PROFIT-MAXIMIZATION

- Technical Supplement -

Contents

In our discussion of Profit & the Firm, we relied upon tables to depict the profit-maximization decision faced by firms and derive the marginal rule.

In these notes, we shall turn to a more direct graphical and mathematical treatment of the same concepts.

Preliminary

Remember that:

total profits = total revenues - total costs

or:

$$\pi = TR - TC$$

where are using the Greek letter π (pi) to denote *total profits* (we shall reserve the Latin letter p to denote price).

There is a very simple rule to find the level of output that maximizes profits at a given price: the *marginal rule*. We will show how we can handle that rule both algebraically and diagrammatically.

TOTAL REVENUE

Total Revenue (TR) Curve

We can plot the total revenue curve as follows. As we know, by definition:

total revenue = price × quantity produced

or, in abbreviated form:

$TR = pQ$

where p = price per unit sold and Q = quantity of output produced. pQ means $p \times Q$.

In a **competitive** market system, firms don't have control over sales prices. Rather, 'the market' does. All a firm has control over is how much they produce and supply. So, *given* a price (p), they must make a quantity decision (Q).

The TR formula implies that at a *given* price (p), there is a *linear* association between the amount produced (Q) and total revenues (TR): the more that is produced & sold, the more the total revenue made - and it is *proportionally* more, e.g. if p = 60, then:

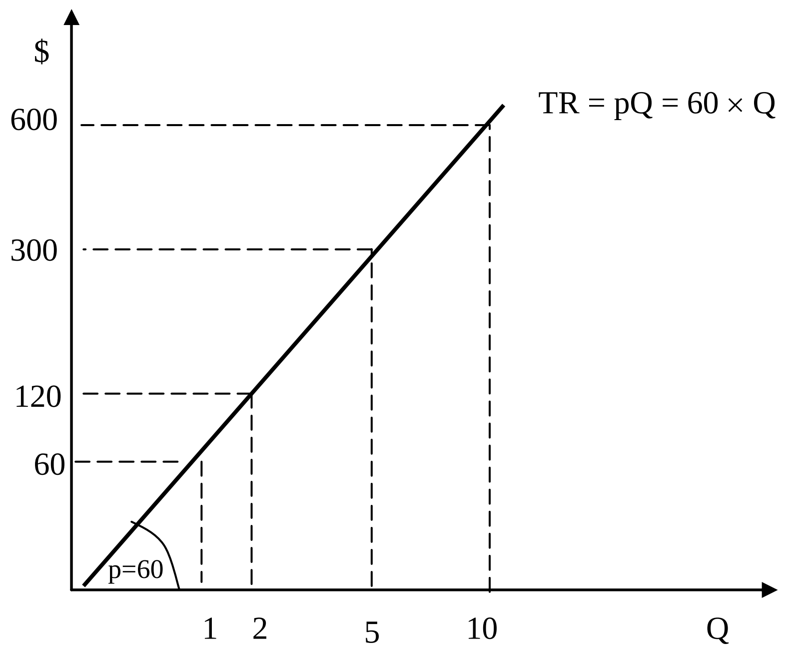
$TR = 60 \times Q$

thus, going by increments:

when Q = 0, TR = 0
when Q = 1, TR = 60
when Q = 2, TR = 120
when Q = 3, TR = 180
when Q = 4, TR = 240
when Q = 5, TR = 300
when Q = 6, TR = 360

and so on.

We can draw a total revenue curve that looks something like this:



(Unlike Supply & Demand curves, this Total Revenue curve is read like normal mathematical functions, that is from the x-axis to the y-axis - so *at* 5 units of production, we obtain $300 of total revenue; *at* 10 units of production, we obtain $600 of revenue and so on.)

Slope of the TR curve ("Marginal Revenue")

You might remember from old math classes that the *slope* (rate of ascent) of any curve is easily calculated as "rise over run".

In our case, the increase in total revenues ("rise", vertical axis) resulting from an increase in quantity produced ("run", horizontal axis).

Let the Greek letter Δ denote "change in". So, ΔTR = change in total revenues ("rise") and ΔQ = change in quantity produced ("run"). Thus slope ("rise over run") is:

Slope of TR curve = rise/run = (change in TR)/(change in Q) = $\Delta TR/\Delta Q$

The slope ($\Delta TR/\Delta Q$) is sometimes called the "marginal revenue" (MR).

> **Marginal Revenue**: the additional revenue a firm gets from increasing quantity produced by an additional unit. This is the slope of the total revenue curve (MR = $\Delta TR/\Delta Q$).

In our diagram, if we move from Q = 0 to Q = 1, then the "run" is merely 1, i.e. $\Delta Q = 1$. What is the associated rise? Well, at Q = 0, TR = 0 and at Q = 1, TR = 60, so the change in total revenues (the "rise") associated with a one unit increase in quantity produced is $\Delta TR = 60$. So the slope of the TR curve, "rise over run", is:

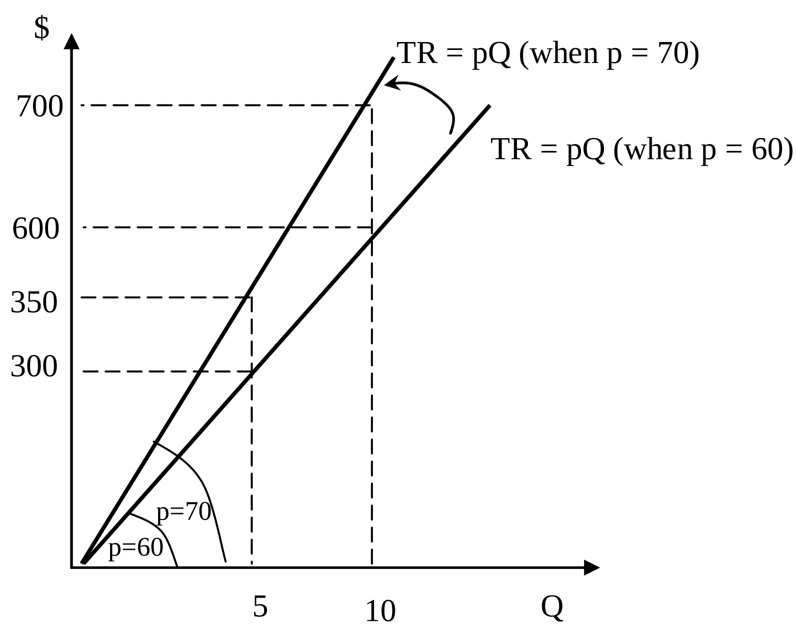
$$\text{slope of TR curve} = \Delta TR/\Delta Q = 60/1 = 60.$$

Notice that in our diagram, the TR curve is a straight line ("linear"). The slope of a linear curve is *constant*. So the slope remains 60 no matter if you calculate your "run" around Q = 0 or Q = 10 or Q = 100. The rate of ascent is proportional and steady all the way through.

Notice an interesting fact: the slope of the TR curve (60) happens to be the *price* we originally chose (60). This is *always* the case when we are dealing with situations of *competitive* markets (but not so when there are monopolies &tc.)

<u>Changing Price</u>

The fact that, in a competitive scenario, the slope of the TR curve (the MR) is the price implies that if we start from a different price to begin with, then we will have a totally differently-sloped TR curve. Specifically, if price rises, then the total revenue curve swings to the new slope. The diagram below show how the total revenue curve changes when we increase the price from \$60 to \$70.



In our example, the price rise from \$60 to \$70 will change total revenues. Now, if we produce 5 units, we obtain revenue of \$350 (= 5 × \$70) and if we produce 10 units, we obtain revenue of \$700 (= 10 × \$70) and so on.

In sum:
An increase in price (e.g. from \$60 to \$70) will make the total revenue curve *steeper*.
A decrease in price (e.g. from \$60 to \$50) will make the total revenue curve *flatter*.

TOTAL COST

The Total Cost (TC) Curve

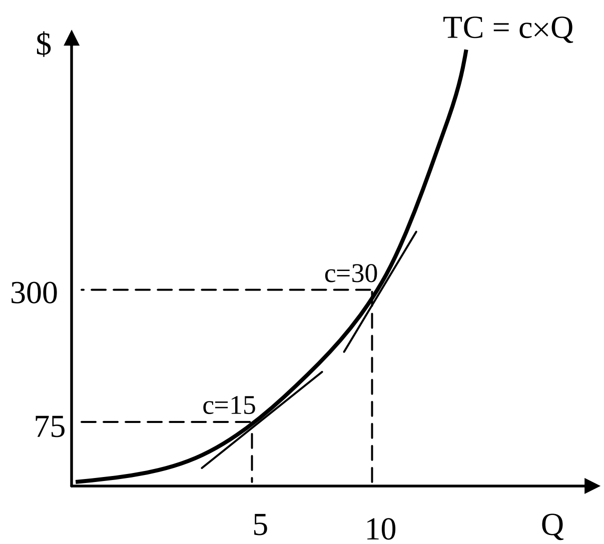
Let us now turn to the total cost curve. By definition:

$$TC = cQ$$

where c = costs per unit produced (e.g. the wage, rent, etc.) and Q = quantity produced. cQ means $c \times Q$.

But, as we know, costs per unit *aren't* constant. Rather, they are subject to the *Law of Increasing Cost*, that is the more we produce (Q), the more it costs per unit (c rises). This is because (as we explained in the earlier notes), producing more requires more factors (inputs, like labor). But more inputs won't be forthcoming unless we pay them higher returns (e.g. higher wages, higher rents). So the more we produce, the more the wage we pay per labor hour, and thus the more it costs per unit per produced.

So, unlike price, cost per unit, c, *changes* as Q rises. So the total cost curve is *not* a straight line with a constant slope. But rather a line with an *increasing* slope, e.g.



The *slope* of the total cost curve increases because the cost per unit of production increases as we increase production volume. (The different slopes are heuristically represented by the tangent lines at those points.)

To capture the law of increasing cost, assume that at Q = 5, the cost per unit is c = \$15 per unit, so total costs are \$75 (= 5 × \$15). Whereas at Q = 10, the cost per unit has *risen* to c = \$30 per unit, so total costs are \$300 (= $10 \times \$30$). Costs are rising at an increasing rate.

So cost per unit, c, is *not constant*, but rather an *increasing function* of Q. Mathematically:

$$c = f(Q)$$

so different levels of Q give us different levels of c (cost per unit). Let us suppose the relationship between the quantity and cost per unit is something simple like:

$$c = 3Q$$

That is:

if Q = 1, then c = 3,
if Q = 2, then c = 6,
if Q = 3, then c = 9,

and so on. So the greater the quantity, the greater the cost per unit. This linear relationship between cost per unit and quantity captures the law of increasing cost.

Now, remember our Total Cost function was

$$TC = c \times Q$$

but since c is not constant but rather itself a function of Q, substituting in our equation c = 3Q, the total cost function becomes:

$$TC = 3Q \times Q$$

or simply:

$$TC = 3Q^2$$

So, $TC = 3Q^2$ is the exact formula of the relationship between total cost and quantity. Notice that $TC = 3Q^2$ is a ***non-linear*** function of Q (that is, Q does not enter linearly, but is 'squared', thus we say the function is 'quadratic'). Diagramatically, a quadratic function will yield a bent curve, like the TC we drew before. Mathematically:

$$TC = 3Q^2 = 3 \times (Q \times Q)$$

(do the exponent ('square') before you multiply the coefficient. It's a good habit)

if Q = 1, then TC = $3 \times (1 \times 1) = 3 \times 1 = 3$
if Q = 2, then TC = $3 \times (2 \times 2) = 3 \times 4 = 12$
if Q = 3, then TC = $3 \times (3 \times 3) = 3 \times 9 = 27$

if Q = 4, then TC = $3 \times (4 \times 4) = 3 \times 16 = 48$
if Q = 5, then TC = $3 \times (5 \times 5) = 3 \times 25 = 75$
if Q = 6, then TC = $3 \times (6 \times 6) = 3 \times 36 = 108$

and so on. Total costs increase as quantity increases, but increasing at an increasing rate.

Slope of the TC curve ("Marginal Cost") :

The slope of the TC curve is called the ***marginal cost***. An extremely important concept. Essentially it gives us the increase in total costs from increasing production by an additional unit.

> **Marginal Cost**: the extra total cost incurred by a firm if it expands production by an additional unit. This is also the slope of the total cost curve (MC = $\Delta TC/\Delta Q$).

It is important to differentiate *marginal cost* from ***average cost***.

> **Average Cost**: total costs incurred by a firm divided by total output produced, i.e. costs per unit produced.

So, a firm that produces 100 stereos at a total cost of $50,000, has an *average cost* of $500 (per stereo). To get the marginal cost, we need to know how much increasing production from 100 stereos to 101 stereos will increase total costs. Suppose 101 stereos cost $51,000. Then the *increase* in total costs by increasing production by 1 stereo is $1,000. That is the marginal cost. So, in this example, at a production level of 100 stereos:

$$\$500 = \text{average cost} \neq \text{marginal cost} = \$1000$$

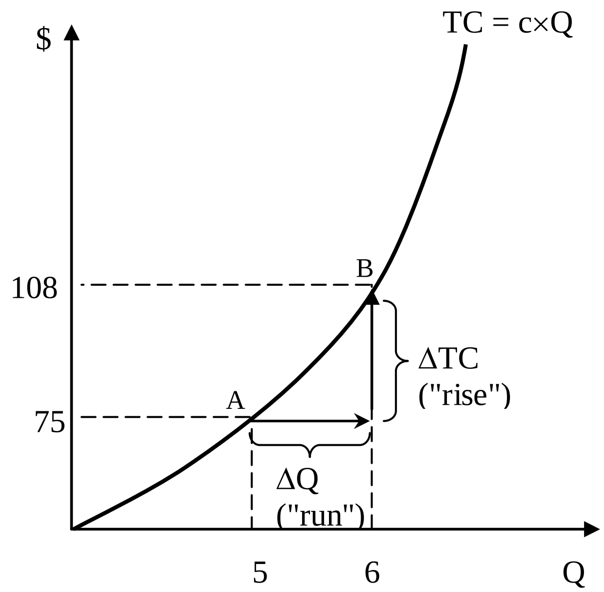
So the average cost and marginal cost are two different concepts. In mathematical terms, one is an average, the other a *derivative* (or 'slope') of a total cost curve

Let us get back to marginal cost. The slope of a curve is calculated as "rise over run", in our case, the increase in total costs ("rise", vertical axis) resulting from an increase in quantity produced ("run", horizontal axis). So:

$$\Delta TC/\Delta Q = \text{(change in TC)/(change in Q)} = \text{rise/run} = \text{slope of TC curve.}$$

But notice that this slope is *not* constant anymore. We get different slopes at different levels.

In the diagram below, we try to calculate the slope of the TC around the Q = 5 mark. Let $\Delta Q = 1$ (raise quantity produced by 1 unit, from Q = 5 to Q = 6), then ΔTC is 33 (TC rises from $75 to $108, that is, by $33). So the slope of the curve around that point - the marginal cost - is $\Delta TC/\Delta Q = 33/1$, or 33:



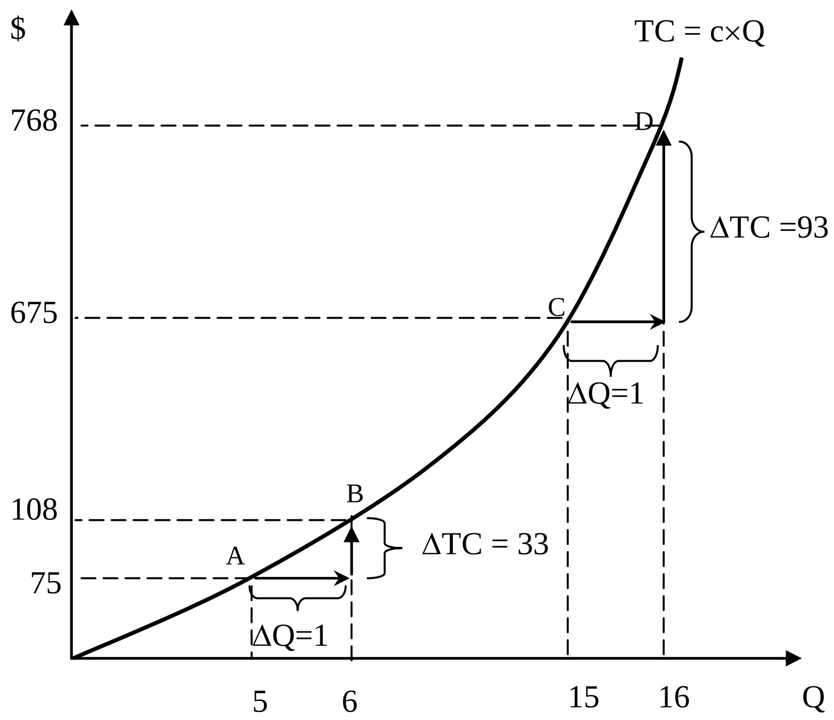
[Caveat: The *proper* calculation of the exact slope at Q = 5 would require that the "run" (ΔQ) be an infinitesimally small amount, a teensy, tiny increase, not a lumpy increase by an entire integer unit. For this we need calculus. We'll get to that in a moment.]

