limiting access by means of a cover charge is not feasible.

Even in the jazz club case, first-degree discrimination by means of a two-part tariff would not be possible if the difference in consumer willingness to pay was attributable to some characteristic that the jazz club owner could not observe. For example, suppose that what differentiates high demand and low demand customers is not age but income. The club will now find that any attempt to implement the first-degree price discrimination scheme of charging high-income patrons an entry fee of \$72 and low-income patrons an entry fee of \$32 is not likely to succeed. Every customer could claim to have low income in order to pay the lower entry charge and there is no obvious (or legal) method by which the club owner can enforce the higher fee.

What about the block pricing strategy of offering entry plus 12 drinks for \$120 and entry plus 8 drinks for \$64? Will that work? Again, the answer is no. It is easy to show that high income customers are willing to pay up to \$96 for entry plus 8 drinks. So they derive \$32 of consumer surplus from the (8 drinks, \$64) package but no consumer surplus from the (12 drinks, \$120) package. They will prefer to pretend to be low income in order to pay the lower charge and enjoy some surplus rather than confess to being high income.

The monopolist could, of course, decide to limit entry only to high-income customers by setting the entry charge at \$72 or offering only the (12 drinks, \$120) package but this loses business (and profit) from low-income customers. Suppose, for instance, that there are N_o older customers and N_y younger customers. The profit from selling to only the older customers is \$$72N_o$. Setting the lower entry fee or offering only the (8 drinks, \$64) package in order to attract both types of customer gives profit of \$$32(N_y + N_o)$. Clearly, the latter strategy is more profitable if $32N_y > 40N_o$. In other words, if the ratio of low-income to high-income customers is more than 1.25:1 (the ratio of the difference in the entry fees to the low entry fee) the policy of setting the higher entry fee or offering only the (12 drinks, \$120) package will generate less profit than offering just the lower entry fee or the (8 drinks, \$64) package to all customers.

The point is that once we reduce either the seller's ability to identify different customers or to prevent arbitrage among them (or both), complete surplus extraction by means of perfect price discrimination is no longer possible. Both the two-part and block pricing mechanisms can still be used to raise profit above that earned by uniform pricing but they cannot earn as much as they did previously. Solving the identification and arbitrage problem has now become costly. It is still possible that the monopolist can design a pricing scheme that will induce customers to reveal who they are and keep them separated by their purchases, but now the only way to do this incurs some cost—a cost reflected in less surplus extraction. Such a pricing scheme is called second-degree price discrimination or menu pricing.

Second-degree price discrimination is most usually implemented by offering quantity discounts targeted to different consumer types. To see how it works, let's continue with our jazz club example illustrated in Figure 6.4. Again, the high demand customers have (inverse) demand $P_h = 16 - Q_h$ and the low demand customers have (inverse) demand $P_l = 12 - Q_l$. Now, however, the jazz club owner has no means of distinguishing who is who because the source of the difference between consumers is inherently unobservable. All the owner knows is that two such different types of consumer exist and they both frequent the club.

Any attempt to implement a differentiated two-part tariff will not work in this case. Both types of customer will claim to be low demand types when entering the club in order to pay

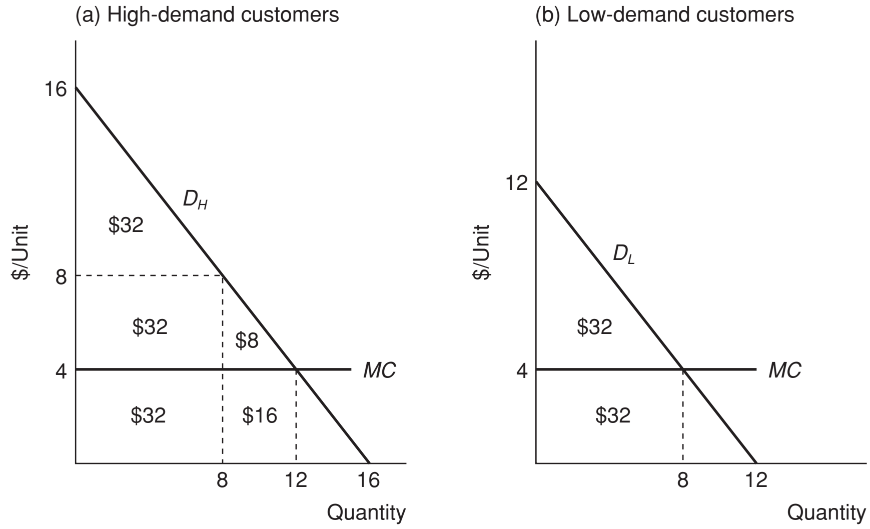

Figure 6.4 Second-degree price discrimination
Low-demand customers are willing to pay $64 for entry plus 8 drinks. High-demand customers are willing to pay up to $96 for entry plus 8 drinks, and so get $32 surplus from the 8-drinks, $64-package. They will therefore be willing to buy a 12-drinks, $88-package, which also gives them a $32 surplus.

