AP Microeconomics Readings

1.2, 1.6 - Marginal Analysis and Consumer Choice

AP Macro and AP Micro have the exact same Unit 1, dealing with scarcity, tradeoffs, PPCs, and comparative advantage, and since we covered those things last semester, we're not going to do so again. There are, however, two things in AP Micro's Unit 1 that Macro does not cover, which we're going to go over today.

The first is that there are different systems for allocating resources. In a **command economy**, the government determines who produces and who consumes what. In a **market economy**, we let the free market determine levels of production and consumption. For a market economy to work, we need a well-defined system of **property rights**. Almost every economy (including the US) is a mix of these two ideas, and we call this a **mixed economy**. Easy.

The second new concept from Unit 1 is a bit more involved, and it comes from the main difference between Macro and Micro. In Macro, we looked at the economy as a whole, whereas Micro is much more interested in individuals and firms and looking at how they make decisions. Unit 1 lays the groundwork for how individuals make decisions, and the rest of the course will look at how firms make decisions.

The main idea is that every item we consume gives us a certain amount of joy / happiness / usefulness, which we call **utility**, and the unit for measuring utility is a **util** (it's ridiculous, I know, but just bear with me). So I could say that eating an ice cream cone gives me 4 utils of happiness.

If I were to eat a second ice cream cone, I wouldn't get nearly as much utility out of it, because we all experience **diminishing marginal utility**. In other words, each item I consume gives me progressively less happiness, so while the first cone gave me 4 utils, the second would only give me 2, while the tenth would make me sick, giving me negative utility.

When I'm at the store, I'm always trying to maximize my total utility with the budget I have. To do that, I have to figure out the marginal utility of each item I'm choosing between so that I know what combination of stuff will give me the most happiness. So in consumer choice problems, your first step is always to figure out the marginal utility of every item you're choosing between. Let's say I'm choosing between buying watermelons and oranges. You'd be given a chart with the total utility derived from each good, and you'd need to figure out the marginal utility by figuring out the difference between each level. In the chart below, you'd be given the information in black, and you'd have to figure out the information in red.

(chart is on next page)

Watermelon			Oranges		
Quantity	Total Utility	Marginal Utility	Quantity	Total Utility	Marginal Utility
1	20	20	1	15	15
2	35	15	2	27	12
3	45	10	3	36	9
4	50	5	4	42	6

If I'm trying to maximize *total utility*, I do that by always consuming the combination of goods that gives me the highest *marginal utility* (MU) at each step of the process. So if I can only buy one good, I'm going to get a watermelon, since MU=20. If I can buy a second good, I'm going to be indifferent between buying either a watermelon or an orange, since MU at these levels is the same (15).

But there's also one last catch: every dollar I spend is painful. Yes, watermelon gives me comparatively more joy than oranges, but if watermelons are $500 each, then I'm going to settle for all oranges. So we have to take into account what each item costs. When consumers are deciding what to buy, they always go for whatever option gives them <u>the most joy per dollar spent</u>. In other words, we have to find the marginal utility of each item and then divide it by the price of that item.

What makes these problems tricky is that the test question isn't going to tell you that you have to find the marginal utility per dollar; they're not going to give you a blank chart to fill out. Instead, they're only going to give you this:

Watermelon (Price = $5)		**Oranges (Price = $3)**	
Quantity	Total Utility	Quantity	Total Utility
1	20	1	15
2	35	2	27
3	45	3	36
4	50	4	42

And you have to know that you need to figure out this:

Watermelon (Price = $5)				Oranges (Price = $3)			
Quantity	Total Utility	Marginal Utility (MU)	Marginal Utility / Price (MU/P)	Quantity	Total Utility	Marginal Utility (MU)	Marginal Utility / Price (MU/P)
1	20	20	4	1	15	15	5
2	35	15	3	2	27	12	4
3	45	10	2	3	36	9	3
4	50	5	1	4	42	6	2

Now let's say that you're given a budget of $19 and asked to figure out how you're going to spend it. To solve this, you always buy whatever gives you the highest MU/P until your budget is gone. So here's what you're going to do:

1. Buy an orange (MU/P=5). Since this cost $3, you now have $16 left.
2. Buy either a second orange or a watermelon. MU/P is the same for both (4), so you're indifferent between the two options and can choose either one. Let's buy a second orange just for fun. Since this cost $3, you now have $13 left.
3. Buy the first watermelon (MU/P=4). Since this cost $5, you now have $8 left.
4. Buy either a second watermelon or a third orange. MU/P is the same for both (3). And as luck would have it, we have exactly $8 left in our budget, so we can afford to buy both, leaving us with $0 left. So the solution is that we are going to buy 2 watermelons and 3 oranges.

So when you get these questions, always 1) set up a chart where you can write MU and MU/P at each quantity, and 2) spend your money on the items with the highest MU/P until your budget is gone.

A corollary to this idea is that *consumers will always buy a combination of goods such that their marginal utility per dollar spent is equal between the two goods*. After all, it would make no sense to buy one orange and five watermelons; MU/P for that first orange is 5, while MU/P for the fifth watermelon is 1. If I had that combination of goods in my shopping cart my total utility would be 65 (50 from the watermelons and 15 from the orange), but I could put a watermelon back and buy an orange instead, giving me a greater total utility of 72 (45 from the watermelons and 27 from the oranges)—giving me more happiness for less money spent.

If we call the two goods Good X and Good Y, then the formula for this idea is:

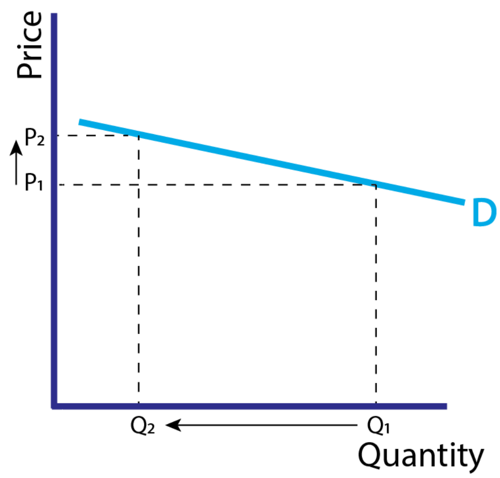
$$\frac{MU_X}{P_X} = \frac{MU_Y}{P_Y}$$

On free-response questions, you'll often see it ask something to the effect of, "Mary buys 5 watermelons and 2 oranges. Explain why this is not an optimal combination of goods, using marginal analysis." In this kind of question, you just have to use this formula and explain that since the two sides are not equal, she should buy more of the good with a higher MU/P and less of the good with the lower MU/P.

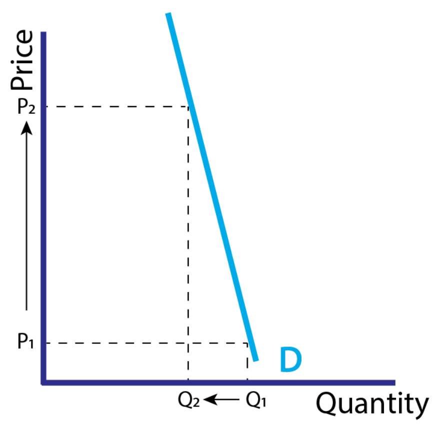
Return to the top of the document

2.1–2.3 Price Elasticity of Demand Part 1

For a firm, one of the most important things to know about the demand curve in their market is how responsive that curve is to changes in price. If stores raise their prices by $10, will they lose all their customers? Or will customers grumble and complain but still buy just as much as they did before? We call this responsiveness "elasticity" (like if we stretch out an elastic band, we want to know how responsive it is to pressure).

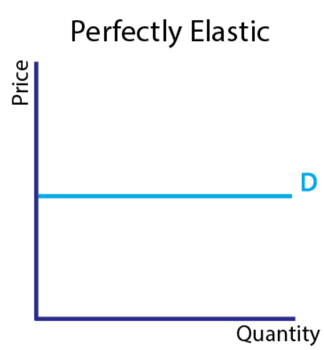


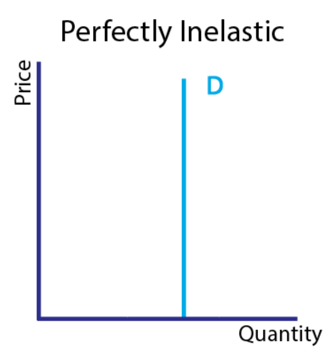
To see how this works, let's look at the market for Twix bars. If you raised the price on Twix bars even a little, the quantity demanded would decrease a <u>lot</u>. Some die-hard fans would still buy them, but most people find Twix to be just as good as Snickers or Skittles, so if Twix went up in price, they'd just switch to buying those other candies. The graph for the market for Twix would be on the right, with a demand curve that is very flat. We would say that this demand curve is very **price elastic**, because consumer behavior is very responsive to changes in price.



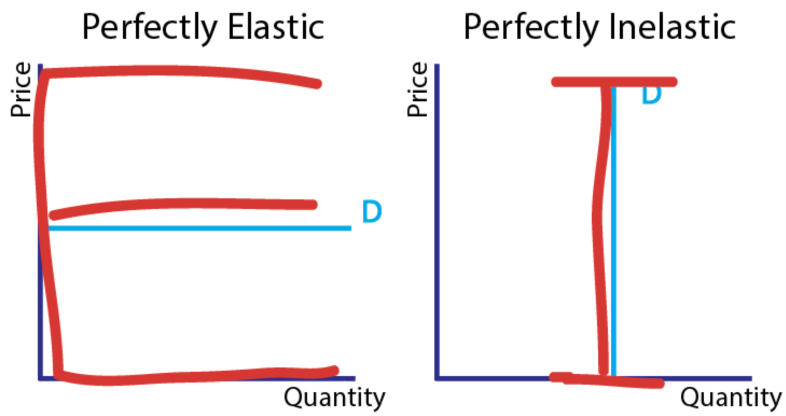
On the other hand, consider the market for insulin. People with diabetes need insulin to live, so I could raise the price of insulin a lot, and I know that people would buy about the same amount they were buying before. The demand curve for insulin would look like the graph on the left, and we would say that this is very **price inelastic**, since consumer behavior is not very responsive to changes in price.

A good that is **perfectly elastic** would be something that people would stop buying entirely if the price increased by even $0.01, and the demand curve would be completely horizontal. A good that is **perfectly inelastic** would be something that people will buy the exact same amount of no matter how much prices change, and that demand curve would be completely vertical.





Thankfully this is easy to remember because the graph for a perfectly **<u>E</u>**lastic good looks like an upper-case E, while the graph for a perfectly **<u>I</u>**nelastic good looks like an upper-case I.



There are two ways to determine whether demand for a good is elastic or inelastic. The first is a formula, which basically just puts what we've been talking about into mathematical terms. The formula is:

$$\text{Price Elasticity of Demand} = \left| \frac{\%\Delta \text{ Quantity Demanded}}{\%\Delta \text{ Price}} \right|$$

(%Δ means "percent change", and the vertical bars mean we're taking the absolute value, so just ignore the negative sign in your answer)

Remember, elasticity is a measurement of how much demand changes *given* a change in price (or *per* a change in price, hence why it is *divided* by %Δ Price). If price goes up by 10% and the quantity demanded goes down by 10%, the formula gives us a value of 1, and we know there is a 1:1 relationship between price and quantity demanded, so we call this **unit elastic**. If price goes up by 10% and quantity demanded changes by more than 10%, now the formula gives us a value greater than 1, so the good is **relatively elastic**. If price goes up by 10% and quantity demanded changes by less than 10%, the formula gives us a value less than 1, so the good is **relatively inelastic**.

Elasticity Value	That Good Is…
<1	Relatively Inelastic
1	Unit Elastic
>1	Relatively Elastic

To remember this, just remember that we're talking about the formula for *elasticity*. So the higher the number is, the more *elastic* it is going to be.

And remember, just like we learned with inflation last semester, the formula for percent change is New – Old / Old, which is easy to remember because it abbreviates to N-O/O, which is what characters in movies shout in slow motion whenever they have to calculate a percent change (that joke gets funnier every time I tell it).

So imagine I were given the following table and told to calculate the price elasticity of demand as price increases from $4 to $5.

Price	Quantity Demanded
$3	20
$4 (old)	16 (old)
$5 (new)	10 (new)

The formula is just New–Old/Old, and you just fill it in with the numbers in the chart. I always write (new) and (old) on the chart to help me keep them straight.

$$\frac{\%\Delta \text{ Quantity Demanded}}{\%\Delta \text{ Price}} = \frac{\frac{\text{Quantity}_{\text{New}} - \text{Quantity}_{\text{Old}}}{\text{Quantity}_{\text{Old}}}}{\frac{\text{Price}_{\text{New}} - \text{Price}_{\text{Old}}}{\text{Price}_{\text{Old}}}} = \frac{\frac{10 - 16}{16}}{\frac{5 - 4}{4}} = \frac{\frac{-6}{16}}{\frac{1}{4}} = \frac{\frac{-3}{8}}{\frac{1}{4}} = \frac{-3}{2}$$

For price elasticity, we always use the absolute value, so we know that the elasticity here is 1.5. Since 1.5 is greater than 1, we would say that this good is relatively elastic (in this case, a 25% increase in price led to a 37.5% decrease in quantity demanded).

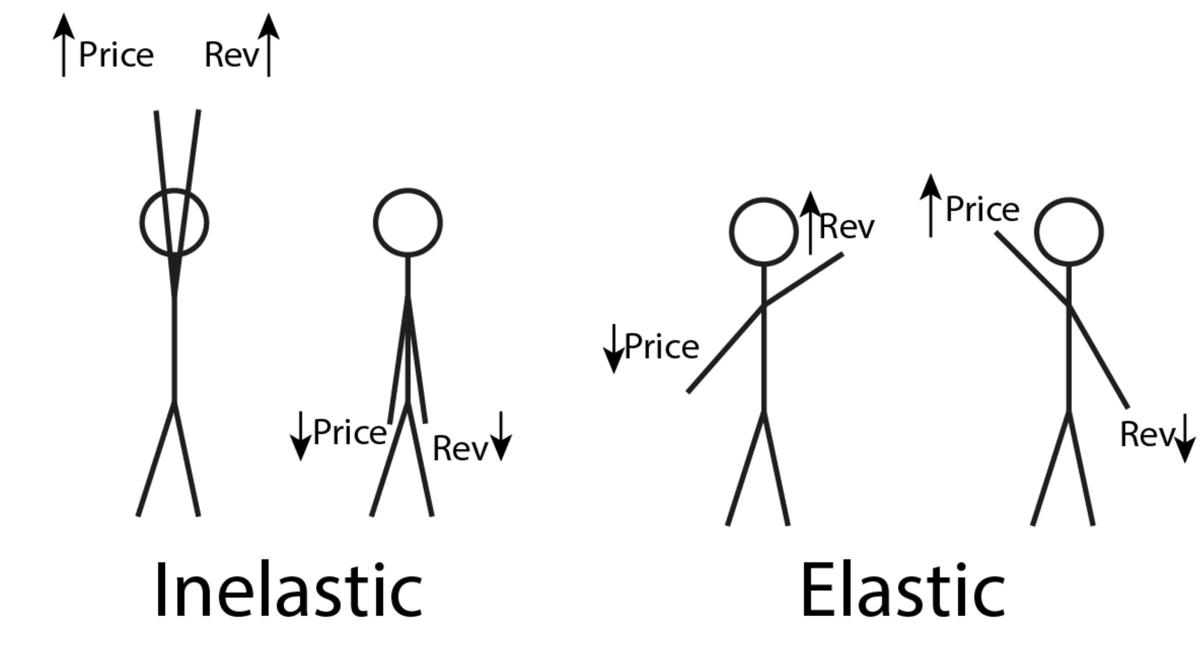
The second way we can determine whether a good is elastic or inelastic is by using the **revenue test**. Imagine that a firm raises prices by 10%. The basic idea of the revenue test is that if there is a 1:1 relationship between price and quantity, then a 10% increase in price will lead to exactly a 10% decrease in quantity demanded, so the firm will end up with just as much revenue as it had before. But if that increase in price leads to a 50% decrease in quantity demanded, then the firm will end up losing a lot of revenue. Alternatively, if that increase leads to only a 1% decrease in quantity demanded, then the firm will end up increasing its total revenue. What this means is that we can use the change in revenue as a shortcut to determine whether a good is elastic or not. The rules for the revenue test are:

If…	Then the good is
↑P → ↑Revenue	Inelastic
↑P → No Change in Revenue	Unit Elastic
↑P → ↓Revenue	Elastic

A decrease in price would have the opposite effects:

If…	Then the good is
↓P → ↑Revenue	Elastic
↓P → No Change in Revenue	Unit Elastic
↓P → ↓Revenue	Inelastic

The way to remember this is to think of one arm representing price and one arm representing revenue. When price and revenue both go up, your body forms a straight "I", so the good is inelastic. The same thing happens when price and revenue both go down. When they move in opposite directions, however, your body isn't straight anymore, and you look more like that horizontal "E" elastic curve.



What makes a good elastic

There are three factors you need to know that contribute to the relative elasticity of a good.

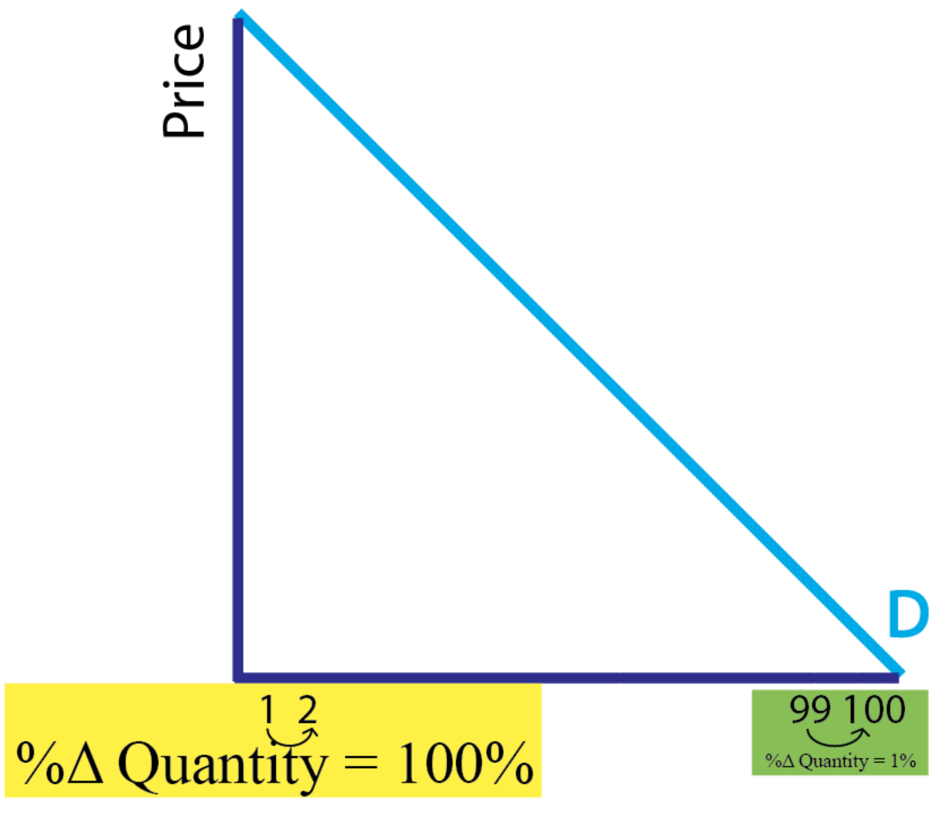
1. *Having close substitutes*. If a good has a lot of close substitutes, it will be more elastic. So if the price of apple juice rises, people will respond to increases in price by switching to grape or cranberry juice.
2. *How narrowly defined the market is*. The market for food is extremely inelastic, but the market for dessert is much more elastic, and the market for ice cream is more elastic still. The market for *vanilla* ice cream is extremely elastic. The more narrowly you define the market, the more elastic it tends to be.
3. *Timeframe*. Everything gets more elastic in the long-run. For example, gas is relatively inelastic; no matter how much prices rise, people will still need to buy gas to get to school and work. But if high prices persist over a few decades, then people will quickly switch to hybrid and all-electric cars, making the demand for gasoline extremely elastic. That increased elasticity in the long-run holds true for all goods and services.