But we noted that the slope of the TC wasn't constant. It changes as we change the quantity.

Check the slope at a different level, say, at Q = 15. We see this in the diagram below. At Q = 15, total costs are $675. Let us do like before and increase the quantity produced by 1 unit (to Q = 16). But at Q = 16, total costs are $768. So keeping the run the same as before (ΔQ = 1), our "rise" is now much larger (ΔTC = $93 = $768-$675). The slope is ΔTC/ΔQ = $93/1 = $93.

So, the *slope* around Q = 5 is 33, but the slope of the same curve around Q = 15 is 93, a much larger number.

This increasing slope - rising 'marginal cost' as we try to produce more - is the mathematical expression of the economic idea of the *Law of Increasing Cost*: that the more you try to produce, the more expensive your inputs get.



Calculating MC

The slope of the total cost curve varies with the quantity we calculate it at.

In other words, the slope, $\Delta TC/\Delta Q$, is *itself* a function of Q.

What kind of function? What is the exact relationship between slope and quantity produced? For this, we need the calculus. Calculus is merely a set of rules which allow us to deduce the "slope function" (or 'derivative function') from the original function. (see the Math Notes)

The original relationship between total costs and quantity produced was:

$$TC = 3Q^2$$

then applying the rules of calculus, that means that the relationship between the slope and quantity produced is:

$$\Delta TC/\Delta Q = 6Q$$

(If you don't believe me, see the Mathematical Notes #2 for a brief review of calculus.)

Notice that this is itself a function:

if Q = 1, then $\Delta TC/\Delta Q = 6$
if Q = 2, then $\Delta TC/\Delta Q = 12$

if Q = 3, then $\Delta TC/\Delta Q = 18$
if Q = 4, then $\Delta TC/\Delta Q = 24$
if Q = 5, then $\Delta TC/\Delta Q = 30$
if Q = 6, then $\Delta TC/\Delta Q = 36$

and so on.

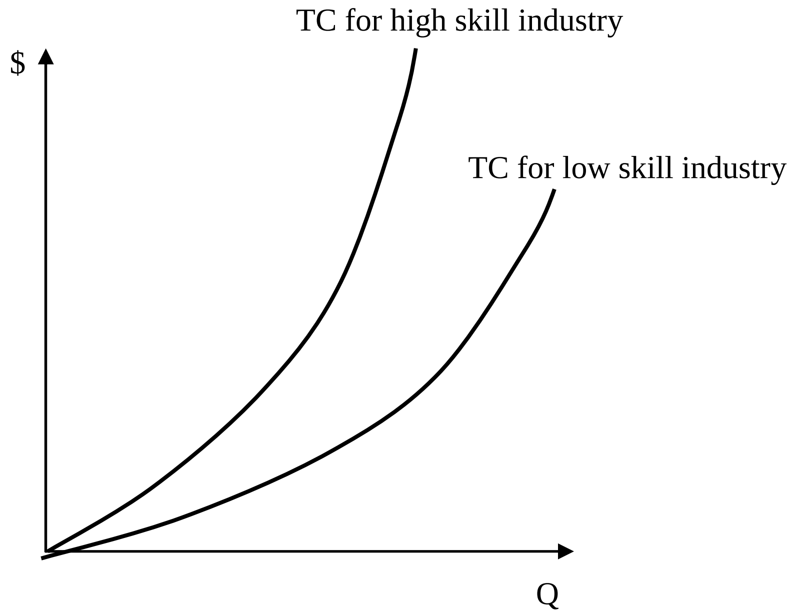
$\Delta TC/\Delta Q$ is, of course, the marginal cost (MC). So the *rising* values (6, 12, 18, etc.) shows that the slope is *steepening* the more we produce (as we see intuitively in the diagram), that is, marginal cost is *rising* as quantity produced increases. That's the law of increasing cost at work.

[Slight note: previously we said the slope around Q = 5 was 33, whereas by the calculus formula, the slope around Q = 5 is 30. Why are they different? That is because in our original graphical calculation, we made the 'run' an entire lumpy unit from Q = 5 to Q = 6 to deduce 33. But is that *really* the "slope at 5"? Or is it the "slope at 6"? Or should we say the "slope between 5 and 6"? The calculus formula is more precise. Heuristically, it measures the slope *really* around 5, as if nudge the run by only a very small infinitesimal unit, from 5 to 5.00001, not all the way to 6. Whereas, by the formula, the slope at 6 is 36. The formula notices that the slope at 5 is different from the slope at 6, whereas the graphical eyeballing method is actually combining the slope at 5 with the slope at 6, that is, giving us the average slope between 5 and 6, not the exact slope around 5. Notice by the formula, the slope at 5.5 - the midpoint between 5 and 6 - *is* 33.].

Changing cost function

Remember that in the TR curve, a price change will lead to a swiveling or swinging of the curve. Is there an analogue in the TC case? There is, but it is much, much more complicated. It is not simply a change in wages - remember, the slope of the marginal cost curve *already* captures the idea of an underlying change in wages as output increases. For the marginal curve to swing, we need something else, a change in the *relationship* between wages and output.

Intuitively, the TC curve will be steeper or flatter depending on what the cost per unit (c = *f*(Q)) function is exactly. If wages rise rapidly as output increase, the MC curve will be steep rate. If wages rise slowly as output increases, the MC curve will be ascending at a flatter rate.

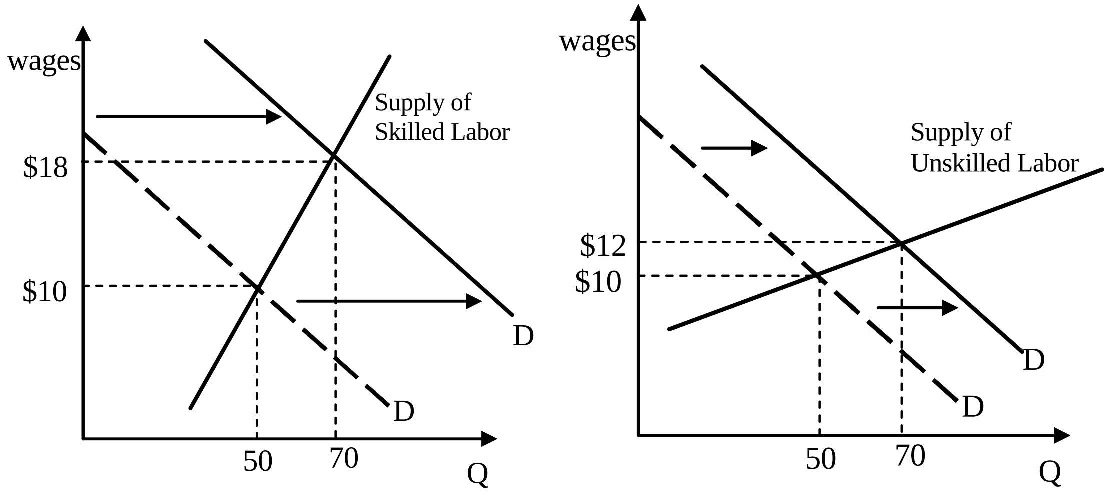


What practical cases might that be? Depends on the industry. We can suppose that industries which rely heavily on very high skilled or specialized labor (e.g. medicine), wages are bound to rise rapidly as output increases. That is because doctors are few, and getting more doctors is hard - you can't just grab a factory worker or a locksmith and give him a doctor's job.

Industries which rely on low skilled or manual labor (e.g. agriculture), wages are bound to rise more slowly as output increases. That is because getting more agricultural laborers is simply a matter of stealing manual laborers from factories and dockyards and throwing them in a field. There's a little change in skill involved, but not as much. A factory worker or locksmith can work a hoe in a field easier than they can conduct a eye surgery. You have to pay them higher wages to get to them to switch, yes, but not that much more.

So what makes for a steep or flat TC curve? In this case, a steep or flat *supply of factors*.

Specifically, skilled labor has a *steep* labor supply curve, while unskilled labor has a relatively *flat* labor supply supply curve, as shown in the diagrams below.



Suppose the industry tries to increase output and thus needs increase the labor hired from 50 workers to 70 workers. If the industry relies on skilled labor, than to get 20 more workers, it needs to increase wages from $10 to $18 per hour. If it relies on unskilled labor, than to get 20 more workers, it need only increase wages from $10 to $12.

In both cases, the law of increasing cost is at work: to get more workers, you need to pay more wages. But wages rises more rapidly in the skilled case than in the unskilled case.

Why labor supply? Again, the pool of skilled labor is smaller. There isn't a bunch of trained medical doctors working in barber shops or factories than you can just draw from instantly. Whereas barbers and factory workers are more easily switched over to manual agricultural labor.

Of course, wages are only one of several costs involved. There are also rents, material inputs, etc. whose costs may rise more rapidly in some industries than others. e.g. textile industries which use rare or less substitutable inputs, say silk or cashmere, may find their costs increasing more rapidly as their output increases than industries which use more easily obtainable inputs like cotton or wool.

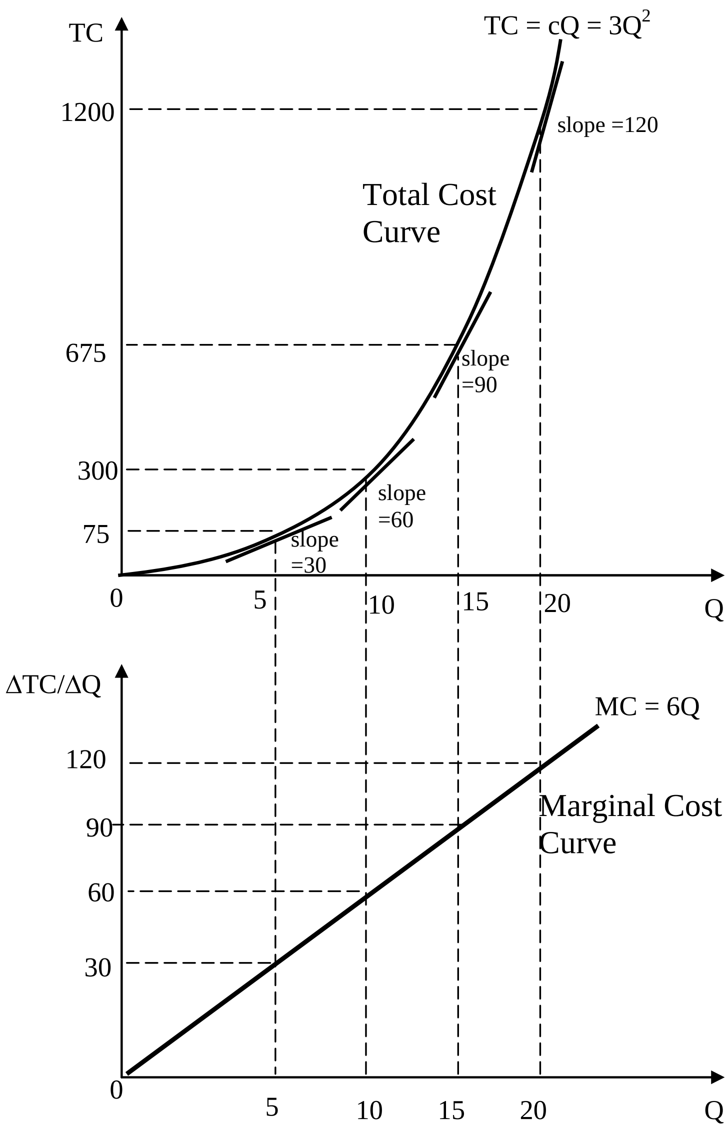
So the increasing cost shape of the TC curve is the same - all must pay inputs more to increase output. That is because factor supply curves - all factor supply curves - are necessarily upward-sloping. But *how* much more they pay - the rapidity of the increase of TC - depends on the underlying steepness or flatness of factor supply curves.

PLOTTING MARGINAL CURVES

Plotting MC

We said ΔTC/ΔQ is a function of Q. If it is a function, then it can be plotted. The following shows the relationship between our *total cost* function ($TC = 3Q^2$) and the *marginal cost* function (MC = 6Q) e.g.

if Q = 5, then TC = 3 × (5 × 5) = 3 × 25 = 75 and MC = 6 × 5 = 30
if Q = 10, then TC = 3 × (10 × 10) = 3 × 100 = 100 and MC = 6 × 10 = 60
if Q = 15, then TC = 3 × (15 × 15) = 3 × 225 = 675 and MC = 6 × 15 = 90
if Q = 20, then TC = 3 × (20 × 20) = 3 × 400 = 1200 and MC = 6 × 20 = 120



You will notice immediately that the marginal cost curve is an upward sloping line. Which should be obvious just from the formula $\Delta TC/\Delta Q = 6Q$, that the relationship between $\Delta TC/\Delta Q$ and Q is linear and increasing. Marginal cost is increasing at a steady rate. This rising marginal cost shows the law of increasing cost explicitly. The more one produces, the more it costs to produce an additional unit.

[You will notice that the marginal cost function *itself* has a slope. That is, the MC is a line with slope 6. Notice that that is simply the application of the calculus again - albeit taking the slope function as the new 'original'. In calculus, the slope of a slope function is called the *second derivative*]

Plotting MR

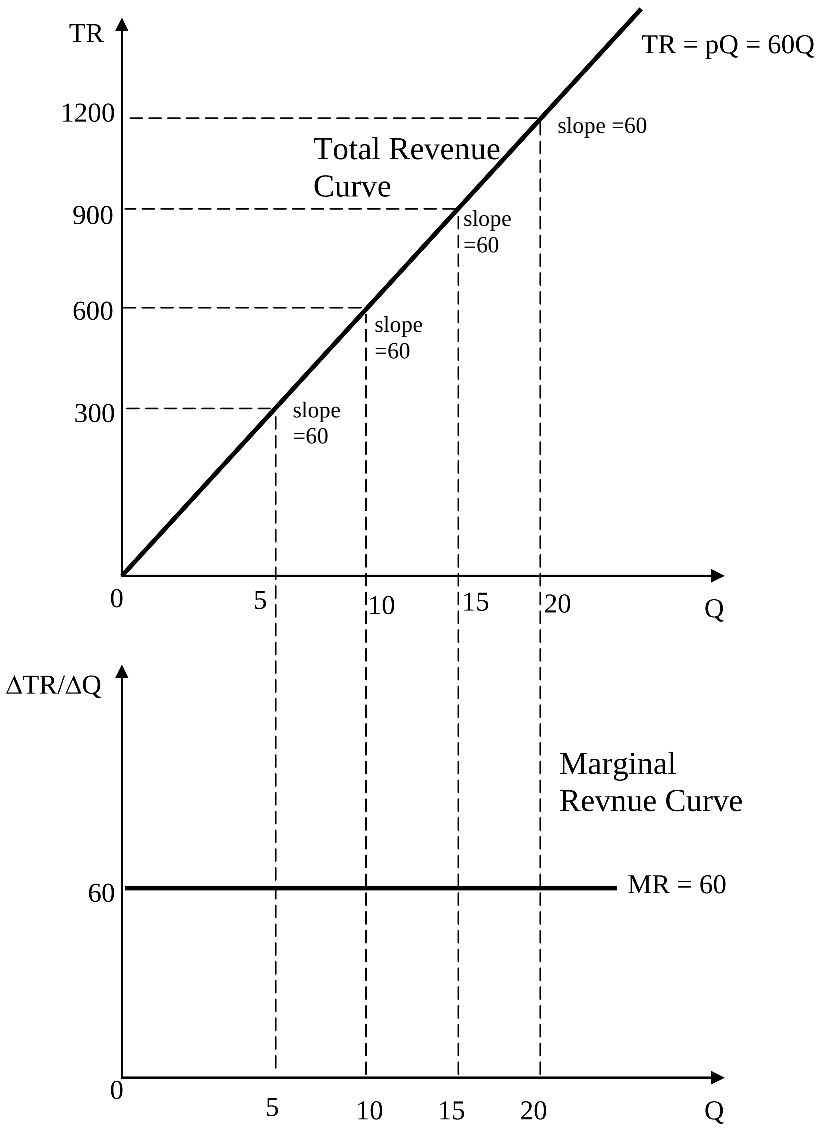
If we can plot marginal cost, can we plot marginal revenue too?

