I. The Definition of Monopoly

Monopoly: a firm that is the only seller of a good or service with no close substitutes.

This definition is abstract, just as the definition of perfect competition is abstract. And just as it’s hard to find a market that really seems perfectly competitive in all respects, it’s hard to find a firm that is a total monopoly.

The source of ambiguity is the term “close substitutes.” How close is close? For example, is Amtrak a monopoly? Yes, if you’re concerned with long-distance passenger rail service. Not really, if you’re concerned with local rail travel (consider SEPTA, New Jersey Transit, etc.), or with long-distance transportation (consider buslines and airlines). Definitely not if you’re concerned with transportation broadly speaking (consider all of the above plus cars).

What counts as a substitute is ultimately a matter of consumer’s preferences, which are generally not in the form of perfect complements or perfect substitutes. Still, there are various situations that seem to qualify for use of this model: local phone companies in most areas, at least until recently, municipal cable monopolies, etc.

Why might there be only one firm in a market? There can be various reasons, some of which we discussed earlier: high costs of entry, patent protection, government licensing or exclusive grants, etc. Later, we’ll talk about another: economies of scale, which can lead to natural monopoly in some cases.

The monopoly model becomes more useful when we realize that any market that is not perfectly competitive will display some features of the monopoly model. Imperfectly competitive firms face downward-sloping demand curve, which means that while raising price will cause them to lose some customers, they will not lose all their customers.

II. The Single-Price Monopolist

The analysis here is pretty much identical to our generic theory of the firm. The main difference is in interpretation: here, we interpret the firm’s demand curve as the entire market demand, not just a fraction of it. As a result, we expect a monopolist’s demand curve to be steeper than a generic firm’s, because consumers lack the ability to substitute the products of other firms.

The picture is identical to that of the generic firm, except we use capital Q and capital D for quantity and demand, respectively, to indicate that we’re talking about the entire market being served by just one firm. Note that, unlike in perfect competition, we draw only a single diagram, rather than side-by-side market-and-firm.
Just as in the generic analysis, we can find whether the firm is making profits or losses by drawing in the ATC curve. And there’s nothing contradictory about a monopoly making losses. If a monopolist is making losses, it means that it doesn’t make economic sense to produce the good in question at all, because even under the most favorable conditions to the producer he cannot get revenues sufficient to cover all opportunity costs of production.

In general, economists are hostile to monopolies, and here is why: It is apparent that the firm could produce more units, at a marginal cost lower than what the remaining customers would be willing to pay. As a result, there is a dead-weight loss (DWL) from mutually beneficial trades not made.

Recall that under perfect competition, firms priced at marginal cost (P = MC). If the monopoly did that, its price would be $P^c$ and its quantity $Q^c$, as shown in the diagram. (The “c” is for competitive.) So we say that the monopoly causes losses in wealth relative to the perfectly competitive benchmark. (However, we will later talk about reasons why this may be somewhat misleading.)

But there’s something fishy here. If the units from $Q^m$ to $Q^c$ cost less to produce than consumers will pay for them, why doesn’t the monopolist choose to produce them?

The answer depends crucially on the single-price assumption, which says that the monopolist must charge a single price for all units. This could be because the firm cannot distinguish “high-price” consumers from “low-price” consumers. The single-price assumption is responsible for MR being below demand. The firm would like to be able to sell from $Q^m$ to $Q^c$ if it could charge $P^c$ for those units. But doing that would require selling the units from zero to $Q^m$ at $P^m$ instead of $P^c$, and it’s not worth it to do that.

Notice that much of this analysis applies to firms that imperfect competitors, even if they are not monopolies. If a firm faces a downward-sloping demand curve, then there will be
some amount of dead-weight loss, just as shown in the graph. The difference is that, to
the extent that a firm faces competition, the dead-weight loss will tend to be smaller than
that of a monopoly.