the lower entry fee of $32. Only *after* they are in the club will the different consumers reveal who they are. Since the price per drink is set at marginal cost of $4, the high demand customers will buy 12 drinks and reveal themselves to be high demanders whereas the low demand customers will buy 8 drinks and reveal themselves as such.

You might be tempted to think that the club owner could implement first-degree price discrimination using the following strategy. When entering the club, patrons are given tickets that allow them to buy drinks. If they pay an entry charge of $32 they will be given 8 tickets while if they pay $72 they will be given 12 tickets. Yet this approach will not work either and for the same reason that the block pricing strategy of offering (12 drinks, $120) and (8 drinks, $64) packages failed. High-demand customers again have every reason to pretend to be low demand customers and pay an entry charge of only $32, thereby getting 8 tickets and buying 8 drinks at $4 each for a total expenditure of $64. Because, as can be seen from Figure 6.4(a), their total willingness to pay for the 8 drinks is $96, high-demand consumers will enjoy a surplus of $32 from this deception. By contrast, they will enjoy no surplus if they pay the entry charge of $72 and get 12 tickets because their total expenditure will be then $120, which exactly equals their willingness to pay for 12 drinks. As a result, it remains the case that the high demand customers are better off by pretending to be low demand even though this constrains the number of drinks that they can buy.

Yet, while unsuccessful, the idea of offering different price and drink combinations as different packages does contain the hint of a strategy that the jazz club owner can use to increase his profit. The point is to employ a variant on the *block pricing* strategy described earlier. The difference is that, since there is no easy way to identify and separate the different types of customers, the block pricing itself must be designed to achieve this purpose. This imposes a new constraint or cost on the owner and so will not yield as much profit as first-degree price discrimination. However, it will substantially improve on simply offering all customers a ($64, 8 drink) package that yields a profit of $32 from each.

To see how one might use block pricing to achieve the identification and separation necessary for price discrimination let us start with the low demand customers. The jazz club owner knows that these customers are willing to pay a total of $64 for 8 drinks. We know that the jazz club owner can offer a package of entry plus 8 drinks at a price of $64 for the package. This package will be attractive to low demand customers, effectively extracting the $32 surplus from each low demand consumer. The problem is that high demand customers will also be willing to buy this package because their willingness to pay for entry and 8 drinks is $96. While the club owner also gets $32 in profit from the high-demand customers buying this package, those customers themselves still enjoy a surplus of $96 − $64 = $32.

The club owner's optimal strategy at this point is to offer a second package targeted to high demand consumers. He knows that the high demand customers are willing to pay a total of $120 for 12 drinks. Yet he also knows that he cannot charge $120 for 12 drinks because the high demand customers will not be willing to pay this much, given that they can buy the (8 drinks, $64) package and enjoy consumer surplus of $32. For an alternative package to be attractive to high demand consumers it has to be what economists call *incentive compatible* with the (8 drinks, $64) package. This means that any alternative package must also allow the high demand customers to enjoy a surplus of at least $32.

A package that meets this requirement but that also generates some additional profit for the club owner is a package of entry plus 12 drinks for a total charge of $88. We know that the high demand customers value entry plus 12 drinks at $120. By offering this deal at a price of $88, the club owner permits these customers to get $32 of surplus when they buy this package, just enough to get them to switch from the (8 drink, $64) package.[6] And while the high-demand consumers get a $32 surplus on this package, the club owner's profit is also higher than it is on the (8 drink, $64) package. On the latter, the owner earns $32, but on the new package, the owner earns $88 − ($4 × 12) = $40. Of course, the low demand customers will not buy the (12 drink, $88) package since their maximum willingness to pay for 12 drinks is only $72. Nevertheless, the club owner still earns $32 from these consumers by continuing to sell them the (8 drink, $64) package. So, the club owner's total profit is increased.

