PROFIT AND THE FIRM

- Simple Version -

In a capitalist system, the decisions of a firm are governed by the owners of capital ('capitalists', 'shareholders'). So the managers of a firm have only one objective: to maximize the return to capital, that is, to *maximize profit*. So the firm's decisions - *how much* to produce, *how many* to hire, etc. - are guided by the profit-maximization criteria.

We already have shown, by means of numerical examples, how the search for profit interacting with the law of increasing costs yields us the upward-sloping *supply curve*. A profit-seeking firm will *not* expand production unless the sales price is high enough to overcome the increasing cost, i.e. unless it is profitable.

Our numerical examples were only a small sample of possible production decisions. We can see which are more profitable than others. But how do we determine the *exact* point at which profit is at a *maximum*?

The most direct way is to calculate profit at *every* level of production and then pick the most profitable level. But that takes a lot of arithmetic and tedious, repetitive calculation. Is there a short-cut? There is. In fact, all a firm has to know is the price and cost structure and they can hone in on the profit-maximizing level of production immediately. This is known as the "**marginal rule**".

These notes explain the foundations of the profit-maximizing decision in a little more detail.

MARGINAL INTUITION

Suppose you have to produce two papers this weekend, one for your Global Flows class, another for Research Methods. Every hour you spend on one paper is an hour you lose on the other.

How should you divide your time? Presuming your weekend has, say, 20 waking hours available for writing, you could lay out a plan to divide your time equally, and spend 10 on one and 10 hours on the other. Or you could spend 18 hours perfecting one, and a mere 2 hours on the other. If you follow the latter option, your grade in the first paper would likely be excellent, but the grade for the second would be quite poor.

Notice the law of increasing cost is at work here. The more your work on your Global Flows paper and reaching for a higher grade there, the less time you spend on the Research Methods paper and thus the lower grade you're bound the receive there.

How do you go about it? You could sit down and figure out ahead of time exactly how much time you ought to spend on each. Or you could do it as most people do it: just start writing. Say, you begin with the Global Flows paper, writing with great care and precision. But as the hours pass, you're likely to get increasingly anxious about how little time is left to write the Research Methods paper. Nonetheless, you will continue on the first until the anxiety about the remaining time has reached such a point that you decide to wind up the first paper, and start the next one.

True, you could always spend more time proofreading and rewording what you've already written for Global Flows, bringing towards ever-greater perfection. But your sense is that the extra perfection that an extra hour of work on Global Flows would achieve is probably simply not worth the sacrifice of one less hour working on the Research Methods paper you haven't even started yet.

This is called "marginal thinking". Marginal thinking is about thinking "along the edge", thinking about the extra, additional little bit, rather than the total. "Should I spend an additional hour?", rather than "How many hours should I spend?"

Associated with marginal thinking is the "marginal rule" of decision-making: keep doing something until the benefit of spending an extra hour on that thing isn't worth the extra cost of doing so.

"Marginal thinking" may seem a little haphazard, undisciplined and disorganized. Surely the "proper" way to organize your time is to sit down, work out schedule and plan it all ahead, e.g. "I shall spend 10 hours on Global Flows and 10 hours on Research Methods". No?

Far from being disorganized and undisciplined, marginal thinking is the very epitome of rationality and efficiency. Marginal thinking tells you to just forget those plans and just go ahead and start doing something until you notice it is no longer worth doing.

If you obey this simple marginal rule of thumb, you'll notice the amount of time you spend on something will be *exactly* right. That is, the amount of time you actually spend on the paper will be the *most efficient* way of dividing your time, probably quite more efficient than you whatever you could work out in a plan ahead of time. The marginal rule ensures you don't spend an hour more on something than it's worth, nor an hour less than it deserves. Which, added up, is exactly how much time you *should* spend on it.

"Marginal thinking" applies in many other contexts. For instance:

- How much cheese should you eat? If you're worried about your slim figure, a nutritionist might advise you to measure total grams and set aside your portion ahead of time. An economist would tell you to just keep eating cheese until you notice the pleasure of an extra mouthful of cheese isn't worth the cost of the extra fat it'll add to your thighs.

- How much output should a firm produce? You can construct huge spreadsheets calculating total profits at every level. But the marginal rule says it is simpler to just start producing and keep producing output until the extra revenues obtained from an extra unit of output isn't worth the extra cost of producing it.

- How much labor should a firm hire? Keep hiring workers until the extra output an additional worker generates isn't worth the extra cost of hiring him.

And so on.

"Marginal thinking", whether about papers, cheese, output or hiring, is the keystone of rational and efficient decision-making. Total calculations are not only cumbersome but also unnecessary. Stick to thinking on the margin – the extra hour, the extra mouthful, the extra output, the extra worker rather than total hours, total grams, total profits - and you'll get it just right.

The concept of the "margin" is so pivotal to modern economics that the mainstream school of thought in economics is sometimes referred to as the "Marginalist School".

PRODUCTION SCALE

The "scale of production" is the amount of output a firm produces. This is a decision that the managers of a firm must make. In a capitalist system, this decision is governed by the goal of maximizing profit.

We can apply the marginal rule to the production scale decision quite easily: firms should produce up until the extra cost incurred by producing an additional unit exactly matches the extra revenue gained from selling it, and no more. Or, more succinctly, defining:

Marginal Revenue: the additional revenue a firm gets from increasing quantity produced by an additional unit.

Marginal Cost: the extra total cost incurred by a firm if it expands production by an additional unit.

then we can apply the marginal rule to the production scale problem as follows:

> **Marginal Rule**: to maximize profits, a firm should find the quantity where the marginal revenue is equal to the marginal cost

Intuitively, a firm should avail itself of every opportunity it has to reap a profit. So, as long as producing an additional unit generates more revenue than it costs, a firm should do it.

But the Law of Increasing Cost tells us that the more we produce the smaller profit gets. So at *some* point producing an extra unit of something begins to cost more than you can sell it for. At the moment, you're making losses, so stop.

Or put another way:

- if MR > MC, that means that by increasing production by one unit, the extra revenue reaped exceeds the extra cost incurred. That is a profitable opportunity. Therefore the firm should take the opportunity and produce one more unit.

- if MR < MC, that means that by increasing production by one unit, the extra cost incurred exceeds the extra revenue received. If the firm insists on expanding production, it will make a loss on that extra unit. In this case, the firm should cut back production.

- if MR = MC, that means that by increasing production by one unit, the extra revenues exactly match the extra costs. It neither makes a profit nor a loss on that extra unit. The firm has no more scope for profitable expansion. It should stop here.

In sum, a firm should produce until MR = MC.

Again, the advantage of the marginal rule is that you don't really need to tediously calculate total profits at every level. You just need to look at the margins, find out if expanding production by one unit yields more revenue than it costs. And keep on going until it doesn't. You can be certain that profits are at their maximum then.

Law of Increasing Cost

The role of the Law of Increasing Cost in making this decision cannot be overstressed.

Suppose that there was no law of increasing cost. Suppose that firms faced the same cost per unit, regardless of how much they produced. Then there is no profit-maximizing point – or rather, it is infinite.

Example: suppose you own a coffee shop in New York City. This is a competitive environment, so you must sell coffee at the going market price of $2 per cup (if you charge more, you'll lose all your customers). You take the selling price as "given". Your only decision is to decide *how much* coffee to produce and sell.

Suppose that you figure that you can brew a cup of coffee for the cost of 50 cents (that includes water, beans, wages for your baristas, rental of coffee machine, and coffee shop space). The cost for every cup added does *not change*, no matter how much you brew and sell. So in this case we have **constant costs**. Going through some calculations:

Qty	Price	Total Revenue (= Price × Qty)	Cost of cup	Total Costs	Profits (TR – TC)
0	$2	$0	$0	$0	$0
1	$2	$2	$0.50	$0.50	$1.50
2	$2	$4	$0.50	$1.00	$3.00
3	$2	$6	$0.50	$1.50	$4.50
4	$2	$8	$0.50	$2.00	$6.00
5	$2	$10	$0.50	$2.50	$7.50
6	$2	$12	$0.50	$3.00	$9.00
7	$2	$14	$0.50	$3.50	$10.50
8	$2	$16	$0.50	$4.00	$12.00
9	$2	$18	$0.50	$4.50	$13.50
10	$2	$20	$0.50	$5.00	$15.00

As you can see, with constant unit costs, the more you produce, the more profit you make. Therefore the profit-maximizing decision is to produce an infinite amount of coffee.

Obviously that is ridiculous.

As you move from 1 to 5 to 10 to 100 to 1,000 to 1 million cups of coffee and beyond, you'll need more beans, more water, more labor, more coffee machines and more space. If you insist on producing 10 million cups per hour, your little coffee shop can't handle it, you have to get larger space, even warehouse-size, or block-size. But rental space in New York is scarce and the more space you demand, the higher the rent per square meter will become. So too will your demand for legions of coffee-brewing workers drive up wages. And your demand for coffee machines drive up machine rental costs. And coffee beans, now gobbled up by the millions per hour by your huge enterprise, will also become more expensive per pound. Costs per every additional cup start rising quickly. The Law of Increasing Cost imposes itself.

So let us introduce a mild form of increasing cost. Let us suppose that every cup of coffee costs 0.25 more to brew than the previous cup of coffee. So, the first cup costs you only 50 cents to make, but the second costs you 75 cents, the third cup costs a full $1, the fourth $1.25, and so on. This is the law of increasing cost at work, raising the costs of brewing as you expanding production.

Total costs are cumulative. If you decide to brew four cups, then the total costs are 50¢ for the first cup + 75¢ for the second cup + $1 for the third cup + $1.25 for the fourth cup = $3.50 in total costs.

The following table gives you the new calculation, with the law of increasing cost in place:

Qty	Price	Total Revenue (= Price × Qty)	Cost of cup	Total Costs	Profits (TR – TC)
0	$2	$0	$0	$0	$0
1	$2	$2	$0.50	$0.50	$1.50
2	$2	$4	$0.75	$1.25	$2.75
3	$2	$6	$1.00	$2.25	$3.75
4	$2	$8	$1.25	$3.50	$4.50
5	$2	$10	$1.50	$5.00	$5.00
6	$2	$12	$1.75	$6.75	$5.25
7	$2	$14	$2.00	$8.75	$5.25
8	$2	$16	$2.25	$11.00	$5.00
9	$2	$18	$2.50	$13.50	$4.50
10	$2	$20	$2.75	$16.25	$3.75

You will notice that profit begins to rise, hits a maximum of $5.25 at around 6-7 cups, then begins to decline again.

The profit-maximizing level of production is thus 6-7 cups.

This is a tedious table to calculate – computing total revenues, total costs and total profits at every level. The Marginal Rule makes it simple. All you have to do is compare marginal revenue with marginal cost. And those two numbers are already shown in the table above.

Marginal revenue – the extra revenue from the sale of an additional cup – is simply the **price** of that cup of coffee on the market. We know that is $2, no matter how much you produce. So we're done there.

Marginal cost is not constant, but increases at 0.25 increments. That is the fourth column ('cost of cup'). So let's pluck that column. So to find the profit-maximizing level, all you need to do is look down that column and find where marginal cost is equal to marginal revenue:

Qty	Marginal Revenue (= price)	Marginal Cost (= cost of additional cup)
0	$2	$0
1	$2	$0.50
2	$2	$0.75
3	$2	$1.00
4	$2	$1.25
5	$2	$1.50
6	$2	$1.75
7	**$2**	**$2.00**
8	$2	$2.25
9	$2	$2.50
10	$2	$2.75

As we see in this abbreviated table, we can hone in on the answer immediately: 7 cups. That's where MR = MC = $2.

Go back to the original table, and you can verify that 7 cups is exactly the quantity that yields the maximum profit ($5.25).

[Of course, 6 cups does too, but notice that's because the 7th unit, where MR = MC, itself does not *add* any profit (and shouldn't, since MR = MC means there's no extra profit to be had at the 7th unit; its extra revenue exactly matches its extra cost.) The important point is we can't do better than 7. Although, 6 also happens to yield the same profit, the important point is that no other level yields more profit than 7 and going beyond 7 reduces it. So 7 is the *maximum*.]

This is *always* true. You don't need to explicitly calculate profit. Once you find the quantity where MR = MC, you can trust that this is the maximum profit you can achieve.

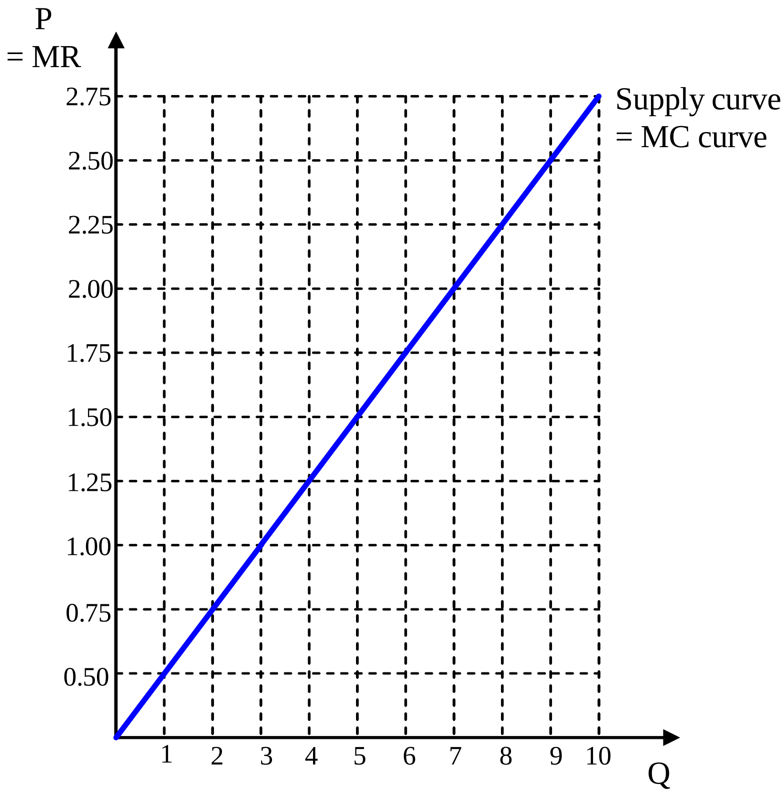
The Supply Curve

One thing we can do with this new concept is to map out the supply curve. *If* price is $2, then the marginal rule tells us to produce 7 cups. What if price of coffee rose to $2.25? Then the marginal rule tells us to produce 8 cups, which is where marginal cost is equal to $2.25. Contrarily, if the price of coffee declines to $1.25, the marginal rule tells us to produce 4 cups. And so on.

Armed with the marginal rule, we can find the optimal or profit-maximizing level of production at any price. Just give us the price, and scan the marginal cost column and you'll immediately find it.