Sure we can. The same way. Except it is not very interesting. That is because marginal revenue, as we noted before, does *not* change with quantity produced. So the MR curve will just be a flat, horizontal line.

To see this, suppose p = 60, then our total revenue TR = pQ = 60Q, but MR = $\Delta TR/\Delta Q$ = 60, which doesn't vary with Q. So:

if Q = 5, then TR = 60×5 = 300 and MR = 60
if Q = 10, then TR = 60×10 = 600 and MR = 60
if Q = 15, then TR = 60×15 = 900 and MR = 60
if Q = 20, then TR = 60×20 = 1200 and MR = 60

and the plot would look something like the following:

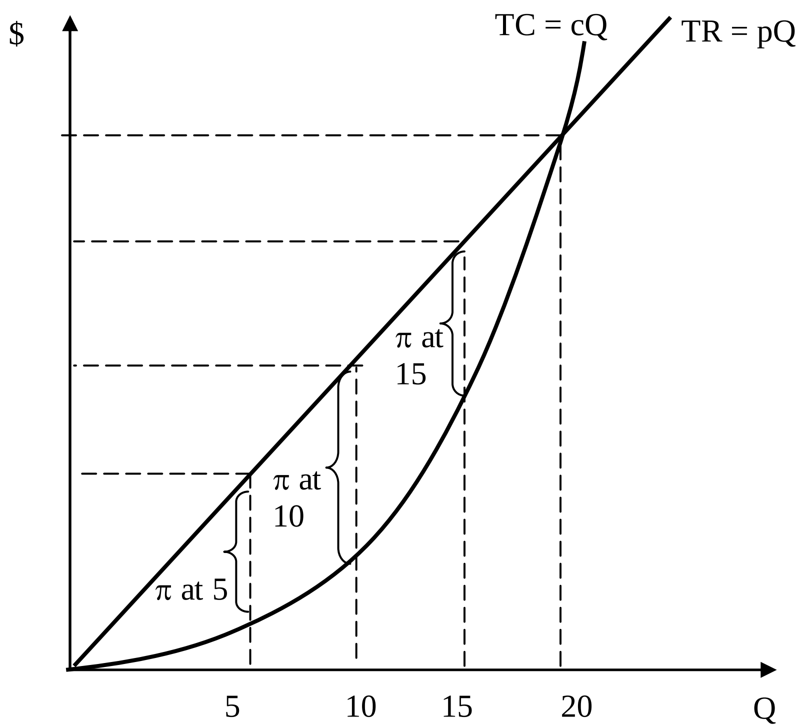


Regardless of what the quantity is (5, 10, 15, etc.), the marginal revenue is *always* 60. So the Marginal Revenue curve is simply a *flat*, horizontal line at 60. Not as interesting.

PROFIT-MAXIMIZATION DECISION

(1) First Approach: The Numerical Example

OK. So what is the level of production that *maximizes* profit?   To see this, we need to combine both the TR and TC functions in one diagram, like so:



Remember that profit (π) is defined as TR - TC. So profit is the *difference* between the TR and TC curves.   So the vertical gaps between the curves denote the different profits at the different levels.

Profits vary depending on how much we choose to produce. We immediately see that the profit at Q = 5 ("π at 5") is different from the profit at Q = 10 ("π = 10") which is in turn different from the profit at Q = 15 ("π at 15") and so on.

Which is highest? Do we make maximum profits when Q = 5, 10, 15 or 20?

Well, Q = 20 is obviously not it. That is because at Q = 20, the TR and TC curves intersect, that is, TR = TC, so profits there are necessarily zero.

But we have three remaining gaps - Q = 5, Q = 10 and Q =15. Which is largest?

Just from eyeballing the diagram, the profit at Q = 10 *looks* largest, so we suspect profits will be highest there. Can we verify this?

To calculate, we need to calculate the profits for the different levels, i.e. subtract TC from TR at every level.

Let's start with total revenue (TR). The total revenue curve is for a single *given* price, p = 60. So TR = pQ = 60 × Q. So, the total revenues for the different levels are:

TR at 5 = 300
TR at 10 = 600
TR at 15 = 900
TR at 20 = 1200

What about total costs (TC)? Our formula says that TC = cQ, but c (cost per unit) is not constant, remember? It varies with the amount produced ("Law of Increasing Cost"). We proposed that this variation in unit costs was represented by the function c = 3Q, with the result that the total cost function is TC = (3Q) × Q or simply $TC = 3Q^2$. As a result, at the different levels considered:

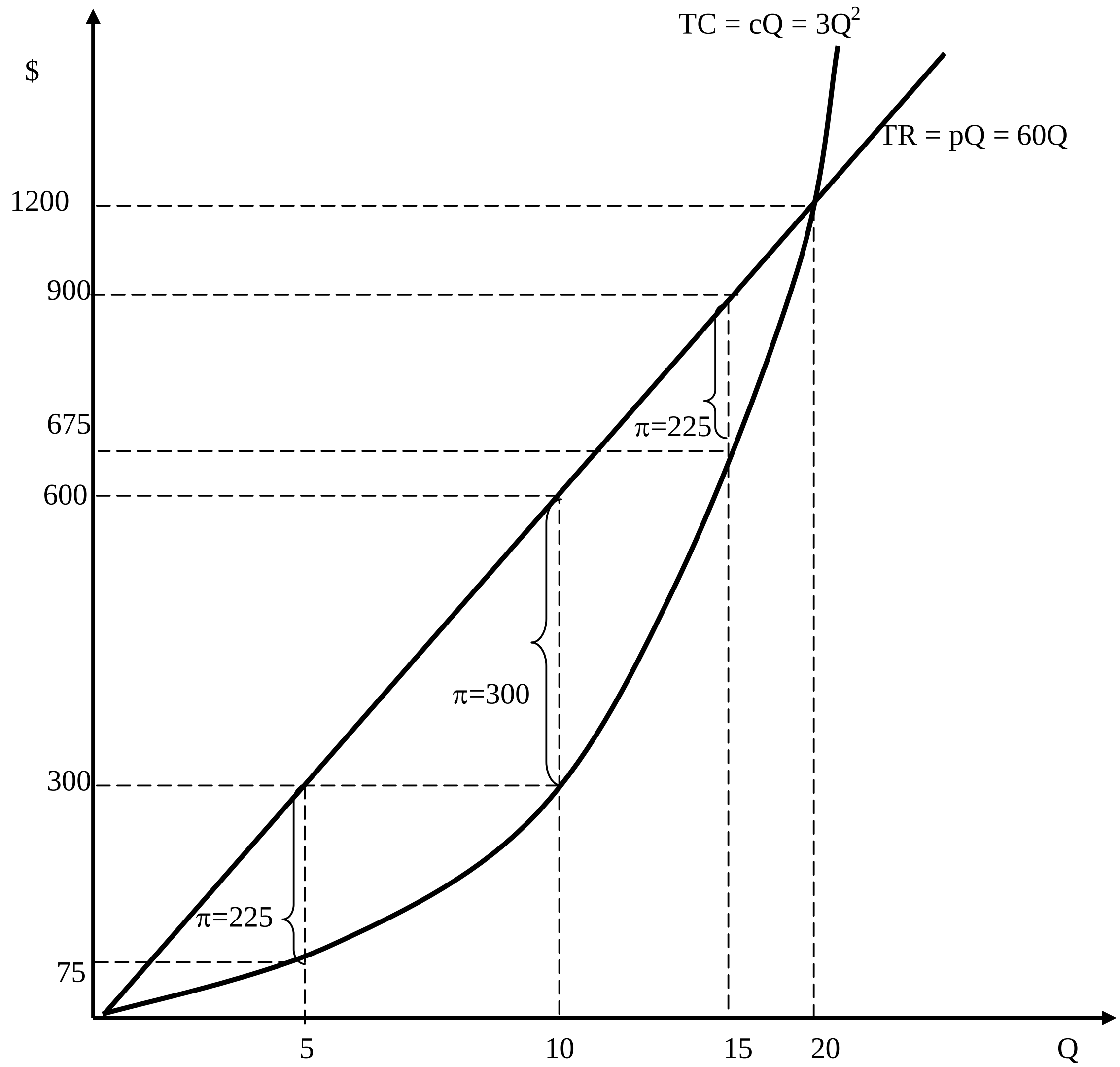
TC at 5 = 75
TC at 10 = 300
TC at 15 = 675
TC at 20 = 1200

Now let us calculate profits at each level. As we know, π = TR - TC. So:

π at 5 = 300 - 75 = 225
π at 10 = 600 - 300 = 300
π at 15 = 900 - 675 = 225
π at 20 = 1200 - 1200 = 0.

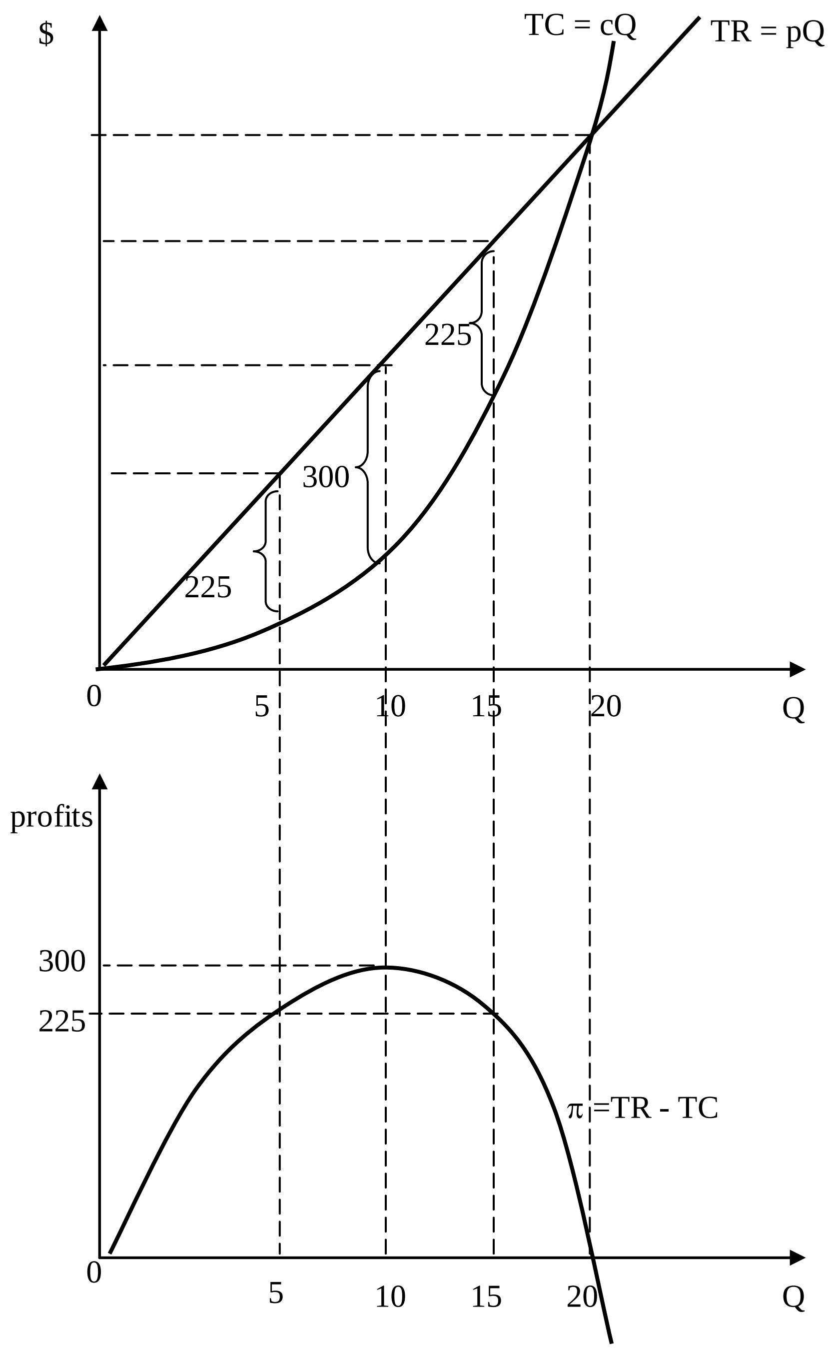
The maximum profit point looks like it is indeed at Q = 10.

The following diagram shows the exact numbers:



(2) <u>Profit Curve</u>

If we'd like, we can actually plot the different profit levels via a **profit curve** plotting the different profit levels at different prices. This is shown below, where we depict the correspondence between the two diagrams. Notice that at $Q = 0$, we have zero profit ($\pi = 0$), so we start on the horizontal axis. As quantity increases from $Q = 0$ to $Q = 5$, profit rises from from 0 to 225. As we go from $Q = 5$ to $Q = 10$, it continues to rise, to $\pi = 300$. *But* as we continue to increase production from $Q = 10$ to $Q = 15$, profit now begins to fall, from $\pi = 300$ to $\pi = 225$ and continues falling until we reach $Q = 20$, where profit is zero ($\pi = 0$) again.



Notice the profit curve is shaped like a hill.   The top of the profit hill is the **profit-maximizing point**.

In this diagram, the profit-maximizing point is where Q = 10 (thus maximum profit is $300). This should be where the firm chooses to produce. If it produces more or less than 10 units, profits will be lower.

(3) <u>The Profit-Maximization Decision: The Marginal Rule</u>

Finding the profit-maximizing level of output from a small array of choices (Q = 5, 10, 15, 20) is merely a matter of arithmetic.

But how do we do it with a much larger array of possible levels. Clearly calculating the profit at *every* level of output between Q = 0 and Q = 20 is tiresome and mind-numbing work. There is a quick and easy "rule" to find the profit maximization level in an instant. We know this from before: it is the **marginal rule**

> **Profit-Maximizing Rule**: (marginal rule): to maximize profits, find the level of Q where the marginal revenue is equal to the marginal cost, MR = MC.

In a competitive market scenario (i.e. *not* monopoly), marginal revenue is just price, and the price is *given* by the market. Remember also that marginal cost (slope of the TC curve) varies with Q. So every different Q yields a different MC. So the rule tells you to choose the Q where the MC at that Q is exactly equal to the market-given price p.

Why? The mathematics is simple. Remember that by definition, at any level of Q:

$$\pi = TR - TC.$$

If we change the quantity produced, profits change. And profits change because total revenues and total costs change. So if we displace Q by some amount (ΔQ), then:

$$\Delta\pi/\Delta Q = (\Delta TR/\Delta Q) - (\Delta TC/\Delta Q)$$

where $d\pi$, ΔTR and ΔTC are the changes in profits, total revenues and costs resulting from the change in Q.

Where does this equation come from? Intuitively, the *extra* profit made from increasing production by one unit is the difference between the *extra* total revenues earned and the *extra* total costs incurred.

[Aside: If you want to be more precise: suppose we only want to find $\Delta\pi/\Delta Q$, that is the change in profit that results from a change in quantity by the amount ΔQ, e.g. in our example, increasing production from Q = 5 to Q = 6, means $\Delta Q = 1$. How much is $\Delta\pi$ then? Well, at Q = 5, $\pi = 225$, as we saw, as TR = 300 and TC = 75. To calculate π for 6, just use the formulas again: TR = 360 (= 60×6) and TC = 108 (already calculated above), so π at 6 = \$360 - \$108 = \$252. So the increase in profits from increasing quantity by a unit is $\Delta\pi$ = \$252 - \$225 = 27.

Notice we could have broken $\Delta\pi$ that down into two components: the *increase* in TR minus the *increase* in TC, and come to the same result. The increase in TR is:

ΔTR = TR at 6 - TR at 5 = 360 - 300 = \$60

ΔTC = TC at 6 - TC at 5 = 108 - 75 = $33

so:

$\Delta TR - \Delta TC$ = $60 - $33 = $27.