III. Welfare Analysis of Imperfect Competition (First Pass)

We can compare monopoly (or imperfect competition) to the perfectly competitive
benchmark in much the same way we compared price floors to the free market: by
finding CS, PS, DWL, and transfer. The main difference is we use MC in place of the
supply curve. (Recall that in PC, the supply curve is just the sum of the firms’ MC
curves for most prices.)

Consumer Surplus (CS) is below demand, above $P^m$, and left of $Q^m$. Producer Surplus is
above MC, below $P^m$, and left of $Q^m$. Transfer is between $P^c$ and $P^m$, and left of $Q^m$.

Note that in this context, the transfer is a shift in wealth relative to what consumers
would have had in a perfectly competitive market. If we imagine a previously
competitive market suddenly becoming monopolized (or just less competitive), the
transfer is the amount of wealth taken from consumers and given to the monopolist (or
remaining competitors) by the act of monopolization.

From a Kaldor-Hicks efficiency perspective, the transfer is irrelevant. What makes
monopoly (or imperfect competition) inefficient is the existence of DWL.

IV. Welfare Analysis of Imperfect Competition (Second Pass)

All of the above seems to imply that monopoly (or imperfect competition) necessarily
leads to higher price, lower quantity, and dead-weight losses relative to perfect
competition. But this may not be true, because larger firms may be able to achieve cost
savings unavailable to smaller firms. Consolidation to serve larger numbers of
consumers can allow the firm to take advantage of greater specialization, moving to a lower ATC and corresponding MC.

Think of it this way. Under perfect competition, we said that every firm would, in the long-run, produce a quantity at the lowest point on the LRATC. Now, take that quantity and multiply by the total number of firms; that will give you the total quantity in the market. But what if there aren’t enough consumers to justify that much production? Then in order to produce only as many units as the consumers want at the lowest possible cost, you might have to have fewer firms. But this could mean the industry becomes less competitive; there might not be enough firms to justify using the perfectly competitive model.

For example, consider this diagram.

In this situation, one large firm operating at SRATC₂ could produce at a much lower cost than three firms operating at SRATC₁. The corresponding MC would tend to push down price and increase quantity, hence passing some of the benefits on to the consumers.

Thus, the picture we drew earlier of CS, PS, etc., might not be valid, because the MC under monopoly (or imperfect competition) is not the same as under perfect competition.

We use the term concentration to refer to how many firms are operating in an industry. An industry that has many small firms is less concentrated; an industry with a smaller number of larger firms is more concentrated. The analysis above shows us two effects of having greater concentration in an industry. First, competition is reduced, which tends to drive prices up. Second, cost savings may be realized from economies of scale, which tends to drive prices down. So which effect dominates? This is an empirical question. It is highly relevant to many policy questions, such as whether the FTC will allow a merger to take place.

V. Price Discrimination (the Multi-Price Firm)
As noted earlier, the monopolist would like to sell to consumers who won’t buy at the monopoly price, if it could do so without lowering price for everyone else. And the same is true of an imperfectly competitive firm: it would sell to more customers if it could lower its price just for those customers.

So if the firm can distinguish between different groups of consumers, it may wish to charge differential prices. This requires, of course, being able to prevent resale of the product by low-price buyers to high-price buyers. This is not a phenomenon limited to monopolists; it can be done by any firm that faces a downward-sloping demand curve and has the ability to distinguish between customer groups.

**Price discrimination** is defined as charging different prices to different consumers on the basis of non-cost-related characteristics of the consumers. Note that it must be the same product or a similar product with the same cost structure. For instance, charging a different price for first-class seats on an airplane is not an instance of (pure) price discrimination; nor is charging different prices for gasoline with different octane levels.

In general, the most important factor for price discrimination is distinguishing different elasticities, to isolate people who are more or less sensitive to price. People with more elastic demand will tend to be charged a lower price; people with more inelastic demand will tend to be charged a higher price.