Return to the top of the document

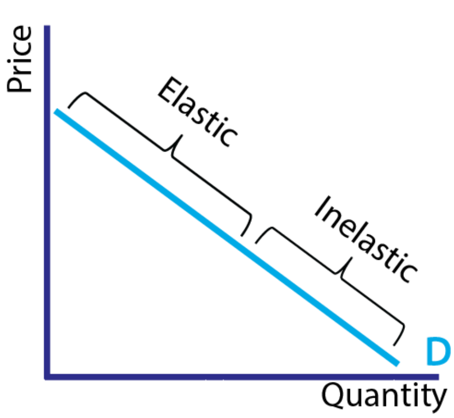
2.3–2.5 Price Elasticity of Demand Part 2; Other Elasticities

In the previous reading, we talked about how the shape of a demand curve—i.e., how horizontal or vertical it is—can tell you whether it is elastic or not. But you have to be very careful here, because slope and elasticity are two very different things. Slope is the change in y over the change in x, but elasticity is dividing the ***percent*** *change* in price and quantity. This difference, it turns out, is really important.

To see why, look at the demand curve below, with a constant slope. Remember, the formula for elasticity is $\frac{\%\Delta\,Q}{\%\Delta\,P}$, and here we're going to focus only on the numerator. If you start at the far left side of the curve (in yellow), going from 1 to 2 represents a 100% increase in quantity ($\frac{2-1}{1} = 1$). As you move further down the line, each increase in quantity becomes a smaller percent change, so by the time you go from 99 to 100 (in green), the percent change is only 1% ($\frac{99-100}{100} = \ .01$). In both cases you're only increasing quantity by 1, but the *percent change* in quantity is vastly different.



This change in the numerator means the elasticity is going to be very high when quantity is low, and it is going to be miniscule (i.e., inelastic) when quantity is high. So elasticity is going to be different at different points on the demand curve.



Price Elasticity of Supply

We can also analyze the elasticity of the supply curve. Just as the price elasticity of demand is a measure of how much consumption changes given a change in price, so the price elasticity of supply is a measure of how much the quantity supplied changes given a change in price. In other words, if price goes up by a lot, will producers really ramp up production, or will they be stuck producing roughly the same amount as before? The formula follows the same pattern as price elasticity of demand:

$$\text{Price Elasticity of Supply} = \frac{\%\Delta \text{ Quantity Supplied}}{\%\Delta \text{ Price}}$$

Just like demand, the more inelastic a supply curve is, the more vertical it will be, while a perfectly inelastic supply curve will be completely vertical. Conversely, a perfectly elastic supply curve will be horizontal. And just like demand, a value of greater than one means that the supply curve is relatively elastic, while a value less than one means the supply curve is inelastic.

There are two main factors you have to know that determine how elastic/inelastic a supply curve is:

1. *The price and availability of inputs*. If I'm making electric cars, I <u>need</u> to have cobalt for batteries—there's no way around it. If the price of electric cars goes up, then I can't really increase production that much, because it will either be really difficult or really expensive to get more cobalt. Alternatively, if I'm making sandwiches and the price of sandwiches goes up, I can pretty easily get more peanut butter and jelly, especially since Smuckers jelly is easy to substitute with another brand if I need to make a substitute. So *the more difficult or expensive an input is to get, the more inelastic supply will be*.
2. *Timeframe*. If I have one factory, I'm not going to be able to increase supply beyond what that factory can make—in the short run. But in the long run, I can build more factories, hire more people, find a substitute for difficult-to-find inputs, etc. So just like demand, supply becomes much more elastic in the long run.

There are two other elasticities you need to know, both of which help us classify goods as normal/inferior or substitutes/complements. These are:

Income Elasticity of Demand

The price elasticity of demand measures how much the quantity demanded changes (numerator) given a change in price (denominator). As you might imagine, the *income* elasticity of demand measures how much the quantity demanded changes (numerator) given a change in a person's *income* (denominator). So the formula is:

$$\text{Income Elasticity of Demand} = \frac{\%\Delta \text{ Quantity Demanded}}{\%\Delta \text{ Income}}$$

If the quantity demanded increases (+) when a person's income increases (+), the income elasticity of demand will be positive, which means the good is a normal good. If the quantity demanded decreases (-) when a person's income increases (+), the income elasticity of demand will be negative, which means the good is an inferior good. So the main thing to know here is:

- Positive = Normal Good
- Negative = Inferior Good

Of course, this makes sense when you think of inferior goods being bad/negative.

Cross-Price Elasticity of Demand

This one is the most intimidating for students, but despite its scary name, it's actually really simple. All we're measuring here is how much the quantity demanded of one good changes (numerator) given a change in the price of *another good* (denominator). The formula is:

$$\text{Cross-Price Elasticity of Demand} = \frac{\%\Delta \text{ Quantity Demanded of Good X}}{\%\Delta \text{ Price of Good Y}}$$

In other words, we want to know: if the price of apple juice increases, how much will people's demand for orange juice change? In short, we're trying to measure the degree to which two goods are related. If people demand more orange juice (+) when the price of apple juice increases (+), that means these two goods are substitutes. If people demand less peanut butter (-) when the price of jelly increases (+), the cross-price elasticity of demand will be negative, which means these two goods will be complements. So the main thing to know here is:

- Positive = Substitutes
- Negative = Complements

Want an easy way to remember this? Think of a "negative compliment," like, "you play basketball well for an old person," or "you got a surprisingly good grade on that test considering that you're an idiot."

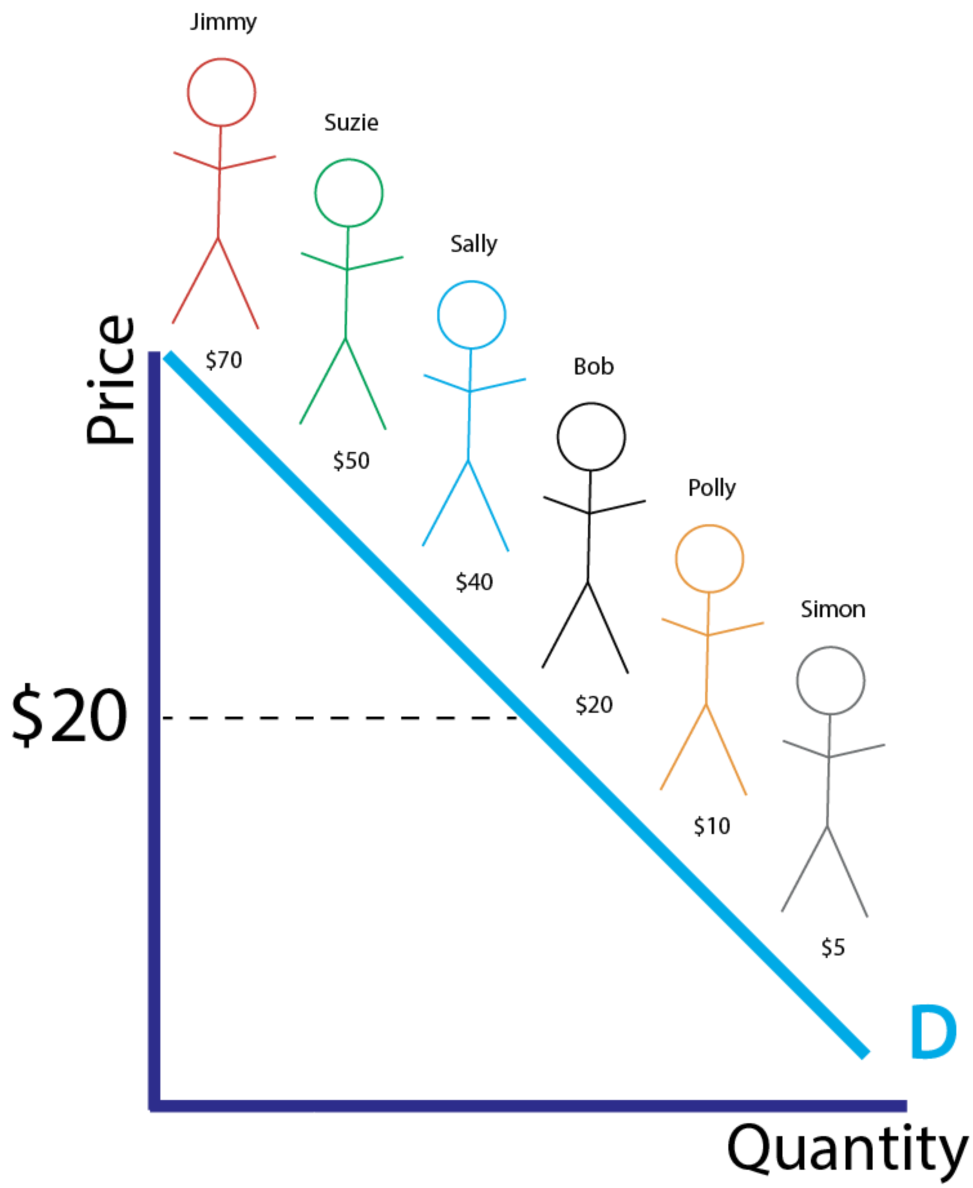
WARNING: be extra careful to keep separate the two things you're checking for with different elasticities.

- For price elasticity of demand and elasticity of supply, you're checking whether the absolute value is greater than or less than one. A value of one means the good is unit elastic.
- For income elasticity and cross-price elasticity of demand, you're checking whether the value is positive or negative. For these two elasticities, a value of zero means the goods are unrelated.

Return to the top of the document

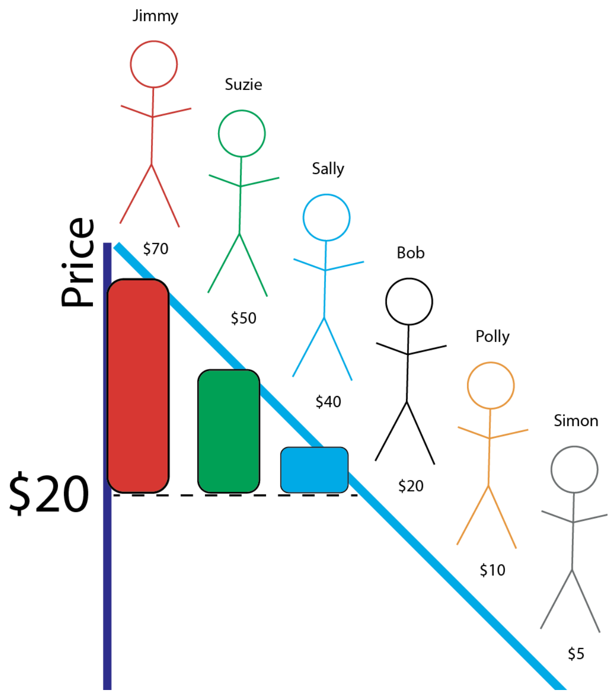
2.6–2.7 Consumer and Producer Surplus

In the market for lamps, there are thousands of people who are willing to buy lamps at various prices. These people all combine to form the demand curve for lamps, and I've picked out six individuals from that curve here: Jimmy, Suzie, Sally, Bob, Polly, and Simon, each of whom is willing to pay different amounts (and thus falls at a different point on the demand curve).

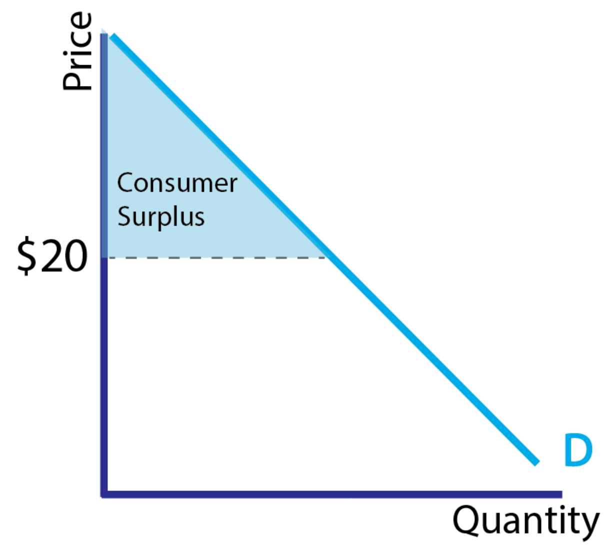


If they all walk into the store and find lamps selling for $20, then only four of them will buy a lamp, and they will all be better off. But importantly, they will all be better off by different amounts. Jimmy was willing to pay $70 for his lamp, but he only paid $20, which means he got an extra $50-worth of benefit or happiness that he wouldn't have otherwise had. Suzie got an extra $30 of benefit, Sally got an extra $20 of benefit, and

Bob is just as well off as he was before, since he valued the lamp at exactly $20 and got it for $20.

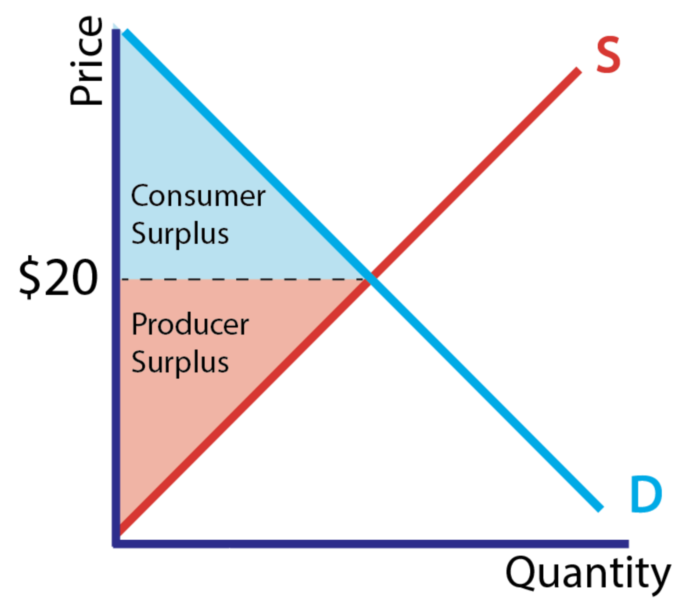


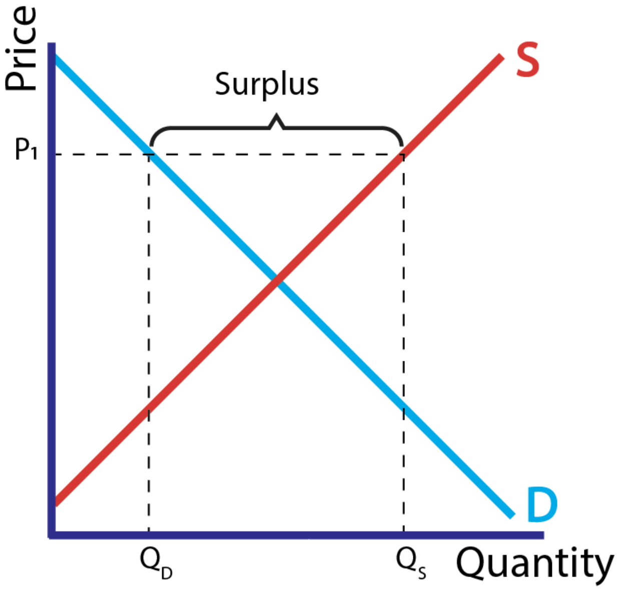
Notice, we can figure out how much better off each person is by taking their willingness to pay (i.e., their position on the demand curve) and subtracting the price they paid, as you see on the left. If we did this for all the thousands of people on this demand curve, we'd end up with a total area of extra wellbeing that looks like graph below:



We call this area **consumer surplus**, because it represents the surplus or extra benefit that all the consumers in the market are receiving. To find the consumer surplus, then, all you have to do is find the area between the demand curve and the price paid.

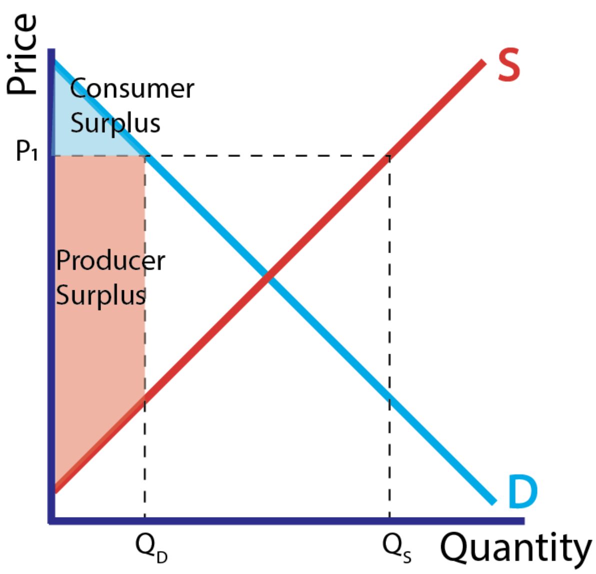
We can run the same kind of analysis for producers. Remember, the supply curve represents the marginal cost of producing each item, so if lamps sold for $20, then every lamp produced at a marginal cost of less than $20 would lead to **producer surplus**. If the first lamp was made for $1, that would be $19 of producer surplus, if the second was made for $2, that would be $18 of producer surplus, and so on. Producer surplus, then, is the difference between the price producers receive and the cost of production (the supply curve). Putting them both together would give us a graph that looks like the one on the right:





Producer and consumer surplus together give us what is known as **market welfare**. Be careful here! Producer and consumer surplus is not the same thing as the surplus that is created when prices are too high (graph on the left). That kind of surplus refers to there being too many products on the shelf, when quantity supplied is greater than quantity demanded. It’s unfortunate that economists use the same term for both.

As long as we’re looking at this graph on the left, let’s figure out what the total market welfare would be. We know that only Q_D amount of stuff is going to be bought and sold (since this is all consumers are willing to buy), and we know that consumer surplus is the area between price and the demand curve, while producer surplus is the area between price and the supply curve. That gives us a total market welfare as seen below.