Which is identical to $\Delta\pi$. So:

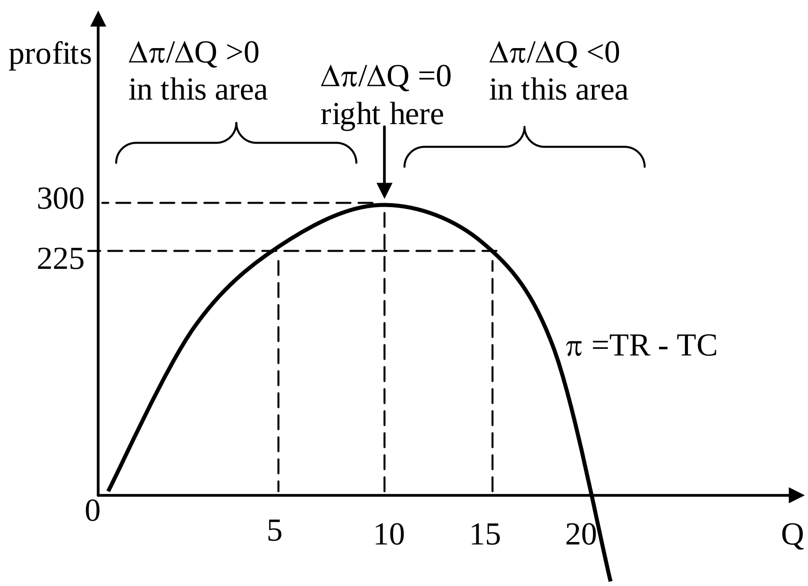
$$\Delta\pi = \Delta TR - \Delta TC$$

Or (remembering $\Delta Q = 1$, as we are only increasing by one unit):

$$\Delta\pi/\Delta Q = (\Delta TR/\Delta Q) - (\Delta TC/\Delta Q)$$

Which is what we proposed.]

Now, $\Delta\pi/\Delta Q$ also happens to be the *slope* of the profit curve depicted earlier. This is instructive. Notice that *below* the peak of the hill (Q = 10), the slope of the curve is positive, that is $\Delta\pi/\Delta Q > 0$, whereas *above* the peak., the slope of the curve is *negative,* that is $\Delta\pi/\Delta Q < 0$. So, somewhere in between, where the slope flips from positive to negative, it must pass through zero, i.e. there is a point somewhere on the curve where $\Delta\pi/\Delta Q = 0$. Where is this point? Why, the peak of the hill *exactly*.



As we know, the peak of the hill is the point of maximum profit. So a simple *condition* to know whether we are at the profit-maximizing point or not is to check whether $\Delta\pi/\Delta Q = 0$. If this is not true, we are not at the maximum point. If it is true, then we know we're there.

What's the implication of this little insight? Well go back to the formula:

$$\Delta\pi/\Delta Q = (\Delta TR/\Delta Q) - (\Delta TC/\Delta Q)$$

If we're at the profit-maximizing point, then it must be true that $\Delta\pi/\Delta Q = 0$, or:

$$\Delta\pi/\Delta Q = (\Delta TR/\Delta Q) - (\Delta TC/\Delta Q) = 0$$

Or, rearranging:

$$(\Delta TR/\Delta Q) = (\Delta TC/\Delta Q)$$

OK. So what? Well, notice from our previous definitions, that $\Delta TR/\Delta Q$ *is* the marginal revenue, or, as we found out, *price* (p). While $\Delta TR/\Delta Q$ *is* the marginal cost. So the profit-maximizing condition $\Delta\pi/\Delta Q = 0$, implies:

$$MR = MC$$

that is the profit maximizing "marginal rule"!

Calculating the Profit-Maximizing Quantity

How does that rule help the firm decide? Well, we know that in a competitive scenario, MR is the price and the price is *given* by the market (p = 60), but marginal cost (MC) is flexible. Marginal costs changes as the firm changes quantity produced (law of increasing cost, etc.). So this rule tells us a firm should adjust Q back and forth until it finds the quantity which has a marginal cost of 60. At *that* quantity, it is can be sure it is maximizing profit.

So what is it? Well, in our example, we proposed that p = 60 and that total costs were governed by the formula $TC = 3Q^2$. To find the profit-maximizing quantity, all you have to do is find the marginal cost formula. As per calculus's power rule. we know marginal cost is:

$$MC = \Delta TC/\Delta Q = 6Q$$

and

$$MR = \Delta TR/\Delta Q = 60$$

Since MR = MC is our rule, then we know that at the profit maximizing point:

$$60 = 6Q$$

So, dividing:

$Q = 60/6 = 10.$

That is, the profit-maximizing point is where Q = 10. Bingo! We got it in an instant. No need for long arithmetic calculations, comparing profits at every different level. Just apply the MR = MC formula, and you'll find the profit-maximizing solution in an instant. We just needed two bits of information - price and the shape of the marginal cost curve - and we got it in a few seconds.

Notice that if the price was *different*, the solution would be different. Suppose price rises to p = 90. What is the profit-maximizing quantity now? Easy. It is still true that MC = 6Q. But MR = p = 90 now. So apply the MR = MC formula:

$90 = 6Q$

so:

$Q = 90/6 = 15.$

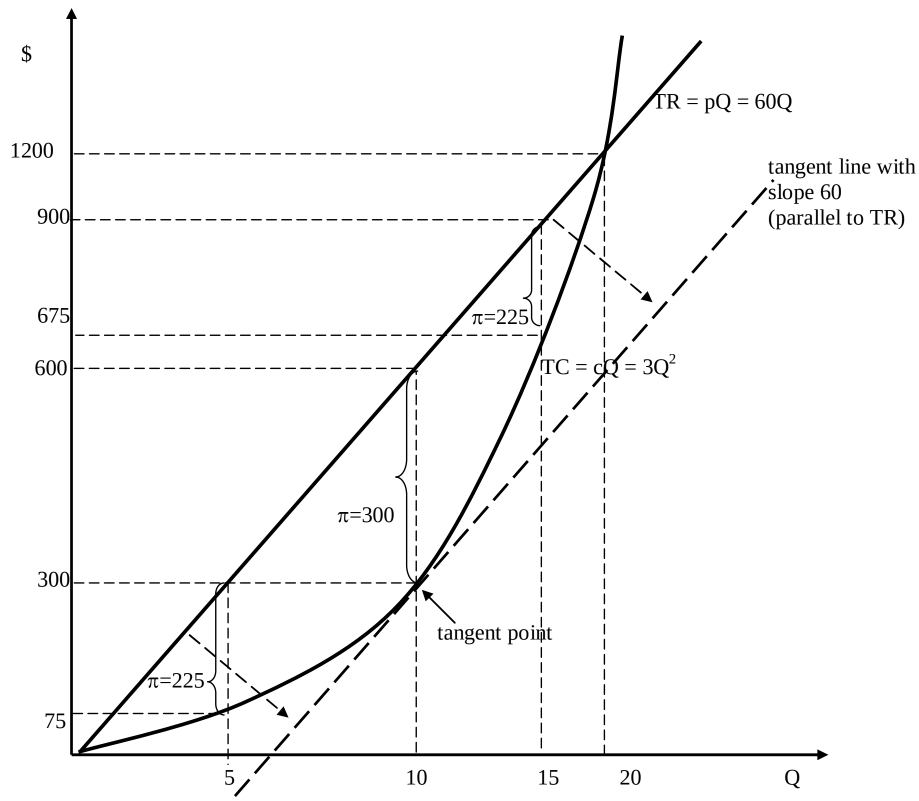
so the profit-maximizing is now Q = 15.

Notice the relationship: at p = 60, profit-maximizing firms produce 10. At p = 90, profit-maximizing firms produce 15. In other words, as price increases, quantity produced increases. This *is* the **supply curve**!

Diagramatic intuition: ugly way

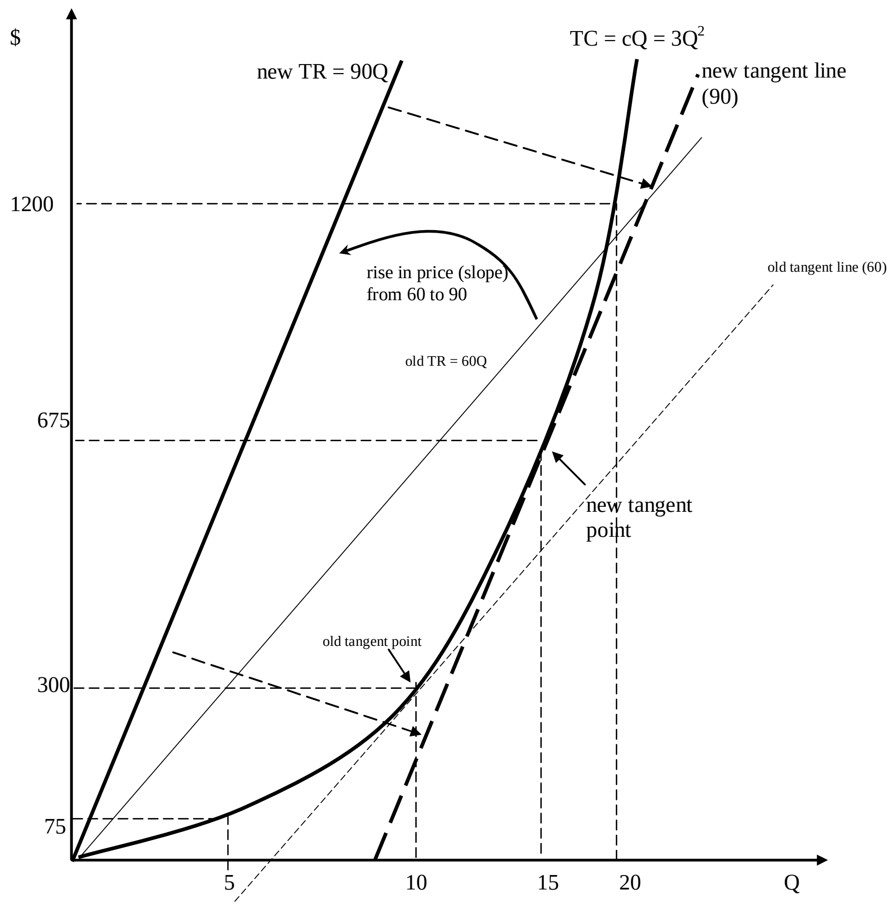
There is a diagrammatic analogue to the MR = MC formula. The analogy is this: we *know* p is the slope of the TR curve and we *know* MC is the slope of the TC curve. So the profit maximizing solution is where the *slopes* of the TR & TC curves are equal to each other.

Diagrammatically, this is done by *shifting* the TR curve in parallel fashion until it is just *tangent* to the TC curve ("tangent" means it touches at one point only). We see this in the figure below, where we show the tangent line (with slope = p = 60) and the tangent point (at Q = 10). At this tangent point, p = MC.



Notice that if price rises, from p = 60 to p = 90 say, then the TR curve (and thus the tangent curve) would be *steeper* and so the tangency point different, in this case at Q = 15. The diagram below shows that.

(sorry for the cluttered look, I wanted to leave the old tangent line on the diagram. Also, I didn't have the patience to try to fill in for the new different profit levels. I hope it makes sense.)



So there is a very simple mechanical way to figure out the profit-maximizing point on any diagram: just do a parallel shift of the TR curve until you reach a tangency point on the TC curve. That's your profit-maximizing level.

<u>Diagramatic intuition: easy way</u>

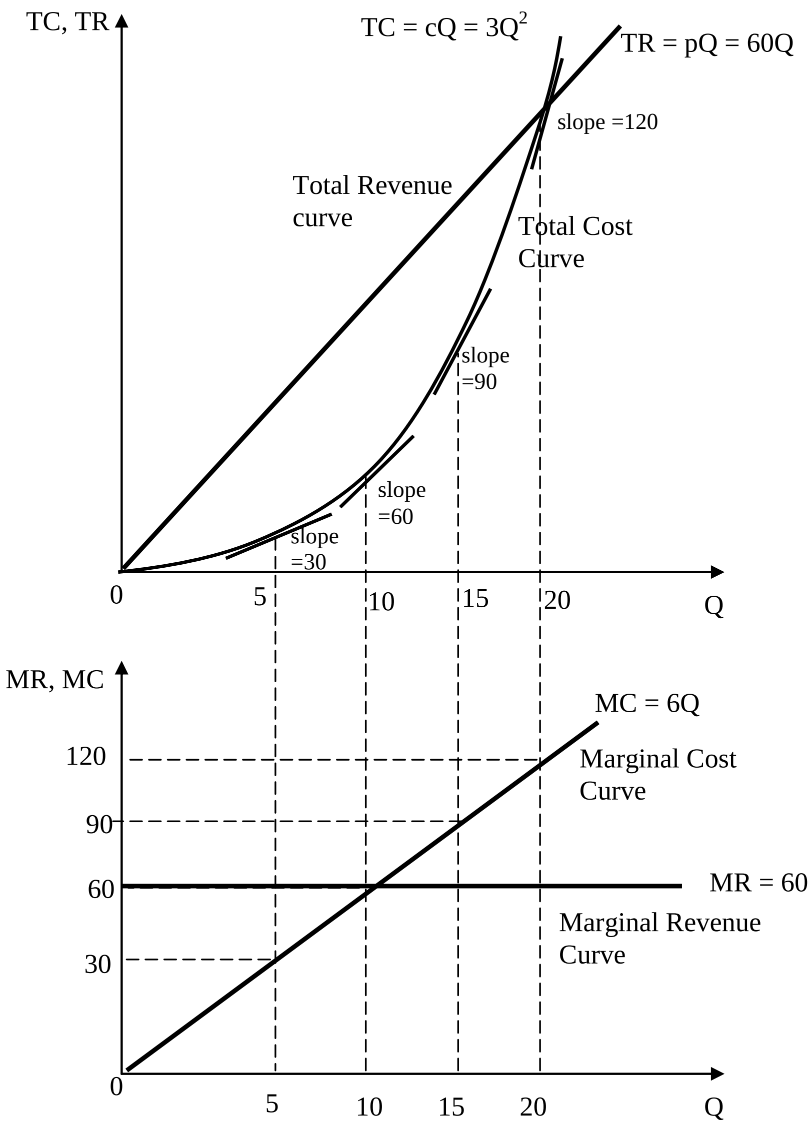
Hold on. Didn't we draw plots for MR & MC curves earlier? Can't we just look at those? Indeed we can. And it's a lot simpler to hone in.

Remember that:

$$MR = \Delta TR/\Delta Q = 60$$

$$MC = \Delta TC/\Delta Q = 6Q$$

which means that the marginal revenue curve, when plotted, is just a flat horizontal line, whereas the marginal cost curve, when plotted, is an upward sloping line. Plotted together, it looks like the following:

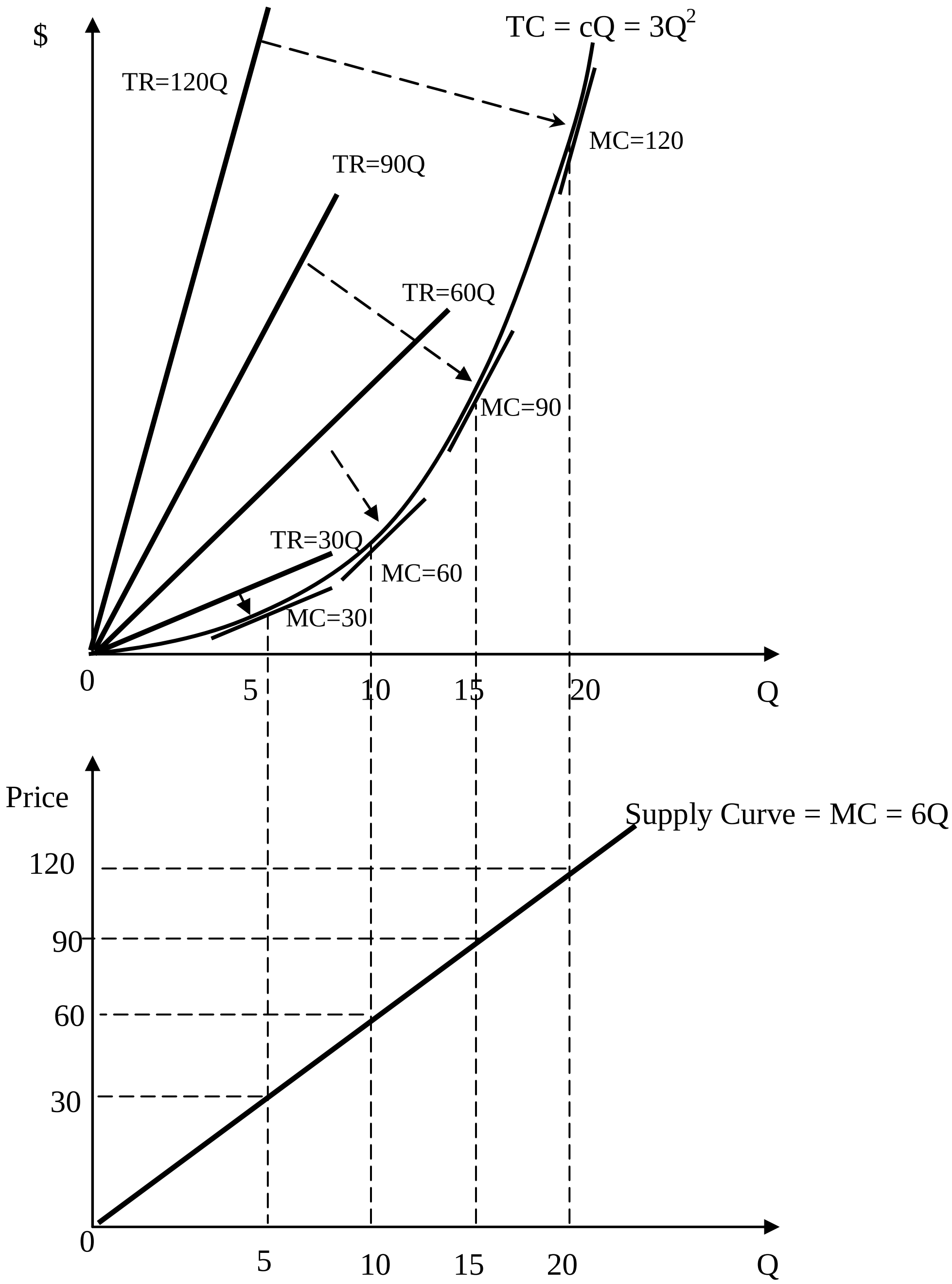


The MR and MC curves are plotted in the lower diagram. So what's the profit-maximizing point? We know the "rule" is that profits are maximized where MR = MC. Diagramatically, that simply means where the flat MR curve intersects the upward-sloping MC. And as we can immediately see, that is where Q = 20. So Q = 20 is the profit-maximizing point (as we have already verified).

