Chapter 29: Natural Monopoly and Discrimination

29.1: Introduction

This chapter discusses two things, both related to the fact that, in the presence of a monopoly, there is less surplus generated in the market because of the existence of the monopoly. You will recall that this was partly as a consequence of the monopolist choosing the price, but more the consequence of a lower quantity of the good being traded. The first thing we examine in this chapter is the issue of whether the lost surplus might be justifiable in some sense because the monopoly has access to a more efficient technology. The second thing is whether it might be possible to recover some of the lost surplus through discrimination of some form. After all the reason why the monopolist can not produce up to the point where the price is equal to the marginal cost is that, in the absence of discrimination, the monopolist has to charge the same price for everyone.

29.2: Natural Monopoly

One very obvious reason why a single large firm might be more appropriate in some industry is simply that a single large firm might have access to a more efficient technology than lots of small firms. This is a plausible case if centralisation of the industry has consequences in terms of increased efficiency – contrariwise, if splitting it up into lots of little units causes losses of efficiency and increased bureaucratic costs. (You might like to ask if that is what happened when British Rail in the UK was split into several operating companies.) If this is the case then it may be better – in some sense which we will discuss shortly – to have a single large firm. Even if it behaves like a monopolist – and there is an associated deadweight loss as a consequence – the end product for both consumers and producer might be better.

The example that follows is extremely simple and is not meant to be realistic. The only point that it wants to make is that the output may be higher and the price lower with a monopoly that has access to a better technology than with a competitive industry with a worse technology. We start by assuming a constant marginal cost function for both the industry when competitive and when a monopoly – the difference being that we assume a much lower (constant) marginal cost for the monopoly. We assume throughout a simple linear aggregate demand schedule in the industry. We start with the competitive industry.

![Graph showing demand and cost with output on the x-axis and demand and cost on the y-axis.]

The demand curve assumed has intercept 100 on both axes – the horizontal axis measures the aggregate quantity demanded and supplied, and the vertical axis the demand and the marginal cost.
The horizontal line at 60 is the assumed constant marginal cost function: for all the firms in this competitive industry the marginal cost of producing an extra unit of output is 60.

In this competitive industry the aggregate output is where the price is equal to the marginal cost—that is, is where the marginal cost curve and the aggregate demand curve intersect. This is at an output of 40 and the price is equal to the marginal cost of 60. Figure 29.2 shows the consumer surplus in the competitive equilibrium. Note that there is no producer surplus.

Now let us suppose that, if all the firms in the industry were merged into one big firm, then the technology would be such that the marginal cost in the merged monopoly would be significantly reduced. Let us suppose that it would be reduced from 60 to 10. We then have figure 29.3.

In this we have inserted the marginal revenue schedule facing the firm: because there is a linear demand schedule with intercepts of 100 and 100 we know that the marginal revenue schedule is also linear—with the same intercept on the vertical axis and half the intercept on the horizontal axis. The horizontal line at the bottom is the monopolist’s marginal cost curve. The monopolist’s optimal output is at the point where marginal cost and marginal revenue are equal—as figure 29.3 shows this occurs at an output of 45. The price that the monopolist charges is then given by the demand curve—as we see from figure 29.3 this is a price of 55. We get the rather obvious result that, if the monopolist has a sufficiently lower marginal cost function then he or she produces a higher quantity at a lower price than the competitive industry.

It is instructive to look at the surpluses under this monopoly and compare them with the competitive case. These are shown in figure 29.4. The triangular area at the top left is the consumer surplus under monopoly—this is obviously larger than under competition because the quantity exchanged is higher and the price lower.
The rectangular area is the surplus going to the monopolist. The sum of these two surpluses is the total surplus generated by the monopoly. So with the more efficient monopolist, even if it chooses the price itself, we could have an industry in which the price is lower, the output higher and the total surplus higher than if the firm were competitive. This kind of situation is what is usually meant by a natural monopoly.

But we can also identify a deadweight loss caused by the monopoly – this is the triangular area to the right of the figure. Why is this the deadweight loss caused by the monopoly? Well, if this industry behaved like a competitive industry in the sense that it set its output so that the price was equal to the marginal cost, then the output would be 90 and the price 10 – and the whole of the area above the marginal cost curve and under the demand curve would go to the consumers as their surplus.

So we get the following conclusion: with the more efficient monopoly replacing the competitive industry then the price is lower and the quantity is higher and everyone is happier. But, if then the monopolist could be made to behave as a competitive firm then things could be even better still in the sense that the price could be lower still and the quantity higher still and the aggregate surplus higher still.

So – you might be asking – why do we not recommend that the government imposes a maximum price in the market (equal to the marginal cost of the monopolist)? This would force the monopolist to produce the competitive output of 90, force it to charge a price of 10, and the consumers would get the surplus that they would have got under competition.