You’ll notice that no matter where you set price, you can’t get more market welfare than when the market is in equilibrium. If equilibrium in a market maximizes total welfare—as it always does in the markets we’ve studied so far—that means our market is **allocatively efficient**, meaning that it allocates resources among producers and consumers as efficiently as possible. The graph on the left is allocating resources inefficiently because at this point, there are still buyers who are willing to buy the good at a higher price than it costs producers to make it. This is what is happening in all that white space where the rest of the producer and consumer surplus used to be. It would be more efficient if we could find a way to get this market to “clear”, i.e., if we could re-establish the original equilibrium.

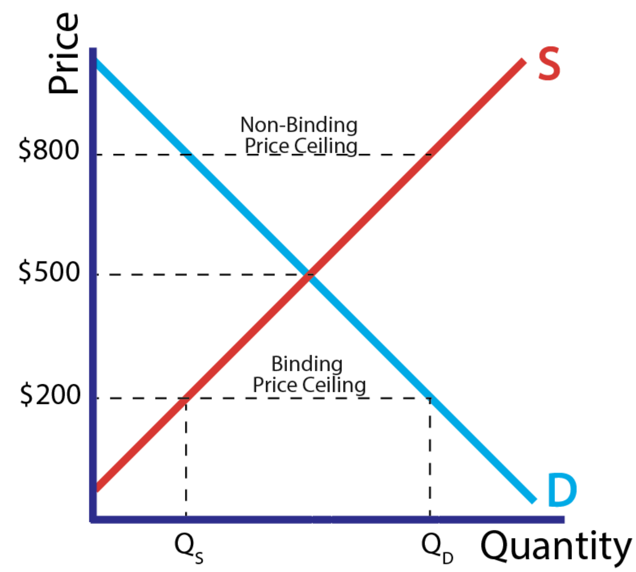
An important side note here is that allocative efficiency happens when our equilibrium is at a quantity where the marginal cost of production (supply) is equal to the marginal benefit consumers derive from that good (demand).

[Return to the top of the document]

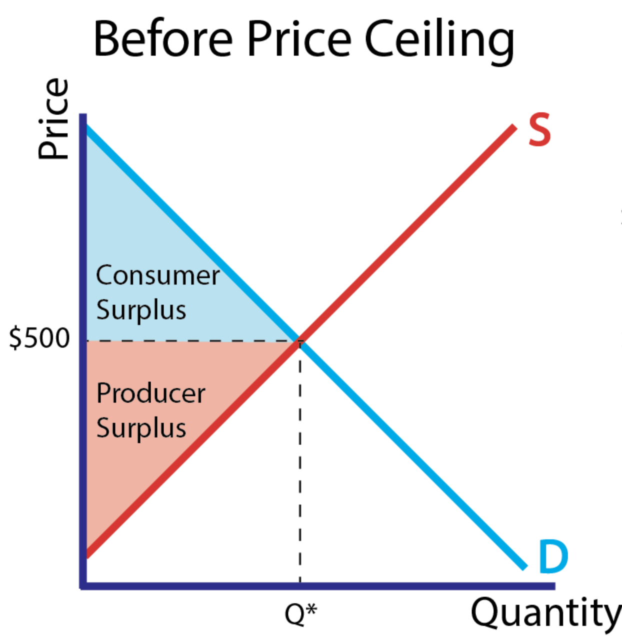
2.8 Government Intervention in Markets, Part 1

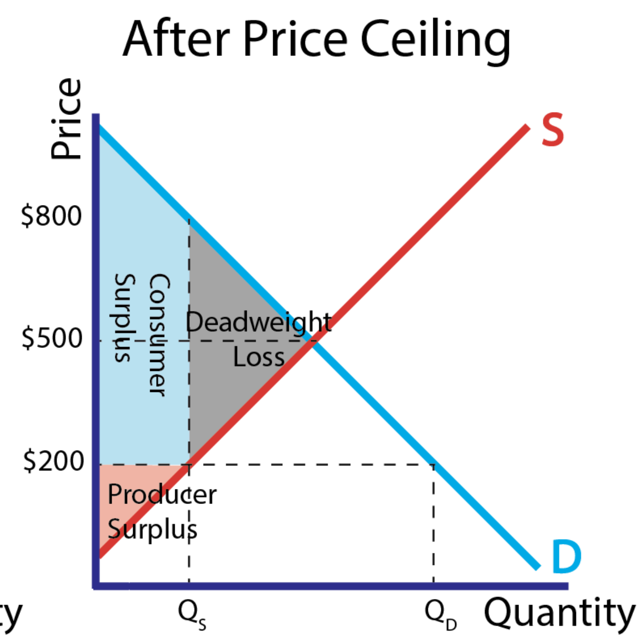
Price Ceilings

Sometimes the government is worried about firms charging too much for certain goods, such as housing, or bottled water during a hurricane, and it steps in to protect consumers. This kind of intervention is called setting a **price ceiling**, i.e., a legal maximum price that firms are allowed to charge. Consider this market for apartments, where the equilibrium price is $500 per unit. If the government sets a price ceiling at $800, this isn't going to have any impact on the market; no one was going to charge that much anyway. We call this a **non-binding price ceiling**. If, on the other hand, the government sets a price ceiling at $200, now our market is going to get out of whack. Using what you know from Unit 1, you know that at a price this low, the quantity demanded is going to be much higher than the quantity supplied, which means we're going to have a shortage of apartment units on our hands.

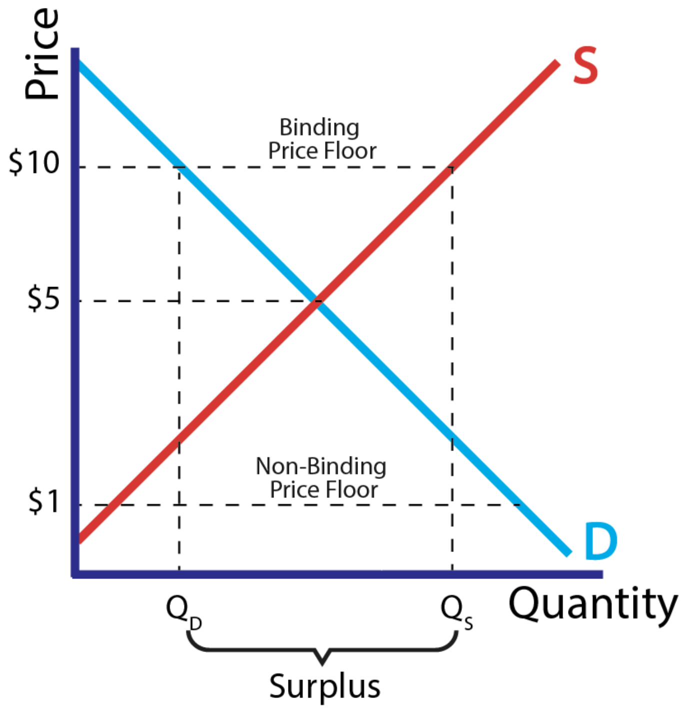


Did this particular intervention make consumers worse or better off? We can tell by analyzing the change in surplus.





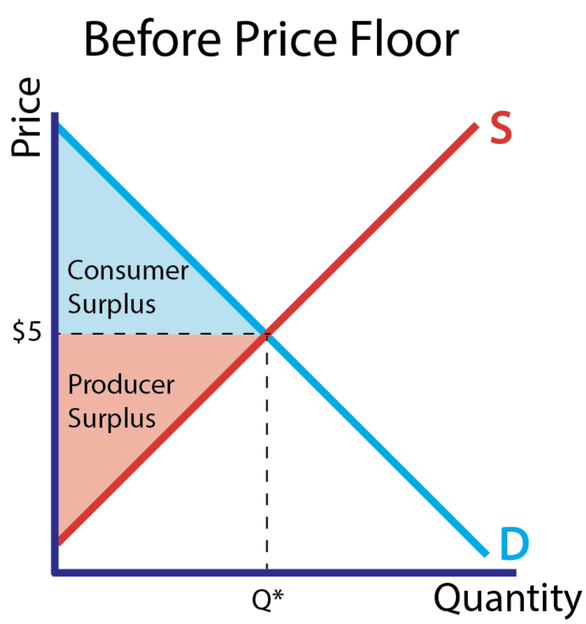
As we saw last lesson, the market surplus only extends as far as the amount of goods that actually trade hands (Q_S in this case). We know that producers are going to be worse off, as they have lost a lot of surplus. Consumers lost some surplus too, but they also gained some of the surplus that producers used to have. So consumers *might* be better off, depending on the shape of the supply and demand curve. But one thing is absolutely clear: our society as a whole is worse off. We have lost a lot of surplus that wasn't given to anyone, and we've moved away from the allocatively efficient point. The difference in surplus between our allocatively efficient point and where we are now is known as **deadweight loss (DWL)**.

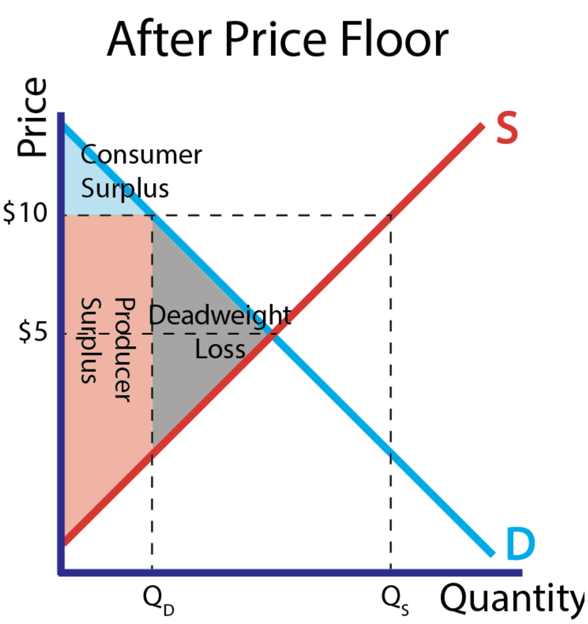


Price Floors

Alternatively, the government could try to help producers by setting a legal minimum price in a market. Similar to price ceilings, if this price floor (legal minimum) falls below the normal equilibrium price, then it is non-binding, and it will not affect the market at all. If it is above equilibrium, however, then it is binding, and it pushes our market into a surplus, where $Q_S > Q_D$.

The analysis of market welfare here is going to work the same as before, only flipped, so now consumers are definitely worse off, producers are *maybe* better off, and society as a whole is worse off due to the deadweight loss.

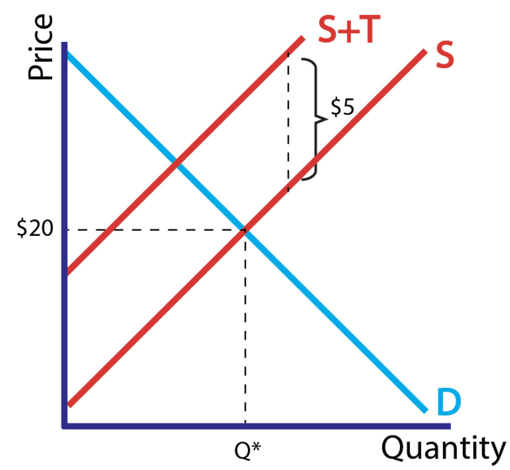




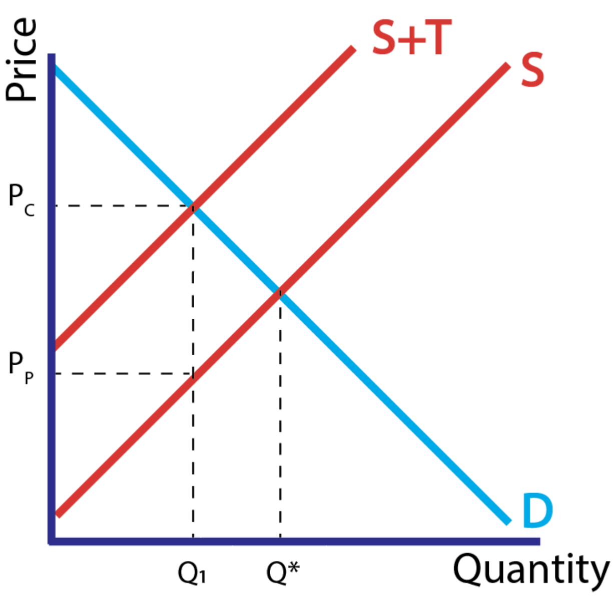
When it comes to price ceilings and price floors, remember: when the ceiling is above you (i.e., above equilibrium) and the floor is below you, everything is great. But when the ceiling is below your feet and the floor is above your head, things get crazy.

Taxes

Another way governments intervene in markets is through taxes. There are two types of taxes the government can impose: **per-unit taxes**, where people have to pay a certain amount for every item bought/sold (like a $3 tax on every carton of cigarettes sold), and **lump-sum taxes**, where companies have to pay a fixed amount to the government to be able to produce a certain good, but each individual good is not taxed (like $500,000 to get a license to run a marijuana dispensary). For lump-sum taxes, companies would pay the same amount to the government if they sold 1 good or 1,000,000 goods.

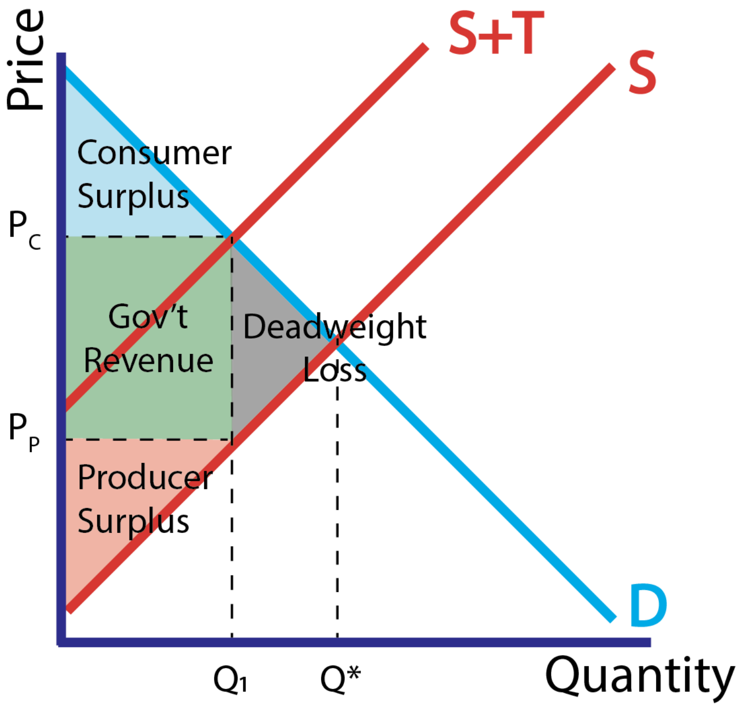


The most common kind of tax is a per-unit tax, and let's assume that lamps sell for $20 and the government imposes a $5 per-unit tax on lamp sellers. Since the tax is per-unit, that means every item produced costs $5 more than it used to, so the supply curve (i.e., the marginal cost curve) shifts up by exactly $5. We'll label this new curve S+T to denote that it is adding the tax. Notice how the vertical space between the two curves is exactly $5—the amount of the tax—and the new supply curve represents a decrease from the original supply curve, just as we saw in Unit 1 that taxes decrease supply.



We have a new equilibrium quantity where D meets the S+T curve, but we're now going to have two different prices. Consumers are paying the full price of the good plus the tax, but producers don't receive all of that money. They only receive the price consumers pay *minus the amount of the tax*. Thankfully, we know that the vertical distance between the two supply curves is the exact amount of the tax, so at any given quantity, if consumers are paying a rate at S+T, then producers are only receiving the amount at S. We label these two points P_C (price that consumers pay) and P_P (price that producers receive). BE CAREFUL: Neither of these prices is the same as the original price. You're always going to draw two new prices when you have a tax.

Now we can figure out everyone's surplus. Consumer Surplus is the area between the demand curve and price the consumers are paying, so it will only extend down to P_C. Producer surplus likewise will only extend up to P_P, since this is the price producers are receiving. The rectangle in the middle represents how much revenue the government receives from this tax. We know this because the height of the rectangle is the amount of the tax, and its length is the quantity of units sold. Importantly, this area is not deadweight loss. Deadweight loss, remember, is the difference in surplus between the allocatively efficient point and where we are now, and even though this government revenue is not going to consumers or producers, it is still going to *someone*. The government is made up of people, so the fact that they get this money means it is still part of our market welfare. The deadweight loss, then, will be the triangle to the right of government revenue.



A lump-sum tax would have no impact on our market. We wouldn't make a new S+T curve, since lump-sum costs don't impact the marginal cost of each unit produced. So if we started at P* and Q* before the tax, we would stay at P* and Q* after it.

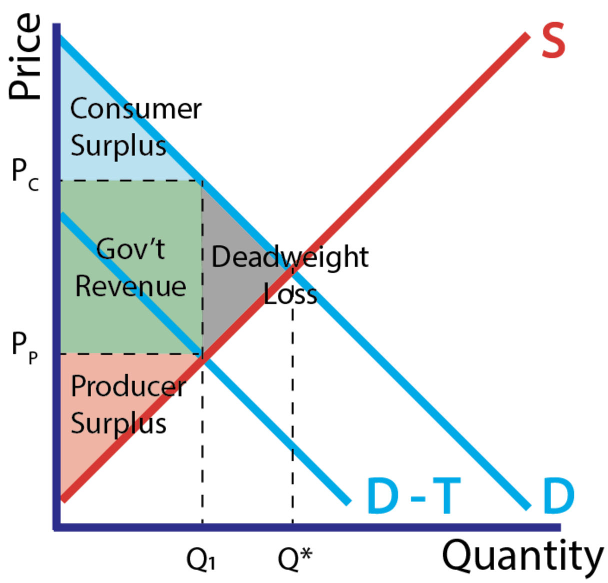
[Return to the top of the document]

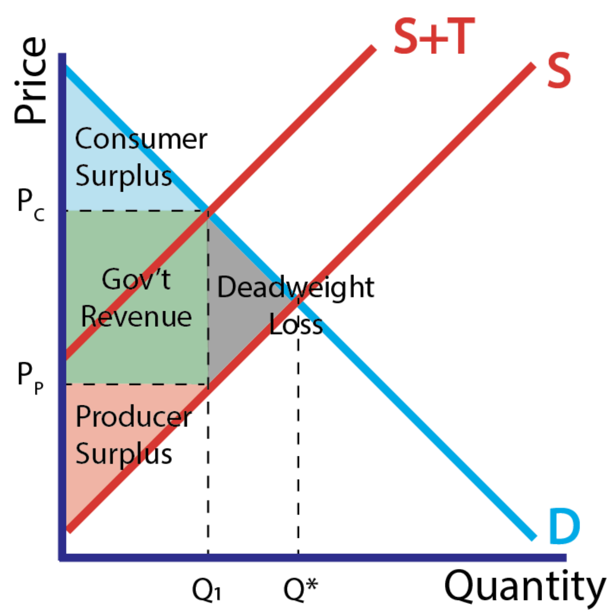
2.8 Government Intervention in Markets, Part 2

Tax Incidence

One of the most surprising things students learn in this course is that it doesn't matter whether you tax producers or consumers: the final result is exactly the same.