This relationship, you will notice, is merely tracing our good old friend: ***the supply curve***.

Indeed, because MR = p and we know profit-maximization requires that MR = MC, then the supply curve *is* a line mapping out the marginal costs! That is why we sometimes call the supply curve the "MC curve", since it traces marginal costs out exactly.



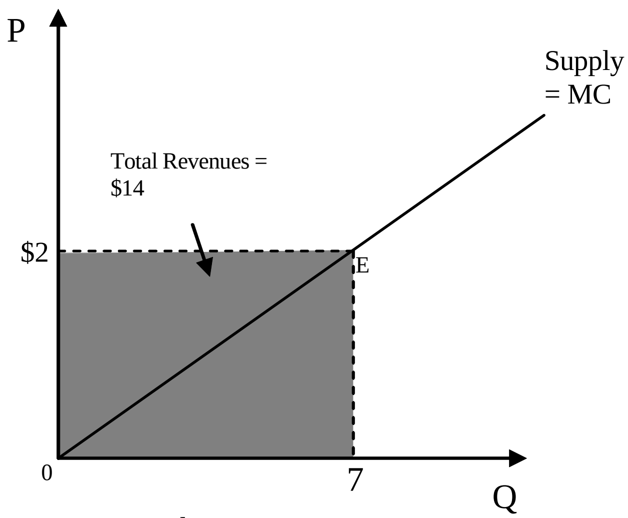
Murray's Coffee Supply

[Not exactly true, though, since this coffee shop isn't the only coffee shop in town. There are hundreds of others, each of them facing the same decision, and each with its own supply curve. The market supply curve – that is, what we see when we draw the picture of the coffee market as a whole - is the *sum* of all the individual supply curves of all the coffee shops. But since *each* individual coffee shop's supply curve is upward sloping, then the market's supply curve will also be upward-sloping – just more quantity at every price.]

Graphical Intuition

For those who like looking at pictures, we can read Total Revenues, Total Costs and Total Profits as the areas of the various geometrical shapes in a diagram depicting the Marginal Cost (Supply) curve.

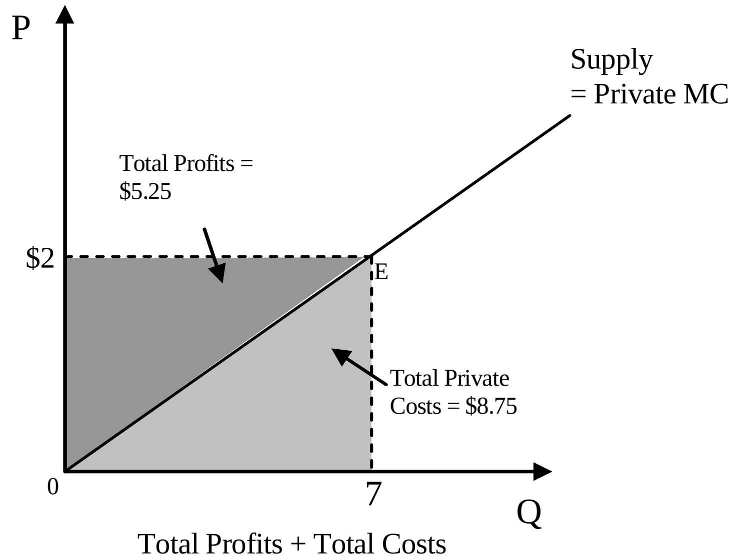
Let's start with Total Revenues. If a firm produces 7 units and sells them at \$2 apiece, then Total Revenues are \$2 × 7 = \$14. Now look at the diagram below where we have depicted the Marginal Cost Curve. Notice that \$2 × 7 is the area of a rectangle with height \$2 and length 7. In other words, Total Revenues represent the area of the shaded rectangle O2E7.



Total Revenues

But we know Total Revenues can be broken down between Total Profits and Total Costs. Remember that cumulating marginal costs adds up to total cost. So it is a matter of simple geometry to show that Total Costs can be measured as the area of the triangle *under* the Marginal Cost curve. Again, if the coffee shop produces 7 units, then marginal costs start from \$0 and cumulate up to \$2. Total Private Costs are consequently the lightly shaded triangle (OE7) below the MC curve. The area of OE7 is \$8.75

Total Profits are the difference between Total Revenues and Total Costs, and consequently can be measured as the area of the darkly shaded triangle (O2E) *above* the MC curve (and below \$2). The area of O2E = \$5.25.



Total Profits + Total Costs

The sum of the areas of the two triangles is the sum of the area of the Total Revenue rectangle. Obvious: $8.75 + $5.25 = $14.

FIXED COSTS

Let us now go a little deeper and introduce a new complication: fixed costs.

So far, we have been assuming costs are *variable*. That is, as output increases, the costs of production increase. This is because of the nature of inputs: the more we want to produce, the more inputs we need, and thus the more we bear down on the factor markets and thus the costlier it becomes per input.

But *some* costs *don't* vary with size of output. An example may be, say, the rent on a particular factory. The rent is established beforehand and it is for a time period (say, a month or a year). Whether that factory produces a lot of output or a little output in that time period, the rent is not affected. The firm has to pay the same rent.

This is *unlike* wages. Wages are a variable cost. To produce more, we *need* more labor (or labor hours). Consequently, we can cut down on labor costs by producing less. But we cannot cut down rent costs by producing less. Rent has to be paid regardless.

So our formula should change to:

Total Costs = Variable Costs + Fixed Costs

where, using the coffee shop example, variable costs are things like coffee beans, water, wages, electricity – costs that rise the more we produce. But fixed costs include things that the coffee owner will spend on even if he produces not a single cup of coffee: e.g. rent, advertising budget, etc. So if a firm produces zero units, there are zero variable costs and the only costs left are the fixed costs.

[Caveat: as in all of economics, no definition is really so clear-cut. Fixed costs sound rock-solid and unavoidable, but they are really only *temporarily* fixed. Over time, they become variable too. So although we're treating rent as fixed in this example, it is really only fixed within one month (or one year, depending on the lease). If the coffee shop decides to keep production at zero, it might as well abandon the shop altogether. But it still has to pay rent on the shop at least until the end of the current lease. But once the lease is over, the firm will stop paying rent too and total costs are zero again.]

How does the existence of fixed costs (like rent) affect the profit maximizing decision?
It doesn't.

To see why, examine the following example for our coffee shop, which is identical to the previous example, with the addition that now we have a fixed cost of $1 (call it rent) that must be paid regardless of scale. Our Variable Costs column corresponds exactly to our previous example with the law of increasing cost raising the cost of brewing an additional cup by increments of $0.25. So if you brew zero cups, variable costs are $0; if you brew

4 cups, variable costs are 50¢ for the first cup + 75¢ for the second cup + $1 for the third cup + $1.25 for the fourth cup = $3.50 in Variable Costs. And so on.

You will notice that the only difference between this calculation and the previous one is that *every* entry in the Total Cost column is now $1 greater (because of the Fixed Costs). If you decide to produce four cups, you have $3.50 in Variable Costs + $1 in Fixed Costs, for $4.50 in Total Costs.

Qty	Total Revenue	Variable Costs	Fixed Costs	Total Costs = VC + FC	Profits (TR – TC)
0	$0	$0	$1	$1	-$0
1	$2	$0.50	$1	$1.50	$0.50
2	$4	$1.25	$1	$2.25	$1.75
3	$6	$2.25	$1	$3.25	$2.75
4	$8	$3.50	$1	$4.50	$3.50
5	$10	$5.00	$1	$6.00	$4.00
6	$12	$6.75	$1	$7.75	$4.25
7	$14	$8.75	$1	$9.75	$4.25
8	$16	$11.00	$1	$12.00	$4.00
9	$18	$13.50	$1	$14.50	$3.50
10	$20	$16.25	$1	$17.25	$2.75

Notice also as a result that the profits of *every* column are $1 less than previously. The maximum profit is 6-7 cups. Just as before.

How about the profit-maximizing rule? It does NOT change. Remember, marginal cost is defined as the increase in cost from brewing an additional cup – that is, the increase in TC for an additional cup. Notice that when going from one cup (TC = $1.50) to two cups (TC = $2.25), we increase total costs by $0.75 That's the marginal cost of the second cup. But notice that is really just the increase in Variable Costs. Since fixed costs do not change with scale, they do not affect the marginal cost number. So calculating the marginal columns:

Qty	Marginal Revenue	Marginal Cost
0	$2	$0
1	$2	$0.50
2	$2	$0.75
3	$2	$1.00
4	$2	$1.25
5	$2	$1.50
6	$2	$1.75
7	**$2**	**$2.00**
8	$2	$2.25
9	$2	$2.50

10	$2	$2.75

which is *exactly* the same as the columns we had in the earlier example without fixed costs. The solution is the same – profits are maximized when MR = MC, which in this case is at $2, and that will be at 7 cups produced. Exactly as we had in the previous example. It doesn't seem as if the inclusion of Fixed Costs have affected the profit-maximizing scale decision at all. How much they produce is determined purely by marginal revenue-marginal cost considerations, which fixed costs do not influence.

In sum, fixed costs do *not* affect a firm's decision on *how much* to produce. Yes, they reduce the amount of profit they end up taking home. But the profit-maximizing level of production is going to be at the same with or without fixed costs.

ECONOMIES OF SCALE

In our section on "Growth", we briefly mention the idea of "**economies of scale**" - that is, the ability to lower unit costs at a large scale of production. This is usually because large scale allows you to either implement more efficient cost-saving production techniques that you couldn't implement at the smaller scale..

The traditional example is, of course, the assembly plant. If you're producing 10,000 cars a year, it makes sense to set up a huge factory with hundreds of workers and a conveyor belt carrying cars down the line, each laborer dedicated to a specialized task in the assembly of the car. But if you're only producing 10 cars a year, the assembly plant method is not very useful or profitable. Better to produce each car individually by craft methods with only a few laborers than undertake the costs of setting up a massive factory and hiring hundreds of workers.

However, *per car*, the assembly plant method is more efficient, that is, it costs less per car than the craft method. The problem is that it isn't worthwhile to implement the assembly plant method unless you achieve a certain large level of production.

Another example is discounts on bulk buying. If you're a producer or retailer of considerable size, you can get discounts on the price of inputs if you buy them on a grand scale, e.g. a small shop which orders only 100 yards of cloth to sell may be charged $1 per yard by the manufacturer; but if you're a gigantic retailer like Wal-Mart who buys 10,000 yards of cloth in one go, you may be able to get a discount and pay only $0.75 per yard. So bulk buying gives a cost advantage that comes with scale, and allows you to *reduce* your costs per unit as you increase in size.

That seems to give the lie to the concept of increasing cost. Because if *cheaper* techniques of production or bulk discounts are made available as you increase production, costs per unit will actually *decrease* as you increase production. So, in a sense, we *can* have a law of *decreasing costs*!

How is this compatible with our story? Well, our story is still true, *eventually*. It is possible for firms to exploit economies of scale and reduce costs for a little while, but it is *not* possible for them to do so as they continue to increase scale. *Eventually* they will hit resource barriers and costs will begin being driven up again.

To see why we cannot have decreasing costs permanently, let us take our coffee retailer again. Suppose that he does indeed take advantage of bulk-buying of coffee beans so that the more he produces the *cheaper* brewing an additional cup gets. Let's start from $0.50 for one cup, and suppose that bulk buying discounts means the cost of every additional cup is *reduced* by increments of $0.05. The table would look like the following:

Qty	Price	Total Revenue (= Price × Qty)	Cost of cup	Total Costs	Profits (TR – TC)
0	$2	$0	$0	$0	$0
1	$2	$2	$0.50	$0.50	$1.50
2	$2	$4	$0.45	$0.95	$3.05
3	$2	$6	$0.40	$1.35	$4.65
4	$2	$8	$0.35	$1.70	$6.30
5	$2	$10	$0.30	$2.00	$8.00
6	$2	$12	$0.25	$2.25	$9.75
7	$2	$14	$0.20	$2.45	$11.55
8	$2	$16	$0.15	$2.60	$13.40
9	$2	$18	$0.10	$2.70	$15.30
10	$2	$20	$0.05	$2.75	$17.25

Notice that Total Costs are increasing, but they're increasing at a decreasing rate. Every extra cup costs less and less to produce.

What's the profit-maximizing point? There isn't any! Profits are rising rapidly with scale. Endlessly so. The profit-maximizing scale of production is *infinite*.

This is a ludicrous proposition. Decreasing cost seems *impossible*.

But we know it is possible – since we have many real world examples of economies of scale – assembly plants, bulk buying, etc. How do we account for that?

We accept it is possible to enjoy decreasing costs, but only for a spell. That is, economies of scale can emerge and decrease unit costs for a while. But if scale *continues* to rise, we will eventually get increasing costs again.

As you increase production from 10 to 10,000 cars, costs per car may fall because you can now introduce an assembly plant. So decreasing costs operate here. But as you *continue* expanding production beyond that, from 10,000 to 100,000 to 1 million or more, increasing costs kicks in again. Simply because there's not many new cost-saving techniques available beyond that. More to the point, resources are limited. If you continue increasing and increasing, your scale will become so large that the entire country's supply of steel and workers will be unable to meet your voracious demand for inputs. When you run up against the resource barrier, you'll be driving up costs rapidly.

In other words, you can decrease costs per unit when going from small to big. But not necessarily when going from big to bigger. So while improved techniques may stave off increasing costs for a while, they cannot stave it off forever. Increasing costs will reassert themselves. And we will be able to identify a profit-maximizing point.

We can see a concrete example of this in the next table. Notice that initially we have decreasing costs, our first few cups become incrementally cheaper – 50 cents, 45 cents, 40 cents, but at the fourth cup, increasing costs kicks in and they start becoming incrementally more expensive again 75 cents, $1, $1.25 and so on.

Qty	Price	Total Revenue (= Price × Qty)	Cost of cup	Total Costs	Profits (TR – TC)
0	$2	$0	$0	$0	$0
1	$2	$2	$0.50	$0.50	$1.50
2	$2	$4	$0.45	$0.95	$3.05
3	$2	$6	$0.40	$1.35	$4.65
4	$2	$8	$0.50	$1.85	$6.15
5	$2	$10	$0.75	$2.60	$8.00
6	$2	$12	$1.00	$3.60	$8.40
7	$2	$14	$1.25	$4.85	$9.15
8	$2	$16	$1.50	$6.35	$9.65
9	$2	$18	$1.75	$8.10	$9.90
10	$2	$20	$2.00	$10.10	$9.90
11	$2	$22	$2.25	$12.35	$9.65

That means that up to four cups, you can benefit from improved techniques/bulk buying and other advantages of increasing scale, but thereafter costs go up again.