The two menu options have been carefully designed to solve the identification and arbitrage problems by inducing the customers themselves to reveal who they are through the purchases they make. The club owner now offers a menu of options, 8 drinks for $64 or 12 drinks for $88 designed to separate out the different types of customers that he serves. For this reason, this strategy is often referred to as *menu pricing*. It has one very important feature. Note that as before, the average price per drink of the (8 drinks, $64) package is $8. However, the average price per drink of the (12 drinks, $88) package is $7.33. The second package thus offers a *quantity discount* relative to the first.

Quantity discounts are common. Movie theaters, restaurants, concert halls, sports teams and supermarkets all make use of them. It is cheaper to buy one huge container of popcorn than many small ones. Wine sold by the glass is more expensive per unit than wine sold by

[6] We are working in round numbers to keep things neat. What the jazz club owner might actually do is price the package of entry plus 12 drinks at $87.99 to ensure that the high demand customers will strictly prefer this to the (8 drinks, $64) package.

the bottle. A 24-pack of Coca-Cola is cheaper than 24 individual bottles. It is cheaper per ticket to buy a season's subscription to your favorite football team's home games than to buy tickets to each game individually. A full-day pass at a ski resort will reflect a lower price per run than will a half-day lift ticket. In these and many other cases, the sellers are using a quantity discount to woo the high-demand consumers.

There is another twist to consider. What if the club owner now decides to offer a lower number of drinks, say 7, in the package designed for the low demand customers? The maximum willingness to pay for entry plus 7 drinks by a low demand customer is $59.50 so this new package will be (7 drinks, $59.50). The profit it generates from each customer is $31.50, which is 50 cents less than the (8 drinks, $64) package. But now consider the high demand customers. Their maximum willingness to pay for 7 drinks is $87.50, so buying this new package gives them consumer surplus of $28. As a result, the jazz club owner can increase the price of the 12-drink package. Rather than pricing it so that it gives the high demand customers $32 of consumer surplus, he can now price it so that it gives them only $28 of surplus. In other words, he can now raise the price the second package (entry plus 12 drinks) to be $120 − 28 = $92, increasing his profit from each such package to $44.

The example illustrates the importance of the incentive compatibility constraint. Any package designed to attract low demand customers constrains the ability of the monopolist to extract surplus from high demand customers. Again, this is because the high demand customers cannot be prevented from buying the package designed for low demand customers, and thus will always enjoy some consumer surplus from doing so. As a result, the monopolist will find it more profitable to reduce the number of units offered to low demand customers since this will allow him to increase the price he charges for the package targeted to the high demand customers. There may even be circumstances in which the monopolist would prefer to push this logic to the extreme and not serve low demand customers at all because of the constraint serving them imposes on the prices that can be charged to other customers. Whether or not the monopolist has an incentive to serve the low demand consumers will depend on the number of low demand consumers relative to high demand ones. The fewer low demand consumers there are relative to high demand ones, the less desirable it is to serve low demand consumers since any effort to do so imposes an incentive compatibility constraint on the extraction of surplus from high demand ones.

For a general case of more than two types of consumers the profit-maximizing second-degree price discrimination or menu pricing scheme will exhibit some key features. In particular, if consumer willingness to pay can be unambiguously ranked by type then any optimal second-degree price discrimination scheme will:

1. extract the entire consumer surplus of the lowest demand type served but leave some consumer surplus for all other types;
2. contain a quantity that is less than the socially optimal quantity for all consumer types other than the highest-demand type;
3. exhibit quantity discounting.

Second-degree discrimination enhances the ability of the monopolist to convert consumer surplus into profit, but does so less effectively than first-degree discrimination. With no costless way to distinguish the different types of consumers, the monopolist must rely on some sort of block pricing scheme to solve the identification and arbitrage problems. However, the incentive compatibility constraints that such a scheme must satisfy restrict the firm's ability to extract the entire consumer surplus. Instead, the firm is forced to make a compromise between setting a high charge that loses sales to low demand buyers, and a low charge that foregoes the significant surplus that can be earned from the high-demand buyers. And contrary to what many consumers may think, the lower price charged for a larger quantity is entirely unrelated to scale economies. If in our example the jazz club owner has no fixed costs and thus no economies of scale. Nevertheless the owner finds it profitable to offer a quantity discount to high demand customers.

6.3 Practice Problem

Assume that a monopolist knows that his customers are of two types, low demand customers whose inverse demand is $P_l = 12 - Q_l$ and high demand customers whose demand is $P_h = 16 - Q_h$. However, he does not know which type of customer is which. His production costs are $4 per unit.

a. Complete the following table for this example.