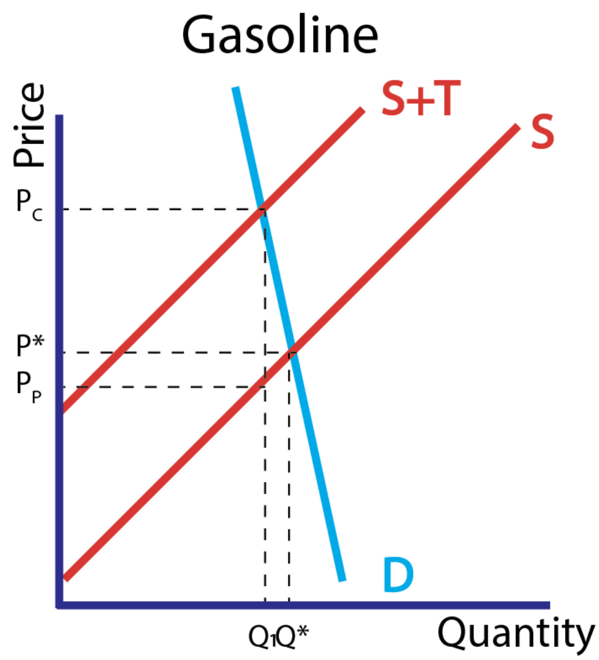
You don't need to be able to produce or even describe anything I'm about to tell you in this paragraph, but you do need to understand the general idea. For the example above on taxing lamps, I imposed the tax on lamp producers, but imagine I imposed the tax on lamp *consumers* instead, making them pay the tax at the cash register rather than making the producer pay the tax before selling it. I would show this by decreasing the demand curve by the amount of the tax, since consumers want just as many lamps as before, but their willingness/ability to pay has decreased by exactly the amount of the tax. We'll call this new demand curve D - T. When you compare this to what happens when you shift the supply curve instead, you'll notice that everything is exactly the same. P_C, P_P, Q_1—everything is in the exact same spot, since in both cases we're shifting the curves by the exact same amount.

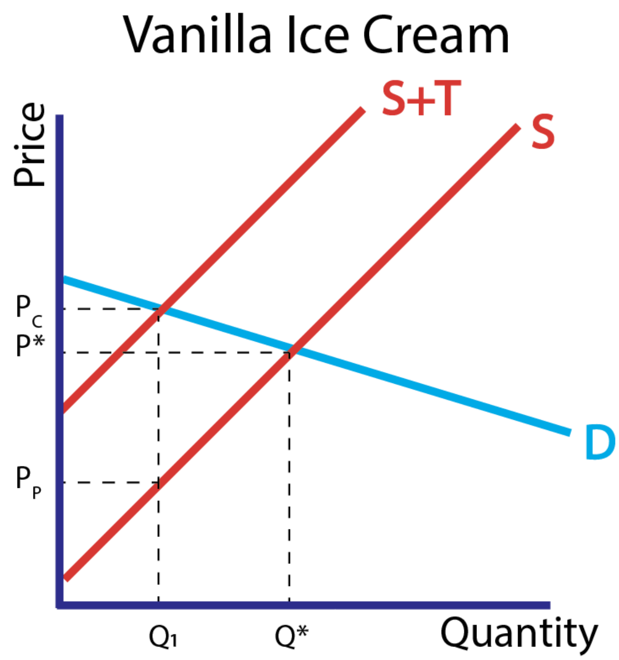




This is why, to make our lives easy, economists show taxes by always shifting the supply curve, no matter whether the government imposes the tax on producers or consumers. As we now know, whom you put the tax on does not matter.

That is not to say that all taxes will harm producers and consumers equally. Look at what happens when you tax two different goods, one that has a relatively inelastic demand (gasoline), and one whose demand is relatively elastic (vanilla ice cream).



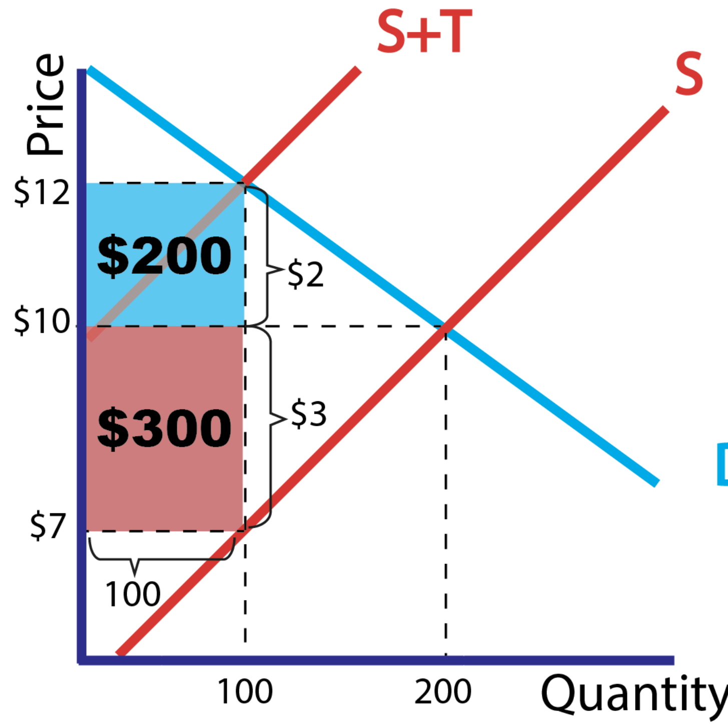


Look at the tax on gasoline. If you compare P_C and P_P to P*, the price before the tax was imposed, you'll notice that P_C is higher and P_P is lower, so producers and consumers are both worse off than they were before: consumers are paying more than they used to, and producers are receiving less than they used to. But notice that they are not hurting equally; consumers are paying *way more* than they used to pay, while producers are receiving just barely less than they used to get. That's because consumers *have* to get gas; they have an inelastic demand curve, so they're willing to pay almost anything for gasoline. This tax, in other words, really hurts consumers.

For the ice cream, on the other hand, the opposite has happened. Yes, consumers are paying more than they used to pay, but only barely. Producers, on the other hand, are receiving way less than they used to get. That's because consumers don't really need vanilla ice cream. If the price goes up, they can just walk away and not buy any. So this tax really hurts producers.

On the exam, you will have to find the **incidence** of a given tax, which is another way of saying that you'll have to determine who bears more of the tax burden, who pays a greater share of the tax, or who is harmed more by the tax. And these graphs illustrate the main point you need to know about tax incidence: *when a tax is imposed, the burden will always fall more heavily on whichever curve is more inelastic.* If demand is more inelastic than supply, then consumers will bear more of the tax burden. If supply is more inelastic than demand, then producers will bear more of the tax burden.

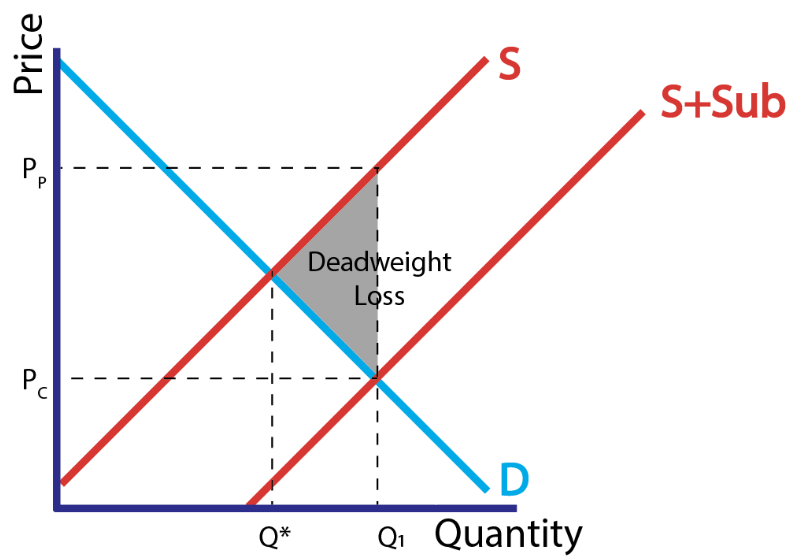
Another way of showing tax incidence is by analyzing the rectangle that represents government revenue. In the example below and to the right, demand is slightly more elastic than supply, so for this $5 tax, the price consumers pay only goes up by $2, while the price producers receive goes down by $3. The new quantity is 100 units, which means the government is receiving a total of $500 in revenue ($5 per-unit tax *x* 100 units), and we can divide this rectangle up to represent how much of that money is coming out of consumers' pockets and how much is coming out of producers'. Remember, consumers are paying $2 more per item, and they are buying 100 units, so in a sense they are contributing $200 of the tax revenue. Producers are giving up $3 per unit, so they are contributing $300 of the tax revenue. Thus we see that the tax incidence falls more heavily on producers.



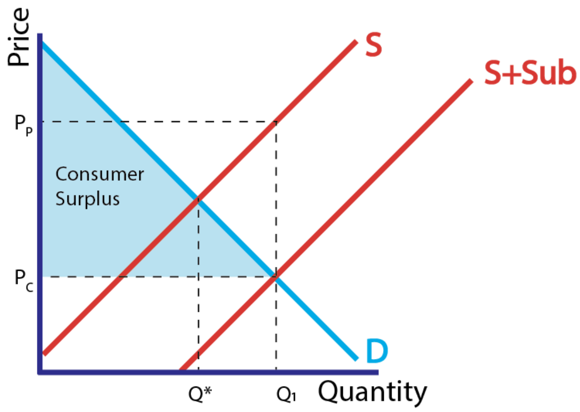
Subsidies

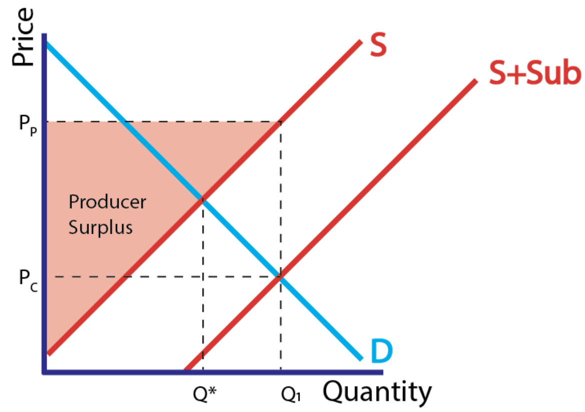
The final type of government intervention is a subsidy, where the government pays producers to make something (in the US, the government subsidizes corn and sugar, to name two examples). Just like taxes, a per-unit subsidy will move the supply curve (down, in this case), while a lump-sum subsidy will not move the supply curve at all.

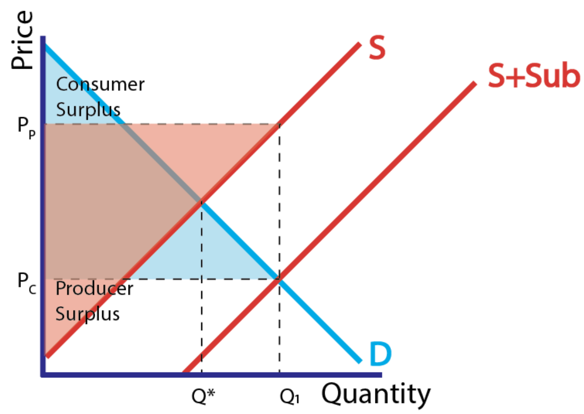
When it comes to subsidies, all you need to know is that a per-unit subsidy moves the supply curve down by the amount of the subsidy, and it creates DWL, which is now to the right of the original equilibrium point. <u>You can stop reading right now and count your reading done for the night</u>, as this is all you're responsible for. If you want to know *why* this happens, then read on. Again, nothing from this point on will be on the AP exam.



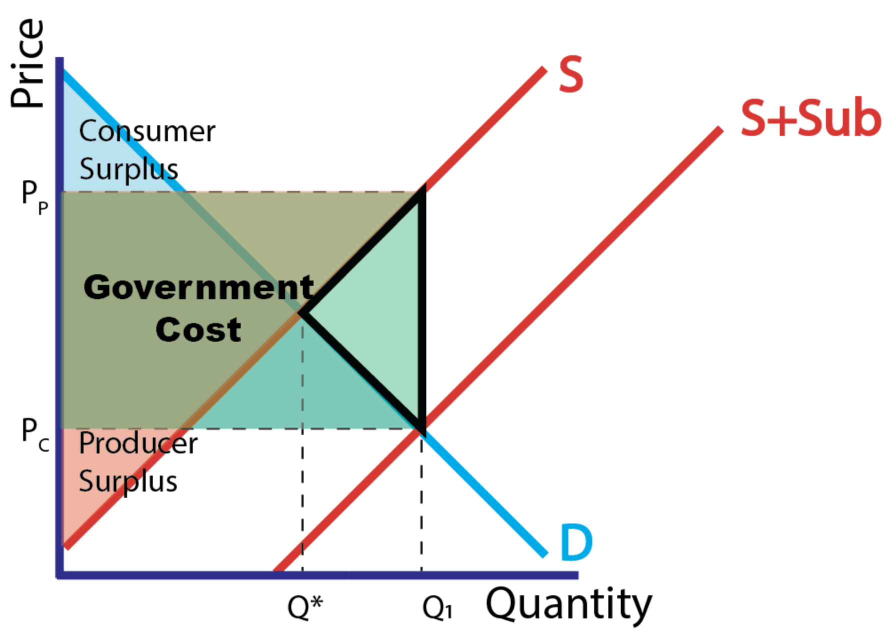
With a subsidy, the government is paying producers for every unit produced, which means that the price producers receive is actually going to be *higher* than the price consumers pay (notice how $P_P > P_C$ on the graph). Consumer surplus is going to go all the way down to P_C, and producer surplus goes all the way up to P_P, which means the two surpluses are now going to be overlapping.







This will cost the government the amount of the subsidy (the vertical distance between the two supply curves) multiplied by Q1, the new amount bought and sold in the market. That means all that extra surplus that consumers and producers gained above is actually being paid for by the government, so we're not actually gaining any market welfare. But that farthest triangle to the right of the original equilibrium (outlined in black) is being paid for by the government without actually being received by anyone as surplus, which means it is DWL.

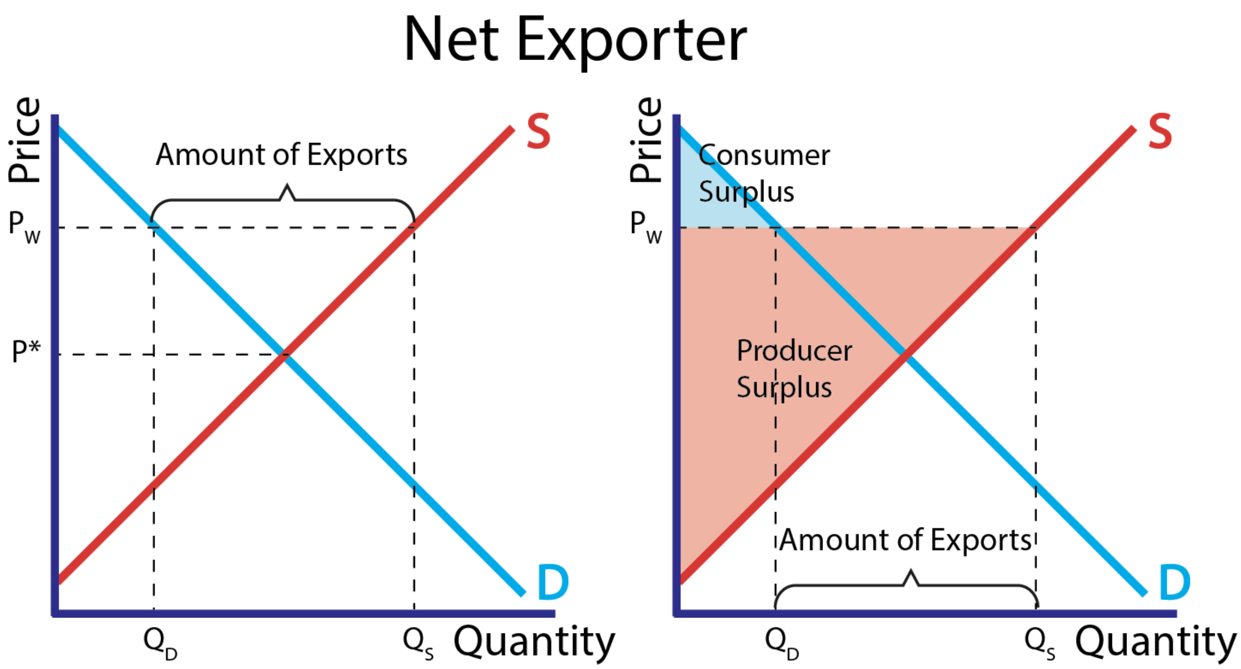


Return to the top of the document

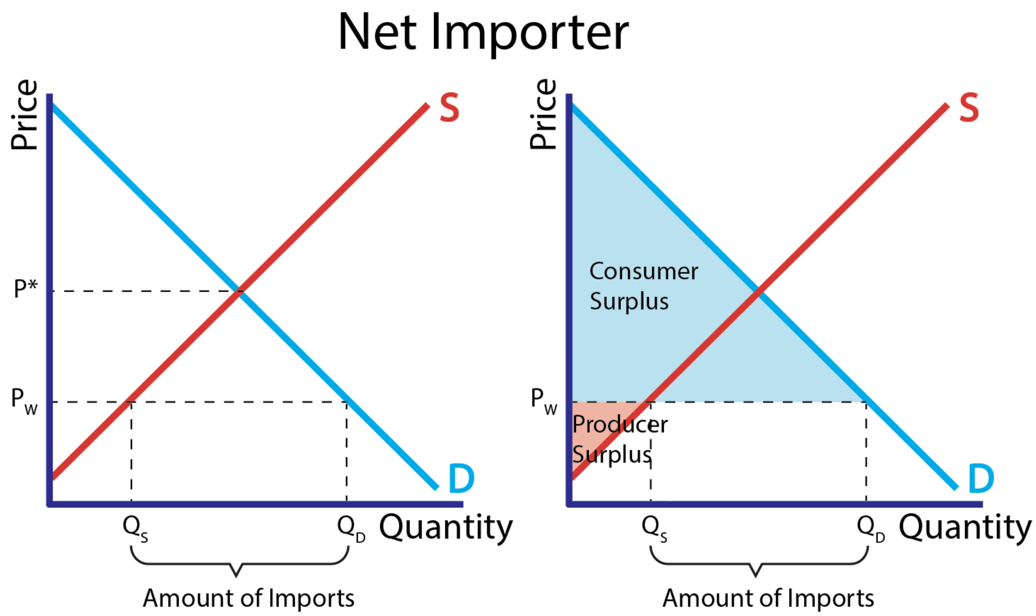
2.9 International Trade and Public Policy

As we saw in Unit 1 with comparative advantage, free trade is generally good for society. In the graphs below, the supply and demand curves represent the domestic supply and domestic demand, while P_W represents the world price. When an economy opens to trade, there are two possible scenarios:

1. World price is higher than domestic price. Once the economy opens to trade in this scenario, everyone (including domestic suppliers) will start selling at that higher world price. This will lead to the domestic quantity supplied being higher than the domestic quantity demanded. This excess quantity is sold on international markets, and the country becomes a **net exporter**. It exports a total amount of $Q_S – Q_D$. Notice that while opening to trade hurts consumers, it helps producers by even more, and there is a net increase in market welfare.