This doesn't change the profit-maximizing rule. Because we have increasing costs reasserting themselves eventually, there will be a profit-maximizing point where MR = MC that is finite. In this example, that the profit maximizing point where MR = MC = $2 is at 10 cups. Further increases will reduce profits.

In sum, while we do recognize that economies of scale exist, and that we may have decreasing costs for a spell, the law of increasing cost will always impose itself eventually.

FACTOR SCALE

Profit-maximizing firms think along margins. But the decision on the scale of output production is not the only decision firms make. Firms also use marginal thinking to determine the most profitable scale of employment of factors of production.

Remember: a 'factor of production' is any *input* into a production process (land, labor, capital).

The language here is slightly different. Economists like to use the terms "Marginal Product of a Factor" or "Marginal Revenue Product" to differentiate it from straight Marginal Revenue. And "Marginal Factor Cost" to differentiate it from "Marginal Cost" The reason is because now we're not measuring per unit of output produced but rather per unit *of input employed*.

Marginal Revenue Product (or **Marginal Product** of a Factor): the additional revenue a firm gets from hiring one additional unit of a factor.

Marginal Factor Cost: the extra total cost incurred by a firm from hiring an additional unit of an input.

which invites a new marginal rule of its own for hiring factors:

> **Marginal Rule** (Factor Version): to maximize profits, a firm should hire inputs until the marginal revenue product is equal to the marginal factor cost

This rule, that MRP = MFC, determines the profit-maximizing factor scale. It is the factor market analogy to the MR = MC rule in the output scale. Succinctly:

-- if MRP > MFC, then the revenue gained from hiring an additional factor is worth more than the hiring cost, so a firm should go ahead and do it.

-- if MRP < MFC, then the revenue gained from hiring an additional factor is worth less than the hiring cost, so a firm should cut back employment.

-- if MRP = MFC, hiring an additional unit of the factor generate as much revenue as it costs to hire him. The firm gains nothing and loses nothing by hiring an additional factor. A firm should stop here.

Let us take a simple example. Suppose you own a car dealership and are thinking about hiring salesmen on your lot. Suppose you experiment a little and find the following:

Number of salesmen	Total sales
1 salesman alone	10 cars sold
2 salesmen together	18 cars sold
3 salesmen together	24 cars sold
4 salesmen together	28 cars sold

Notice that the more salesmen you hire, the greater the number of car sales. But notice that it is increasing at a decreasing rate. One salesmen can sell 10 cars in a day. Two salesmen can sell 18 cars. That averages to about 9 cars each.

The reason the average declines is *not* because the second salesmen is a poorer salesman than the first, but rather that two salesmen working on one lot tend to get in each other's way a bit and end up not individually making as many sales as they would if they worked the whole lot by themselves.

Why? Well, maybe they get confused about who's serving whom, and waste time fighting for the same customer. Or maybe they have to jostle for space in the back-office, e.g. to finalize his sale, one salesman has to wait until the other one is done with the office computer. Or maybe they have to wait to talk to you, the owner, to relieve a doubt or approve something, while you are busy chatting with the other guy.

These confusions, jostles and little wastes of time the two salesmen impose on each other means that they don't do as well together as they would alone, with full and unhampered access to the office and your attention. Yes, with two salesmen, more is sold as a whole - 18 cars. But the jostling means they only manage about 9 sales each. Whereas if they other guy wasn't getting in the way, they would have been able to manage 10 sales each.

Similarly with three salesmen. Again, more as a whole is sold - 24 cars - than with two salesmen. But the extra jostling, confusion and little wastes of time you get with three workers implies each sells less - 8 cars sales each.

The diminishing sales as you increase staff is called the ***Law of Diminishing Returns*** (or the law of "diminishing productivity").

> **The Law of Diminishing Returns**: The more of an input you add to a production process, the less revenue each additional unit of that input generates.

You can consider it a flip version of the law of increasing costs, expressed from the point of view of adding inputs.

So how many salesmen should you hire? Depends on the value of the cars and the wages of the salesmen. Suppose each car sells for $1,000 (these are cheap cars) and hiring a salesman costs you $7, 500. How many should you hire?

You could do the total profit calculations at every level. But suppose you're lazy and want to take a short-cut. You figure that so long as a salesman makes more sales than he costs to hire, then you shall keep him.

Your first temptation may be to calculate how much each salesman sells. To do this, you divide the total sales volume by the number of salesmen, i.e. calculate the ***average* product**.

Number of salesmen	**Total revenues**	**Average revenue Product**
1 salesman	$10,000	$10,000
2 salesmen	$18,000	$9,000
3 salesmen	$24,000	$8,000
4 salesmen	$28,000	$7,000

You notice that in all cases except the last, the salesmen seem to be making more sales than they cost to hire.

You remember very vaguely from your economics class many years ago something about equating costs to revenues for adding workers and decide to focus on 3 salesman. They each sell $8,000, and they cost only $7,500. Maybe that's the profit-maximizing point?

Of course, you're wrong. Because the marginal rule is not about averages but about *margins*. Not how much each actually sells, or sells on average, but how much an *additional* salesman sells.

This is the table you want:

Number of salesmen	**Total revenues**	**Marginal revenue Product**
1 salesman	$10,000	$10,000
2 salesmen	$18,000	$8,000
3 salesmen	$24,000	$6,000
4 salesmen	$28,000	$4,000

The second salesman lowers the average to $9,000, but he *adds* only $8,000 to what the first one was doing. The third salesman lowers the average to $8,000, but *adds* only $6,000 to what the first two salesmen were doing.

The Marginal Rule tells us we should equate marginal cost to marginal revenue (or marginal revenue product, to be a stickler for nomenclature), not average revenue.

Evidently it is worth adding the first salesmen - he produces $10,000 and only costs $7,500. Clearly profitable. Adding a second worker increases sales revenues by $8,000. Again, profitable, since he only costs us $7,500. But adding a third worker increases sales revenues by only $6,000 – that is *less* than what he costs to hire.

So we should hire only two salesmen. The full profit calculations bear this out:

Number of salesmen	Total revenues	Total Costs (wages × salesmen hired)	Total Profits (total revenues - total costs)
1 salesman	$10,000	$7,500	$2,500
2 salesmen	$18,000	$15,000	$3,000
3 salesmen	$24,000	$22,500	$1,600
4 salesmen	$28,000	$30,000	-$2,000

Maximum profit is clearly at two salesmen, not three. Our marginal rule has not failed us.

Factor Demand Curve

Suppose wages increase to $8,500, the marginal rule tells us to only hire one salesman since the second salesman only adds $8,000 and that is less than what he costs. If you are not convinced, check the profit calculations when wages are $8,500:

Number of salesmen	Total revenues	Total Costs (wages × salesmen hired)	Total Profits (total revenues - total costs)
1 salesman	$10,000	$8,500	$1,500
2 salesmen	$18,000	$17,000	$1,000
3 salesmen	$24,000	$25,500	-$1,500
4 salesmen	$28,000	$34,000	-$7,000

Maximum profits are indeed at one salesman.

What if wages decline to $5,500? The marginal rule tells us to hire 3 salesmen since the first three salesmen generate enough additional revenue each (10,000, 8,000 and 6, 000) to overcome their hiring cost of $5,500, but the fourth (who adds only 4,000) does not. Again, if you want to double-check, just look at profit tables when wages are $5,500:

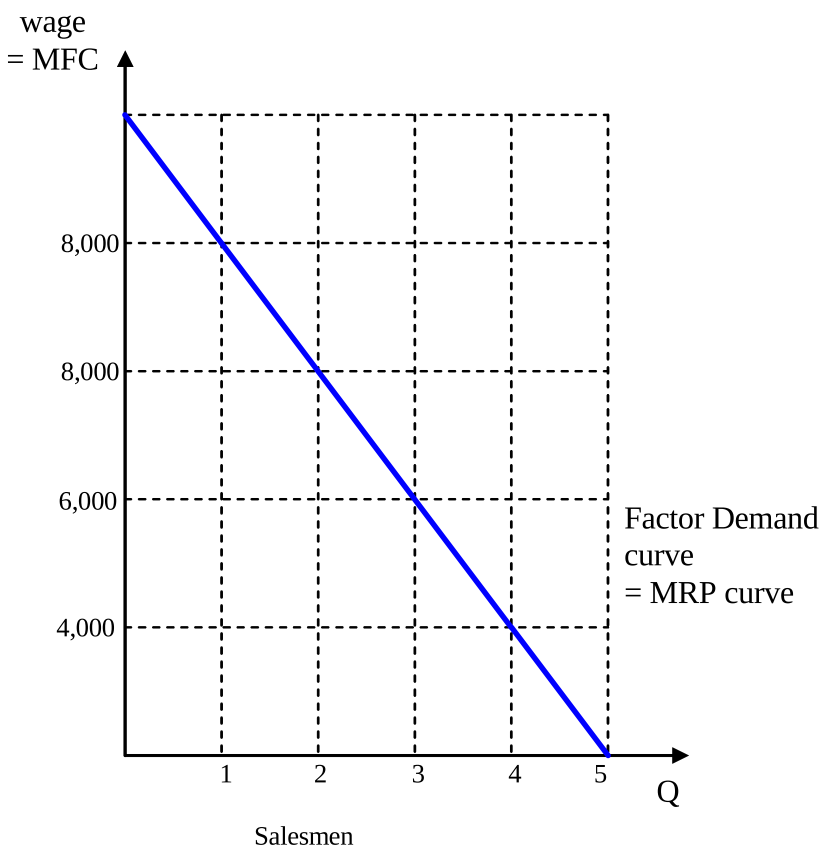
Number of salesmen	Total revenues	Total Costs (wages × salesmen hired)	Total Profits (total revenues - total costs)
1 salesman	$10,000	$5,500	$4,500
2 salesmen	$18,000	$11,000	$7,000
3 salesmen	$24,000	$16,500	$7,500
4 salesmen	$28,000	$22,000	$6,000

In summary, when wages were $5,500, we hire three salesmen; when they are $7,500 we hire two, when they are $8,500, we hire only one. In other words, when wages rise, the firm decides to hire *less* salesmen.

As wages rise, the firm reduces the amount of workers hired; if wages decline, they increase the amount of workers demanded.

Does this sound familiar? It should.   This explains the downward-sloping shape of the labor **demand curve**. We had left that unexplained before. Now it may make more sense.

In fact, the MRP = MFC rule the marginal revenue product of a factor exactly traces the firm's factor demand curve.



In sum: Language can sometimes be tricky. With so many margins flying around, it is sometimes easy to get confused. So remember:

- when we are dealing with output scale, that is, thinking about the output of *goods* (stereos, coffee, etc.), profits are maximized by comparing the given price of the good (e.g. price of stereos), with the *marginal cost* of output ('marginal cost of stereos'). This yields the output supply curve. The Law of Increasing Cost is at work here.

- when we are dealing with inputs of *factors* (labor, land, capital), profits are maximized by comparing the *marginal product* of the factor ('marginal product of labor') with the given factor price (wage, rent, etc.) This yields the factor demand curve. The Law of Diminishing Returns is at work here.

MONOPOLISTIC SITUATIONS

Let us return to the issue of output scale. The profit-maximizing rule was to adjust quantity produced until Marginal Revenue = Marginal Cost.

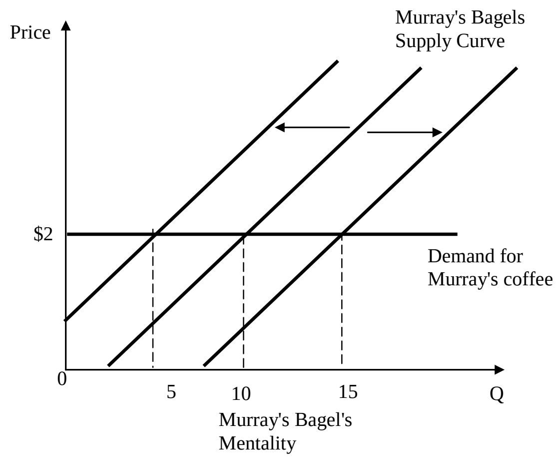
Up until now, we always assumed that marginal revenue was just the market price. That may be true for situations where the market is perfectly competitive, e.g. our coffee shop. An individual coffee shop in NYC really doesn't have a choice but to take price of coffee as "given" by the market. There are so many competing coffee shops in the city, that a single shop does not really have the market power, individually, to push the price of coffee across the city up or down.

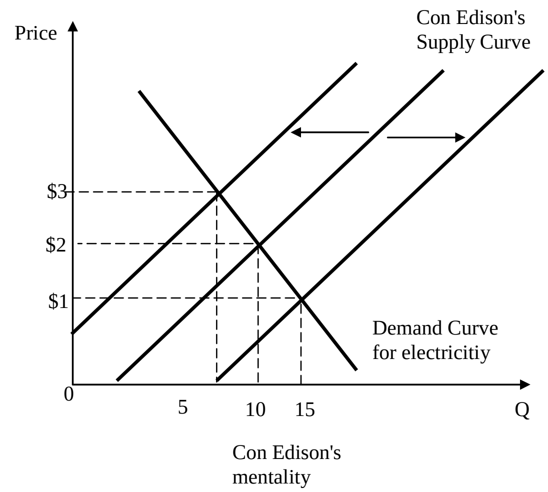
But what if the market is less-than-perfectly competitive? Suppose it is case of **monopoly** (market with only one producer/seller) or an **oligopoly** (market dominated by only a handful of firms, say two or three). Clearly such large firms do have market power and whatever they do will end up pushing prices up or down. We're thinking of the likes of companies like Verizon in telephone service or Con Edison in energy, firms with sufficient size to influence the price of telephony and energy.

Clearly, monopolistic companies like Verizon and Con Edison are operating in a different environment than little coffee shops like Murray's Bagels. While Murray's Bagel's has a miniscule and negligible effect on the market supply curve of coffee in NYC, Con Edison's supply *is* practically the total market supply of electricity.

As such, the individual decisions of monopolistic firms like Verizon and Con Edison have an impact on price. What kind of impact? Well, if Con Edison increases the production of electricity, then the price of electricity drops. It must. Since Con Edison is the only supplier, its production decision is simultaneously the market's entire supply. And, by the laws of supply & demand, we know when market supply increases, prices fall.

So Con Edison's supply decision affects market prices. Consequently, when deciding how much to produce, it must realize whatever it does will have an effect on price, it *must* take market demand curve into account. It cannot simply consider a quantity decision and assume it will sell at the current price. It must realize that its supply decision will also affect price.





[Pictures contrasts the "mentality" of Murray's Bagels and Con Edison. Murray's assumes coffee prices are given, that is, that however much coffee it decides to supply – 5, 10 or 15 units - it can sell it all at the same price ($1.95). So it "thinks" of the demand for its coffee as a perfectly flat demand curve. No matter how much it supplies, Murray's Bagels imagines prices won't fall.