<u>The Supply Curve</u>

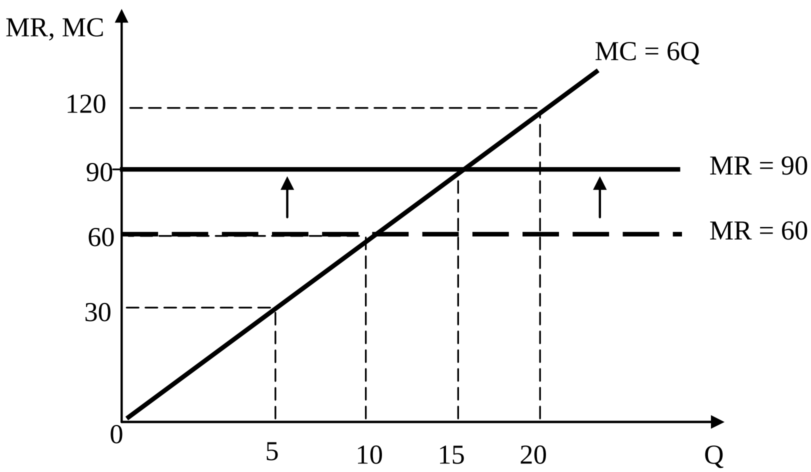
For every price, we can obtain a different profit-maximizing outcome by the p = MC rule. And if we vary price continuously from 0 to infinity, we will trace out the profit-maximizing output. This will be nothing less than our familiar friend, the **supply curve**.

An example is shown heuristically below, where we consider a whole range of total revenue curves, each for a different price: p = 30, p = 60, p = 90 and p = 120. For each TR curve, we can get their equivalent tangencies (MC) on the TC curve and consequently the profit-maximizing outputs for each price. In the lower quadrant, we thus can trace the supply curve by plotting these quantities against the prices which yielded them:



Notice a curious and very important fact. Since p = MC at *every* one of these points, then we can think of the firm's supply curve as the ***marginal cost curve*** plotted out again. That is, the curve in the bottom quadrant is merely mapping the *slopes* of the TC curve in the upper quadrant. This readily gives us the equation for the firm's supply curve, e.g. if the TC curve is $3Q^2$, then the MC curve - i.e. supply curve - is 6Q. That is, in this example, the firm's supply curve is a linear curve with slope 6.

Another way of seeing this is by looking directly at MR & MC curves. When prices change from 60 to 90, all that changes in the MR & MC diagram is the MR curve. MC doesn't budge. But the MR shifts *up*. That is, the whole flat horizontal line moves from MR = 60 to MR = 90. The new profit-maximizing point is the new intersection of the MR = MC curve. So profit maximizing quantity increases from Q = 10 to Q = 20.



Proceeding like this for different prices (that is, for different MRs), we are effectively just "tracing out" the shape of the marginal cost curve. So the supply curve of the firm *is* its marginal cost curve.

Nota Bene : Of course, to be accurate, this is only the supply curve for a *single* firm. The supply curve of the *market* as a whole is not this. Lots of firms supply to the market and we have to add it all together. But we can get the market supply curve just by adding up the supply curves of all the individual firms.

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.

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

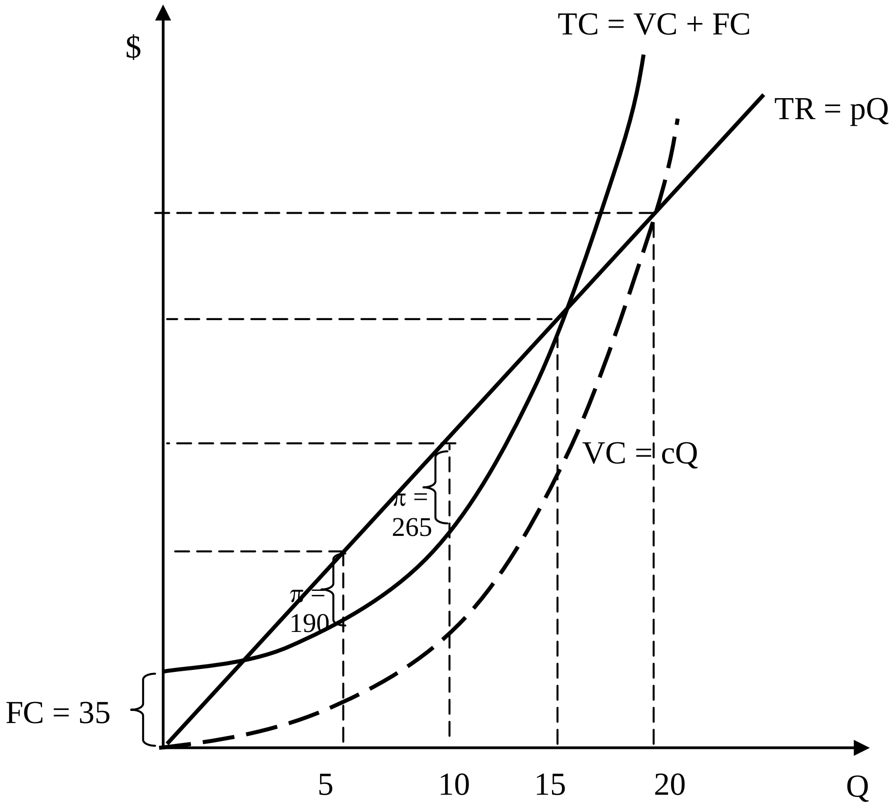
The reason is that fixed costs are *not* subject to the law of increasing costs. The more we produce, the more variable costs (like wages) rise, but fixed costs don't rise at all.

Diagrammatically, the only difference fixed costs make is that the *entire* total cost curve "shifts upwards". That is, instead of TC = cQ, we now have:

$$TC = cQ + FC$$

where FC = "fixed costs" (we can consequently call cQ = "variable costs", VC).

An example is shown below, where assume fixed costs FC = 35.



Notice that Fixed Costs is the 'intercept" of the new curve. The logic is simple. Look at the origin. When Q = 0 (we produce nothing), variable costs are zero (VC = cQ = 0) because the factory is producing nothing, it hires no labor and thus pays no wages. *But* the factory still have to pay rent (FC = 35) to the landlord *regardless*.   So, at Q = 0,

TC = VC + FC = 0 + 35 = 35

Thus the TC curve intercepts the vertical axis at 35.

[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 firm is really going to maintain Q = 0 over a long time, it might as well abandon the factory too. It will still have to pay rent on the factory land until the end of the lease. But once the lease is over, the firm will stop paying rent too and total costs are zero again.]

The introduction of Fixed Costs obviously affects the TC curve. So why won't it affect the profit maximizing decision?

We can see this intuitively in the diagram. All FC does is shift the TC curve *up*, vertically. Profits, remember, are the vertical distance between TR and TC curves. So, with FC, profits are squeezed. *But* the profit levels are squeezed by the *same* amount at *every* level.

Remember that, before fixed costs were introduced, at Q = 5, profits were $225 and at Q = 10, profits were $300. Well, all the FC does is subtract the *same* amount ($35) from both. So with fixed costs, profits at Q = 5 are $190 and profits at Q = 10 are $265. So it is *still* true that profits at Q = 10 are greater than profits at Q = 5. There is no reason to adjust quantity produced.

Are we sure that Q = 10 remains the profit-maximizing level even after we introduced fixed costs? Yes. It is easy to prove mathematically. Remember that FC is a fixed amount (like $35), so its just adding a constant to our total cost formula. FC doesn't affect the slope, so marginal cost doesn't change at all.

If you want to make sure, let us use our old example:

$$TC = VC + FC = 3Q^2 + 35$$

Marginal cost is the derivative of this entire function. But we know from calculus that the derivative of a constant (FC) is *zero*. So:

$$\Delta TC/\Delta Q = 6Q$$

just as before. So *marginal cost doesn't change* at all. The profit-maximizing decision is still where MR = MC, and so it is *still* Q = 10.

In sum, fixed costs may squeeze profits, but they don't affect the profit-maximizing quantity decision of the firm.

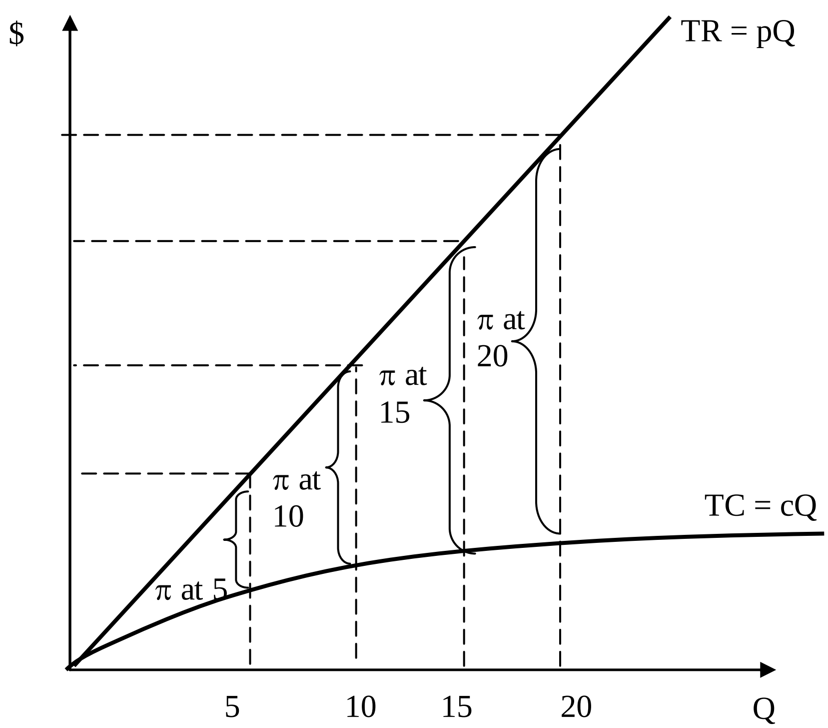
ECONOMIES OF SCALE

In our simpler notes, we briefly covered the idea of "**economies of scale**" - that is, the ability to implement a more cost-efficient methods of production at a large scale of production.

The traditional example is, of course, the assembly plant. Or discounts for bulk buying. And so on.

That throws a wrench into the concept of increasing cost. Because the availability of *cheaper* techniques of production or discounts if you produce on a large enough scale implies that costs per unit will actually *decline* as you increase production. So, in a sense, we *can* have a law of *decreasing costs*!

We said the slope of the TC curve represented increasing costs. That gave it its exponential or u-shape. Decreasing costs would be represented by a logarithmic or n-shape. It would look something like the following:



Notice the TC curve is still upward-sloping, but at a decreasing rate. That gradual flattening of the slope represents decreasing costs per unit as production increases. (c declines as Q increases).

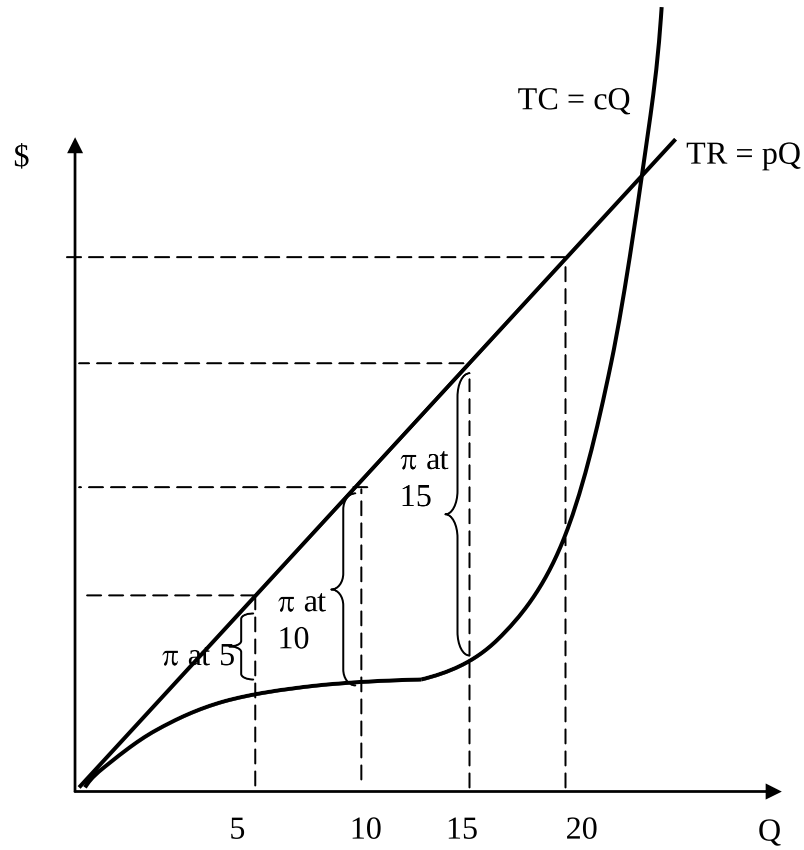
What's the profit-maximizing point? There isn't any! Profit is still the gap between TR and TC. And notice that the gap between the TR and TC is widening all the time. So profits are higher the more you produce. The profit-maximizing point is *infinity*.

This, of course, is a ridiculous notion. There aren't infinite resources out there - infinite laborers, infinite raw materials, etc. So both logically and empirically, there can't be decreasing costs.

At least not everywhere along the curve. But there *can* be decreasing costs for a spell. As you increase production from 10 to 10,000, costs per unit 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 and resources are limited, so you'll be driving those costs up eventually.

That is, 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.

This implies that our total cost curve will combine both decreasing and increasing costs. It could look something like this:



Where, notice, there are increasing costs until about Q = 12, and then decreasing costs after that. That means that up to 12, you can benefit from improved techniques and 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 that is finite and rational. And it is still where MR = MC.

MONOPOLISTIC SITUATIONS

We said the profit-maximizing rule was to adjust Q until MR = MC. We also said that p = MR. That didn't change when we considered fixed costs or even accounted for the possibility of economies of scale. But it *does* change tremendously when we consider situations of **monopoly** (market with only one producer) or any situation with *less* than perfect competition (e.g. an **oligopoly**, with only a few firms, two or three, dominating the entire market).

That is because in a less-than-perfect competition situation, MR is *not* equal to price. That is because the firm doesn't take the market price as "given". As, in a monopoly, the firm is the *only* firm, that means that whatever quantity it produces is the *entire* supply on the market. And if market supply increases, as we know, market prices fall.

In short, in a monopoly, a firm has enough heft to influence the price on the market. So *price varies* with the quantity produced by a monopolistic firm.