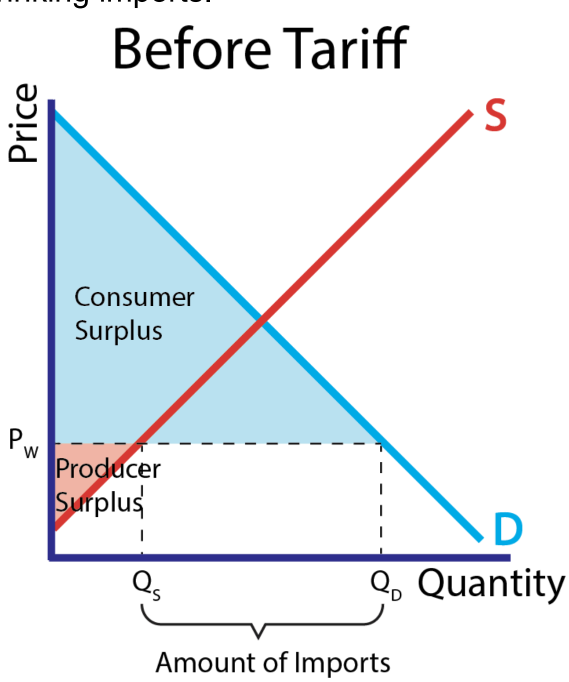
2. World price is lower than domestic price. When the economy opens to trade in this scenario, domestic suppliers can no longer sell at their previous price, since international sellers will undercut them with lower prices. So price will drop to P_W, and the difference between Q_S and Q_D will be made up by imports, so the country will become a **net importer**. The amount of imports is the difference between Q_S and Q_D, and while this policy hurts producers, it helps consumers by even more, and total market welfare rises.

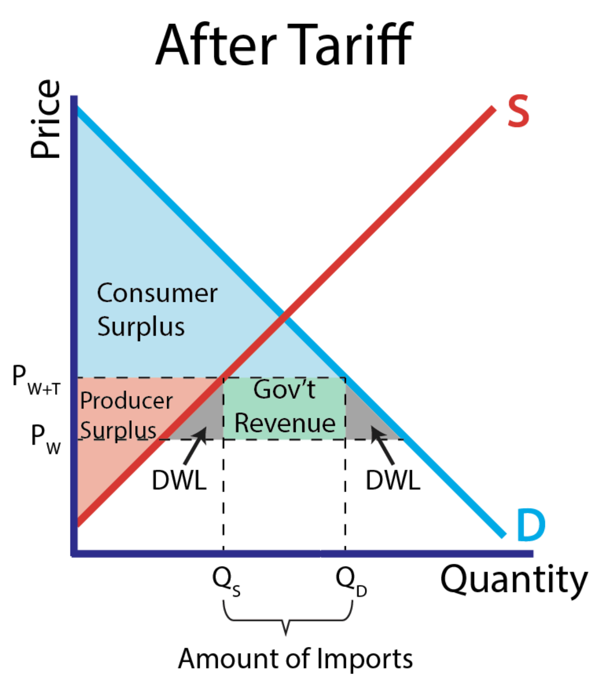


Tariffs and Quotas

When countries become net importers, governments often feel pressured to help domestic producers. They do this in two ways. One way is by imposing **quotas**, or caps on the amount of imports. All you need to know about this is that it decreases total market welfare, as does all the government intervention we've studied in this unit.

The second way that governments intervene is by imposing **tariffs**, or taxes on imports. A tariff effectively raises the price that people pay for the imported good, and we mark this new price as $P_{World + Tariff}$, or P_{W+T}. This higher price allows more domestic producers to enter the market, and it reduces the difference between Q_S and Q_D, thus shrinking imports.





The green rectangle is government revenue, which makes sense when you think about it. The height of that rectangle is the vertical distance between PW and PW+T, which means it is exactly the amount of the tariff. And the length of the rectangle is the amount of imports. So the amount of imports multiplied by the tax on each import would be the amount of money the government gets. The two DWL triangles represent the lost market welfare—the area that used to be consumer surplus but that now doesn't belong to anyone.

As you can see, every government intervention we've studied this unit leads to DWL, and thus a decrease in total market welfare. No matter how well intentioned governments may be, their intervention in these markets leaves society as a whole worse off. As we'll see in the following units, however, there are a number of circumstances where markets fail, and in these scenarios economists believe that government intervention can increase market welfare.

Return to the top of the document

3.1 The Production Function

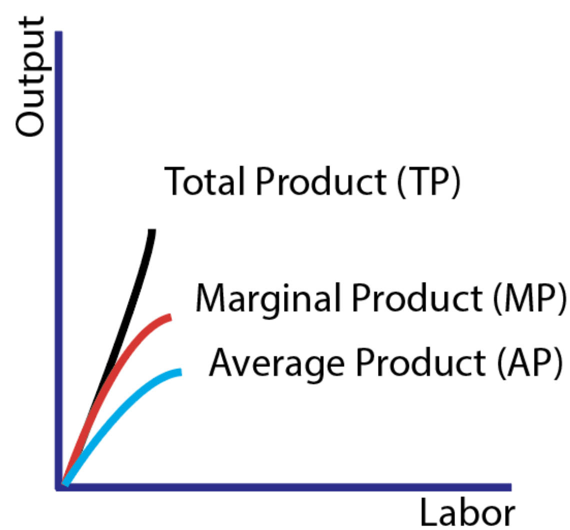
Now that we have the tools to analyze markets in depth, this unit begins our study of firms and how they make decisions. And the number one decision we focus on in this course is how much output a firm should produce, since this will ultimately determine the firm's profitability.

When producing any good, firms have to balance the amount of inputs (land, labor, and capital) they use with the amount of output they want. And the relationship between inputs and outputs is shown by the **production function**. The name "production function" is a little misleading, since in this class we're not going to study the actual equation/function. Instead, we're going to look at a graphical representation of this function.

The production function looks at only one input at a time, so for today, we're going to look at labor while we hold land and capital constant. We could, of course, repeat this analysis for the other two factors of production as well, but we're not going to walk through that today.

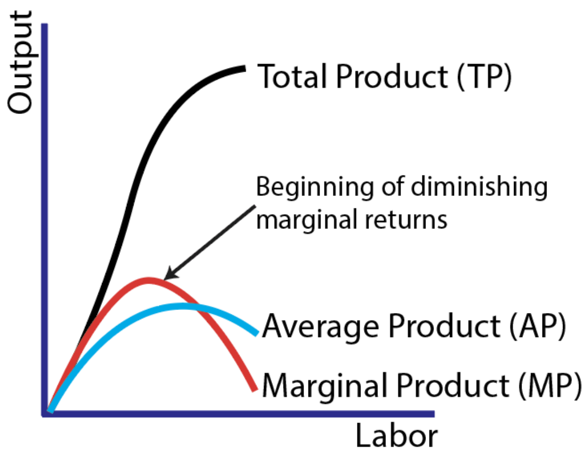
For this graph, we're going to put output on the y-axis and the input we're studying on the x-axis, which in this case is labor. We're trying to figure out what happens to output as we increase labor.

For essentially all firms, as you first begin to increase labor, each additional laborer you add allows you to become more efficient and more productive. That's because one person working a factory on his own would have a hard time running back and forth between every machine, but as you add more workers, they are able to take care of different parts of the process and allow it to run more smoothly.



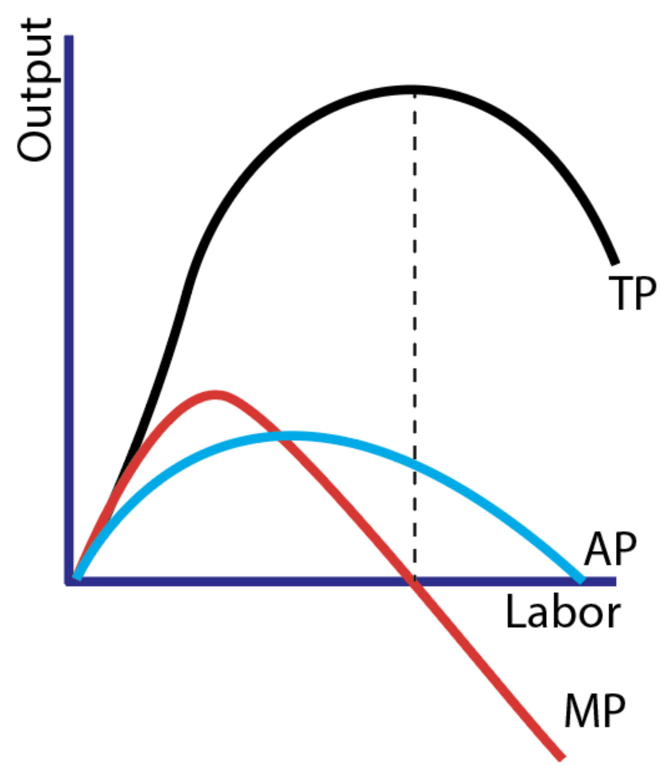
With the production function, we're going to graph out three different things: the **total product (TP)**, which is how much output the firm produces; the **marginal product (MP)**, which is how much output the most recently added unit of labor produces; and the **average product (AP)**, which is how much the average worker produces. In the graph to the right, you can see how these three lines work for the first few workers: each additional worker is more productive than the last, so the marginal product is rising. And with the addition of each high-productive worker, they bring up the average productivity of the factory as a whole. Meanwhile, the total product continues to climb.

Eventually, however, each additional worker stops adding as much benefit. Once you've manned every machine, then additional workers might help a little, but they're not nearly as helpful as the first few workers were. That means marginal product will start to fall. This is where our firm runs into **diminishing marginal returns**. Eventually,



marginal product will fall below the average product line and start dragging the average product down with it. Notice in the graph on the right that the average product line starts to fall at the exact point where marginal product crosses it. That's because these two lines work just like your grades do. If you have an 80 average and you score higher than an 80 on a test (i.e., your marginal grade is higher than your average grade), this will bring your average up. But if you score anything less than an 80 (i.e., your marginal grade is lower than your average grade), this will bring your average down. So average product will rise or fall based on whether the marginal product at that point is higher or lower than the average is. Meanwhile, total product is still rising, but as marginal product starts to fall, average product isn't climbing quite as quickly as it did before.

Eventually, adding additional workers is actually going to hurt the production process. If you have too many workers crammed into one tiny factory, everyone will be bumping into each other and getting in each other's way. So there will come a point where the marginal product of every laborer added is actually negative. At this point, total product will actually start to decrease. This is important: when the marginal product crosses the x-axis, total product starts to decline. That's because—if you want to use a fancy math term—the MP curve is the derivative of the TP curve. Don't worry; you don't need to know what a derivative is for this class. If you haven't taken calculus, just think of the MP curve as representing the slope of the TP curve. When TP is increasing (i.e., has a positive slope), the MP curve will be positive. Conversely, when MP is negative, TP will have a negative slope (i.e., it will be decreasing).



Remember, these graphs, showing output as a function of labor, assume that we're holding all other factors of production constant. So we're not adding any additional factories here—the only thing we're changing is the amount of labor.

[Return to the top of the document]

3.2 Short-Run Production Costs

I hate to break it to you, but you're going to have to memorize a lot of formulas for tonight's reading. My advice: focus on the general categories, such as the general rule for how to find all average formulas. That will be a lot easier than trying to memorize each one individually.

There are two main types of costs that firms face: **Fixed Costs (FC)** do not change with production levels, and they include things like rent and insurance. Whether you make 0 items or 1 million, you have to pay your fixed costs. A **Variable Cost (VC)** is any cost that does change with production levels, such as the cost of materials and wages. These two added together give us a firm's **Total Cost (TC)**.

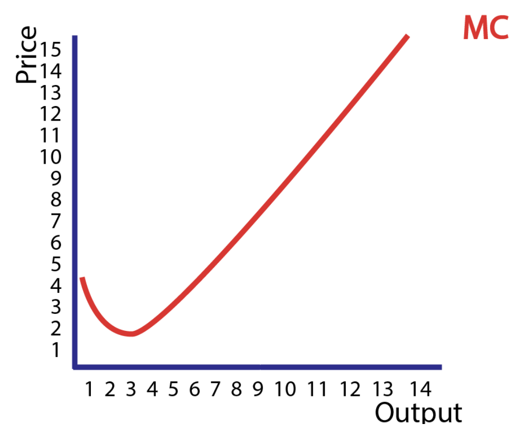
$$FC + VC = TC$$

When we analyze these costs, there are two different ways we can approach them: either looking at average costs or at marginal costs.

Let's start with **Marginal Cost (MC)**. To find MC, you simply need to figure out the change in Total Cost as you increase output. So the formula is

$$\mathrm{MC} = \frac{\Delta \mathrm{TC}}{\Delta \mathrm{Q}}$$

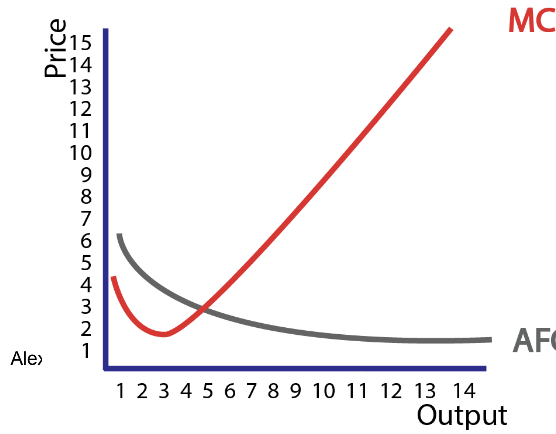
The shape of the MC curve is going to look just like the Marginal Product curve from the previous lesson, only mirrored upside down, and it looks like this for the exact same reason: as you add workers, they initially become more productive, so the marginal cost of production falls, but as you add more and more, you encounter decreasing marginal returns, so MC is going to rise.



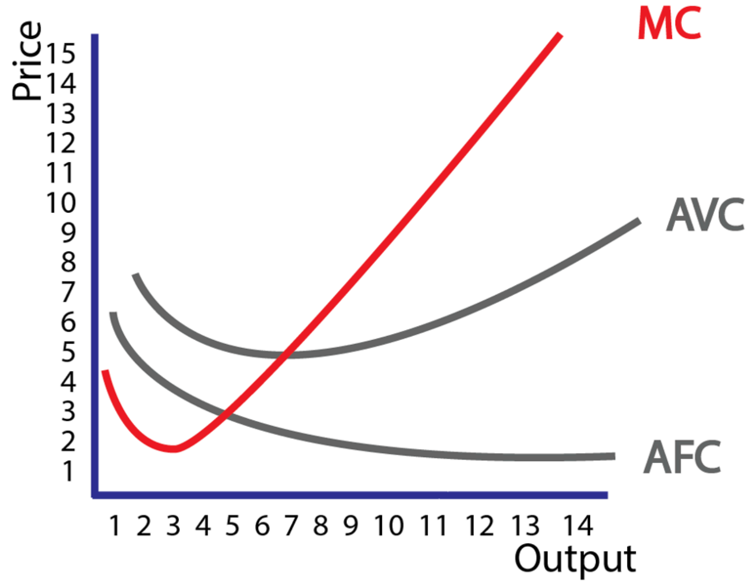
If we turn to average costs, there are three formulas you need, but thankfully they all follow the same pattern. The average of any cost is that cost divided by the amount produced. So **Average Total Cost (ATC)** is TC/Q, **Average Variable Cost (AVC)** is VC/Q, and **Average Fixed Cost (AFC)** is FC/Q.

$$\mathrm{AFC} = \frac{\mathrm{FC}}{\mathrm{Q}} \qquad \mathrm{AVC} = \frac{\mathrm{VC}}{\mathrm{Q}} \qquad \mathrm{ATC} = \frac{\mathrm{TC}}{\mathrm{Q}}$$

The shape of the Average Fixed Cost curve is simple: since fixed costs never change, the more you produce, the lower your AFC is going to be. If you have $100 in fixed costs, then at an output of 1 your AFC is 100, at 2 it is 50, at 3 it is 33.3, at 4 it is 25, and so on. So the AFC curve is always decreasing.

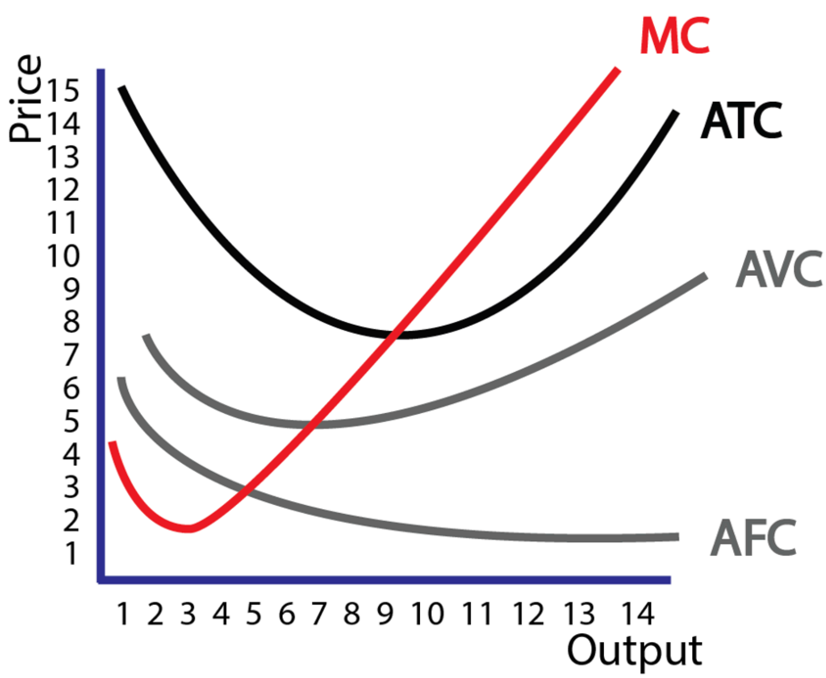


Average Variable Costs will follow the same pattern as the Average Product curve from last lesson, only flipped upside down. They will initially be higher than MC and sloping downward (since falling MC means variable costs are falling on average), but rising MC will eventually surpass Average Variable Costs, at which point they will start to slope upwards. The same idea holds true here with grades and averages that we talked about last lesson; if your grade is lower than your average, then your average is decreasing; if it is higher than your average, then your average is increasing. So MC will cross the AVC curve at its lowest point.



Just like Fixed Cost + Variable Cost = Total Cost, so Average Fixed Cost + Average Variable Cost = Average Total Cost. So the formula is:

$$AFC + AVC = ATC$$



And this means that the Average Total Cost curve is going to be the sum of the AFC and AVC curves, so if you just added them together, you'd end up with something that looks like a mix of these two. MC will once again cross the ATC curve at its lowest point, for the exact same reason as above (just think of grades and averages).