Con Edison, however, knows it is the only supplier of electricity and recognizes that the demand for electricity is like a regular market demand curve. If it supplies only 5 units, then it can sell for the price of $3 per unit, if it increases supply to 10 units, the price at which it can sell falls to $2 per unit, and so on. It recognizes that its supply decision will affect market prices.]

Economists like to say, unlike firms in a competitive environment, firms in a monopolistic or oligopolistic environment have "**market power**", their supply decision influences the market price of the output.

What implications does this have?

Firstly, marginal revenue is no longer simply the current price.

Remember, marginal revenue is defined as the additional revenue received if you increase production by one unit. A small coffee shop assumes it can sell that additional unit at the same price as the current price. So if current prices are $2, then the additional unit will also sell at $2. But a monopolist has market power. It is the sole producer facing a downward-sloping demand curve. So every extra unit it produces will push market price down a little. It knows it can't sell its output for the same price.

Let us suppose our coffee shop is actually a monopoly – the only seller of coffee in the city. Recognizing this fact, it realizes that price isn't constant but now that the more coffee it produces, the lower the selling price becomes. As such, its table of calculations is as follows:

Qty	Price	Total Revenue (= Price × Qty)	Marginal Cost	Total Costs (cumulative)
0	$3.00	$0	$0	$0
1	$2.75	$2.75	$0.50	$0.50
2	$2.50	$5	$0.75	$1.25
3	$2.25	$6.75	$1.00	$2.25
4	$2.00	$8	$1.25	$3.50
5	$1.75	$8.75	$1.50	$5.00
6	$1.50	$9	$1.75	$6.75
7	$1.25	$8.75	$2.00	$8.75
8	$1.00	$8	$2.25	$11.00
9	$0.75	$6.75	$2.50	$13.50
10	$0.50	$5	$2.75	$16.25

The cost structure is the same as in our earlier example: increasing costs, by increments of $0.25. In this, the monopolist is no different from a competitive firm. The difference lies in the *price* column. Now the price is *not* constant, $2 at every entry. Rather price starts at $3.00 and declines by increments of $0.25. If it produces only one cup, then market price for coffee is $2.75. If it produces two cups, then market price decline to $2.50. If it produces three cups, market price declines to $2.25. And so on.

(This is just supply tracing the shape of the market's coffee demand curve: the more coffee our monopolist produces, the lower the price of coffee becomes. Again, this is because the monopolist is the only supplier, so whatever it produces is the total market supply, interacting with demand in the conventional demand-and-supply manner.)

The important thing is that the monopolist *realizes this*. It realizes that if it increases production, price will fall. And that changes its calculation of marginal revenue at every entry.

Remember, marginal revenue is defined as the change in Total Revenue from the production of an additional unit of output. Consequently, it calculates the following:

Qty	Price	Total Revenue (= Price × Qty)	Marginal Revenue
0	$3.00	$0	$0
1	$2.75	$2.75	$2.75
2	$2.50	$5	$2.25
3	$2.25	$6.75	$1.75
4	$2.00	$8	$1.25
5	$1.75	$8.75	$0.75

6	$1.50	$9	$0.25
7	$1.25	$8.75	-$0.25
8	$1.00	$8	-$0.75
9	$0.75	$6.75	-$1.25
10	$0.50	$5	-$1.75

So increasing production from 1 to 2 units increases total revenues from $2.75 to $5, that is, an increase of $2.25. So marginal revenue of the second cup is $2.25. Increasing from 2 to 3 units increases revenues from $5 to $6.75, so the marginal revenue of the third cup is $1.75. And so on.

So marginal revenue is no longer constant. It is declining. Indeed, as we see, it is declining at a rate of $0.50 per additional cup produced.

[How is it that price declines at a rate of $0.25 per cup, but marginal revenue declines at a faster rate of $0.50 per cup? That is because we assume our monopolist is *not* a price discriminator, that it cannot sell different cups at different prices. If it produces 2 cups, it sells *both* at $2.50, for a total of $5. If, however, it decides to produce three cups, the market price declines to $2.25. That $2.25 is not merely the price of the third cup, but rather the price of *all* three cups on the market. It is not merely the third cup that sells for less, but all three cups sell for less. So the decline in marginal revenue falls *faster* than the price. If it was a price discriminator it could charge different prices for different cups, and marginal revenue wouldn't fall that fast. But it isn't.]

What's the implications of this for the firm. Well, the profit-maximizing decision of the monopolist is the same as that of any firm: profits are maximized where marginal revenue equals marginal cost. So putting those two columns alongside each other:

Qty	Marginal Revenue	Marginal Cost
0	$0	$0
1	$2.75	$0.50
2	$2.25	$0.75
3	$1.75	$1.00
4	**$1.25**	**$1.25**
5	$0.75	$1.50
6	$0.25	$1.75
7	-$0.25	$2.00
8	-$0.75	$2.25
9	-$1.25	$2.50
10	-$1.75	$2.75

We can see MR = MC = $1.25 at 4 units, so the profit-maximizing decision of the monopolistic firm will be to produce 4 units.

The interesting point arises in the next step. Because we know that in the case of a monopolistic firm, marginal revenue is *not* equal to market price, but rather something less, that means the prices are no longer equal to marginal cost. Take a careful look at the complete table:

Qty	Price	Total Revenue (= Price × Qty)	Marginal Revenue	Marginal Cost	Total Costs (cumulative)	Profits (TR – TC)
0	$3.00	$0	$0	$0	$0	$0
1	$2.75	$2.75	$2.75	$0.50	$0.50	$2.25
2	$2.50	$5	$2.25	$0.75	$1.25	$3.75
3	$2.25	$6.75	$1.75	$1.00	$2.25	$4.5
4	**$2.00**	**$8**	**$1.25**	**$1.25**	**$3.50**	**$4.5**
5	$1.75	$8.75	$0.75	$1.50	$5.00	$3.75
6	$1.50	$9	$0.25	$1.75	$6.75	$2.25
7	$1.25	$8.75	-$0.25	$2.00	$8.75	$0
8	$1.00	$8	-$0.75	$2.25	$11.00	-$3
9	$0.75	$6.75	-$1.25	$2.50	$13.50	-$6.75
10	$0.50	$5	-$1.75	$2.75	$16.25	-$11.25

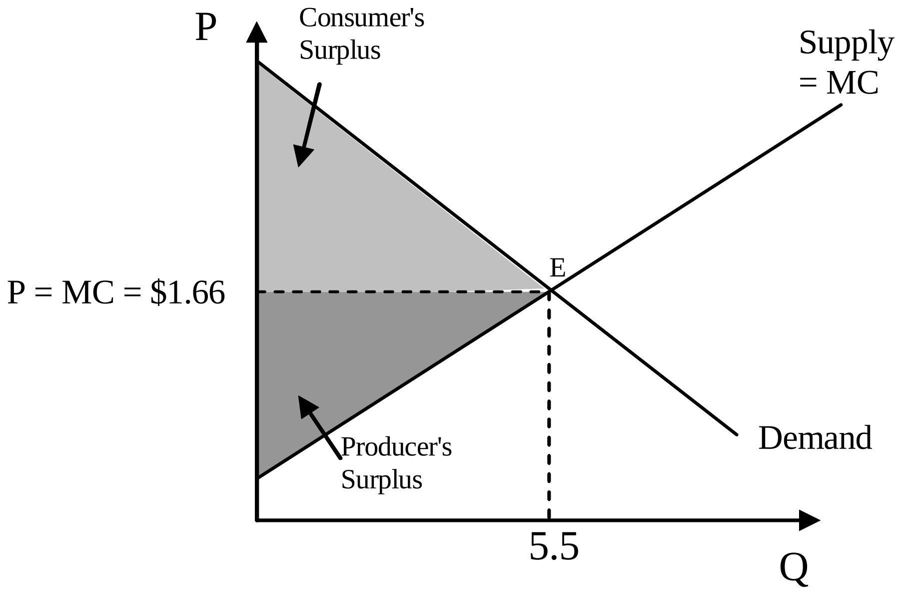
At the profit-maximizing quantity (4 units), the market price is $2.00, but the marginal cost is $1.25.

<u>Monopoly vs. Competition</u>

This is the big contrast between monopolistic and competitive environments. In competitive environments, Price = Marginal Cost at the solution. But in a monopolistic environment, Price > Marginal Cost at the solution.

What does this mean? It means that in monopolistic situations, the market no longer settles where demand and supply curves intersect, but somewhere to the left of it.

To understand why, look at the simple market below:



Competitive Market

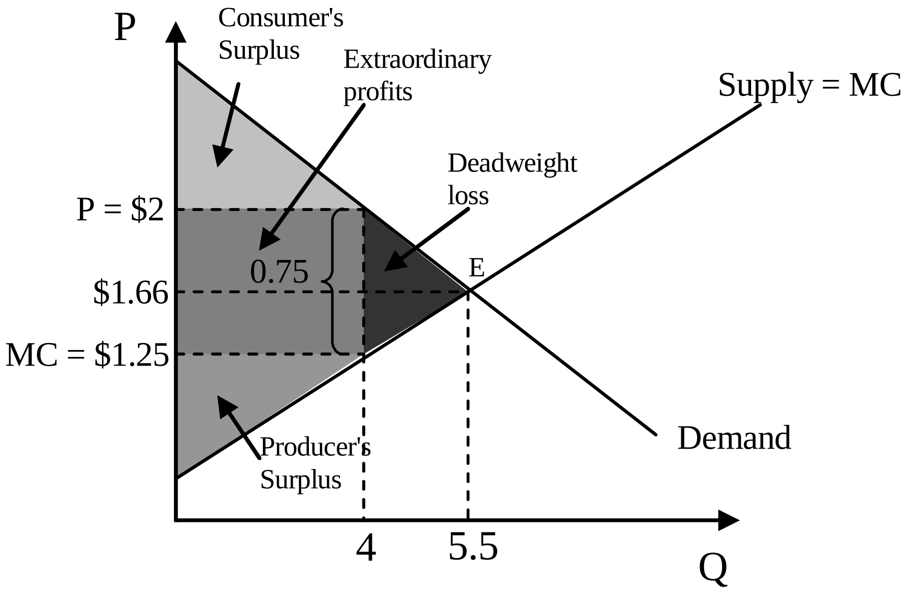
If equilibrium settles at the intersection of the D & S curves, then market price will be $1.66 and quantity is 5.5. But remember that we said earlier that the firm's supply curve *is* the marginal cost curve. That is, every point on the supply curve represents MC. So, if quantity is 5.5, then marginal cost (bounce off the supply curve) is $1.66. So $1.66 is both the market price and the marginal cost, i.e. P = MC = $1.66.

So it is *necessarily* true than in a competitive market situation, price is equal to marginal cost (P = MC).

[Note: that equilibrium entry of 5.5 units isn't explicitly shown in the earlier numerical table. But it is approximately there. That is because you will notice that at 5 units, P = $1.75 and MC = $1.50 (so P > MC), while at 6 units, P = $1.50 and MC = $1.75 (so P < MC). So P = MC somewhere in between 5 and 6, maybe around 5.5 units, with price somewhere around $1.66 or so. I don't feel like reworking my entire numerical example just to show it explicitly. Take it on faith.]

In the monopolistic situation, we know that the solution is *not* where P = MC, but rather where P > MC, or, explicitly, where P = $2.00 and MC = $1.25. There is a $0.75 gap between the market price consumers face ($2) and the marginal cost to the monopolistic firm ($1.25).

So the monopolistic solution *can't* be at the intersection of the D & S curves, since at the intersection it is necessarily the case P = MC. Rather, the monopolistic solution can be found by fitting in a "wedge" of size $0.75 between demand & supply curves (sort of like the excise tax). It will look like the following:



Monopolistic Market

Production is at 4 units, price (bouncing off the demand curve) is at $2, while marginal cost (bouncing off the supply curve) is $1.25.

Compare the monopolistic to the competitive situations.

- output in a monopolistic situation (4) is *lower* than in a competitive situation (5.5)
- the market price in a monopolistic situation ($2) is *higher* than the price in a competitive situation ($.166).
- the consumer's surplus in a monopolistic situation is *smaller* than in a competitive situation.
- the monopolistic firm not only takes the producer's surplus, it also makes extraordinary profits (rectangle formed by 0.75 wedge), what is sometimes called "monopoly rents".
- there is deadweight welfare loss (dark triangle).

Noticed we obtained this result not by assuming the monopolist *behaved* any differently. Both the competitive firm and the monopolist are attempting to maximize profits, both the competitive firm and the monopolist find their output decision where MR = MC. The difference is that the monopolist is operating in a different environment. The competitive firm faces a constant MR (its output decision has a negligible influence on price), while the monopolist faces a declining MR (its output decision has a significant influence on price). For that reason, and for that reason alone, we reach two completely different results on the market: output is lower and prices higher under a monopoly than in a competitive situation. As a result, consumers are worse off and the result is less efficient (deadweight loss).

The detrimental effect of monopoly upon social welfare can thus be precisely measured by the smaller consumers' surplus *and* the deadweight loss which no one is getting.

Barriers to Entry

The first thing you should probably ask is: if monopolies make such extraordinary profits, how can they remain monopolies? Those gains are great. Wouldn't dozens of other firms be set up to try to make extraordinary profits like that too? What is stopping them?

There are many types of barriers to entry, which can be roughly classified as technical or legal.

(1) For instance, it may be that there are **economies of scale** that make entry unprofitable. A newcomer breaking into a market usually has to start at low levels of production. But at those low levels, rudimentary, expensive techniques may be the only ones available, while the large monopolist, already large, has been enjoying economies of scale and *can* consistently under-price the newcomer and run him out of business.

That is why Wal-Mart, a large chain-store with stupendous economies of scale (large warehouses, bulk suppliers, etc.) to keep down costs, can monopolize a regional market. Its economies of scale allow it to keep costs low and under-price any newcomers (and old mom-and-pops) which are producing at a smaller (and costlier) scale.

(2) Another barrier may be **network externalities**. Some products enjoy "network externalities", in the sense that the more people that use them, the more useful they are. Things like operating systems, credit cards, etc. have network externalities. If everyone is using MS Windows or Visa, there is little point for consumers using anything else. For reasons of technical compatibility, you want the OS that everyone is using, you want the credit card which most vendors accept.

A newcomer trying to launch a new operating system or a new credit card will have great difficulties selling it as people don't want to end up with a product they can't use anywhere. As such, it is unprofitable for a newcomer to break in.

(3) **Patents** are a type of legal barrier to entry. If a company owns the patent for a very unique product, it may be impossible for a newcomer to break in. e.g. for a long time, Xerox and Polaroid were the only producers of their kind of product and newcomers couldn't break into the market without violating patent law.