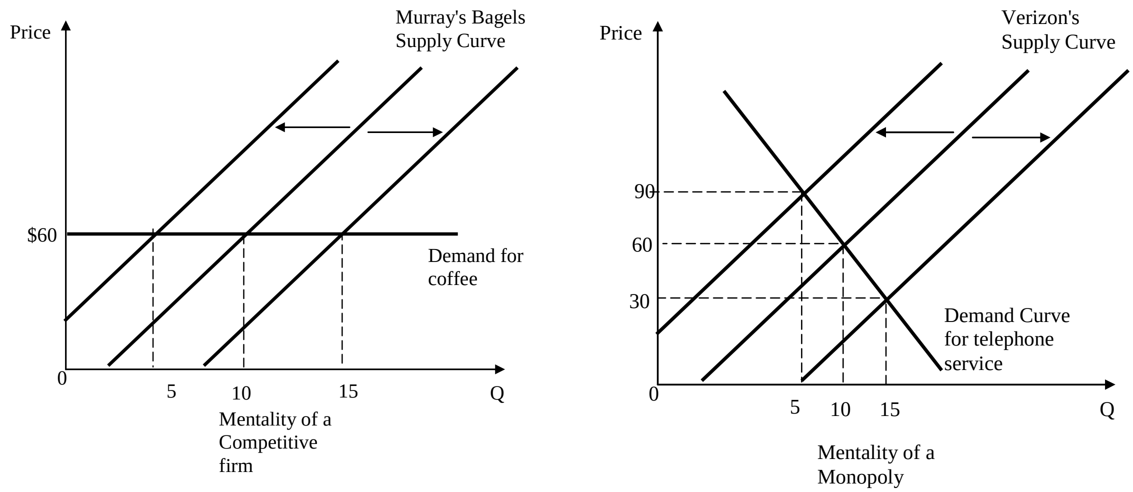
You can think of the difference between competitive and monopolistic firms by thinking of the difference between, say, a small coffee shop like Murray's Bagels and a large utility company, like Verizon or Con Edison.

Small coffee shops operate in a highly competitive market. The amount of coffee sold by Murray's is miniscule when compared to the total amount of coffee sold in New York City. Murray's doesn't have much influence on market price of coffee. It's contribution to total NYC coffee output will *not* affect the average price of coffee in NYC. Whether it produces more or less, the NYC price will remain the same. Consequently, we might as well say it has no choice but to consider the price of coffee as "given" by the market

Big monopolies, like Verizon, Con Edison, and the like, are a whole different kettle of fish. They are the sole suppliers. Their supply *is* the entire market supply. While Murray's Bagel's has a miniscule and negligible effect on the market supply curve, Verizon's supply *is* practically the total market supply of telephone service.

That means that if Verizon supplies *too much* telephone service, then price of telephone service must drop in order to bring demand up to Verizon's higher supply. If Verizon supplies *too little*, then the price of telephone service rises, cutting demand down to the lower supply.

So Verizon's supply decision affects prices. Consequently, when deciding how much to produce, it must realize its affect on price, it must take market demand 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 Verizon.]

In sum, the monopolist firm takes the demand function into consideration. Competitive firms didn't. Competitive firms just *assumed* the market price was "given" and fiddled with their quantity decision (i.e. chose Q) *regardless* of what impact that decision might ultimately have on market price. But when monopolists fiddle with their quantity decision, they *know* that decision has an impact on market price, and they take that impact into consideration.

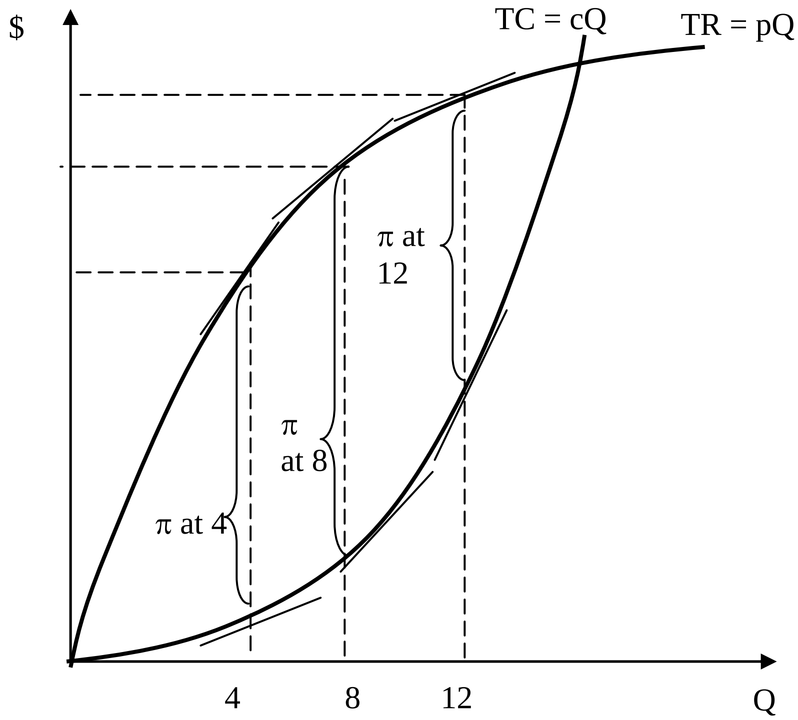
To understand why, remember that in a highly competitive environment, there are *many* firms trying to supply a product. Think of all the coffee shops in New York City, each trying to supply coffee. Each individual coffee shop has only a tiny share of the total New York coffee market. So, relative to the total size of the market, an individual firm is miniscule. An individual firm's decision of whether to produce more or less, or even if it goes out of business altogether, has a *negligible* impact on the total market. Yes, all taken together, they have an impact. But individually, they don't.

So what's the point of a single coffee shop taking total market demand into consideration when making its quantity decision? Its contribution to the market is a drop in the bucket. In a competitive scenario, an individual firm doesn't believe its particular quantity decision is going to influence the market price for coffee in the city. *Thus* it takes market prices as "given" and acts *as if* it can't influence them. This is the basic economic definition of "**perfect competition**": namely, that the size of the firm is so small relative to the whole market that firms take prices as given in their decision-making.

Not so with a **monopoly**. The monopolist is the only supplier. It knows that its individual quantity decision *is* the total quantity supplied on the entire market. So it *knows* that if it increases supply, it will drive down market price. It takes that into account.

Diagramatic intuition: TR-TC with monopoly

How does this translate into our diagrams? Well, in a monopoly, the total revenue curve (TR) isn't a straight line anymore. Rather it has a curved shape and looks something like the following:



Notice that that the TR curve now has a *declining* slope. Again, this is because in a monopolistic environment, the firm has a direct *influence* on the market price. More precisely, the more a firm produces, the *lower* the market price for the good (because the market has to clear).

Now, it remains true that, by definition:

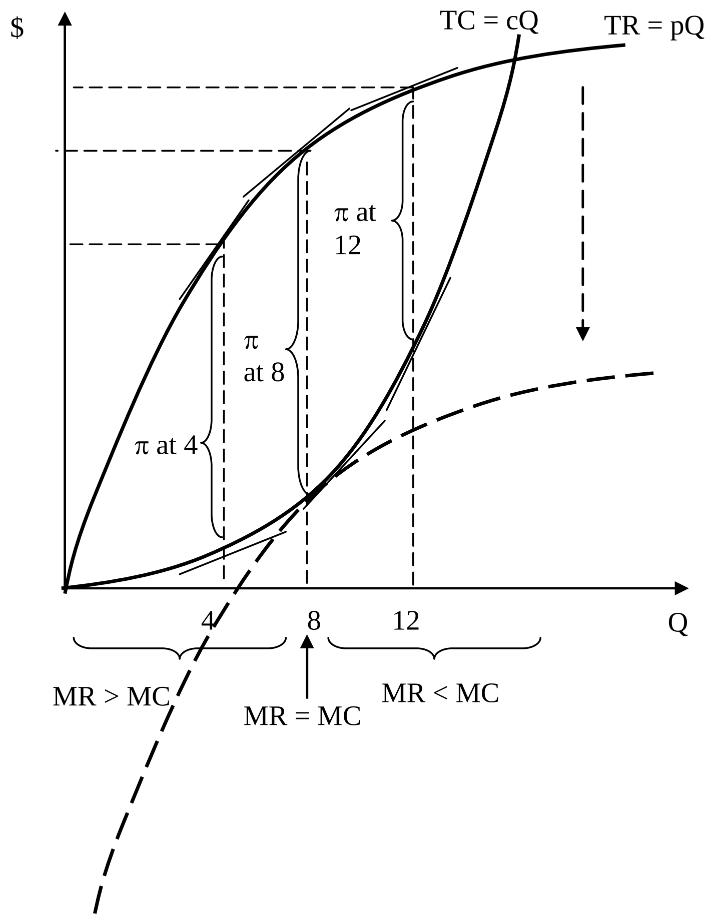
$$TR = pQ$$

what is different is that p is *no longer constant* Rather, price varies with the amount of quantity supplied by the firm (since firm supply = market supply in the case of a monopoly). This depends on the **market demand** curve. Let us use the example drawn in "Verizon's mentality" diagram. That is just a regular demand and supply diagram for the market, but now seen from the monopoly's perspective:

If Q = 5, then market price = \$90
If Q = 10, then market price = \$60
If Q = 15, then market price = \$30

The more the monopoly produces, the *lower* the market price will become. (why? Demand. By increasing its scale of production, the firm is increasing market supply (since all supply on the market is the monopolistic firm's supply decision). And if market supply increases, that necessarily means market price must fall. Just the old Law of Demand & Supply at work.

How about profit-maximization? Well, profit is still the distance between TR and TC. So the firm still wants to find the point of greatest difference. And that point is still found by the rule of tangencies. That is profit is maximized where MR = MC. We can find this diagrammatically by "shifting" the TR curve down until we hit a tangency point (in this case, where Q = 8). This is the *only* point where MR = MC and thus is the profit-maximizing point.



[Notice that everywhere below Q = 8, the TR is steeper than the TC curve, that is, MR > MC. While above Q = 8, TR is flatter than the TC curve, that is MR < MC. Only at Q = 8 do we have it that MR = MC]

Monopoly Solution: An Explicit Example

OK, that diagram is probably not the most intuitive one. Let's turn to a more explicit example.

Deriving MR

Remember, by definition, it is still true that:

$$TR = pQ$$

except p is no longer constant, but itself a function of quantity produced by the monopolist. What kind of function? The shape of a market demand function! After all, diagrammatically, by "increasing supply" we simply "trace out" the demand function.

So let us take a market demand function like the following:

$$Q^d = -4P + 240$$

This is a simple, regular, linear demand function, that expresses the quantity demanded as a function of price. We've seen demand functions like this before:

if price = 0, quantity demanded = 240,
if price = 1 , quantity demanded = 236
if price = 2, quantity demanded = 232
if price = 3, quantity demanded = 228

and so on.

The monopolist (unlike the competitive firm) is aware of this demand function. And thinks of it ***inversely.*** The monopolist *knows* that if market supply is 228, then for the market to clear, price will have to be \$3. It *knows* that if he increases market supply to 232, then price will drop to \$2. It takes this demand function into consideration. And thinks "strategically": *because* the amount the monopolist supplies is going to be total market supply, then the quantity decision will influence the price that clears the market. So how much *should* the monopolist supply?

In this respect, the monopolist's decision is different than a competitive firm's decision. A competitive firm cannot think strategically - it assumes it won't have an impact. But the monopolist knows it has an impact, and thus tries to tailor its decision, taking that into consideration. Mathematically, that means the monopolist looks at the *inverse* of the demand function. To see why, remember that for markets to clear:

$$Q^d = Q^s$$

quantity demanded must equal quantity supplied. Quantity demanded is given by the market demand function, $Q^d = -4P + 240$. Quantity supplied is what the monopolist has to decide. So calling Q the monopolists' decision variable, then the market clearing condition $Q^d = Q^s$ implies that:

$$-4P + 240 = Q$$

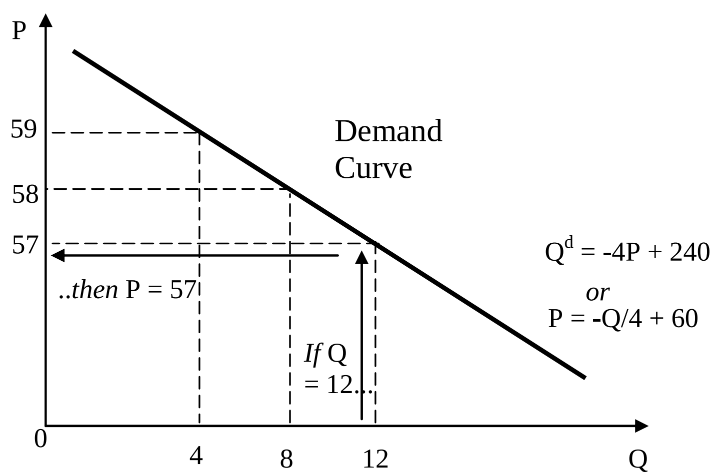
where Q is the quantity the monopolist has to determine. Well, rearranging, to express it inversely, with P as a function of Quantity:

$$P = -Q/4 + 60$$

that's how the monopolist thinks strategically. *If* I, the monopolist, produce quantity x, what will the market-clearing price be? e.g.

if the monopolist produces Q = 4, then P = -4/4 + 60 = -1 + 60 = 59
if the monopolist produces Q = 8, then P = -8/4 + 60 = -2 + 60 = 58
if the monopolist produces Q = 12, then P = -12/4 + 60 = - 3 + 60 = 57

So the monopolist is thinking ahead of the impact its quantity decision will have on price. And it traces that decision by mapping out the (inverted) demand function. That is, it reads the demand curve *inversely* from the normal way we do when we think of markets. Instead of starting from the vertical price axis, it starts from the horizontal quantity axis. It doesn't ask *if* price, *then* quantity supplied, but rather *if* quantity supplied, *then* what will price be?



[Historical Note: And if you ever wondered why economists customarily and unintuitively flip the axis when drawing normal Demand & Supply diagrams, putting price on the vertical and quantity on the horizontal, here is your answer: because economists started analyzing the theory of monopolistic firm. And that meant reading causality the other way, with quantity as the independent variable and price as the dependent. So then they decided to stick with this flipped diagram when talking about markets in other contexts.]

This inverse demand function is our **price function** $p = f(Q)$. It tells the monopolist how market-clearing prices change when the quantity it supplies changes.

What about Total Revenues? Well TR = pQ. But p is no longer a constant ("given"), but rather itself a function of Q. Instead of taking p as a given parameter, the monopolist will substitute the inverse demand function in place of p:

$$TR = (-Q/4 + 60) \times Q$$

or, if you prefer to open the brackets:

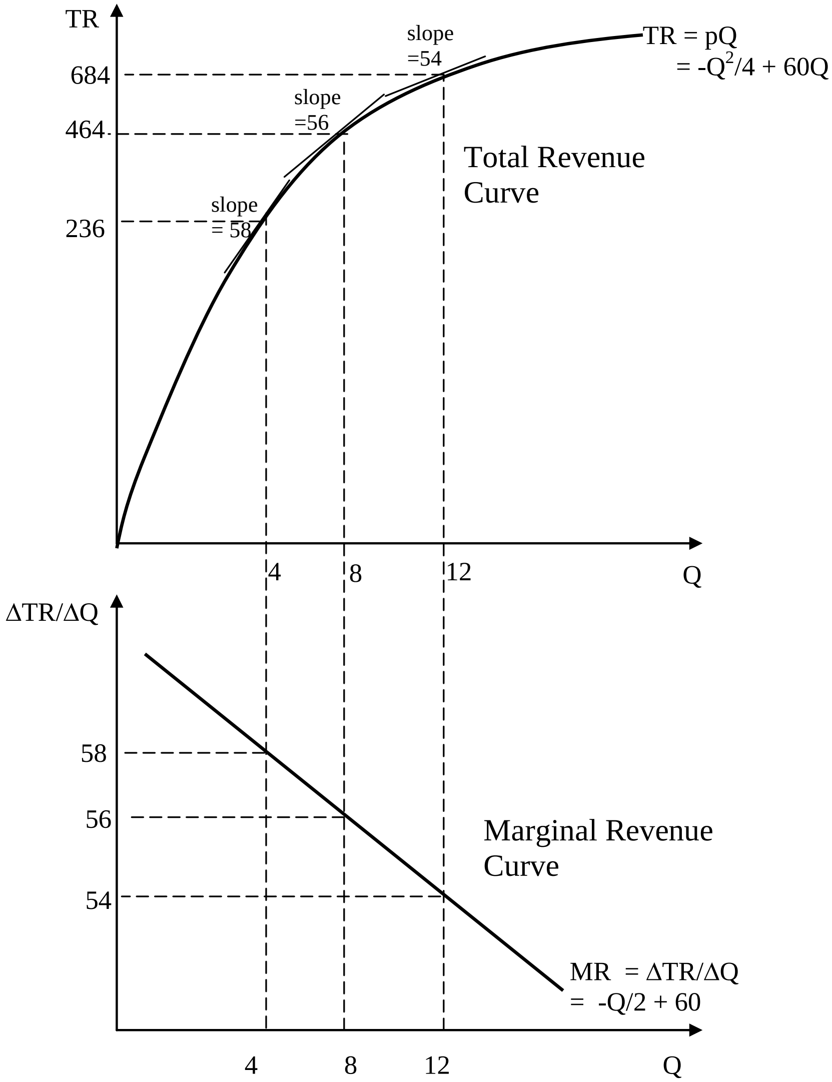
$$TR = - Q^2/4 + 60Q$$

So the total revenue formula is now more complex than before. It is a non-linear function - a concave quadratic function, to be precise.