Once the firm starts selling things, we also have to figure out revenue. Thankfully, the revenue formulas follow the exact same pattern. Total Revenue is just the Price of the good multiplied by the Quantity sold, so

$$TR = P \times Q$$

The “average” anything will be the total amount divided by quantity, so the formula for Average Revenue is

$$AR = \frac{TR}{Q}$$

The Marginal Revenue will be the change in revenue divided by the change in quantity, just like we saw for Marginal Cost. Its formula is

$$MR = \frac{\Delta TR}{\Delta Q}$$

So here’s what you need to know from today. You need to be able to:

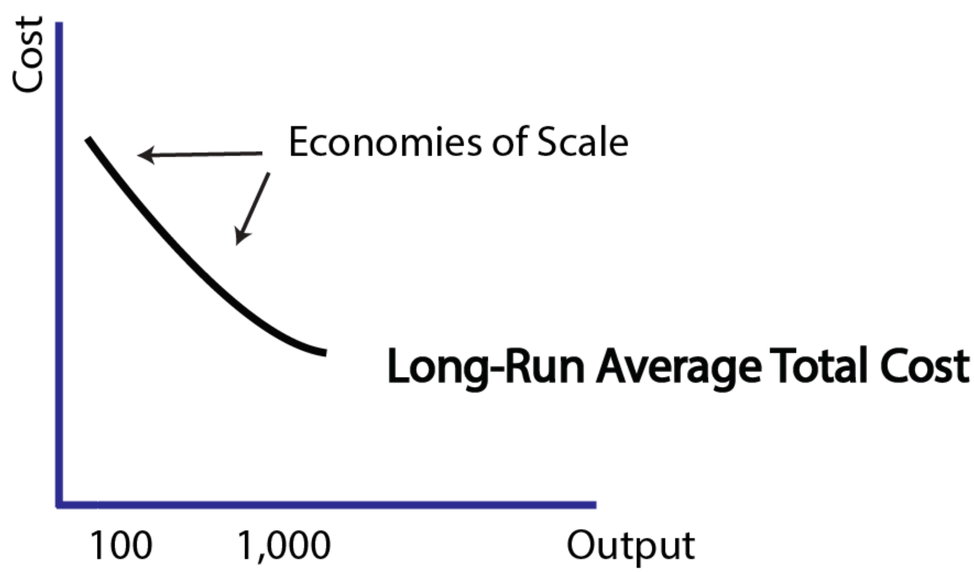
- Define fixed and variable costs
- Know the formulas for TC, AFC, AVC, ATC, and MC
- Be able to graph the MC, AFC, AVC, and ATC curves. What matters for these curves is that
    - MC is shaped like a Nike swoosh, and it starts increasing when you encounter diminishing marginal returns
    - AFC is downward sloping
    - AVC is bowl-shaped, and MC crosses at its minimum
    - ATC is the highest curve. It is also bowl shaped, and MC crosses at its minimum
- Know the formulas for revenue (TR, AR, and MR)

[Return to the top of the document]

3.3 Long-Run Production Costs

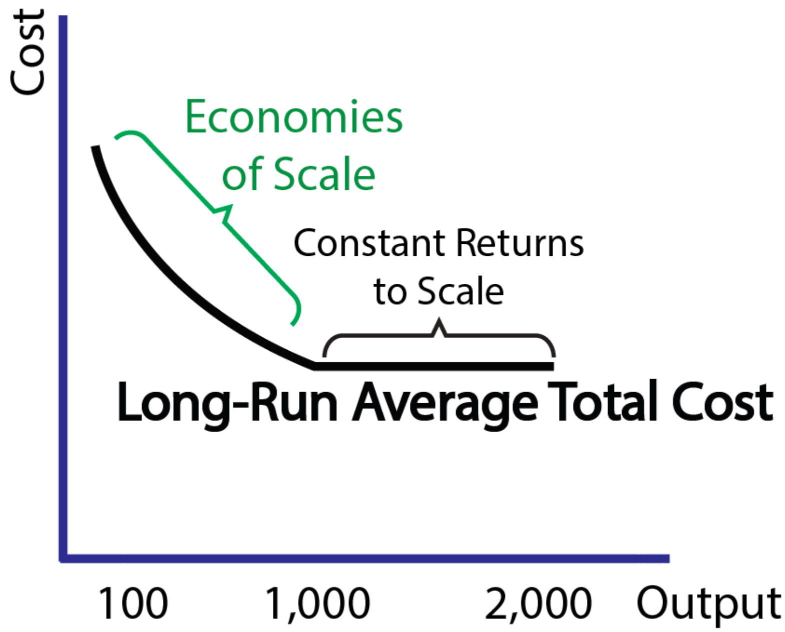
In the short run, all fixed costs are fixed. Once you've signed a contract on your factory, you're stuck in your factory until that contract is up. But in the long run, all these fixed costs become variable: you can close factories or open new ones, restructure your corporation, and buy or sell land and capital. So in the long run, what determines whether your company should be tiny (with only one factory) or huge (with hundreds of factories and hundreds of thousands of workers)?

The answer depends on what the **long-run average cost curve** in your industry looks like. Let's look at farming as an example. If you run a small family farm, growing 100 ears of corn, you can make a profit in the short term, but your farm will be pretty inefficient. You can't buy industrial-sized farm equipment, since it would be wasted on your small land, so you're stuck with the smaller, more inefficient equipment. If you scaled up your operation by buying more land and increasing the size of your farm to 1,000 ears of corn, then the average cost of growing corn would decrease. This is called **economies of scale**, because when you scale up, your production is more economical/cheap on average. Graphically, this means that your long-run average total cost curve is decreasing as you increase quantity.

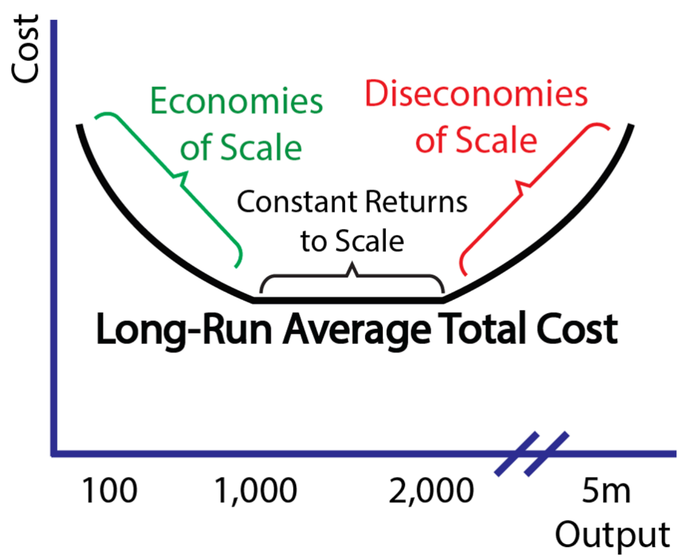


The most important characteristic of a market that is in economies of scale is that if you doubled your inputs (i.e., you spent twice as much money on farm equipment), you would more than double your outputs, since you are becoming more efficient.

Once you have an industrial-sized farm, you could keep on expanding, but you're no longer increasing in efficiency. Going from 1,000 ears of corn to 2,000 ears is simply a matter of doubling how much money you're spending on inputs. If you buy twice as much equipment, you can grow twice as much corn. This is called **constant returns to scale**, because as you scale up your business, your efficiency is constant. That means your average costs are flat.



At some point, continuing to expand is going to start decreasing your efficiency. Running a farm with 2,000 ears of corn is fairly easy, but trying to organize a farm with 5 million ears of corn would be a logistical nightmare—you'd have to hire so many middle managers and upper-level managers that your average cost at that output would be much higher, so it no longer makes sense to keep expanding. We call this **diseconomies of scale**, because as you scale up your business (by, for example, doubling the amount of money you spend on farm equipment), you become less economical/efficient (i.e., you less than double the amount of output you receive). That means your average cost curve at these quantities is rising.



In any given industry, then, firms will tend to fall in the constant returns to scale portion of the long-run ATC curve.

When it comes to questions on the exam, if you ever see anything about a long-run ATC curve, you know that the answer will have to do with economies of scale, constant returns to scale, or diseconomies of scale, depending on whether the slope is down, flat, or up. Also keep in mind that you can tell where you are on the LRATC curve if you increase inputs and you get an increase in output that is lower, equal to, or greater than the increase in inputs.

Return to the top of the document

3.4–3.5 Perfect Competition: Characteristics and Types of Profit

The cost curves we've looked at over the past few lessons are the same for every firm in every type of market. In this lesson, we're going to look at one particular type of market known as **perfect competition**. Perfect competition is characterized by

- *Identical products*. In the market for milk, there's no real difference between Walmart-brand milk and Dairy Farms milk. In these markets, we can assume that the product sold by every firm is identical.
- *Firms are price takers*. There are so many firms in the milk market that no individual farm can influence or set the price of milk. No matter how much milk I produce, I'm just a drop in the bucket compared to the market as a whole, so my firm will not impact the broader market.
- *No barriers to entry/exit*. If a firm is making losses, they can close shop and exit the market. If other firms see that there is profit to be made in this market, they can open up a factory and join the market without major barriers. By "major barriers," I mean that there are no laws barring them from entry (like patent protection), nor are start-up costs so high that it would be prohibitively expensive for firms to join. The market for supplying water to households, for example, has a major barrier to entry: if you wanted to compete, you'd have to lay an entirely new system of pipes to every household in the area, which would be crazily expensive. The market for household water, therefore, is not perfectly competitive.

The most important thing to know is that <u>if there are no barriers to entry/exit, profit will always equal zero in the long run</u>. If people could make a profit by starting dairy farms, then people would start opening dairy farms, thus increasing supply (and decreasing price) until there was no more profit to be made. If firms were making losses running dairy farms, then farms would shut down, thus decreasing supply (and increasing price) until farms started breaking even again.

But when economists say that there is no profit in the long run, they're kind of cheating. What they mean is that there is no **<u>economic</u> profit** in the long run. Remember, in economics the "cost" of something is what you give up, and that includes your opportunity cost. So let's say that I have a choice between running a shoe store, with $100,000/year in revenue, or being a janitor, with a salary of $40,000 a year. The only catch here is that if I decide to run the shoe store, I have to pay $60,000/year in operating costs, which economists call **explicit costs** (explicit because I have to actually pay someone this money). So running the shoe store would make me $40,000/year in **accounting profit**.

Which job should I prefer? From a money standpoint, I would be indifferent between the two. Either way I'd be making $40,000/year. If I chose to run the shoe store, I'm giving up the opportunity to earn $40,000/year as a janitor. This is an **implicit cost**, since I'm not actually paying anyone anything, but it's still a real cost, because by running my shoe store I'm giving up the chance to earn that $40,000. So in running this shoe store, my accounting profit is positive, but my economic profit (which accounts for both implicit and explicit costs) is zero. If my economic profit were positive, then people would start opening more shoe stores, thus driving down price until my economic profit was zero. If my economic profit were negative, I'd quit my job and go work as a janitor, since I could make more money that way. So the main thing to know here is:

$$\text{Accounting Profit} = \text{Revenue} - \text{Explicit Costs}$$

$$\text{Economic Profit} = \text{Revenue} - (\text{Implicit Costs AND Explicit Costs})$$

Perfect Competition Market Structure

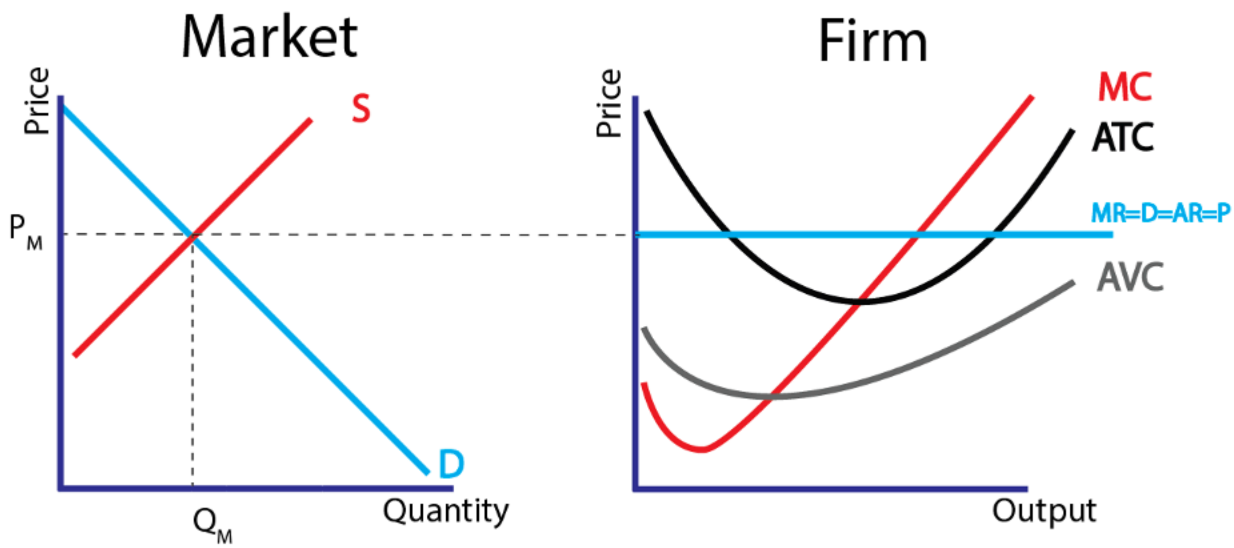
In perfect competition, we divide our analysis into two parts: the market and the firm. The market is made up of everyone, while the firm represents the choices that any particular firm faces. On the market side, we have supply and demand as usual, and these together determine market price and quantity, which we mark as P_M and Q_M. On the firm side, we have all the costs discussed in the previous lessons, and the firm faces the price that the broader market has determined. Remember, the firm is a price taker, so it faces a perfectly elastic demand curve—it can produce as much or as little as it wants, but it can't change the price people pay.

Because price never changes for the firm, that means the average amount of money I bring in per apple (i.e., the Average Revenue) is going to be equal to the market price. If I sell 1 apple for $5 or 100 apples for $500, the average amount of money I bring in per apple is going to be $5—it never changes. We can even see this in the formula. Remember, AR = TR / Q, and TR = P x Q, so Q will cancel out, leaving us with AR = P

$$AR = \frac{P\,x\,\cancel{Q}}{\cancel{Q}}$$

Similarly, Marginal Revenue is also going to be $5. My first apple sold will bring in an additional $5 in revenue, and so will the 100th apple I sell. So on the firm side, I can label this curve as either Demand, Price, Marginal Revenue, or Average Revenue; they are all the same thing. Because economists enjoy lame jokes, we call this "Mr. Darp," to remember that MR=D=AR=P.

BE CAREFUL: If you label this curve as "MRDARP" or "Mr. Darp", you will not get credit on the exam. You have to put the equal signs in (MR=D=AR=P), showing that you know these are four different measurements and yet all equal to the same thing. So here is the full market structure in perfect competition:



Notice, we're not putting AFC on our graph, since we don't actually use that curve for anything in the analysis we're about to do.

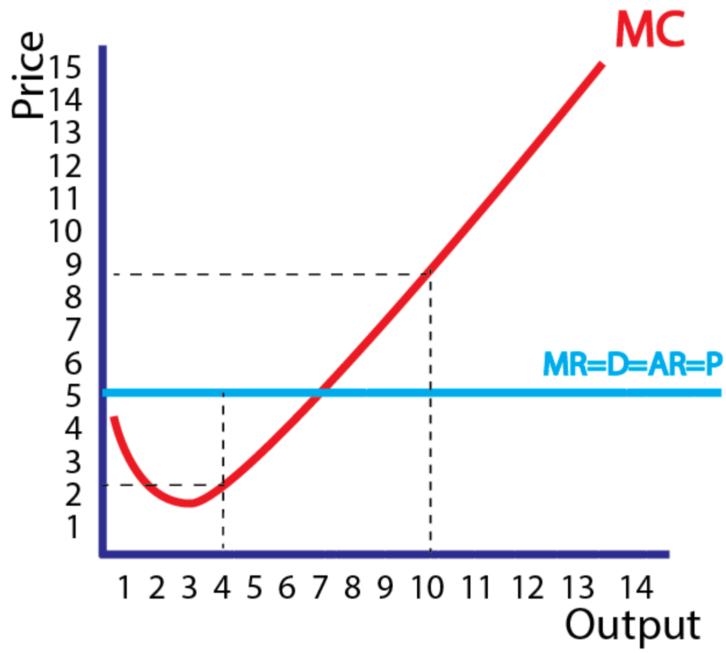
[Return to the top of the document]

3.5,3.7 Perfect Competition: Profit Maximization and Long-Run Equilibrium

Maximizing Profit / Minimizing Loss

Now that we have the model, let's figure out how to use it. When you're trying to figure out what a firm will do, there are always two steps.

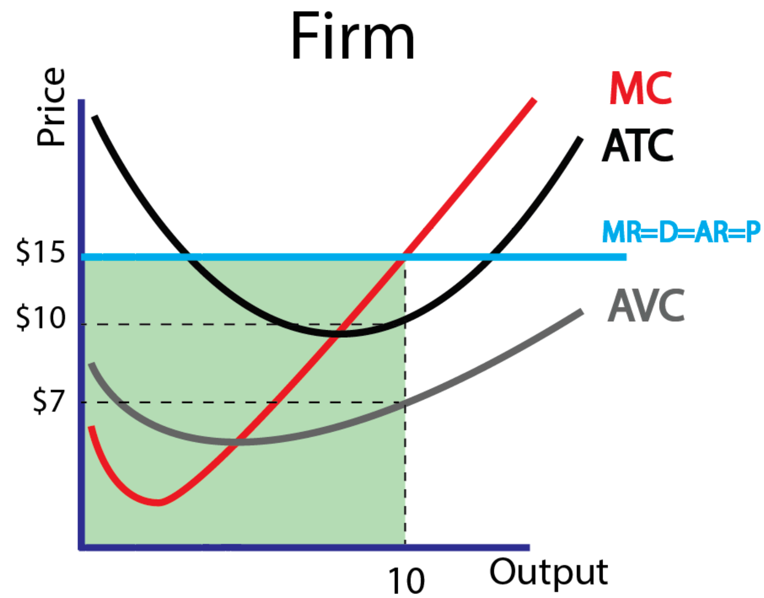
1. *Figure out the profit-maximizing (or loss-minimizing) quantity to produce.* When you do this, you only look at your marginal values: you want to produce up until the point where Marginal Cost = Marginal Revenue, and no further. In the example shown here, if I produced 4 units, then at that price my MC is only $2, while my MR is $4. Making one additional unit here would only cost me $3, but I'd be making $4 on that extra unit, so I should definitely increase production. If, on the other hand, I were making 10 units, the marginal cost of that 10th unit is $9, but it's only bringing in $4 in revenue, so it makes no sense to produce that much. So set your quantity where MC=MR. This, by the way, is an ironclad law of economics: t<u>o maximize profit, you always produce at a quantity where MC=MR</u>. The quantity the firm produces is labeled Q_F.