(4) Sometimes governments may just grant an **exclusive franchise** to a company by legal means. This is usually in the case of **natural monopolies** (see below), like public utilities, like gas and electricity service, communications services, airline routes, railways, television/radio markets and variety of other situations. Government limits entry because, in this case, competition can be socially costly. If, for instance, the radio spectrum was wide open, and anyone could set up a radio station, the signals would all be crossed and transmission would be garbled. So, it has partitioned the spectrum into a

few channels and sells them to only a small number of companies and forbids new entrants beyond that number.

(5) Some kinds of barriers to entry are not naturally technical or naturally legal, but are falsely and artificially created by monopolistic firms (what is called "**rent-seeking behavior**"). This can be through all kinds of tricks. The De Beers cartel on diamond supply, and the OPEC cartel on oil supply, is an artificial creation, made possible only by **collusion** of major suppliers, preventing entry of newcomers by taking over all the major sources of raw material. A firm or cartel may also lobby the government and obtain a legal warrant of monopoly for little or no good reason.

Anti-Trust Policy

How big is the welfare loss from monopoly? Back in 1954, Arnold Harberger did a celebrated study which measured the total loss deadweight loss (the dark triangle) from all the monopolies and oligopolies in the US economy. The total amount was $150 million - that is, about 0.1% of GDP. The proportion hasn't really changed much since, which makes some economists believe that perhaps the pernicious effect is minor and that there is little point in pursuing too-vigorous anti-trust policy. However, using slightly different methods, some other economists (e.g. Cowling and Mueller, 1978) have calculated that the deadweight loss to be about as great as 13% of GDP. The truth is probably somewhere in between.

The welfare loss from monopoly is the principle reason why governments have pursued **anti-trust policy**. The first anti-trust law in the US was the **Sherman Act** of 1890 and has been followed up by others since. The law was designed to prevent the creation of monopolies and encourage competition.

Anti-trust law also authorizes the government to break up already-existing monopolistic companies if it is found to be too detrimental to the public interest. Famous break-ups by the federal government include the dismantling of John D. Rockefeller's Standard Oil company in 1911 into smaller separate companies (Mobil, Exxon, Chevron, Conoco, etc.) and the 1982 break up of AT &T ("Ma Bell") into the "Baby Bells" (NYNEX, Bell Atlantic, Bell South, Ameritech, Qwest, Southwestern Bell, Pacific).

However, break-ups are used only rarely, and anti-trust law usually focuses on pre-emptive strikes, catching firms using tricks to bully competitors out, before it is too late.

Regulation of Natural Monopolies

But it might not be possible - or wise - to break up or prevent all monopolies. Some monopolies, called "**natural monopolies**", should be left to stand because competition would just waste resources.

A typical example is, say, water supply. To compete properly, different water companies would have to lay down competing pipes and infrastructure. Pipe-laying in a city is disruptive to city life and traffic. If there are ten competing pipes for every household, it is a bit of a replicated effort and a waste of resources. Isn't there are more efficient way to get the competitive results without breaking up the monopoly into multiple companies?

(a) Nationalization

One solution is **nationalization** of the industries which are natural monopolies (like water, electricity, railways, etc.), that is government seizes control of the industry. The problem with monopoly, after all, was that the profit-maximization decision was that they aimed for MR = MC, and that meant *less* quantity would be produced than otherwise. A nationalized company is not required to maximize profit. We can propose a different rule for it.

One obvious rule is to force the nationalized company to follow a "**marginal cost pricing**" rule, that is, find the level of output where P = MC, not where MR = MC. That is we *force* the company to price at marginal cost, rather than allow it to maximize profits. In that manner, we force the nationalized monopoly to behave *as if* it were in a competitive market, even though it isn't.

Example: using the numbers of our previous numerical example for our coffee monopoly, we can compare the monopolistic solution (MR = MC, with price at $2 and quantity at 4 units) and the competitive solution (P = MC at 5.5 units and price at $1.66 – we have to make numbers up a bit, but they're approximately correct if you compare the earlier table):

Qty	Price	Total Revenue	Marginal Revenue	Marginal Cost	Total Costs	Profits (TR – TC)
4	$2.00	$8	$1.25	$1.25	$3.50	$4.5
5.5	$1.66	$9.13	$0.50	$1.66	$6.00	$3.13

Clearly the profit at P = MC ($3.13) is less than the profit at MR = MC ($4.5). But it is profitable enough. So forcing firms to price at 5.5 rather than 4 wouldn't be forcing firms to swallow a loss. They make profit enough.

The advantage of a marginal cost pricing rule is that you get the same *result* as if you broke it up into competing firms. Except that you don't really break it up. You end up with the competitive solution in price & quantity, but save on the unnecessary trouble and waste of resource of replicating pipelines.

But government ownership brings with it a whole slew of other problems - principally that nationalized companies are not very good at staying efficient or even sticking to P = MC. They tend to start chasing other objectives, such as maximizing employment. Unconstrained by profit motives, they can accumulate higher costs, out of inefficiency or incompetence and then just raise the price to "keep up" with rising MC. (currently, the

Metropolitan Transit Authority, MTA, is the poster boy of such misbehavior by a publicly-owned monopoly.)

(b) Price Controls

Nationalization was a popular solution between the 1950s and esp. 1970s for the natural monopolies. But since the 1980s, the mood has changed. It is generally recognized that disciplinary whip is often better held by private shareholders, who demand profits and thus have a vested interest in efficiency, than government bureaucrats who care not an iota and are happy enough to pile on costs and pass it on. Consequently, since the 1980s, the trend has to been to find a better solution.

An alternative to outright nationalization is to allow the monopoly to remain private, but impose a government **price-control board** to oversee its pricing. The board can set the selling price to ensure that the private monopoly replicates a P = MC solution. And the shareholders would make sure it does that (since it is the profit-maximizing solution in that context), and not stray off into other objectives. Again, this achieves the competitive solution without actually breaking anybody up.

(c) Common Carrier

Another alternative is **common carrier** schemes to regulate the delivery, but allow competitive supply. That is the monopoly *is* broken up by anti-trust law on the supply side, but to avoid replication of pipes, the owner of the pipe is forced to rent its pipes to competitors at a government-set price. Again, we have a competitive P = MC result, but without unnecessary replication of infrastructure.

This is what we see with a lot of telephone and electricity supply companies in New York. Verizon maintains the phone lines, Con Edison maintains the electrical wires, but the suppliers of telephone and electricity service are many.

Alternatives to MC Pricing

Forcing natural monopolies to price at marginal cost may be a laudable objective. But it may not always be possible for technical reasons. For instance, it *may* be that the P = MC solution may nonetheless yields *negative* profits, that is average costs (cost per unit) are higher than price. This is not impossible. Indeed, it is quite common for natural monopolies with huge fixed costs that eat up a lot of profit and drag its profits into the negative.

e.g. suppose our coffee monopoly has huge fixed costs of, say, \$4. Let us compare again the profit-maximizing MR = MC solution against the forced P = MC solution using our earlier numbers:

Qty	Price	Total Revenue	Marginal Revenue	Marginal Cost	Variable Costs	Fixed Costs	Total Costs	Profits (TR – TC)
4	$2.00	$8	$1.25	$1.25	$3.50	$4.00	$0.50	$0.50
5.5	$1.66	$9.13	$0.50	$1.66	$6.00	$4.00	$10.00	-$0.87

If allowed to maximize profits freely (MR = MC), the firm would produce 4 units and achieve profit of $0.50. If forced into marginal cost pricing (P = MC), the firm would produce 5.5 units and make a loss of $0.87. So if we force it to price at marginal cost, the firm becomes unprofitable.

In this case, the firm may petition the government to grant it monopoly status and allow it to price at the higher monopoly price. The government is loathe to allow it to act like a monopolist (again deadweight loss), but the government cannot really force it to price at marginal cost since it force the company to run at a loss and go out of business entirely.

There are several solutions to such a condundrum:

(a) Government enforces P = MC pricing but also gives **subsidies** to make up for the negative profits. An example of this is Amtrak, the US railway passenger carrier monopoly, which has the huge fixed costs of railway track maintenance across the United States. It follows marginal cost pricing rules, but receives subsidies to make up for its losses.

(b) Another solution is **two-tier pricing**. That means, it allows the monopoly to charge a high monopolistic price for some users, while charging the marginal cost price for others. In effect, the higher-paying customers subsidize the lower-paying ones. This kind of legally-permitted discrimination is common.

While price discrimination may seem unfair, keep in mind it is little different from direct subsidies, except handing over subsidies has tax-payers funding railway-riders, which is unfair to tax-payers who don't ride at a all. Whereas a two-tier system has some railway-riders subsidize other railway-riders, which at least keeps it within the railway-riding population.

(c) "**Fair rate**" is another scheme often followed, in that is it allows the monopoly to charge more than MC, but only a bit more (not the full-blown MR = MC solution). Deciding what is the "fair", however, is a whole other story.

Contestability

Finally, let us point out that there are many objections to anti-trust law, arguing that the pernicious effect of monopoly are not all that bad, and that the alternatives are worse. The most famous argument is of "**contestable markets**". That is, even if there is a natural monopoly, the monopolistic firm may be forced to behave competitively (price at marginal cost) because of the prospect of someone else swooping in.

The most famous case is air routes. In some cases these are natural monopolies because often not more than one airline can connect two relatively small destinations (not enough travelers to fill competing regular planes, at least not enough to make the flight profitable). But if the airline with a monopoly over the route starts "behaving badly" and pricing way above marginal cost, it is easy for another airline to set up the same route and steal all its customers.

So the mere *prospect* of being contested by a new entrant is often enough to get the monopolist to set prices & quantities closer to the competitive market solution. (e.g. it *will* price at 5.5 and content itself with a less-than-maximum profit). No need for anti-trust break-up, nor government, nor pricing boards and such schemes.

Which brings up the final point: it is not enough to *be* a monopoly to come under the hammer of anti-trust regulation. You can be a monopoly, but so long as you *act like* a competitive industry (that is, price at marginal cost), anti-trust law doesn't apply. Anti-trust actions must *prove* you are abusing your monopoly position and engaging in "**restraint of trade**". Microsoft, which is practically a monopoly, has escaped the anti-trust eye largely because it has "behaved" as if it were competitive and (as it likes to point out in its defense). It claims it is working in a "contestable" market (e.g. Linux, it is claimed, can "swoop in" and take over if MS overcharges its customers).

EXTERNALITIES

Economists define "**externalities**" as benefits or costs that are imposed on other people which are not captured in the market.

Externalities can be good. For instance, the beekeeper and the neighboring orchard-owner benefit from each other: bees pollinate the orchard and increase its apple production, while providing the nectar that helps the bees create honey. But this is not a market activity. Bees do it on their own and the beekeeper and orchard-owner don't charge each other for the "services" they provide each other.

Some externalities can be quite bad. The most famous is **pollution**. Con Edison produces electricity at its Long Island City. While Con Edison's own costs are merely the expenses of running the plant, the total costs are wider, since a chunk of it is borne by Long Island City residents who have to spend resources to clean up their homes from the polluting plant.

Externalities mean the total cost to the firm ends up being different from the total cost to society.

Example: Let us take our coffee shop example again. Suppose that the coffee shop disposes of its used coffee by dumping it in sidewalk. Neighborhood stores who want to avoid their customers slipping on the dumped coffee have to take the expense of cleaning the coffee waste off the sidewalk. These are "external" costs which are paid for not by the coffee shop but by the neighbors. The more coffee is dumped, the more it costs to clean up. So there are increasing external marginal costs, say, starting from $0.30 to clean up the first dumped cup, and increasing by increments of $0.15 thereafter. So the table looks something like this:

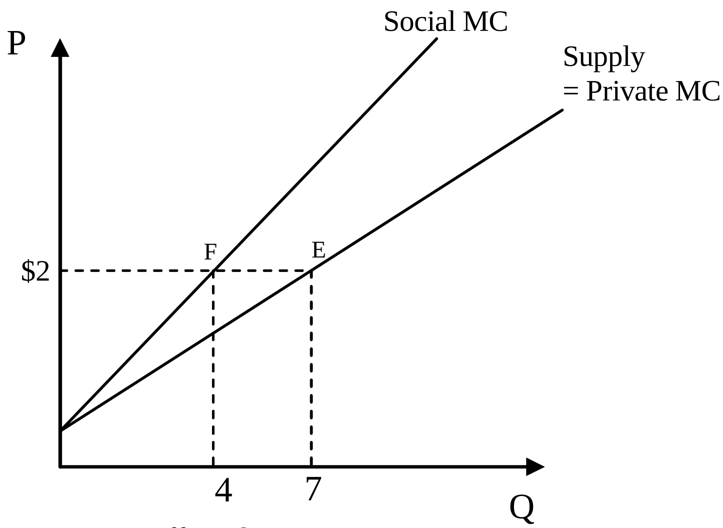
Qty	Price	Total Revenue (= Price × Qty)	Marginal Cost (Private)	Marginal Cost (External)	Total Social MC (private + external)
0	$2	$0	$0	$0	$0
1	$2	$2	$0.50	$0.30	$0.80
2	$2	$4	$0.75	$0.45	$1.20
3	$2	$6	$1.00	$0.60	$1.60
4	$2	$8	$1.25	$0.75	$2.00
5	$2	$10	$1.50	$0.90	$2.40
6	$2	$12	$1.75	$1.05	$2.80
7	$2	$14	$2.00	$1.20	$3.20
8	$2	$16	$2.25	$1.35	$3.60
9	$2	$18	$2.50	$1.50	$4.00
10	$2	$20	$2.75	$1.65	$4.40

If the coffee shop bases its production decision on *private* marginal cost (that is the marginal cost it pays itself), then the profit-maximizing solution will be set where MR = MC (private) = $2.00, that is, produce an output level of 7 cups.

But if the coffee shop was somehow forced to be responsible for cleaning up its own dumped coffee, it would base it on total social marginal cost, that is where MR = MC (social) = $2.00. In this case, the coffee shop's profit-maximizing solution would be to produce only 4 cups.

So the profit-maximizing solution is quite different depending on whether or not the coffee shop is responsible for its' own clean-up or whether it leaves it to its neighbors.

Diagrammatically, we can depict the situation as follows:



Coffee Shop PMC & SMC

Notice the Social Marginal Cost curve is *steeper* than the Private Marginal Cost curve. That is because Social Marginal Cost includes both the private and the external marginal costs. But since a firm only pays attention to the marginal costs it has to pay, then the supply curve of the firm is the private marginal cost curve (PMC). At $2, it will produce 7 units. But *if* the firm was made responsible for the external clean-up costs, then its supply curve would be the SMC curve, and it would produce 4 units.

Which *should* be the solution?

Moral considerations might tempt us to argue that the coffee shop should clean up its own mess. But economists are not moralists. Our concern here is what is the *efficient* solution, that is, whether resources are being efficiently allocated.