Examples of price discrimination:
- Movie prices (for students and seniors)
- Airline prices (for business and personal)
- Grocery coupons (“smart” versus casual shoppers)
- 10-10-321 and other long-distance services (price sensitive v. not price sensitive)

What are the effects of price discrimination? On the one hand, it tends to decrease dead-weight loss for both consumers and producers. This is because the firm can charge a lower price for units it would otherwise not have sold.

On the other hand, it tends to transfer wealth from consumers to producers. It does this because the firm can take a larger chunk out of CS from customers who are willing to pay more than the single-price monopoly price. A firm that finds it can price discriminate won’t just lower the price for the more elastic customers; it will also tend to raise price for the less elastic customers.

So consumers as a whole may or may not benefit from price discrimination; it depends on which consumers one has in mind, and how the firm discriminates. If you’re a consumer in the range from zero to the monopoly quantity, you either lose or remain the same. If you’re a consumer in the range from monopoly quantity to competitive quantity, you tend to get better off.
Note again that from an efficiency perspective, the DWL is what’s important. To the extent that price discrimination reduces the size of DWL, it’s more efficient than monopoly.

**Perfect price discrimination** occurs when the firm can charge a different price for each consumer, with the price exactly equal to the maximum price the consumer will pay. A firm that perfectly price discriminates can extract the entire CS from the consumers and add it to PS. Note that a perfectly price-discriminating monopolist has a MR that’s identical to the demand curve, and it chooses the competitive quantity. There is no DWL under perfect price discrimination.

### VI. Natural Monopoly

A natural monopoly exists when one firm can produce at a lower ATC than can two or more firms because of economies of scale.

One extreme example is an industry in which LRATC is always decreasing. Why might this happen? Suppose that once certain large capital investments have been made, the marginal cost of production is close to zero. For example, once the electric company has built its generators and electrical grid, the cost of using them to produce units of electricity is very low. As a result, the average cost keeps getting smaller as you divide the cost of the capital investments by larger and larger numbers. (Eventually, of course, you’d need to build more generators for a large enough number of customers. But that could happen far to the right of the graph we’ve drawn.)

In a situation like this, we expect a single firm to emerge as a monopolist. Why? Because a firm can lower its price to just below others’ average cost. This will attract more consumers and allow this firm to get a lower ATC. Meanwhile, the other firms will make losses because they can only sell at a price below average cost.

Example: Suppose your ATC is $5 and mine is $4 because I’m already serving more customers. If I set my price at $4.99, I’ll attract your customers, causing you to make losses – either because you get no customers or because you match my price and lose a penny per unit. This practice is called limit pricing.

The incentives in this case drive firms to move further and further out on their LRATC curves. But since there is only so much demand to be satisfied, some firms will go out of business when others lower price and raise output. Eventually, there will be only one firm remaining.

The problem here is that, once only one firm remains, it can raise price and lower quantity in typical monopoly fashion, as shown by Q<sub>m</sub> and P<sub>m</sub> below. Note: we have to introduce a new concept here, the long-run MC. This is simply the added cost of producing one more of a product in the long run, with the ability to adjust the input combination optimally.
Often, governments have stepped in to regulate such monopolies. Examples include most utilities (water, electricity, gas). There are two basic regulatory approaches:

- The $P = MC$ rule. The firm is forced to set a price equal to its marginal cost. This is the competitive benchmark, and it allows no dead-weight loss. But look at what happens to the firm’s profit. Since $LRMC$ is necessarily below the $LRATC$, the firm must be making losses. This is why the $P = MC$ rule is generally not used.

- The $P = ATC$ rule. The firm is allowed to set a price equal to its average cost. This allows the firm to break even. Usually, the firm will actually be allowed a small amount of profit, by setting price equal to $ATC$ plus a small amount. The problem with this approach is that it reduces incentives to lower costs through technological efficiency and innovation. The firm can incur whatever costs it wants, knowing that the regulatory authority will include those costs in calculating the price the firm can charge and allow the firm to make a profit regardless.