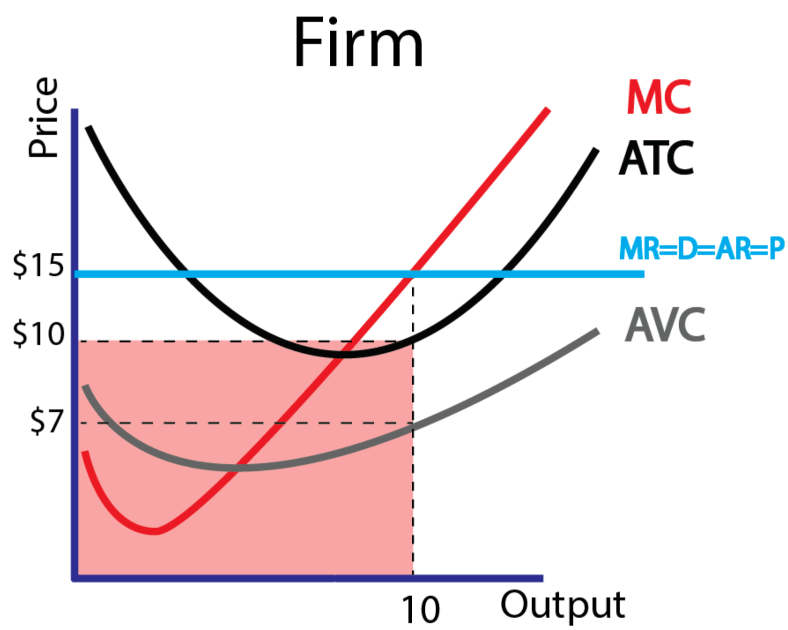
2. *Figure out if you are making profit or loss*. You used the <u>marginal curves to figure out how much to produce</u>, but now that you have this quantity set, you use the <u>average curves to figure out profit/loss levels</u>. The reason we do this is because of a quirk in the way the average curves are set up. Remember, AR = TR/Q, and ATC = TC/Q. In both cases, if we multiply by quantity, Q will cancel out of the formulas and we'll be left with Total Cost and Total Revenue. Since profit is just Total Revenue minus Total Cost, the average curves give us an easy shortcut to getting the profit (plus an easy way to visualize profit graphically).

   Let's start with total revenue. Now that we have Q_F (where MC=MR), we can take the Average Revenue at that quantity and multiply it by Q_F, which will give us TR (remember, $AR = \frac{TR}{Q}$, so if we multiply AR by the quantity produced, the Q cancels out: $AR = \frac{TR}{\cancel{Q}} x \cancel{Q}$)

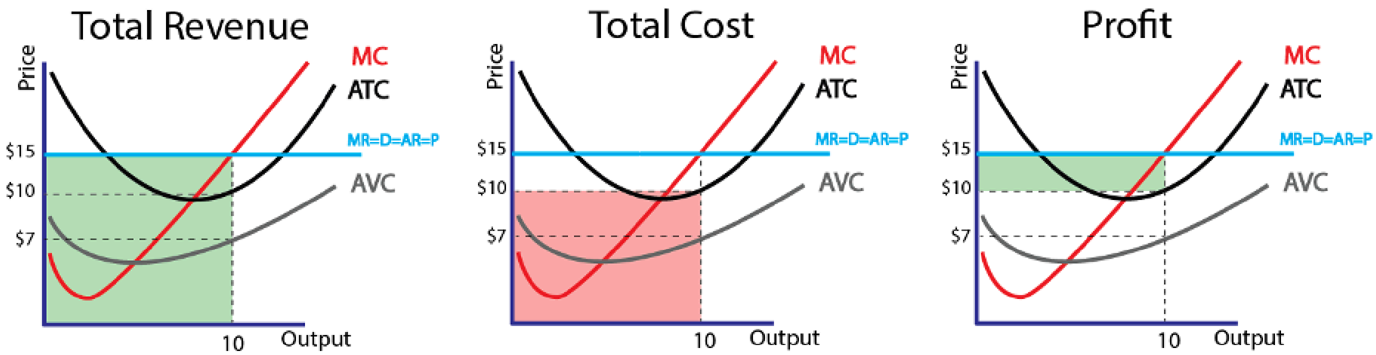
That means the rectangle on the right, with a height of AR and a length of Q_F, will be equal to the total amount of revenue the firm is bringing in. Checking our work, $15 x 10 is indeed $150, which is the total revenue we'd expect if 10 units were sold for $15 each.



We can do the same thing with Average Total Cost. If we multiply ATC by Q_F, the Q's will once again cancel out and we'll have our Total Cost. At this quantity, ATC is $10, so $10 x 10 gives us a total cost of $100. We can graph this out below:



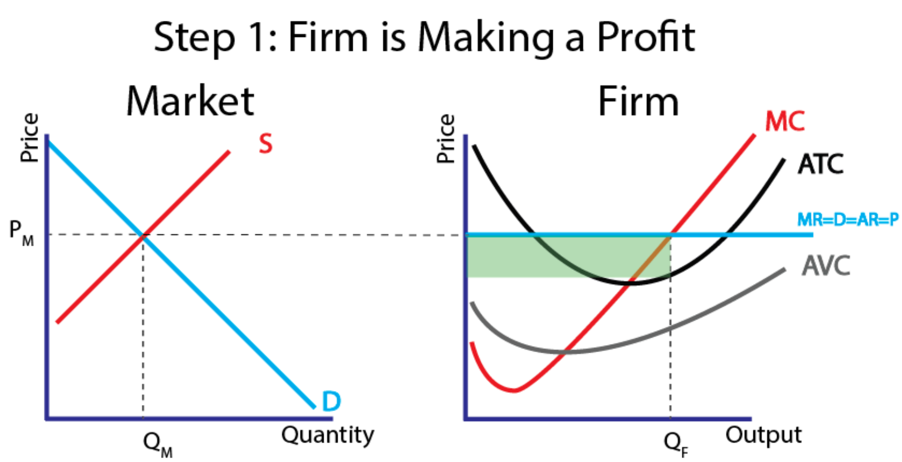
Profit, remember, is Total Revenue minus Total Cost. So to figure out whether this firm is profitable, all we need to do is look at the difference in area between the green (TR) and red (TC).

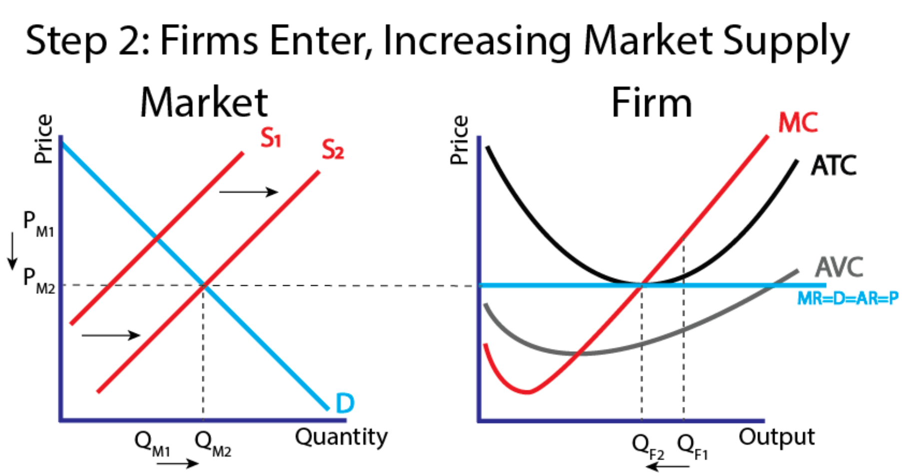


In this case, since AR is greater than ATC at this quantity, the firm is profitable, with a total profit of $5 x 10, or $50.

Long-Run Equilibrium

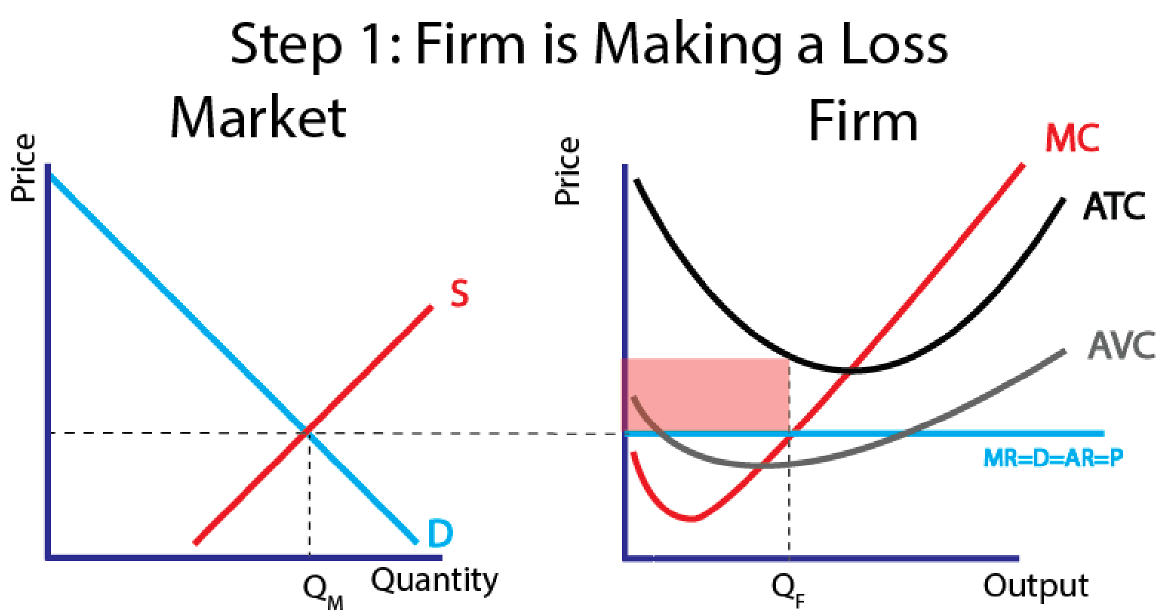
We said earlier that firms cannot be profitable in the long run, and that's still true. This firm can earn a profit in the short run, but once other firms see that there is profit to be made in this market, they will start to join. This mass entrance of other firms increases the supply curve for the market, which pushes prices down, and prices will fall until there is no more profit to be made.

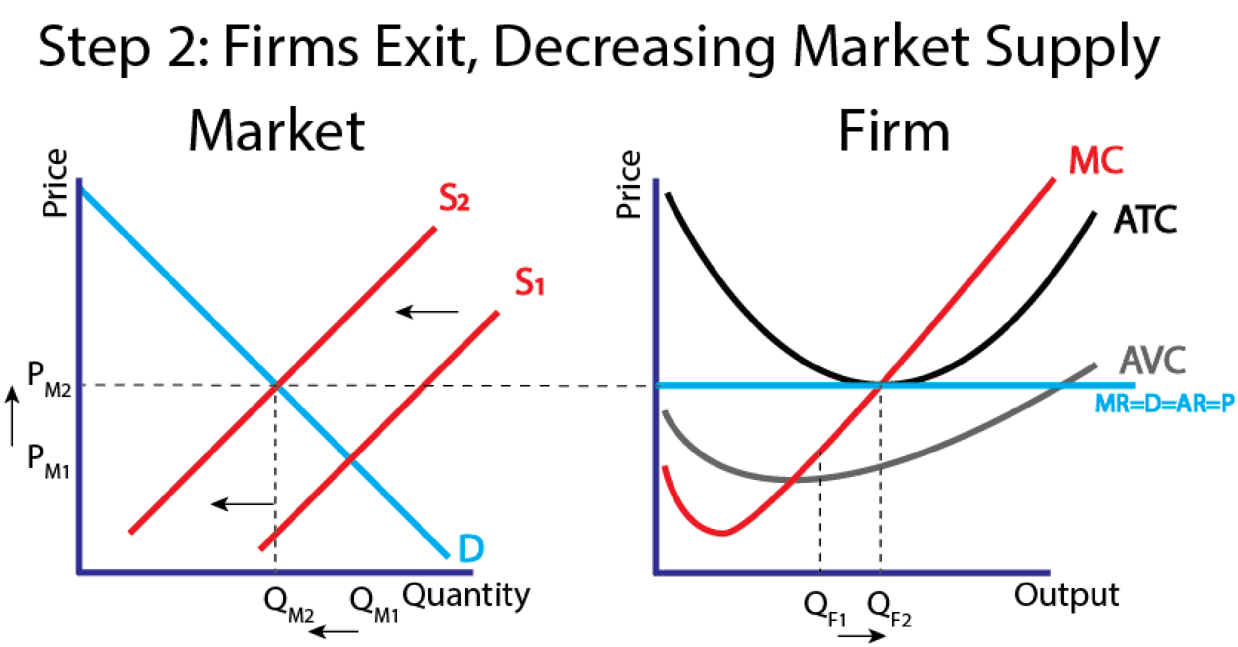




As price falls, the firm will continue to set quantity where MR=MC. Once we get to Q_{F2}, now our Average Revenue and Average Total Cost will be equal, which means there will be no more profits. Now this market is in long-run equilibrium. Notice that in getting here, the market quantity increases (since more firms have joined), but the quantity for each firm has decreased.

Let's walk through this process if the firm is making a loss, looking at the graphs below. Notice how at this price, firms are still going to set MC=MR, but now at this quantity ATC is higher than AR, which means the firm is making a loss. Firms will ultimately decide to leave this market, since they're making a loss, which decreases market supply. This raises price, and firms will continue to leave / price will continue to rise until firms are breaking even again at our long-run equilibrium point, where MC=MR=AR=ATC.

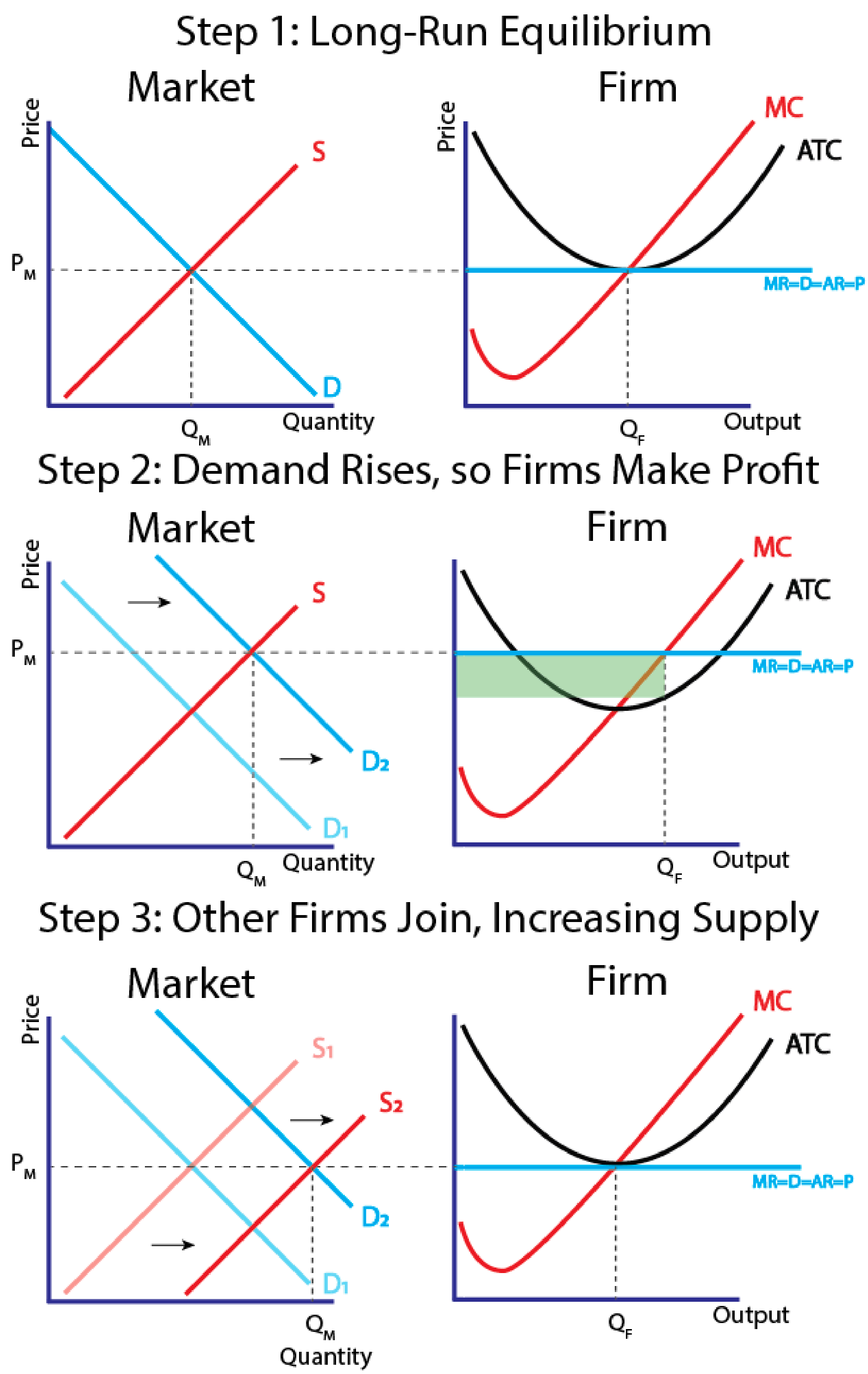




Changes in Demand

In working with this model on the exam, one of the most common problems you'll have to work through is a change in market demand. In this scenario, imagine that the market starts in long-run equilibrium, but then something happens to make market demand increase. This leads the firm to start making a profit, which in turn leads to other firms joining the industry. Supply will then increase until we are back at long-run equilibrium.

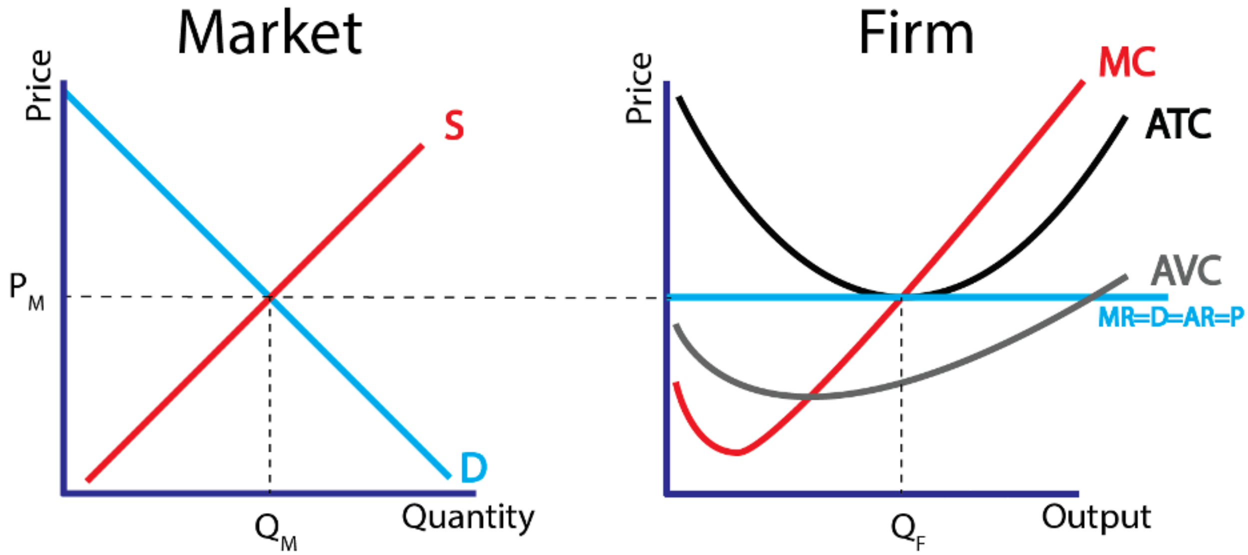
[Return to the top of the document]



ıy, and modify.