Indeed this is the best solution – if the situation we have pictured is the long run and the surpluses are the true surpluses. Indeed this is often the situation in nationalised natural monopolies in many countries. But suppose that this is the short run position. Suppose further that it masks the fact that the monopolist has to pay a fixed cost - $K$. Let us add this into the analysis – by inserting into the figure a curve representing the average total costs of the firm. These are simply $K/y + 10$, where $y$ is the output of the firm and the ‘10’ is the marginal cost (already portrayed in the figure). This average total cost function is obviously downward sloping and is the convex curve in figure 29.5.

What do we notice? At the monopoly output (45) and price (55), the price quite clearly is bigger than the average total cost – so the firm makes a positive profit. But at the competitive output (90) and price (10) the price is lower than the average total cost – so the firm is making a loss. (Rather obviously, in fact, for any fixed cost, because the price is covering only the marginal cost – because it is equal to the marginal cost!) This loss is shown in the figure – it is the rectangular area bounded by the quantity produced, the price and the average total cost.
So the monopolist can legitimately reply to the government – “we cannot charge the marginal cost price and produce the competitive quantity, because if we did we would make a loss. We need to cover our fixed costs.”

This situation is typical of many industries we would consider as ‘natural monopolies’ – the electricity, gas, water industries and the transport industries, for example. What is the solution? Leave them as competitive and miss out on the technological advantages of large-scale operation? Make them state monopolies, force them to charge marginal cost – and then pay them their fixed costs?

This latter seems the best thing to do – but there remains the problem: if the state is going to pay the monopoly its fixed costs, how is the state sure that the monopoly is stating them correctly? There is a clear incentive for the monopoly to be lazy and inefficient. Does that remind you of anything?

### 29.3: Discrimination

At this point it may be useful to remind ourselves of why there is a deadweight loss with monopoly. Let us take a specific example, which we can then use to suggest a way of eliminating this deadweight loss through price discrimination. This example will use a discrete good but the point can also be made with a continuous good. Consider the aggregate demand curve pictured in figure 29.6. Up to 10 units may be bought – with reservation prices 100, 90, 80, 70, 60, 50, 40, 30, 20 and 10.
Suppose that the monopolist has a cost function such that the marginal cost of production is constant and equal to 36. If the monopolist is restricted to a unique price for every unit sold what price should be chosen? A little calculation reveals that, if the monopolist’s objective is to maximise the surplus or profit then a price of 70 should be chosen, Figure 29.7 illustrates this and the table gives some more detail. I should note that I have assumed throughout that, if a buyer is indifferent between buying and not buying, then he buys. Note also that the consumer surplus is the area between the price of 70 and the demand curve - that is 60. Note further that there is a deadweight loss caused by the monopoly – equal to the area (42) between the monopoly quantity (of 4) the marginal cost curve and the demand curve. The total surplus under competition would be the total area between the marginal cost curve and the demand curve – this is of size 238 – this all goes to the consumers under competition.

<table>
<thead>
<tr>
<th>Price</th>
<th>Surplus</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>64</td>
</tr>
<tr>
<td>90</td>
<td>108</td>
</tr>
<tr>
<td>80</td>
<td>132</td>
</tr>
<tr>
<td><strong>70</strong></td>
<td><strong>136</strong></td>
</tr>
<tr>
<td>60</td>
<td>120</td>
</tr>
<tr>
<td>50</td>
<td>84</td>
</tr>
<tr>
<td>40</td>
<td>28</td>
</tr>
</tbody>
</table>

This shows that the best unique price is 70. So the buyers with reservation prices of 100, 90, 80 and 70 buy – all paying 70, and thus giving a total revenue of 280 – and each unit costing the monopolist 36 and thus costing 144 in total – leading to a surplus of 280 minus 144, that is, 136. But notice that the buyers with reservation prices 60, 50 and 40 do not buy even though the monopolist could make a profit out of them, given that the cost per unit is only 36. So this raises the interesting question – why does the monopolist not sell to the buyers with reservation prices 60, 50 and 40 at – say – a price of 40? He or she would make a profit of 12 out of them.

The problem is this: if the monopolist has to choose a unique price at which to sell all the units sold, then he or she cannot charge a price of 70 to four of the buyers and charge a price of 40 to three of the buyers. Obviously it is in the monopolist’s interests to sell up to 7 units because the first 7 reservation prices are all greater than the cost of production, so a price can be chosen for each unit which would make both the monopolist and the buyer happy – but if the monopolist is to charge a low price for the later units (in the absence of discrimination) then the monopolist has to charge a low price for the early units. The fact that the monopoly output falls short of the output at which marginal cost is equal to the price is not a consequence of the fact that the monopolist is a monopolist, but a consequence of the fact that he or she charges a unique price for each unit of the good sold.