That looks ugly. The values will be calculated like:

if Q = 4, then TR = -16/4 + 60(4) = -4 + 240 = 236
if Q = 8, then TR = -64/4 + 60(8) = -16 + 480 = 464
if Q = 12, then TR = -144/4 + 60(12) = -36 + 720 = 684

What kind of shape is *that*? Well, if you plot it out, it will yield a **concave**-shaped TR curve. Something like the TR in the diagram below.



The TR is now a concave-shaped curve. Concave means that as the quantity increases, the slope declines. At Q = 4, the slope of the TR curve is very steep, but it is less steep at Q = 8, and becomes increasingly flatter as quantity increases.

Marginal revenue, by definition, is the slope of the Total Revenue curve. We can figure out the precise formula for that easily. The total revenue curve, in our monopoly example, is the ugly:

$$TR = -Q^2/4 + 60Q$$

So MR is simply the derivative of that. Mathematically, using our power rule:

$$MR = \Delta TR/\Delta Q = -2Q/4 + 60 = -Q/2 + 60$$

Notice that marginal revenue is no longer simply "price", but rather a function of quantity produced. What kind of function? A downward-sloping function, as we see in the diagram.

when $Q = 4$, then slope $MR = -4/2 + 60 = -2 + 60 = 58$
when $Q = 8$, then slope $MR = -8/2 + 60 = -4 + 60 = 56$
when $Q = 12$, then slope $MR = -12/2 + 60 = -6 + 60 = 54$

So now all we need is marginal cost to figure out our solution.

The Monopolist's Decision: MR = MC

e.g. suppose we have the following functions:

$TR = - Q^2/4 + 60Q$
$TC = (2.75)Q^2$

(sorry for the decimals, it's rather hard to come up with neat examples that yield round numbers)

calculating the marginal revenue and marginal cost, as we have seen:

$MR = \Delta TR/\Delta Q = -Q/2 + 60$
$MC = \Delta TC/\Delta Q = 2(2.75)Q = 5.5Q$

Since for profit-maximization, it must be that:

$$MR = MC$$

then:

$$-Q/2 + 60 = 5.5Q$$

rearranging:

$$5,5Q + Q/2 = 60$$

or:

$$(5.5 + 1/2)Q = 60$$

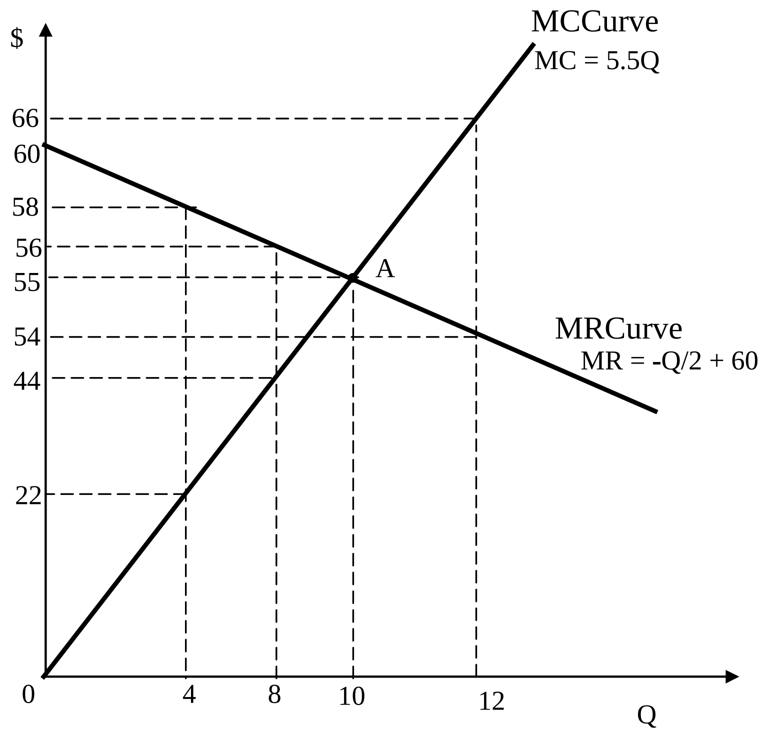
$$(6)Q = 60$$

$$Q = 60/6 = 10$$

So Q = 10 is the profit-maximizing level of output

Having the marginal revenue curve at hand makes life easier than trying to "eyeball" TR & TC diagrams and finding tangencies. That is because we know the profit-maximizing solution will be where MR = MC. So just plot the downward-sloping MR on top of the upward-sloping MC, and their intersection marks the spot where MR = MC.

If you plot it out, it will look something like this: (not perfectly drawn)



Or if you want to verify the points:

Quantity	MC = 5.5Q	MR = -Q/2 + 60
0	0	60
4	22	58
8	44	56
10	55	55
12	66	54

Point A in the diagram marks the intersection of MR & MC, and thus the profit-maximizing quantity Q = 10.

Now comes the tricky bit. We've solved for the quantity, but what is the *price* that consumers will be charged?

For this you need to plug the quantity Q = 10 back into the inverse *demand function* (***not*** the MR function). Remember we stated this whole thing off with the demand function function, $Q^d = -4P + 240$. When we inverted we saw the inverse demand function:

$$P = -Q/4 + 60$$

So to find the price, just plug in Q = 10, so:

$$P = -(10)/4 + 60 = -2.5 + 60 = 57.5$$

The market price is P = 57.5.

Step back for a moment. At quantity Q = 10, we saw the marginal revenue (= marginal cost) was 55, but the *price* is 57.5. In other words, P > MR (= MC).

So we have it here that price exceeds marginal cost! This is *always* true in a monopolist case and stands in stark contrast to the competitive scenario where we found that it was always P = MC. So the profit-maximizing solution is qualitatively and quantitatively *different* in the monopoly scenario in contrast to the competitive scenario.

Step back for a moment. Look at the MR & MC diagram. What does it look like? Demand & Supply diagram, doesn't it? Remember, we saw before that the MC curve was truly the firm's supply curve. And that sort of remains true. The temptation is to make a similar analogy for the MR curve, and call that some sort of demand curve, right?

Wrong. The MR curve is *not* the demand curve. The demand curve is different. The inverse demand function has shape:

$$P = -Q/4 + 60$$

while the marginal revenue curve has shape:

$$MR = -Q/2 + 60$$

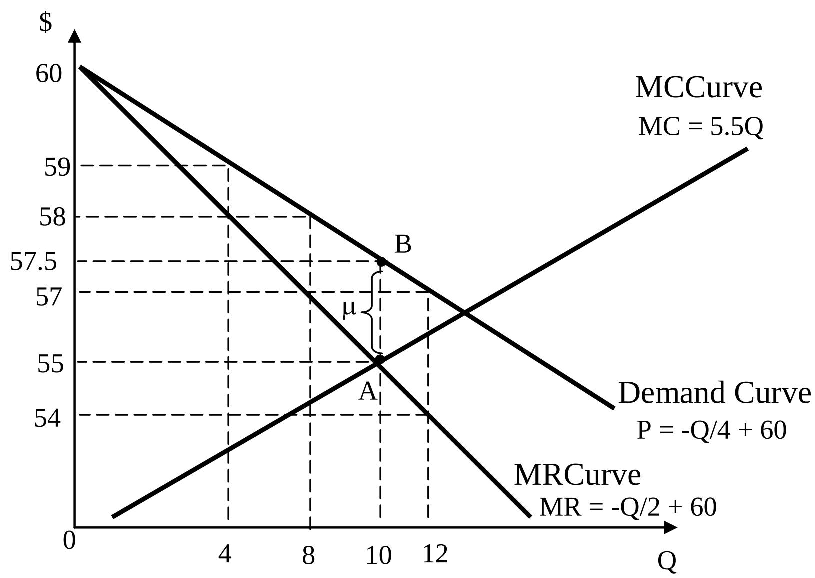
from a glance, the inverse demand function *looks* similar to MR curve. They both have the same intercept (60), and they both are downward sloping with respect to quantity. But they have *different slopes*! The MR declines at the rate -1/2, while the inverse demand curve declines at the rate -1/4. The MR curve thus declines at a *steeper* gradient than the demand curve.

If we were to tabulate some of the points:

Quantity	Marginal Cost MC = 5.5Q	Marginal Revenue MR = -Q/2 + 60	Demand Price P = -Q/4 + 60
0	0	60	60
4	22	58	59

8	44	56	58
10	55	55	57.5
12	66	54	57

we can see immediately that MR drops at a faster rate (60, 58, 56, etc.) than demand price (60, 59, 58, etc.). Diagramatically, introducing the demand function into the MR & MC diagram, it would look like the following:



(the MC curve is horribly drawn - it is in reality much steeper than that. But we want to focus on the different slopes of the MR & Demand curves).

The solution quantity will be point A - that is the profit-maximizing decision, where MR = MC = 55. But the solution *price* on the market - the price consumers face - will be on the demand curve, at point B, where P = 57.5.

We have denoted the gap between the point A (MR = MC) and point B (the price) by the Greek letter **μ** (pronounced "mu"). This is what is commonly called the "**monopoly mark-up**", the amount by which the price exceeds the marginal cost, i.e.

$$P = MC + \mu$$

In our example, $\mu = 2.5$. That is:

$$57.5 = 55 + 2.5$$

In sum, under a monopoly scenario, Q = 10, P = 57.5 and the mark-up is 2.5 (i.e. MC = 55).

Comparing Monopoly vs. Competitive Solutions

Can we see exactly how exactly this compares to the competitive market equilibrium exactly? Well, we know the competitive solution will be where P = MC. We already have the formula for MC = 5.5Q. Our price formula (from inverse demand function) is P = -Q/4 + 60. So we know that in equilibrium:

$$-Q/4 + 60 = 5.5Q$$

or:

$$(5.5 + 1/4)Q = 60$$

since 5.5 + 1/4 = 5.75, then:

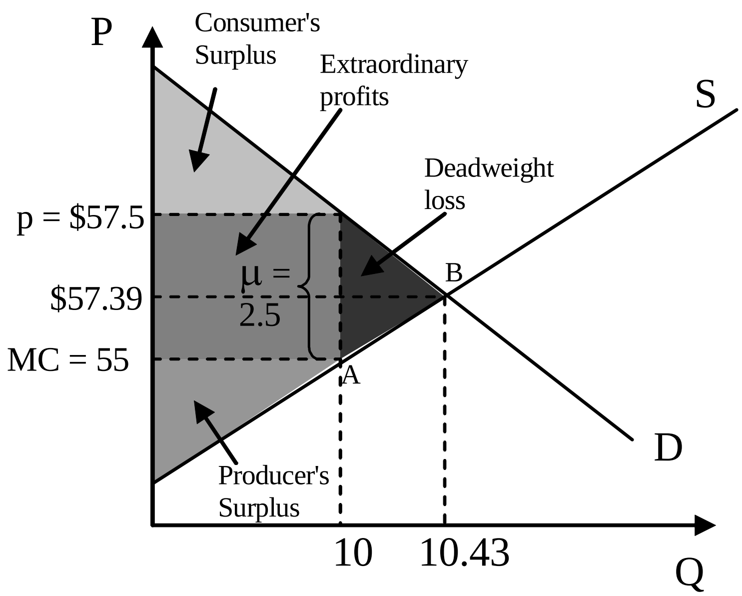
$$Q = 60/5.75 = 10.43.$$

So if we had a competitive market scenario with the same underlying cost & demand, the quantity produced on the market would be 10.43 (greater than the quantity produced under a monopoly, 10). The price consumers would face would be:

$$P = -(10.43)/4 + 60 = -2.61 + 60 = 57.39$$

which is *lower* than the price they face under the monopoly scenario (57.5).

OK, the numbers don't seem that dramatically different (10 vs. 10.43 and 57.5 vs. 57.39). But the general point stands: under a monopoly scenario, the equilibrium quantity is *less* and the price consumers face is *higher* than in a competitive scenario. Diagramatically, we could draw it comparatively as follows:



Notice the area called "Extraordinary Profits". These are the profits reaped by the monopolist simply by virtue of being a monopolist. It is a rectangle of length 10 and height 2.5, so extraordinary profits are 2.5 × 10 = $250.

(No, I'm not going to ask you to calculate the deadweight loss - although it can be done!)

General Formula

Remember that TR = pQ by definition. And by the definition of marginal revenue:

$$MR = \Delta TR/\Delta Q$$

In a competitive scenario, price was "given" by the market, it was a constant, so just applying the constant rule:

$$MR = \Delta TR/\Delta Q = p$$

marginal revenue is necessarily equal to price.

However, in a monopolistic scenario, p is *not* "given", it not no longer a constant, but rather varies with the quantity produced, so we can write the TR formula as:

$$TR = p(Q) \times Q$$

where p(.) is itself a formula, denoting that price is a decreasing function of Q (just by the downward-sloping shape of the demand curve). When we had an explicit demand formula, we obtained an explicit TR formula (e.g. $TR = -Q^2/4 + 60Q$) and figuring out the MR was easy. But let's do it more generally. Remember, now we have Q entering twice into the TR formula - directly and by affecting price. Figuring out the marginal revenue requires a little more calculus - specifically the "product rule", which states: that when we have a compound like we have, the derivative of the whole can be figured out by rule "derivative of the first component times the original second plus derivative of the second component times the original first.". In the formula $TR = p(Q) \times Q$, the first component is p(Q) and the second component is Q. So derivative of the first component is:

$$\Delta p(Q)/\Delta Q = \text{slope of demand curve} = \Delta p/\Delta Q$$

Since we don't have an explicit formula for the slope of the demand cure, let us just denote it by $\Delta p/\Delta Q$ simply. The derivative of the second component is simple:

$$\Delta Q/\Delta Q = 1$$

So now we're ready to use use the calculus rule:

$$\text{derivative of the first times the original second} = \Delta p/\Delta Q \times Q$$

$$\text{derivative of the second times the original first} = 1 \times p.$$

So the derivative of the whole:

$$MR = \Delta TR/\Delta Q = (\Delta p/\Delta Q \times Q) + (1 \times p)$$

or:

$$MR = \Delta TR/\Delta Q = (\Delta p/\Delta Q)\times Q + p$$

Yuk.

The important thing is to notice is that MR is *no longer* equal to p, but rather p *plus* some ugly term $(\Delta p/\Delta Q)\times Q$. This expression contains the expression $\Delta p/\Delta Q$, that is the change in price that results from an increase in quantity - i.e. the demand relationship. But we know that inverse demand curve is downward sloping: if quantity rises, prices fall. So, $\Delta p/\Delta Q$ is a *negative* number. So this is not price *plus* something, but rather price *minus* something. Let us call this something μ ("mu"), and refer to is as the "**mark-up**". Specifically, let:

$$\mu = -(\Delta p/\Delta Q)\times Q.$$

So plugging in:

$MR = \Delta TR/\Delta Q = p - \mu$

Now, let us go back to the generalized profit-maximization rule. That tells us MR = MC, or:

$p - \mu = MC$

Or, rearranging:

$p = MC + \mu$

The profit-maximizing decision will be where price (p) is marginal cost (MC) plus the mark-up (μ).

So, under a monopoly, the price will be *greater* than marginal cost by the amount of the mark-up (μ). That is, $p > MC$. You can think of this mark-up is the excess profit it makes per unit sold.