Low-demand customers			*High-demand customers*			
Number of units in the package	*Charge for the package**	*Profit per package*	*Consumer surplus from low-demand package*	*Maximum willingness to ay for 12 units*	*Charge for package of 12 units*	*Profit from each package of 12 units*
0	0	0	0	$120.00		$72.00
1	$11.50		$4.00	$120.00	$116.00	
2		$14.00	$8.00	$120.00		$64.00
3						
4	$40.00	$24.00		$120.00		
5	$47.50	$27.50	$20.00	$120.00	$100.00	$52.00
6	$54.00			$120.00		$48.00
7	$59.50	$31.50	$28.00	$120.00	$92.00	$44.00
8	$64.00	$32.00	$32.00	$120.00	$88.00	$40.00
9						
10	$70.00	$30.00	$40.00	$120.00		
11						
12	$72.00		$48.00	$120.00	$72.00	

*This is the low-demand customer's maximum willingness to pay for the number of units in the package.

b. Assume that there are the same numbers of high demand and low demand customers. What is the profit-maximizing number of units that should be offered in the package aimed at the low demand customers?
c. Now assume that there are twice as many low demand customers as high demand customers. What is the profit-maximizing pair of packages for the monopolist?
d. The monopolist is considering offering two packages, one containing 6 units and the other 12 units. What are the charges at which these packages will be offered? What is the ratio of high demand to low demand customers above which it will be better for the monopolist to supply only the high demand customers?

6.3 SOCIAL WELFARE WITH FIRST- AND SECOND-DEGREE PRICE DISCRIMINATION

One way to understand the welfare effects of price discrimination is to consider a particular consumer group i. Suppose each consumer in this group has inverse demand:

$$P = P_i(Q) \tag{6.13}$$

Assume also that the monopolist has constant marginal costs of c per unit. Now let the quantity that each consumer in group i is offered with a particular pricing policy be Q_i. Then the total surplus—consumer surplus plus profit—generated for each consumer under this pricing policy is just the area between the inverse demand function and the marginal cost function up to the quantity Q_i, as illustrated in Figure 6.5.

The pricing policy chosen by the firm affects the quantity offered to each type of consumer, and it alters the distribution of total surplus between profit and consumer surplus. The first effect has an impact on welfare, whereas the second effect does not imply a change in total welfare, but rather a transfer of surplus between consumers and producers. As a result, *price discrimination increases (decreases) the social welfare of consumer group i if it increases (decreases) the quantity offered to that group*.

It follows immediately that first-degree price discrimination always increases social welfare even though it extracts all consumer surplus. With this pricing policy we have seen that the monopoly seller supplies each consumer group with the socially efficient quantity (the quantity that would be chosen if price were set to marginal cost). Hence first-degree discrimination always increases the total quantity to a level [$Q_i(c)$ in Figure 6.5] that exceeds that which would have been sold under uniform pricing.

With second-degree price discrimination matters are not so straightforward. As we have seen, this type of price discrimination leads to high demand groups being supplied with quantities "near to" the socially efficient level. However, we have also seen that the seller will

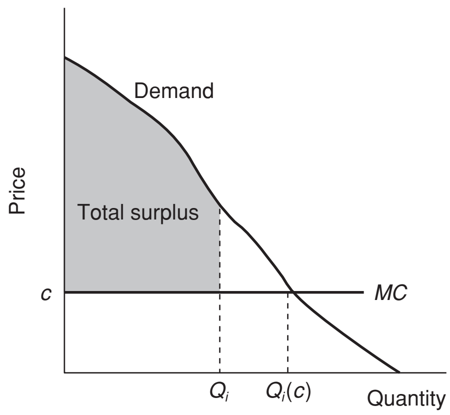

Figure 6.5 Total surplus
When the total quantity consumed is Q_i total surplus is given by the shaded area. Total surplus is maximized at quantity $Q_i(c)$.