Cleaning up the coffee on the sidewalk is an expenditure of resources & time. But whether those resources are expended by the coffee shop or by the neighbors doesn't really matter from an efficiency point of view, at least at first glance. The resources will be spent, whether by the coffee shop or the neighbors, so who exactly ends up doing it doesn't seem to make a difference.

But notice that it *does* make a difference to the production decision. If the coffee shop is responsible for cleaning up, it will produce 4 units. If it shunts the clean-up job to its neighbors, then the coffee shop will produce 7 units.

So what? Well, the extra three units it produces are *not worth it*. Their marginal benefit to society is *lower* than the marginal cost to society of producing them. Or put another way, they are "worth" only $2 each (the marginal revenue), but each of the three extra cups costs *more* than $2 in resources to produce (social marginal cost is $2.40, $2.80 and $3.20 respectively). It is *inefficient* to expend more resources on something than it is worth.

Consequently, from an *efficiency* point of view, never mind fairness, whether the coffee shop or the neighbors clean up begins to matter. The *efficient* solution is to produce 4 cups, it is *inefficient* to produce 7. So correcting the externality isn't merely a matter of morality or fairness, it is a matter of economic efficiency. If the externality remains unaddressed, we have overproduction and waste of resources.

This is known as a "**market failure**". That is, the market has produced an inefficient allocation of resources.

Solutions to Externalities

Market failures are unusual. If nothing else, markets can usually be relied upon to produce efficient outcomes. You cannot rely upon it for justice, you cannot rely upon it for fairness, but at least you could normally rely upon it for efficiency. But not this time. Externalities are one of those situations when the market is inefficient. Consequently, there may be a role for the government or outside agency to "correct" the failure and make it efficient.

But how exactly *should* the government go about "fixing" a market failure like this? There are several solutions.

(A) Regulatory Limits

The simplest and most straightforward is for the government to prohibit the firm from producing more than the efficient level and levy fines if it breaches that level. In the coffee shop case, the shop would be prohibited from producing more than 4 cups. In this manner, the efficient level is achieved instantly and brutally.

This may seem the simplest solution. The problem is that such limits are hard to enforce. The coffee shop has a strong incentive to produce 7 units and will do what it can to try to dodge the regulatory limit of 4 cups. Whenever the government isn't looking, it will try to breach the limit. Which means the government has to deploy an army of inspectors to keep an eye on the shop around the clock. And inspectors are not only expensive to hire, they are also human – they may be fooled by coffee shop trickery, fall asleep on the job, or be bribed into turning a blind eye to violations.

In the end, the government may end up spending more resources to fix the externality, than the externality itself costs.

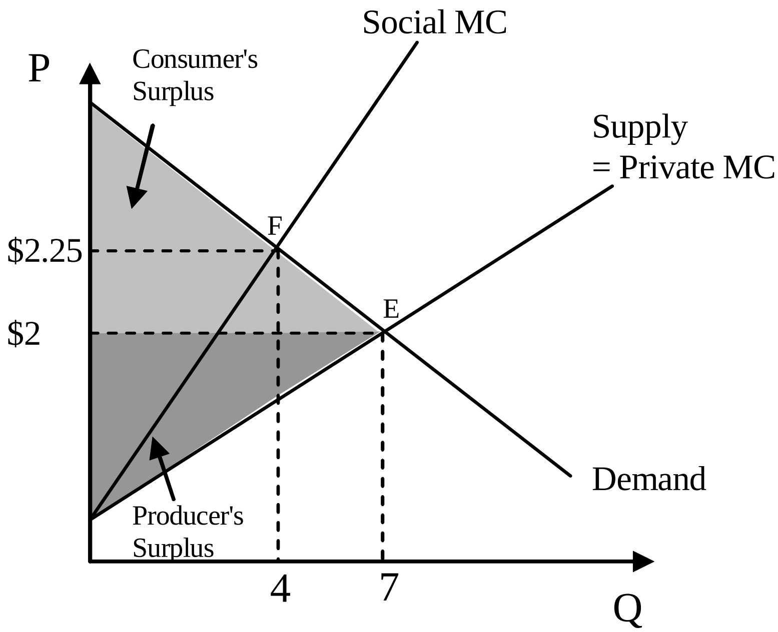
(B) Pigouvian Taxes

More ingenious and simpler is to introduce **excise tax** on the output of the firm. The point is not to raise funds or punish the firm, but to induce the firm, by its own profit-maximizing motives, to produce at the lower socially-optimal level.

This way of dealing with externalities was first suggested by the British economist Arthur Cecil Pigou (1912) and thus are known as "Pigouvian taxes".

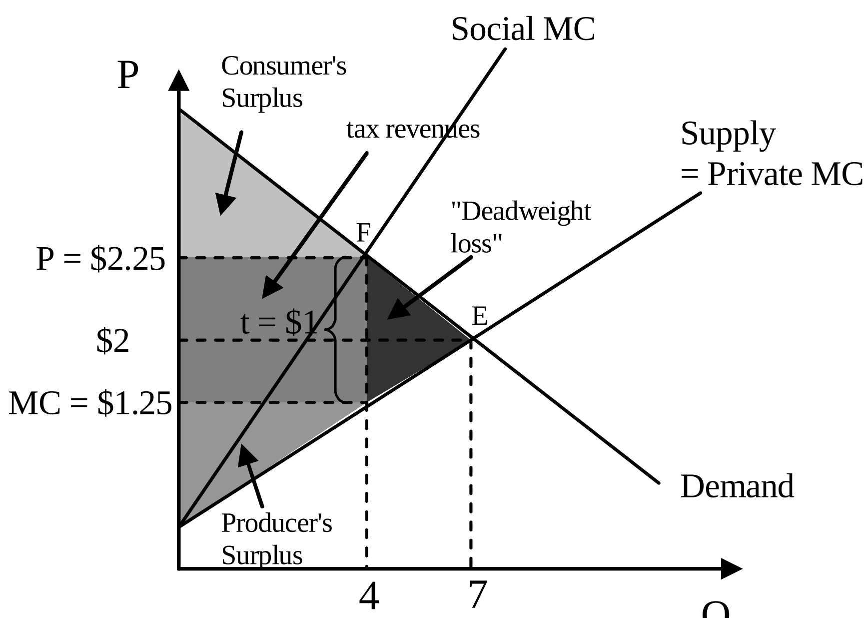
Remember from our discussion of welfare, the impact of an excise tax on a market is to *raise* the price of the good to consumers, *reduce* the price received by producers and, more importantly, reduce the level of production.

Consider the following uncorrected pre-tax situation depicted in the diagram below. Here we have the market depicted, superimposing demand on the private and social marginal cost curves. If left to its own devices, the market will hone in on point E, the intersection of the demand and supply (private MC) curves, the market price will be $2 and 7 units will be produced. This is the inefficient outcome, since the social marginal cost curve – that is, the real resource costs of production - is steeper. The efficient equilibrium point is point F, the intersection of the demand curve and the social marginal cost curve. Output should be 4 and market price $2.25. But since the firm's supply curve is the private marginal cost curve, point F cannot be achieved by the free interaction of the market.



Market with Externality

So let us impose a Pigouvian tax – a sales tax of $1 per cup. The impact of the sales tax on this market will look like the following:



Market with Pigouvian tax

The $1 tax will create a "wedge" between the demand and supply curves. The sale price to the consumer of coffee is now $2.25, but what is received by coffee shop is only $1.25. A profit-maximizing firm, as you know, will equate marginal revenue to (private) marginal cost. Since marginal revenue to the firm is now $1.25, then looking at the table above, it will equate (private) marginal cost at 4 units. Thus the Pigouvian tax has twisted marginal revenue to make 4 units the profit-maximizing level of output for the firm. In other words, with the tax in place, the firm has no private incentive to produce any more than 4 units. We have achieved the socially optimal output level.

What about the deadweight loss? Well, in truth, the little dark triangle above isn't really a "deadweight loss" since E was an inefficient equilibrium to begin with. The efficient equilibrium is F. And that's where we're at *now*. Far from being a deadweight loss, that little dark triangle was a waste of resources from overproduction – resources paid for by neighbors, that is now eliminated.

Tax revenues are $1 × 4 = $4. This goes to the government. How it spends that tax doesn't really matter. The government is not imposing this tax to raise revenues for any purpose. It is imposing the tax *functionally*, that is, in order to force the equilibrium level of output from the inefficient level of 7 to the efficient level of 4 units. That it happens to get some pennies in the process is a windfall, but not the purpose of the tax.

The burden of the tax is shared between the consumers of coffee and the producers of coffee. Which may seem "how it should be" because it was the production and sale of coffee that was imposing the external costs of society. Both consumers and producers of coffee were "at fault" for the coffee spillage, consequently the burden of the tax falls upon both of them.

[In this particular case, you'll notice that the burden falls mostly on producers – producers pay 0.75, consumers pay 0.25 of the $1 tax, but that's just happenstance from the numerical example I am using. It is not necessarily the case that producers pay more than consumers.]

The advantage of the Pigouvian scheme over regulation should be obvious. With a regulatory limit, the firm continues to have every incentive to dodge it, and its compliance needs to be enforced by armies of regulators, inspectors and watchdog groups, on top of which we must add the cost of punitive fines. The Pigouvian scheme leaves the firm to its own devices. All the government needs to do is impose the proper excise tax and then step out of the way. The firm will go to the socially-optimal level of production by itself, on the basis of its own profit-maximizing criteria.

(C) <u>Zoning</u>

The principal disadvantage of the Pigouvian scheme is that it automatically imposes the cost entirely on one side – that is on the producers and consumers of coffee. That is not always the most just or efficient solution.

Consider the following scenario. Suppose that in order to serve New York City customers, Con Edison decides sets up an electrical plant in an obscure corner of the city where nobody lives. It causes pollution, but no one is affected.

But then one fine day someone decides to build their home near the plant. Now we have a problem. The homeowner proclaims Con Edison's exhaust is costly for him to clean up and demands the plant's output of electricity be curbed. The city intervenes with a Pigouvian tax on the sale of electricity.

Is this the most just and efficient solution really? Should Con Edison and millions of its customers be forced to pay higher tax bills and receive less electricity because one person decides to live near the plant? It's not as if the homeowner didn't *know* ahead of time that he was moving into a polluted area, or that he freely chose to do so. Should millions be forced to pay for the stubborness or lack of foresight of this single homeowner?

A far more efficient and less costly solution for this case is to simply move the homeowner out of the area. Indeed, Con Edison might be willing to foot that bill, if that costs them less than the Pigouvian taxes.

Many cities recognize these kinds of problems may arise and try to pre-empt them by a variety of "zoning" laws, that is designating certain areas for residential or industrial areas, prohibiting people or businesses from moving across zones. The heavy hand of the law may seem awkward and blunt at times, but if zoning is done properly, it can forestall these kind of conflicts.

(D) <u>The Coase Theorem</u>

The problem of the single homeowner moving near a Con Edison plant illustrates a more general problem with the way externalities are handled. It may seem harsh to put it in this manner, but blame for pollution is not always one-sided. Just as it takes two to tango, it takes two to make a pollution problem – one to pollute, and another to be polluted. It is not always certain that the blame falls entirely on the polluter, or that one side is necessarily "evil" and the other is necessarily an "innocent victim". In this example, it seems clear that the 'pollutee' (is that a word?) seems to be at least partly blame for the pollution "problem" – he knowingly moved near the plant. Before he moved, there was no pollution problem. After he moved, the problem emerged.

Pigouvian taxes do not have this flexibility. It puts the blame firmly on one side (the polluter's), but does not have a way to handle the possibility the pollutee might be partly to blame.

An alternative to the Pigouvian solution is the "Coase" solution, introduced by the English economist Ronald Coase (1960). Coase recognized that externality problems are more ambiguous than Pigou declared it to be, that it necessarily involves two parties, and putting the burden of compliance simply on one side might not always be the best way to handle it.

More importantly, Coase also proposed that it might be *unnecessary* for the government to come in and correct it. He proposed that private negotiation between the affected parties can achieve the socially optimal solution.

The intuition is a bit subtle. Consider the following thought exercise. If Con Edison plant is causing so much damage to the neighborhood, why don't the neighbors simply *bribe* the Con Edison plant into cutting back its production to a more tolerable level? Of course, the LIC residents must come up with a bribe that is *greater* than the profits that Con Edison foregoes. But it is possible that the pollution *cost* to residents of continued Con Edison production levels may be greater than the extra profits Con Edison makes by insisting on producing at a very high level. In this case, it is *possible* for residents to come up with a bribe which is *cheaper* for them than cleaning up the pollution costs.

So it is in the self-interest of LIC residents to pay Con Edison to cut back production, and it is in the self-interest of Con Edison to accept the bribe. Bribe is paid, Con Edison cuts back, bingo, pollution problem solved.

The objection to this that can be immediately made is why should the residents shell out their own hard-earned money to bribe the polluting firm to cut back output? Shouldn't it be the firm who is imposing the cost the one to be paying?

Depends on who owns the "property rights" to the air. If the firm "owns" the air, then certainly it is incumbent on the residents to bribe the firm to cut back.

But what if the residents "own" the air? In this case, the bribe goes the other way. If the residents own the air, they might demand Con Edison shut down production altogether. But Con Edison might be interested in paying the residents for a license to pollute. After all, a little bit of pollution isn't too hurtful or expensive to clean up. And Con Edison is running a profitable business. So Con Edison is willing to forego part of its profits to pay a license fee to pollute. And so long as that license fee more than compensates LIC residents for their clean up costs, the residents are bound to accept. Again an agreement can be reached which is to mutual benefit and a socially optimal level of pollution will be agreed upon.

Whether residents bribe Con Edison to cut back, or Con Edison pays residents for a license to pollute, the result is the same: pollution will be brought to an efficient level that satisfies both polluters and pollutees and achieves the socially efficient level of production.