Notice that by the formula, $\mu = -(\Delta p/\Delta Q)\times Q$, indicates that the exact *size* of the mark-up depends on ($\Delta p/\Delta Q$), which is, remember, the *slope* of the inverse demand curve. The *steeper* the slope (i.e. the more negative ($\Delta p/\Delta Q$)), the *greater* the mark-up. So the degree of excess profits a monopoly reaps - and the degree of hurt it imposes on the economy - depends on the elasticity of the demand curve. If the demand is very elastic (that is, relatively flat), the mark-up μ will be small. If demand is very inelastic (that is, relatively steep), the mark-up will be very big.

Remember: goods with good substitutes (e.g. coffee, bananas) will generally be quite elastic, whereas goods with poor substitutes (e.g. transportation, water, energy) tend to be inelastic. That is why government get more nervous when they see monopolies emerge in inelastic sectors like transportation, petroleum and utilities, and rush to regulate them, while they might be more tolerant of monopolies emerging in more elastic sectors, .

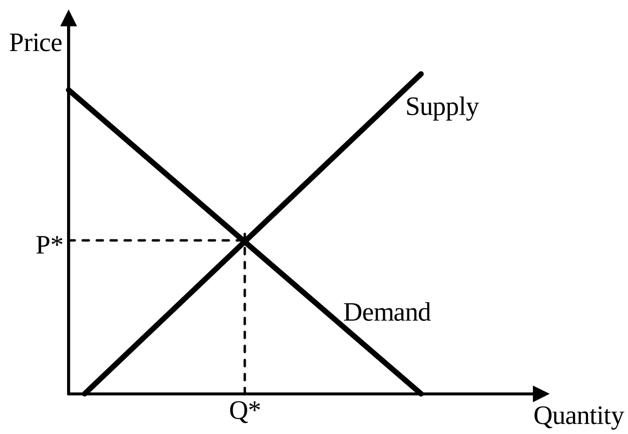
POSTSCRIPT:

THE FIRM IN ECONOMIC THEORY

The theory of the firm we have been outlining here was originally proposed as far back as the 1890s by English economist Alfred Marshall, and was developed and elaborated upon by other economists through the course of the 20th C.

Assumptions of Economic Theory

The whole pain of this exercise is partly to drive home one enormously important point. Namely, that every time you draw this picture,



every time you shift the curves around, every time you discuss how this, that and the other raise or reduce prices or change quantities on the market, and so on, you are making several enormously important ***assumptions*** about the organization of economic society.

Whenever you see draw or, play with that nice, upward sloping supply curve and talk about the equilibrium:

(1) You are assuming ***capitalism***. That is, a society where the decisions of productive entities (i.e. firms) are governed by the owners of *capital*. Because if you were not in a capitalist system, then firms would not necessarily seek to *maximize profits* as the aim of their supply decisions. And if they don't seek to maximize profits, then we can't say they are following the MR = MC rule. And if they are not following MR = MC rule we don't know what their supply decision is based on. It could be to maximize size, or to flatter the entrepreneur's ego, or to meet consumer needs, or to meet welfare goals of workers, or some other personal or social objectives. But the decision will not be MR = MC, and you will not be able to derive the supply curve from that.

(2) You are assuming ***scarcity*** or ***limited*** factor resources. That is, that the available productive resources - land, labor, capital - that exist in the economy are *limited*, that if

you try to increase your production, you will necessarily have to bid factor prices *up*, pay more for those resources, and raise the costs of hiring them against yourself. If resources were not limited, then the *Law of Increasing Costs* would not impose itself. And if the law of increasing costs did *not* impose itself, then the supply decision when firms follow the profit-maximizing rule - MR = MC - will not have a mathematically-determinate solution. The solution is infinite, the supply curve would be completely flat, or something like that. You won't be able to draw the simple S & D diagram and talk about how prices and quantities are determined by it. You'll have to come up with a different theory of price.

(3) You are assuming ***competition***, that is, a market environment where there are many firms producing the same good. If you were in a monopoly, then the supply decision will not be governed simply by the supply curve, the solution will not be *on* the intersection of the demand & supply curve, but *off* it. You can still say there is a determinate solution, but it is no longer simply where "X" marks the spot. Quantities will be below equilibrium, prices above equilibrium, and calculating exactly where will be a whole lot more painful.

You may raise your hands in protest and note that "in the Real World" these assumptions are not really "True" - the system is not really just pure focused capitalism (other kinds of firms exist, capital owners are often neglectful, decisions are often quite muddle-headed, lazy and vain), resources are not really so dramatically constrained (there are technical improvements, new ways to overcome increasing costs, and so on), competition is not really that perfect (firms don't just take pries as given, they have some little degree of market power, at least locally).

All that may be true, all those objections have merit. But it doesn't change the fact that *everytime you draw that diagram* you are implicitly assuming away those objections.

How scandalous is it? This depends on you attitude towards scandals. By and large, it is not too shocking. Wherever you look, you will find the system is largely capitalist, the law of increasing costs does tend to impose itself, and competition does tend to exist in most markets. Not purely, not always and not perfectly. But it is roughly true. And as such, Supply & Demand diagrams remain *useful* in showing how the forces of the market work and the direction things move when this or that happens.

Don't make a fetish of out of the diagram. It is just a rough approximation of how markets work and tends to be a decent one at that. But you should always be aware of the assumptions underlying it.

Scandal Uncovered: The Sraffa Critique

(Don't read this in front of children nor bring it up in polite conversation)

I have waved my hands and told you there is nothing to worry about. And you may go on assuming that everything is approximately fine. But there are some significant and shocking scandals in the theory we have been presenting that perhaps you ought to be aware of.

The scandal is simply this - two assumptions seem to contradict each other: the law of increasing cost is not really compatible with competition.

The law of increasing cost tells you that the more a firm produces, the more it raises factor prices (wages, rents, etc.) against itself. Ergo, increasing cost.

The assumption of competition tells you that firms take output prices as "given" by the market, and do not (or believe they cannot) affect market prices.

The contradiction should be evident. Competition *assumes* firms are very small relative to the market - a single firm is such a small operation that it doesn't affect prices in the output market. But at the same time, the Law of Increasing Cost implicitly *assumes* a firm is very big relative to the market - so big, that the hiring decisions of that same firm will raise wages throughout the economy!

How can the same firm have such huge power in the market for its inputs, but be so hopelessly powerless in market for its output?

So which is it? Does the firm have price-changing market power or doesn't it?
.
It is amazing how often this story is told and taught and how often this contradiction isn't spotted. It seems rather obvious, once it is pointed out. But it very rare that someone notices it on their own. Generations of students continue to be taught this theory without noticing it (and their professors are not keen to point it out.)

The contradiction was first noticed in 1926, in an article by an Italian economist Piero Sraffa, and thus is known as "**Sraffa's Critique**" (although really, just the first of several contradictions and inconsistencies Sraffa went on to point out about Neoclassical economics; the man was a troublemaker.)

How do we get out of it? We can't. There simply isn't any easy hand-waving we can do here. The theory of the firm has this really troubling, inherent, internal flaw.

Solving the Scandal

> "I am trying to find what are the assumptions implicit in Marshall's theory; if Mr. Robertson regards them as extremely unreal, I sympathize with him. We seem to be agreed that the theory cannot be interpreted in a way which makes it logically self-consistent and, at the same time, reconciles it with the facts it sets out to explain. Mr. Robertson's remedy is to discard mathematics, and he suggests that my method is to discard the facts; perhaps I ought to have explained that, in the circumstances, I think it is Marshall's theory that should be discarded."
>
> (Piero Sraffa, 1930, "Symposium", *Economic Journal*, p.93)

Sraffa himself saw only two solutions to the 'scandal' he had uncovered: (i) dump the assumption of perfect competition; (ii) dump the assumption of increasing cost.

(1) "Imperfect Competition" Solution Dump the competition assumption. That means, dump the price-taking theory of the firm as we've presented it. Assume firms have market power both on the output and input side. No more simple demand-and-supply diagrams for you. You have to work through the pain of more complicated MR = MC derivations (as we did for the monopoly case). You don't have to assume outright monopoly itself, but you have to allow a degree of market power. e.g. one of Sraffa's students, Joan Robinson, constructed a new theory of the firm in the 1930s she called *imperfect competition*, an intermediary theory which showed how to model firms which are not quite monopolies but not just price-taking either. The downside, is that analyzing markets with imperfect competition is quite uglier and more difficult than the easy simplicity of simple D & S diagrams.

(2) "General Equilibrium" Solution Dump the law of increasing cost. Have perfect competition in both the output and input sides. That is, assume that firms take output prices as given by output markets (as we have assumed), but *also* that they take input prices as "given" by input markets. Firms have no market power anywhere, they have no capacity to influence their own costs themselves.

This solution is the one favored by modern Neoclassical economists, since it 'saves' Demand & Supply diagrams. It is also more subtle to interpret. So we need to be clear about they are saying.

The GE solution is saying the law of increasing cost does not apply to a *single* firm, but it still applies to the economy *as a whole*. That is, we can continue to draw demand & supply curves as we normally do - that is, for the *market*. But we cannot "derive" or "explain" market supply curve by resorting to the theory of a single firm and then adding them up.

Rather, we must explain the foundations of the supply curve based on a much more ambitious theory of "general equilibrium", a treatment that takes *all* firms and *all* sectors into consideration simultaneously.

Or put another way, the law of increasing cost still works, but only in the economy as a whole. When we take demand for labor by *all* firms from *all* sectors together, bearing down on the limited supply of labor, yes there is "increasing cost" and market supply functions will be upward sloping. But not if only a single firm increases output, and the other firms don't. An individual firm is a teeny, itsy bitsy entity which has no market power by itself, but only when taken together with all other firms in the economy.

The consequence of assuming perfect competition on the input side is that there is no rising marginal cost for a firm, and as a result the firm's MC and thus the firm's supply curve is *flat*. Since the MR curve is also flat (by the assumption of perfect competition on the output side), that means the only way MR = MC is if the "flat" MC curve lies right on top of the "flat" MR curve, which means MR = MC at *every level* of output. The quantity decision of the single firm is indeterminate - it can be anywhere, everywhere. All quantities are profit-maximizing.

This should be shocking. Hello? What? No law of increasing cost? Then the firm decision has no solution! Yes, exactly. The firm decision has no solution. So what? Why should we care what the individual solution of the firm is? What we care about is wages and prices, and wages are determined by the labor market *as a whole* and prices by the output market *as a whole*, not by any particular firm.

So, stop bothering yourself about the firm. Ignore it. The firm isn't all that interesting. It just a technical conduit, an institutional production arrangement, just a tool that turns inputs into outputs. The details of how it goes about making its decisions don't really matter. The GE solution says that what is of interest in the linkage between input markets and output markets, how demand for *all goods* in the economy *as a whole* bear down on the supply of factors in the economy *as a whole*. Don't bother analyzing a firm in isolation. You should analyze it simultaneously with all other firms, in all other sectors, taken together *at once*. Then, yes, you will have the law of increasing cost - but only on the whole, not individually.

As you can imagine, the mathematics of analyzing "general equilibrium" - that is, all the output markets for apples, oranges, computers, stereos, coffee, tea, jackets, shirts, wool as well as all the factor markets, labor, land, ovens, welding machines, altogether, all at once, finding *all* their prices simultaneously - is an *immensely* complicated mathematical feat.. And so we shall not be analyzing it here. But rest assured it can be done (with a lot of pain).

But mathematics aside, what is the intuitive message of "general equilibrium" approach? One economist tried to put it colorfully:

> "The firm fits into general equilibrium theory as a balloon fits into an envelope: flattened out! Try with a blown-up balloon: the envelope may tear, or fly away: at best, it will be hard to seal and impossible to mail....Instead, burst the balloon flat,

> and everything becomes easy. Similarly with the firm and general equilibrium - though the analogy requires a word of explanation."
>
> (Jacques H. Drèze, "Uncertainty and the Firm in General Equilibrium Theory", *Economic Journal*, 1985, p.1)

(Frankly, I have no idea where he was going with that analogy....)

Let's try it another way. It goes something like the following: a firm is an entity that demands factors and supplies output, right? But is that decision, by itself, really that interesting? After all, a firm demands factors from factor owners (that is workers, landlords, etc.). And it supplies output to consumers. But who is really making the decision? Who is the human here with the power of subtlety and decision? Firms are programmed to "maximize profits". That is more like a technical command than an independent, human decision. The firm's "opinion" doesn't really matter. It is just the application of a technical rule ("Lo! I command! Maximize Profits!"). A computer can be programmed to make the "decision", there's no human factor. Once you know prices and technology, maximizing profits is a mathematical problem that can be done by an algorithm (or in a homework exercise). So why are we fussing so much over it?

The interesting part, the subtle part, the human part, the part which really requires careful "decision-making" is not the firm, but (1) the demand for goods by consumers and (2) the supply of factors by laborers, capitalists, etc. Consumers demand coffee, tea, computers, etc. and they made that decision of how much to demand of each good carefully with a whole host of considerations involved. Similarly, workers have also a subtle decision to make - how many hours of leisure versus how many hours of work do I want to put in? How do I balance my desire for money to buy nifty things with the amount of time I have to enjoy them? These are the decisions GE theorists want to emphasize, these are the *real* decisions of the economy, the ones that really matter. The firm's decision is blind and bland by comparison.

I can't make the "flattened balloon" analogy work, so I prefer another one: the firm is like a toaster. Toasters make toast out of plain bread. Who makes the decision of how much toast to make? Sure as heck not the toaster itself! It it just a machine, a tool, with no capacity for independent thought. *You* order it to make toast, it makes toast. But the decision of *how much* toast to make is made by *you*. And *you* make that decision in light of your desire for toast, counterbalanced by the amount of soft bread you have available and your desire to retain some soft bread for finger sandwiches. The toaster is just an intermediary tool you *utilize* when you're balancing your decision between toast and soft bread. So what is the ultimate determinant of how much toast is made is not the toaster's decision, but (1) you, the consumer's, taste for toast (relative to soft bread); (2) the amount of bread you got to begin with (supply of resources). The toaster itself is really just "technology" to enable you to conduct that swap of toast for bread. But it doesn't, itself, determine *how much* toast to make.

General equilibrium theory treats the firm like a toaster. Assume perfect competition on both sides. It takes output prices as given *and* it takes input prices as given and order it to profit-maximize. Yes, the firm's decision is "indeterminate". As it should be. How much toast the toaster "decides" to make is also indeterminate. What determines the amount of goods produced in the market is the demand for computers, jackets and coffee as whole, bearing down on the supply of resources (labor and land) *as a whole*. That is, consumers decision of how much output to buy and the factor owners' decisions of how much factors to sell. The firm is just an intermediary tool between the consumer's demand for goods and the laborer's supply of factors. But the firm's decision in itself is simply not interesting.

Another way general equilibrium theorists like to get their point across is to emphasize that production is a just an "**indirect form of exchange**".

Think about it this way. Consumers consume jackets, workers make jackets. Firms are just a way of connecting these two. The way we have dealt with it up to now is to deal separately with two markets: one the one hand, we have the output market, where we have exchange between demanders for jackets (consumers) and suppliers of jackets (firms), and then, in another diagram, we have the factor market, where we have exchange between demanders of labor (firms) and suppliers of labor (workers). It looks like its two markets with two exchanges. But it is really only one. Just cut out the middleman, the firm, from the analysis. It's just a distraction. The real exchange is ultimately between consumers of jackets and suppliers of labor. Firms are just a mechanical step in between, that converts labor into jackets, it is what allows laborers and demanders of jackets to exchange *with each other*.

So the real exchange going on in the economy is between one human (who demands jackets) and another human (who supplies labor). "Production" is just a faceless technical intermediary that allows this exchange act between consenting human adults to happen.

Caveat:

Treating firms as simplified "toasters" is not very helpful if you want to analyze "real world", that is, if you want to examine case studies of corporations, monopolies, externalities, regulations and all that. To analyze these 'real world' questions, you do need to pay attention to how firms make decisions, which means modelling firms as a decision-making, living, breathing entity ("blown-up balloon"?), not merely as a mechanical toaster ("flattened balloon"?).

So general equilibrium (GE) theory, while beautiful and logically coherent, is not very practical or helpful - it has a terrible time accounting for real world questions where answers are required. It is for this reason that I bothered to teach you the old "Marshallian" theory of the firm. Despite its internal logical contradiction, it is at least "useful", in the sense of being able to help us analyze "practical" cases and make some predictions and suggestions. And for that reason, it is still taught and used.

The conclusion? Not a happy one. We are faced with the choice between picking a theory that is logically coherent, or a theory that is practically useful. I'd love to tell you that economics has figured it all out, and can give you something both coherent *and* useful. But it hasn't. And we just have to live with that.