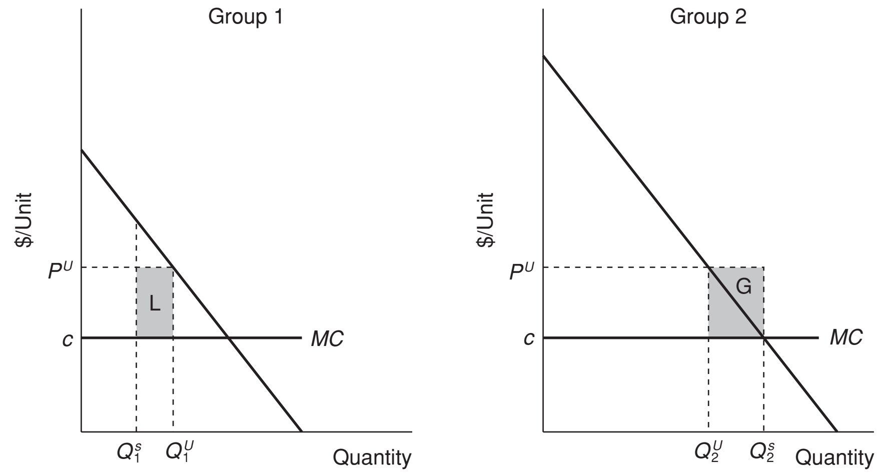

Figure 6.6 Impact of second-degree price discrimination on welfare
An upper limit on the change in total surplus that arises from second-degree price discrimination is the upper limit on the gain, *G*, minus the lower limit on the loss, *L*.

want to restrict the quantity supplied to lower demand groups and, in some cases, not supply these groups at all. The net effect on output is therefore not clear a priori.

The impact on social welfare of second-degree price discrimination can nevertheless be derived using much the same techniques that we used in Chapter 5. By way of illustration, suppose that there are two consumer groups with demands as illustrated in Figure 6.6 (i.e., Group 2 is the high-demand group). In this figure P^U is the non-discriminatory uniform price, and Q_1^U and Q_2^U are the quantities sold to each consumer in the relevant group at this price. By contrast, Q_1^s and Q_2^s are the quantities supplied to the two groups with second-degree price discrimination.[7] We define the terms:

$$\Delta Q_1 = Q_1^s - Q_1^U;\ \Delta Q_2 = Q_2^s - Q_1^U \tag{6.14}$$

In the case illustrated we have $\Delta Q_1 < 0$ and $\Delta Q_2 > 0$. This tells us that an upper limit on the increase in total surplus that follows from second-degree price discrimination is the area G minus the area L. This gives us the equation:

$$\Delta W \leq G - L = (P_U - MC)\Delta Q_1 + (P_U - MC)\Delta Q_2 = (P_U - MC)(\Delta Q_1 + \Delta Q_2) \tag{6.15}$$

Extending the analysis to n markets, we then have:

$$\Delta W \leq (P_U - MC)\sum_{i=1}^{n} \Delta Q_i \tag{6.16}$$

It follows that for $\Delta W \geq 0$ it is necessary that $\sum_{i=1}^{n} \Delta Q_i \geq 0$. In other words, a necessary condition for second-degree price discrimination to increase welfare is that it increases total output.

7 Since group 2 is the high demand group, we know that $Q_2^s = Q_2(c)$.

We know from last chapter that this requirement is generally not met in the case of third-degree price discrimination and linear demands because then the monopolist supplies the same total quantity as with uniform pricing, so third-degree price discrimination does not increase welfare. By contrast, it could be the case that second-degree price discrimination leads to an increase in the quantity supplied to both markets and this would increase social welfare. In the jazz club owner case, for example, this will be the case if there are an equal number of high-demand and low-demand customers. (You are asked to show this in the end-of-chapter problem 6.)

Summary

In this chapter we have extended our analysis of price discrimination to cases in which firms employ more sophisticated, non-linear pricing schemes. Our focus has been on commonly observed examples of such non-linear pricing schemes. These are: (1) two-part pricing in which the firm charges a fixed fee plus a price per unit; and (2) block pricing in which the firm bundles the quantity being offered with the total charge for that quantity. Both schemes have the same objective, to increase the monopolist's profit either by increasing the surplus on existing sales or by extending sales to new markets, or both.

The most perfect form of price discrimination, first-degree price discrimination or personalized pricing, can only be practiced when the firm can costlessly solve the identification and arbitrage problems. The firm needs to be able to identify the different types of consumers and must also be able to keep them apart. If this is possible, then two-part tariffs and block pricing can, in principle, convert *all* consumer surplus into profit for the firm. The positive side to this is that the firm supplies the socially efficient level of output to each consumer type. The negative side is that there are potentially severe distributional inequities in that all social surplus takes the form of profit.

If the requirements necessary to practice perfect price discrimination are not met, then the monopoly seller cannot achieve such a large profit. The monopolist may then rely on second-degree price discrimination or menu pricing, another form of non-linear pricing. Second-degree price discrimination differs from both first- and third-degree, however, in that it relies on the pricing mechanism itself—usually some form of quantity discount—to induce consumers to *self-select* into groups that reveal their identity or who they are on the demand curve.