Externalities & the Coase Theorem – Graphical Illustration

Let us consider the single coffee shop again and its profit-maximizing decision to produce 7 units, and leave the neighbors to clean up the mess. We omitted Total Cost calculations before, but now let's calculate them (also toss in the Total Profits column for reference)

Qty	Marginal Cost (Private)	Total Private Costs	Total Profits	Marginal Cost (External)	Total External Costs	Total Social MC (private + external)	Total Social Costs
0	$0	$0	$0	$0	$0	$0	$0
1	$0.50	$0.50	$1.50	$0.30	$0.30	$0.80	$0.80
2	$0.75	$1.25	$2.75	$0.45	$0.75	$1.20	$2.00
3	$1.00	$2.25	$3.75	$0.60	$1.35	$1.60	$3.60
4	$1.25	$3.50	$4.50	$0.75	$2.10	$2.00	$5.60
5	$1.50	$5.00	$5.00	$0.90	$3.00	$2.40	$8.00
6	$1.75	$6.75	$5.25	$1.05	$4.05	$2.80	$10.80
7	$2.00	$8.75	$5.25	$1.20	$5.25	$3.20	$14.00
8	$2.25	$11.00	$5.00	$1.35	$6.60	$3.60	$17.60
9	$2.50	$13.50	$4.50	$1.50	$8.10	$4.00	$21.60
10	$2.75	$16.25	$3.75	$1.65	$9.75	$4.40	$26.00

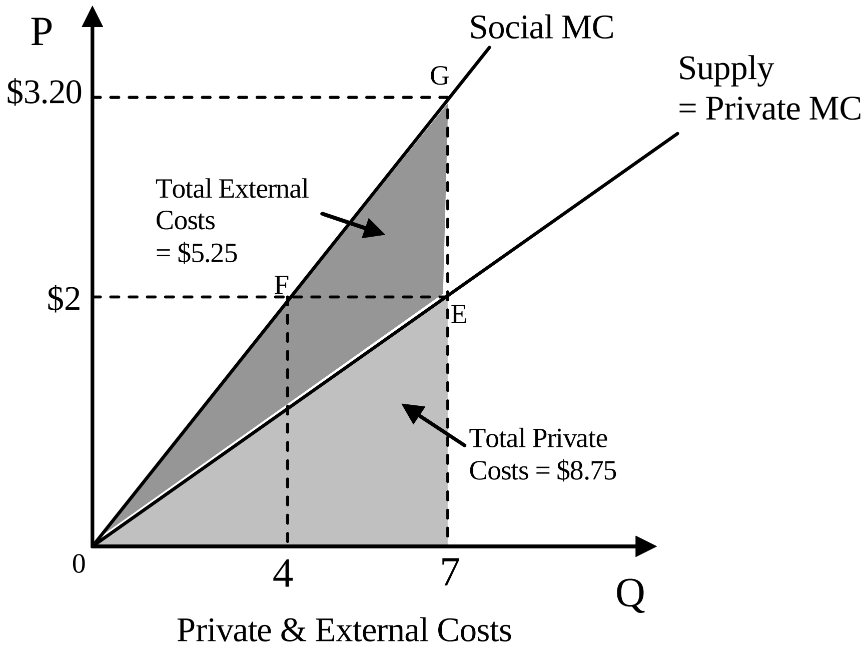
Graphically, we can analyze the relative burden of the externality by comparing areas under the curves. Remember that cumulating marginal costs adds up to total cost. So, geometrically, Total Private Costs can be measured as the area of the triangle under the private marginal cost curve. Total Private Costs are consequently the lightly shaded triangle (OE7) below the PMC curve.

But at 7, the social marginal cost is $3.20. So Total Social Costs is the area of the large triangle (OG7) under the Social Marginal Cost curve. Now since:

Total Social Costs = Total Private Costs + Total External Costs

then dark-shaded area (OGE) that covers the difference between the area under the SMC curve and the area under the PMC curve is precisely Total External Costs.

Using the numbers from the table, Total Private Costs (area of OE7) = $8.75, Total External Costs (area of OGE) = $5.25 which, added together, yield Total Social Costs (area of OG7) = $14.00.

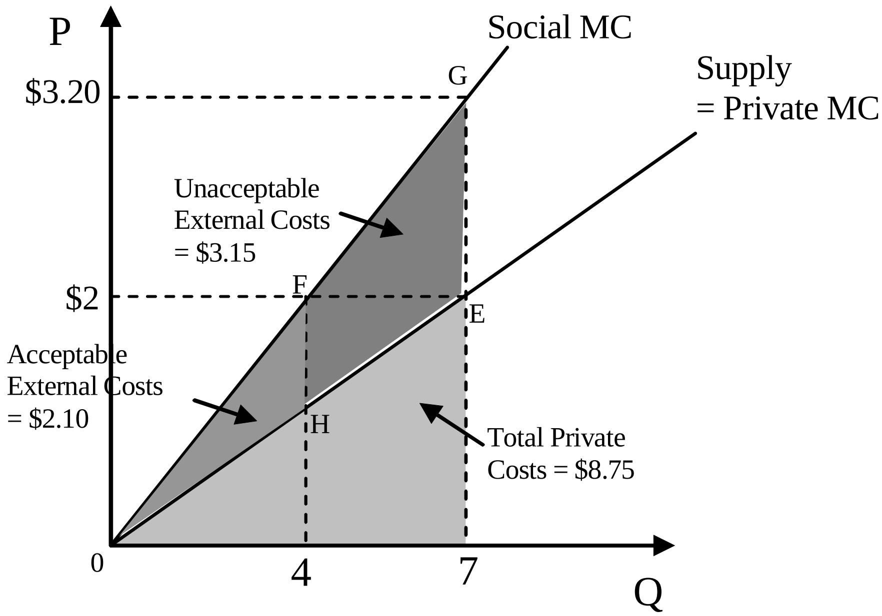


Now let us consider room for negotiation. The socially optimal level of production is 4. Now it is important to notice that even at 4 units, there remain *some* external costs. As you can read from the table, at 4, external costs are $2.10. Thus we can consider $2.10 the level of "acceptable" external costs.

[Note: optimal pollution is *not* zero. Pollution is intrinsic to production. You only obtain zero pollution if you don't produce anything at all. But that's not economically efficient. And our interest here is economic efficiency, not social justice. Economic efficiency dictates that output should be 4, and thus that *some* external costs ($2.10 to be exact) be incurred.]

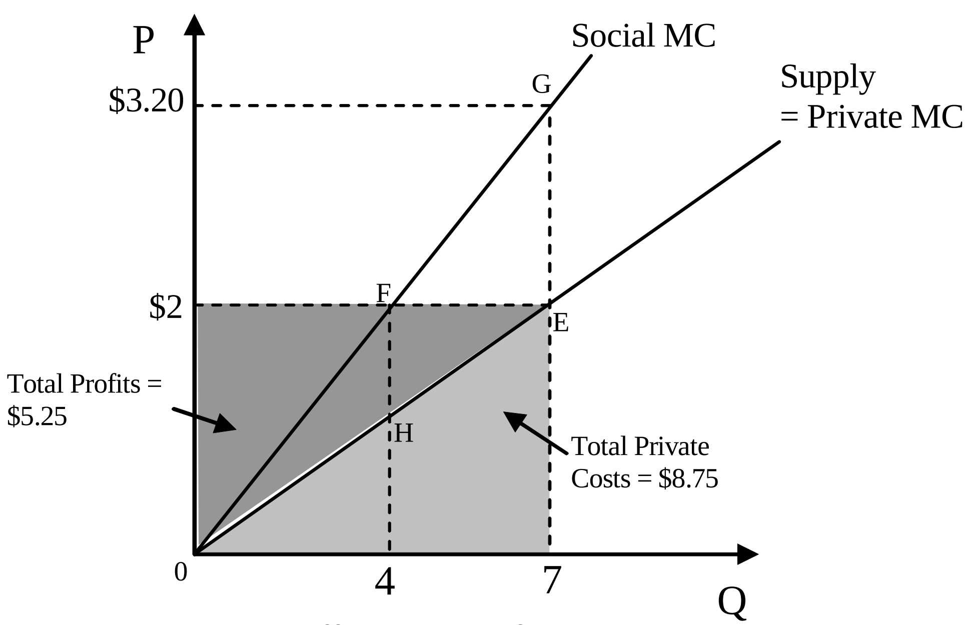
But if the firm *insists* on producing 7, it is imposing $5.25 in external costs. So it is creating imposing an additional $3.15 of excessive or *unacceptable* external costs on top of the $2.10 in acceptable external costs. What we want to get rid of is this $3.15 burden.

The size of the acceptable and unacceptable external costs can be depicted on the diagram by partitioning the Total External Costs triangle into two portions, as we see below. The lightly shaded triangle OFH is the acceptable external costs ($2.10), while the strangely shaped polygon HFGE is the unacceptable external costs ($3.15).



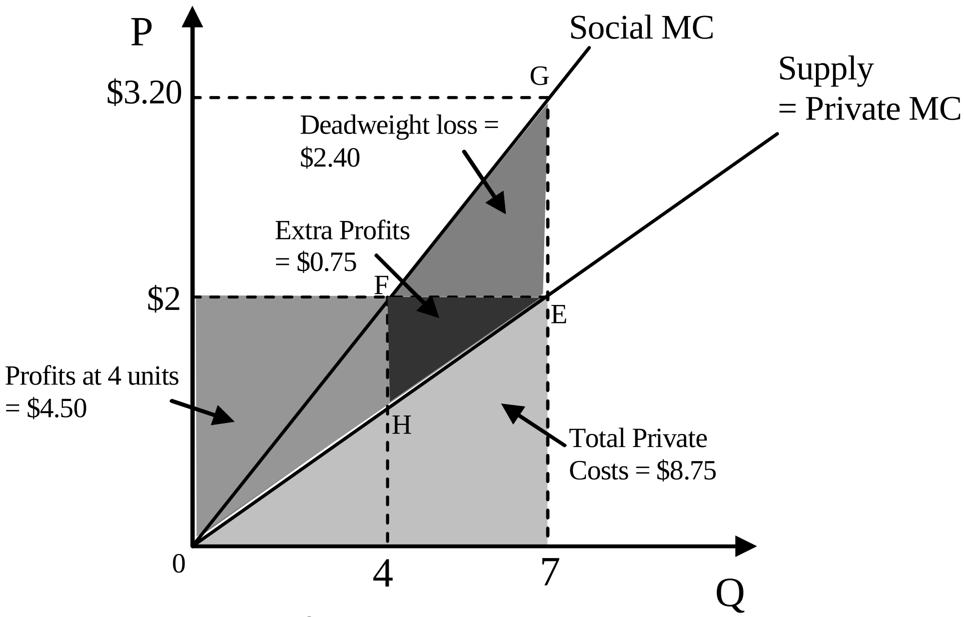
Acceptable & Unacceptable External Costs

But we are still not done. Remember the private firm's total profits can be depicted on a diagram by a triangle *above* the Private Marginal Cost curve. This is shown in the diagram below by the shaded triangle O2E. The area of this large triangle is $5.25 (check earlier table; total profits at 7 units is $5.25).



Coffee Shop Profits

Now comes the important bit. *If* the firm was forced to produce at 4 units, total profit would be $4.50 (again check earlier table). Diagrammatically, we can represent that by the area of the strange polygon O2FH in the diagram below.   The darkly shaded triangle HFE represents the *extra* profits the firm can make by expanding production from 4 units to 7 units. This extra profits is $0.75, the difference between profits at 7 (= $5.25) and profits at 4 (= $4.50). That is the extra amount of profit the firm is getting by overproducing by three units.



Extra Profits & Deadweight Loss

But that overproduction costs neighbors, as we saw, $3.15 (the area of polygon HFGE). So the difference between the unacceptable external costs to neighbors ($3.15) and the profit gains to the polluting firm ($0.75) is the *deadweight loss* of pollution.   That is, producing at 7 units imposes an additional external cost of $2.40 which is sheer waste. It is the extra cost imposed on neighbors that the firm is not taking in as profits for itself. It is complete waste of resources to society. The deadweight loss is depicted in the diagram by the triangle FGE.

Here is where we have room for Coase negotiation. The excessive production above 4 is costing the neighbors an additional $3.15, but the firm only gets $0.75 in benefit. Or seen from the opposite perspective, if the firm cut back production to 4, the firm would lose only $0.75 in profit, and the neighbors would gain $3.15 in savings for clean-up costs they don't have to incur.

Consequently the firm is willing to accept any bribe greater than $0.75 to cut back production to 4 units, whereas the neighbors are willing to pay any bribe up to $3.15, to get it to cut back. The maximum bribe the neighbors are willing to offer is greater than the minimum bribe the firm will accept to cut back. There is room for bargaining. If the neighbors pay the firm a $1 bribe, the firm will happily accept to cut back production to 4, since it makes more money that way than if it insisted on producing. Conversely, the neighbors are happy to pay $1, since that bribe costs them far less then the $3.15 they'd have to spend in time, mops & pails to clean up the mess.

The pollution problem is thus "solved" by simple private negotiation. For a modest bribe of $1, output will be reduced to 4 units and we will have the "efficient" outcome.

Consider now the opposite scenario. Suppose the city gives the neighbors property rights over the sidewalk. That is, the coffee shop cannot pollute without the neighbor's permission.

Now, if the neighbors refuse to grant the coffee shop a license to pollute the sidewalk, then the coffee shop is forced to produce at a level of 0, since 0 is the only level production where there is absolutely no pollution whatsoever.

Compare this to the efficient solution of 4 units. At 4 units, the neighbors would suffer $2.10 in external costs. If the firm pays the neighbors $3 for a license to pollute 4 units, they will readily accept, for it more than compensates them for their clean-up costs. They make a $0.90 gain. If they refuse and insist on zero pollution then they are foregoing that gain, and return to making $0. Would a firm be willing to pony up a $3 license fee? Sure it would. By producing 4 units, it makes $4.50 in profit. After paying the neighbors the $3.00 license, it has $1.50 left over in profits, which is more than it would have otherwise ($0).

So long as the extra profits the polluter makes exceeds the extra clean-up costs, there is room for negotiation of a license fee where both the coffee shop & neighbors will be better off.

[**Coase payments in detail**: Will the negotiations *necessarily* settle on the socially-efficient output level of 4 units? How are we sure they won't settle on 5? Or 3?

The answer is to think of it marginally. Suppose the coffee shop "owns" the sidewalk, so the neighbors have no choice but to bribe. Suppose the neighbors decide to just come up with enough of a bribe to get the firm to cut back by one unit – that is, to go from 7 to 6. Cutting back from 7 to 6 reduces external costs from $5.25 to $4.05, so it is saves the neighbors $1.20 in clean up costs (marginal external cost in reverse), whereas profits will decline by $0 (since profits at 6 are also $5.25). So the neighbors only need to offer the firm a bribe of 1 cent to get the firm to cut back from 7 to 6. Certainly worthwhile for the neighbors.

But let's keep going. What will it take to cut output from 6 to 5? Neighbors will see their clean up costs fall from $4.05 to $3.00, a savings of $1.05. Profits will be reduced from $5.25 to $5.00, a loss of $0.25. So, again it is worthwhile for the neighbors to pay a bribe of, say, 26 cents to get the firm to cut back, and the firm will happily accept.