The obvious solution then is for the monopolist to discriminate. The most obviously efficient way is when the monopolist is allowed to discriminate completely by charging a different price for each unit sold. In this example it is easy: the monopolist sells the first unit at a price of 100 to the buyer with a reservation price of 100, the second unit at a price of 90 to the buyer with a reservation price of 90, ..., the seventh unit at a price of 40 to the buyer with a reservation price of 40. In this way the monopolist gets a surplus or profit of \((100 - 36) + (90 - 36) + \ldots + (40 - 36) = 238 – the entire area between the marginal cost of 36 and the demand curve.\) If this happens, there is no deadweight loss – the whole surplus available under competition is realised. The main difference, in this example, is that instead of going all to the consumers it now all goes to the monopolist. This is not
necessarily bad – as the government can always tax it away again. What is important is that the surplus is *generated*. As long as it is generated, it can always be re-distributed. If it is not generated there is nothing to be re-distributed. For that reason, price discrimination might be a solution to the problem caused by the deadweight loss of monopoly.

The discrimination described above is called *discrimination of the first degree*. It is complete discrimination in that, for each unit sold, there is a different price. So the monopolist discriminates not only between individuals but also between different units bought by the same individual. For rather obvious reasons this kind of perfect discrimination might be difficult to implement and enforce, and so in practice we observe less complete forms of discrimination.

Two kinds of discrimination can be observed in practice – one where there is discrimination of price depending upon the quantity purchased and another where there is discrimination of price depending upon the purchaser. The first of these is called *discrimination of the second degree* and the second *discrimination of the third degree*.

The most obvious example of discrimination of the second degree is when the monopolist offers discounts for bulk purchase – for example, supermarkets give discounts if large quantities are bought. The most obvious discrimination of the third degree is when the monopolist offers different prices for different types of people – for example, train companies charge less to young people and to old people than to middle-aged people. The obvious reason why they do this is to increase their surplus – otherwise they would not do it. In a sense these types of discrimination are part way between no discrimination and discrimination of the first degree – in which all the surplus is extracted and the situation – as far as the total extracted surplus is concerned, though not its distribution – is back to the situation under competition. Perhaps it is obvious that it is in the interests of the monopolist to extract as much of this total surplus as possible and take as much of it as possible. So the greater the degree of discrimination the better – as far as the extraction of the total surplus is concerned.

### 29.4: Summary

This chapter has been concerned with the deadweight loss caused by a monopoly. We have investigated whether it might be justified or eliminated.

*If a monopoly can take advantage of a more efficient technology it may be the case that the price could be lower and the output higher under monopoly than under competition – and the total surplus higher (though there still a deadweight loss caused by monopoly).*

*A monopoly, forced by government legislation to charge a price equal to the marginal cost may make a loss – which could be covered by the government so as to maximise the surplus extracted from the market.*

*Discrimination of price could reduce the deadweight loss – the extent of the reduction depending upon the extent of the discrimination.*

1. The monopolist needs to know all the reservation prices.
2. The monopolist would have to somehow stop individuals with low reservation prices buying the good and then re-selling it to individuals with high reservation prices.
29.5: What kinds of price discrimination are there in the real world?

There are innumerable examples of price discrimination in the real world. They are all illustrations of ways in which one side of the market tries to appropriate more of the surplus available in the market. In those markets in which the seller sets the price (such as most goods markets) the discrimination takes the form of different prices of the good sold to different people under different circumstances. In those markets in which the buyer sets the price (for example, in most labour markets) the discrimination takes the form of different prices of the ‘good’ bought from different people under different circumstances. The purpose of this simple application is merely to point out these numerous forms of discrimination in the real world.

The simplest form is discrimination over the quantity bought or sold. If the price per unit for a bulk purchase is less than the price per unit for a single purchase, then we have this form of discrimination. Many sellers do this. At the moment my supermarket is full of offers of the form “buy 2, get the second at half-price”. They really annoy me – I end up with food that goes off in the fridge or drinking too much beer (which is exactly what the supermarket wants).

Another form is when different people pay different prices for the same good or service. As the law in many countries is very hostile to certain forms of discrimination (for example sex discrimination) it is rare to see different prices for men and for women (though note the habit of apparently the same garment being sold in the men’s and women’s section of a store at the same price). We often see, however, different prices for students, or for the elderly or for the unemployed. This is clearly the seller extracting more of the surplus than they would do in the absence of discrimination.

More subtle forms of discrimination exist: mobile phone companies charge a lower tariff to calls on the same network (you should ask yourself why); the new generation of cheap airlines charge lower prices to those people who book early (so a plane-load of people can show any enormous heterogeneity in the prices paid); clubs offer lower per-visit prices to members; the National Trust offers a free visit to those who join the Trust at the time of the visit; employers pay a higher salary to those who have been employed longer (though these could be different, more efficient, workers); universities and other public sector employers, have a pay scale such that they salary rises every year (though the workers could be getting more efficient).

The list is almost endless. You could try to add to it yourself.