The use of a quantity discount to sort or screen consumers must always satisfy an *incentive compatibility* constraint across the different consumer types. This constraint affects in a negative way the monopolist's ability to extract consumer surplus. Because the incentive compatibility constraint adversely affects profits, the monopolist may choose to avoid it by refusing to serve low demand markets, with the result that the low demand type consumers are clearly worse off. As a consequence, the welfare effects of second degree price discrimination are not clear. Yet unlike the case of third degree discrimination (with linear demand curves at least), second degree pricing strategies do have some positive probability of making things better.

Problems

1. Many universities allocate financial aid to undergraduate students on the basis of some measure of need. Does this practice reflect pure charity or price discrimination? If it reflects price discrimination, do you think it lies closer to first-degree discrimination or third-degree discrimination?

2. A food co-op sells a homogenous good called groceries denoted g. The co-op's cost function is described by: $C(\boldsymbol{g}) = F + cg$; where F denotes fixed cost and c is the constant per unit variable cost. At a meeting of the co-op board, a young economist proposes the following marketing strategy: Set a fixed membership fee $\boldsymbol{M}$ and a price per unit of groceries $\boldsymbol{p_M}$ that members pay. In addition, set a price per unit of groceries p_N higher than p_M at which the co-op will sell groceries to non-members.

a. What must be true about the demand of different customers for this strategy to work?
b. What kinds of price discrimination does this strategy employ?

3. At Starbuck's Coffee Shops, coffee drinkers have the option of sipping their lattes and cappuccinos while surfing the Internet on their laptops. These connections are made via a connection typically provided by a wireless firm such as T-Mobile. Using a credit card, customers can buy Internet time in various packages. A one-hour package currently goes for an average price of $6. A day pass that is good for any time in the next 24 hours sells for $10. A seven-day pass sells for about $40. Briefly describe the pricing tactics reflected in these options.

4. A nightclub owner has both student and adult customers. The demand for drinks by a typical student is: $Q^S = 18 - 3P$. The demand for drinks by a typical adult is: $Q^A = 10 - 2P$. There are equal numbers of students and adults. The marginal cost of each drink is $2.
   a. What price will the club owner set if he cannot discriminate at all between the two groups? What will his total profit be at this price?
   b. If the club owner could separate the groups and practice third-degree price discrimination what price per drink would be charged to members of each group? What would be the club owner's profit in this case?

5. If the club owner in #4 can "card" patrons and determine who among them is a student and who is not and, in turn, can serve each group by offering a cover charge and a number of drink tokens to each group, what will the cover charge and number of tokens be for students? What will be the cover charge and number of tokens given to adults? What is the club owner's profit under this regime?

6. A local phone company has three family plans for its wireless service. Under each of these plans, the family gets two lines (phones) and can make local and long distance (within the U.S. and Canada) for free so long as the total number of minutes used per month does not exceed the plan maximum. The price and maximum minutes per month for each plan are: plan 1: 500 minutes for $50; plan 2: 750 minutes for $62.50; and plan 3: 1,000 minutes for $75.00. Assuming that there are equal numbers of consumers in each group and that the value of a marginal minute for each group declines at the rate of $0.0004 per minute used, work out the demand curves consistent with this pricing. What surplus will each consumer group enjoy?

7. Now return to our club owner in the text in which low-demand consumers have an inverse demand of: $P = 12 - Q$; while high-demand consumers have an inverse demand of: $P = 16 - Q$. Marginal cost per drink is again $4. Assume that there are N_h high demand customers and N_l low demand customers. Show that under these circumstances the firm will only serve low demand customers, i.e., will only offer both packages if there are at least as many low-demand consumers as high-demand ones. In other words, $\frac{N_h}{N_l} \leq 1$ in order for low consumers to be served.

References

Ekelund, R. 1970. "Price Discrimination and Product Differentiation in Economic Theory: An Early Analysis." *Quarterly Journal of Economics* 84 (February): 268–78.

Oi, W. 1971. "A Disneyland Dilemma: Two-part Tariffs for a Mickey Mouse Monopoly." *Quarterly Journal of Economics* 85 (February): 77–96.

Shapiro, C. and H. R. Varian, 1999. *Information Rules: A Strategic Guide to the Internet Economy*. Boston: Harvard Business School Press.

Varian, H. 1989. "Price Discrimination." In R. Schmalansee and R. Willig, eds., *The Handbook of Industrial Organization*. Vol. 1. Amsterdam: North-Holland, 597–654.