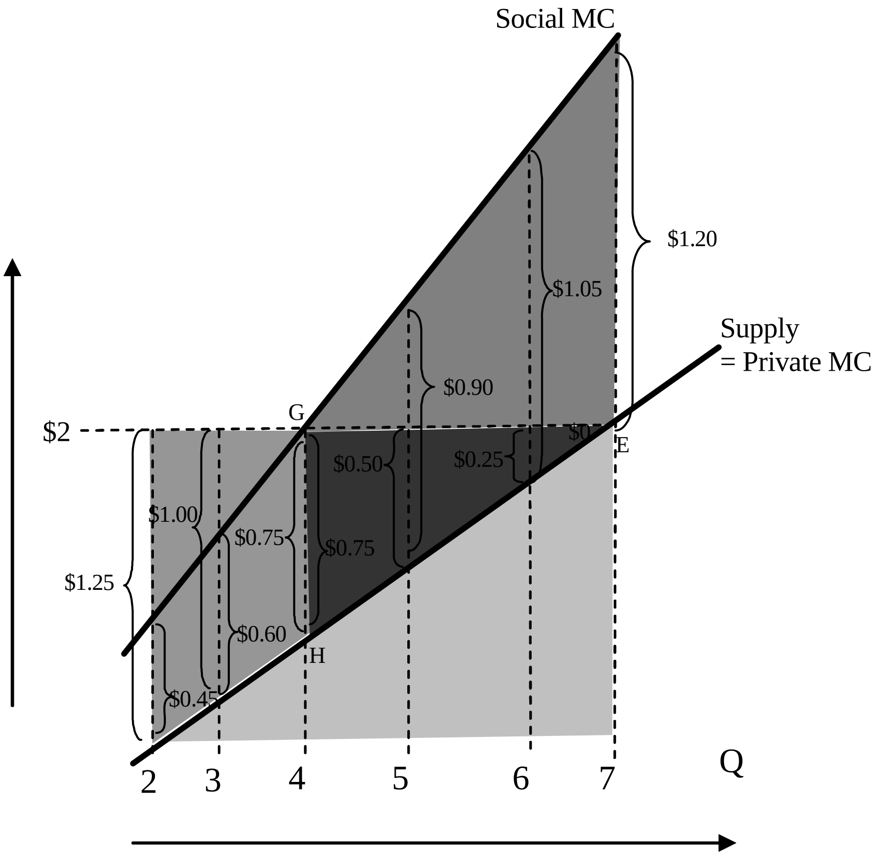
Keep going: cutting back production from 5 to 4, neighbors see their clean up costs decline from $3.00 to $2.10, a savings of $0.90. Profits decline from $5.00 to $4.50, a loss of $0.50. Again worthwhile for the neighbors to offer a 51 cent bribe, and worthwhile for the firm to acccept.

Can we go further? Can we come up with a bribe to cut production from 4 to 3? No. That is because going from 4 to 3, the external costs of the neighbors decline from $2.10 to 1.35, that is a saving of $0.75, but the profits of the firm will decline from $4.50 to $3.75, a loss of $0.75. The only way to persuade the firm to cut back is to offer it a bribe of 76 cents. But that is 1 cent more than the savings they'd be making from the external costs.

And if 1 extra cent doesn't quite impress you, certainly you'll notice that going from 3 to 2 becomes even less possible, since the profits foregone will be $1.00 whereas the external costs saved would $0.60. So the minimal $1.01 bribe the firm would accept is definitely not worthwhile for the neighbors to offer.

So, approaching it incrementally unit by unit, the bribe would take us down to 4 units and then stop.

Graphically this is obvious. The diagram below shows a magnified view (a little messy) of what's going on around the center of the earlier picture, unit by unit. Compare the vertical distances: the height between the SMC and PMC is the external cost of that unit, while the height between the $2 horizontal line and the PMC is the profit from that unit. It is evident that everywhere to the right of 4 units, the external cost is greater than the extra profit, whereas everywhere to the left of it, the external cost is smaller than the extra profit. Only at 4 units is the extra profit *exactly* equal to the extra profit.



If the situation was reversed, and the neighbors "owned" the sidewalk. Now we start from the left, at 0 production. Producing one unit will give the coffee shop a profit of $1.50, and impose an external cost to the neighbors of $0.30. If the shop offered to pay the neighbors, say, $0.31 for permission to produce that first unit, the neighbors would accept since it more than covers the cost of cleaning up. And the shop is willing to pay since the fee is less than the profit.

Same thing for the second unit: the coffee shop will net a profit of $1.25 on the second unit, and impose an external cost of only $0.45. Again, an offer to pay the neighbors $0.46 for permission to produce a second unit will be beneficial for both parties.

Again for the third unit: coffee shop's extra profit on that will be $1.00, the while it will only cost the neighbors $0.60, so a 61 cent license fee will be accepted.

And it finally comes to a halt at 4 units. Here the extra profit is $0.75, and the external total cost imposed $0.75. Trying to push it beyond that to 5 will impose a greater external cost ($0.90), for which the neighbors will demand at least 91 cent fee to give permission. But the firm only nets an extra profit of $0.50 at 5, so it is not worthwhile for them to pay the 91 cent fee.

So whether we're coming incrementally from the right (neighbors bribe shop) or coming from the left (shop pays license to neighbors), we will gradually approach and come to a halt at the same point: 4 units, the efficient allocation. **]**

In sum:

Whether the neighbor bribes the coffee shop, or the coffee shop bribes the neighbors, they can privately negotiate a side-payment that takes them to the efficient output level. There is no need for the government to intervene with regulations or Pigouvian taxes. Let the affected parties talk and negotiate privately with each other, and the pollution "problem" will be solved.

The difficulties arise only if no one "owns" the sidewalk and/or transactions costs make negotiations too expensive to conduct. In this case, the pollution problem cannot be privately resolved. The shop keeps producing at 7 and the neighbors are saddled with the deadweight loss. We have an inefficient outcome, and the only thing that is exchanged between them are bad feelings and angry words.

Problem #1 - Property Rights

The two schemes - residents bribing polluter vs. polluter bribing residents - reach efficient outcomes, even though they differ in who pays whom. So, from an economic efficiency point of view, it *does not matter who owns property rights* on the air or the sidewalk or whatever. All that matters is that *someone* owns the commons. Once that is the case, the socially optimal outcome can be realized by private negotiation.

Yes, one party receives money, and the other party pays money. But remember the paying party would also lose money from clean-up costs or foregone license fees/bribes or profits. *Both* parties are better off in *either* case. Doesn't matter who owns it.

But somebody *must* own them. If the sidewalk "belongs" to the coffee shop, neighbors recognize their only solution is to bribe and bribe they will, for it is in their self-interest to do so. If the sidewalk "belongs" to the neighbors, then the coffee shop recognizes it must pay a license fee, and pay they will since it is in their self-interest to do so. But if the sidewalk belongs to nobody, then neither feels they should have to pay the other, both imagine they have the right to the sidewalk (whether to dump waste on it or to keep it clean), and they will do nothing but quarrel endlessly with each other. The problem goes unresolved.

And if it unresolved, the outcome is inefficient – the coffee shop continues stubbornly producing at a high level and the neighbors grumbingly and needlessly waste time & resources on mopping the mess up.

Problem #2 - Transactions Costs

Besides ensuring somebody owns property rights on the commons, there is one further condition necessary to make the Coase Theorem work: namely, low transactions costs.

What is that? "**Transaction costs**" are defined as the arranging and processing costs necessary to negotiate a deal.

There are hardly any transactions costs between the coffee shop and its neighbor. One can just walk next door and have a talk.

But consider the Con Edison-LIC resident quarrel. Here it is a little more complicated since one of the sides (the residents) is composed of many people dispersed over a large area. Con Edison's transactions costs are the fees it has to pay a stable of lawyers, accountants and negotiators to draft up the proposals, hire agents to go around knocking on doors, finding the residents and communicating them their proposals. The transactions costs to the LIC residents are the time, effort and costs it takes to elect and organize a resident's committee, draft its own proposals, communicate with the company and report back to its constituents.

Transactions costs may not seem like a big deal - after all, some costs exist in every negotiation or trade. But in many actual cases of externalities, they can often be lopsided or prohibitive. In particular, the larger the group of people involved, the costlier it becomes to just overcome the logistical difficulties of striking a deal.

It is relatively difficult to mobilize and organize a collection of disparate people into a negotiating body. Besides just the communications problem, there will be **incentives problems**. For a single firm to reach an agreement with a few hundred different residents each with their own individual interests can be quite difficult and costly.

For instance, suppose that 100 residents "own the air" and the Con Edison decides to offer them a $1,000 license to fee to allow them to pollute. Suppose it drafts a proposal breaking down that fee so every resident receives $10. But a few residents may turn down the $10 offer in the hope for a larger share of the total. They may say they want $15 or $20, even though $10 should be enough. If enough do that, it quickly becomes unprofitable for the company and the fee negotiations break down.

Similarly, if the firm owns the air and the residents must arrange for a collective bribe, that is also difficult to organize. Some residents may refuse to contribute their share of the bribe, figuring the deal can be done without him, so let the others pay and he'll get the benefits anyway ("**free rider problem**"). If enough do so, the residents may be unable to come up with a sufficient bribe to persuade Con Edison to cut back.

All these things come under the rubric of "transactions costs", that is costly impediments to a deal. If the transactions costs are cumbersome enough, then the Coase solution to an externality problem may not be practically achievable.

Summary

In Coase's view, externalities aren't the problem. The root of the problem is unclear property rights and high transactions costs. Otherwise, there are mutual gains to be made from a deal and "internalizing" the externality with some sort of side-payment. And such a deal *will* be made - because for residents, refusing to let the firm pollute has the opportunity cost of foregoing a license fee from the firm, whereas for the firm, refusing to cut back production has the opportunity cost of foregoing the bribe. A deal *will* be reached. And that deal *will* be socially efficient.

If property rights are unclear or transactions costs are too high, then a deal might not be achieved, even if it is in the self-interest of both parties to reach it. The result is that externalities remain unaddressed and firms will continue producing at inefficient levels and there will be deadweight loss for society. In this case, as Coase himself admits, there may be a need for government intervention.

But the kind of government intervention isn't necessarily heavy-handed regulation or one-sided Pigouvian taxes. Rather, the Coast theorem suggests it may be enough: (1) for the government to assign or clarify who owns property rights over the commons/air/sidewalk/noise, etc. (2) if need be, the government can help the parties overcome transaction costs (e.g. use public funds to help disseminate information and facilitate communication, even supply experienced organizers, negotiators to work on behalf of the affected communities).

Once property rights are clear and the parties are assembled at the negotiating table, the government should step back and let the parties hammer out the deal on their own. As the theorem suggest, they will necessarily hone in the efficient solution.

The Coase Theorem, originally articulated by the Anglo-American economist Ronald Coase in 1960, is one of the most celebrated insights in economics and has revolutionized not only the economics of public policy, but has even revolutionized legal theory and practice (esp. tort law).

PROFIT MAXIMIZATION IN THE REAL WORLD

We have assumed that a firm is dead-set on maximizing its profits. We asserted such was necessarily the case in a capitalist economy, as the owners of capital (that is, those who's income will be the profits) are in control of the firm and insist that the managers maximize profits rather than any other objective.

Economics has given us a rather technical rule to find this solution: that is, expand production so long as additional revenues exceed additional costs, cut back whenever additional costs exceed additional revenues. In short, adjust production until marginal revenue equals marginal cost.

Is this what firms *really* do in the real world? There are three criticisms:

(a) **Operationality:** the MR = MC rule doesn't provide many details of how exactly it is to be implemented. In the real world, the necessary information to calculate marginal revenue and costs is often hard to obtain. Moreover, managers and accountants have varying measures and definitions of costs which is not exactly identical with the economist's. Which should they use? Do we include depreciation and taxation? What is the time frame? How do we account for risk and uncertainty?

Economists have gone to great lengths to fill in the gaps, and they have developed associated "rules" for profit-maximization in different situations. But the general "marginalist" principle remains.

(b) **"Mangerialism"**: Granted that shareholders control the board of directors which orders the managers about. But let's be realistic: managers don't always obey, and may go and do something else other than maximize profits, and then bamboozle the capitalists with smoke, mirrors and tales of derring-do. Managerial egos may drive them to expand the *size* of the firm for the sheer glory of it, rather than to actually make profits for the owners. They may want to acquire lots of new divisions, or maximize sales revenues, or want the quiet life and avoid all risky activities, or just aim to maximize their own perks and pay, damn the firm.

Boards have tried to give managers incentives by paying a good part of their salaries in stock or stock options (an option is a pieces of paper which become valuable only if the firm achieves a certain stock price). As stock price is theoretically tied to dividends - that is, profits - then the managers' pay depends on how much profits they make. But that just takes the dog-and-pony show to another level. Stock markets are themselves quite susceptible to managerial hoopla, bells and whistles, and stock prices may be driven up, way up, without waiting for profits to justify it.

(c) "**Satisficing**" ("**Bounded Rationality**") Most actual business case studies do confirm that a large majority of firms do try to make profits and managers do (eventually) get fired if they get off track.

But are they really profit-*maximizing*? Or are they doing "just enough" to keep the shareholders happy? We see managers giving their shareholders a 20% profit return that pleases them greatly. But maybe they could have achieved 30% if they *really* tried. To use a buzz-phrase, might they just be profit-*satistficing* rather than profit-*maximizing*?

Profit-satisficing (or what has sometimes been called "bounded" as opposed to "full" rationality) in firms has been studied by a lot of economists of the '**behavioral**' school. Through studies of the internal organization of firms, they have dug up mountains of evidence that firms *don't* really seem to obey the MR = MC rule, but just use a whole assortment of rules of thumb & routines, e.g. a standard % mark-up over average costs, maintain a target inventory as a percent of sales, or target a market share percentage, or set a benchmark level of profit performance, etc. None of that sounds like the MR = MC rule in the proper sense.

Behavioral economists have crowed that this proves the whole MR = MC exercise is nice in theory, but bunk in practice. But Marginalist economists have given several replies:

(a) The behavioralist evidence proves nothing about motivation. Yes, managers use a myriad of rules of thumb, but these rules of thumb may be *part* of a wider picture with an ultimate profit-maximization objective.

(b) Who cares what firms actually do? We are interested in what firms *should* do. If firms use rules that don't maximize profit, then they are making a mistake, perhaps because they're badly informed or are deliberately and mischievously misleading their shareholders. If they - or their boards - were aware of the error, they'd correct it.

(c) Even if it the rule isn't followed in a particular firm, it will *de facto* be followed in an industry by a natural evolutionary process. This is a trickier argument. Suppose there are a hundred firms in an industry and each are assigned a random output level - that is, they produce a particular for no good reason at all. The ones where Q happens to be where MR = MC will be making more profits than the ones whose Q happen to be at positions where MR < MC or MR > MC. Consequently, over time, the profitable MR = MC firms will grow in relative share and market size, while the firms in the less profitable positions will gradually decline or go bankrupt. Eventually, only the firms that happen to have the MR = MC position will persist. So MR = MC will become the rule we *observe* in the industry, even if no one firm is consciously aware of it or pursuing it.