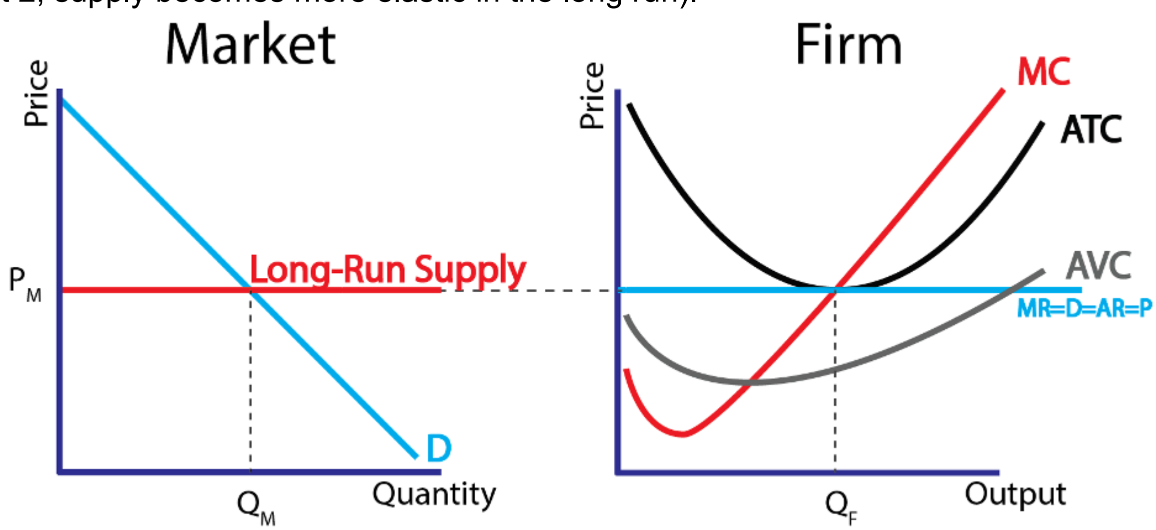
3.7 Long-Run Equilibrium and the Long-Run Market Supply

When you look at a market's long-run equilibrium, there are a few things to note. First, this market is **allocatively efficient**. Every firm is producing at a point where marginal cost = marginal benefit (i.e., where supply meets demand), and we've maximized market welfare.

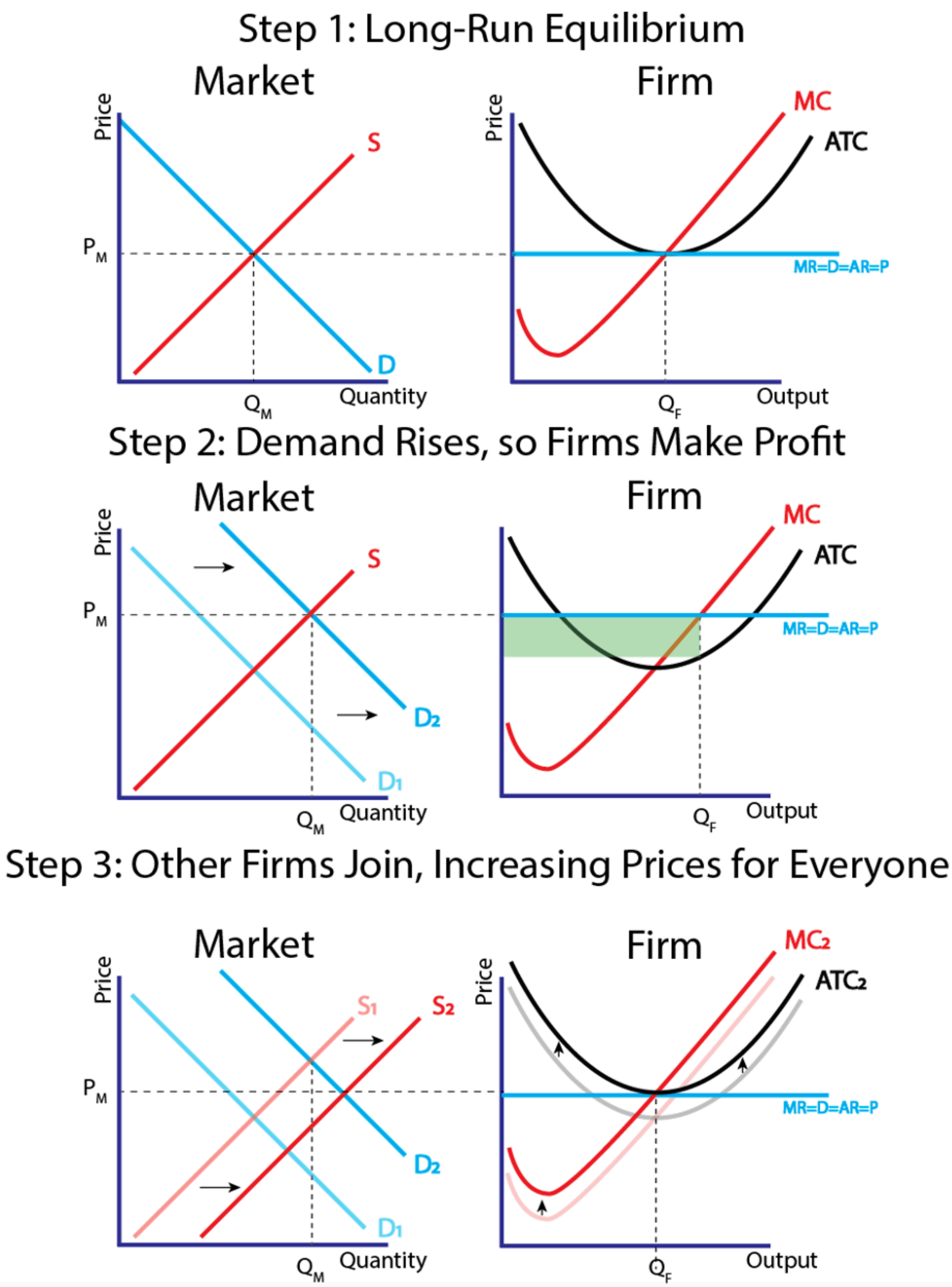


Second, this market is also **productively efficient**. Productive efficiency means that the firm is producing at the minimum of the ATC curve, i.e., average costs are as low/efficient as they could possibly be.

Finally, if you look at where P_M is right now, you'll notice that in the long run, price can never deviate from this point. The market quantity may rise and fall as demand changes and as firms enter or exit the market, but we will always return to PM in the long run because this is the point of zero economic profits, where each firm ends up at MC=MR=ATC=AR. So if we wanted to draw the **long-run market supply curve**, it would be perfectly horizontal at that line (which makes sense, because as we saw in Unit 2, supply becomes more elastic in the long run).

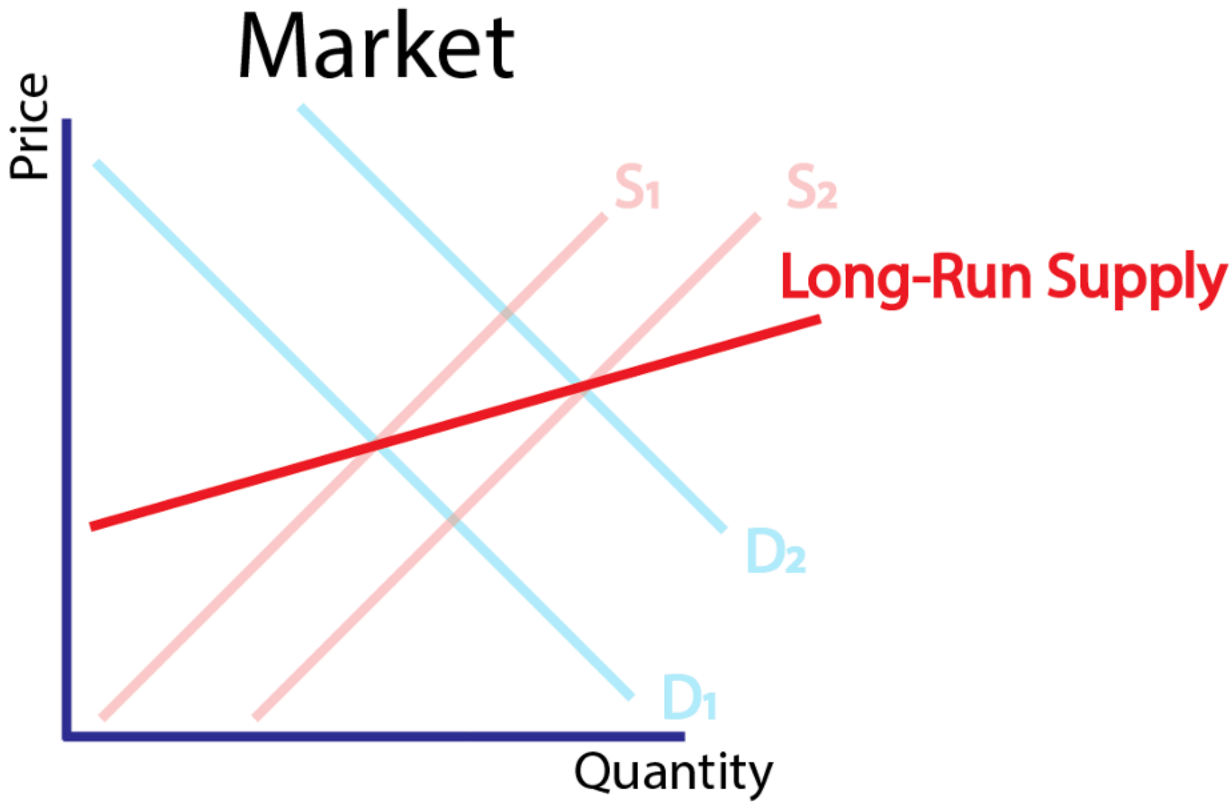


This long-run supply curve assumes that we are operating in a **constant-cost industry**, i.e., as more firms enter or exit the market, this does not impact the costs that other firms face. There are some industries, however, where this is not true. Take for example the electric car industry. One of the most important inputs in electric cars is cobalt for the batteries. But if firms are making a profit in this industry and more firms decide to join, this increased demand for cobalt will drive prices up for everyone.



ify.

If you connect these two market equilibrium points (and all possible market equilibrium points), you'll end up with a long-run supply curve that looks like this:



This is called an **increasing-cost industry**, because as more firms join, costs increase for everyone.

Alternatively, firms could be operating in a **decreasing-cost industry**, so named because—you guessed it—as more firms join, costs decrease for all firms. This is extremely rare in practice, though it could happen in something like farming, where the more crops people grow, the more plentiful (and therefore cheaper) seeds become. The long-run market supply curve for a decreasing-cost industry is downward sloping.

[Return to the top of the document]

3.6 The Shut-Down Point and the Decision to Exit the Market

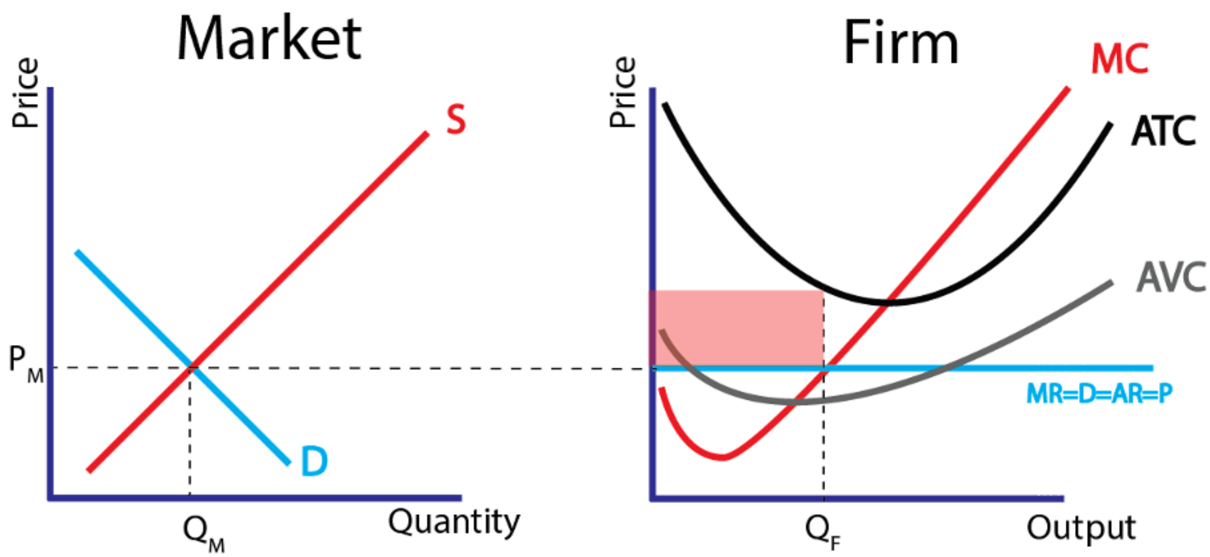
As we've seen, if the firm is producing at a quantity where AR is below ATC, that firm is making losses, and it eventually must exit the market. But of course, firms aren't going to go out of business the minute that demand falls and prices slightly dip. Most firms will continue to operate at a loss in hopes that prices will eventually rise again and they will return to making zero economic profit. They only exit the market if price stays below ATC *in the long run*.

In the short run, firms will stay open as demand and market prices fluctuate, but at a certain point, if losses are too great, they will shut down in the short run as well. Today's topic is about figuring out where that shut-down point is.

The main thing to keep in mind is that in the short run, fixed costs cannot be changed. Economists refer to these as **sunk costs**, because you have already sunk that money into your business, and you can't get it back. If you've paid your rent on the factory for this month, then you need to recognize that no matter what you do, that money is gone, so you should ignore it. Sunk costs should never factor into a firm's decision about how much to produce or whether to keep their doors open in the short run. Obviously firms care about sunk costs in the long run—after all, at the end of the year the firm needs to be breaking even—but there's no use crying over spilled milk in the short run. That is why we don't even draw the AFC curve on our graph; we always ignore fixed costs.

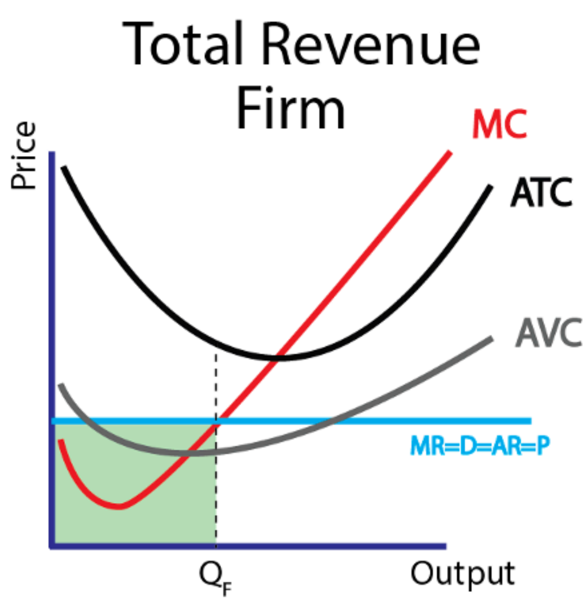
So let's go back to our firm that's making a loss. If this firm is deciding whether to stay open or shut down *in the short run*, then the only thing that matters is variable costs. Think of a restaurant that mainly serves lunch and dinner but that is trying to decide whether it should open or shut down during breakfast time. The restaurant owner doesn't care whether the revenue from selling breakfast will cover the rent or not—we already know that demand is too low during the morning to make the restaurant profitable. The only thing that matters is whether the revenue from breakfast will cover the costs the owner will incur by opening a few hours earlier. If she opens the restaurant during breakfast time, she'll have to pay the servers and busboys, plus the bill from the extra electricity and water usage. So the question is: can the revenue from opening for breakfast cover those variable costs? If it can, the restaurant should stay open. If it can't, the restaurant should temporarily shut down until demand rises to the point where it *can* cover its variable costs.

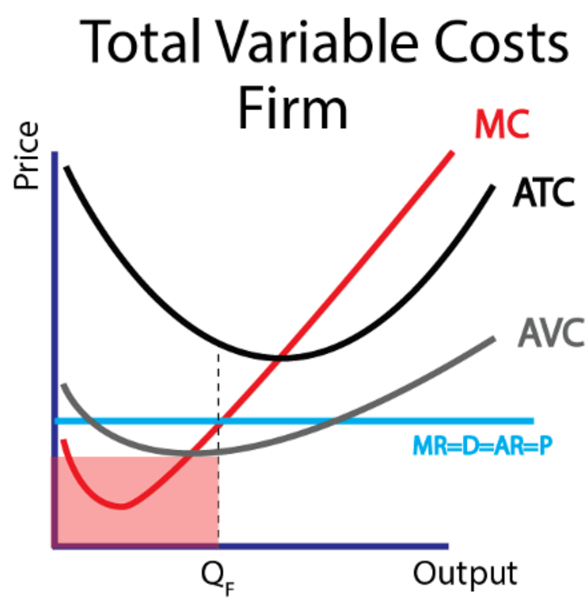
To see this graphically, let's look at the scenario below.



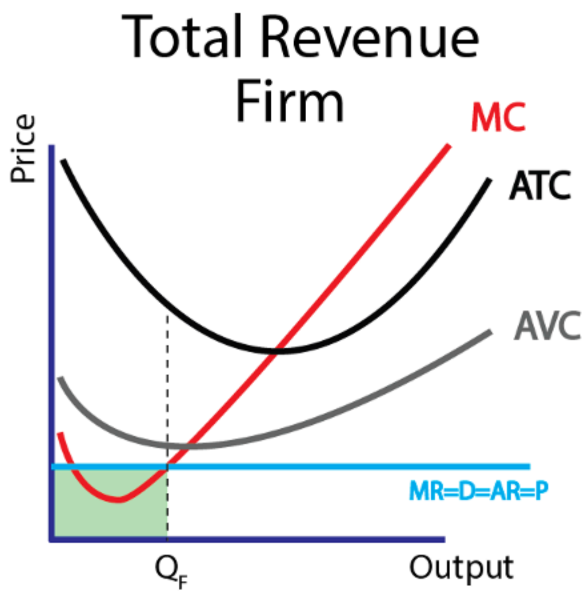
In this case, market price is low, so at the firm's profit-maximizing quantity, ATC>AR, which means costs are higher than revenue, so the firm is making a loss. But if you look closely at this quantity, you'll notice that AR is higher than Average Variable Costs. We can use this curve to find the total Variable Costs in the same way we did earlier with the ATC curve, namely by multiplying it by the quantity.

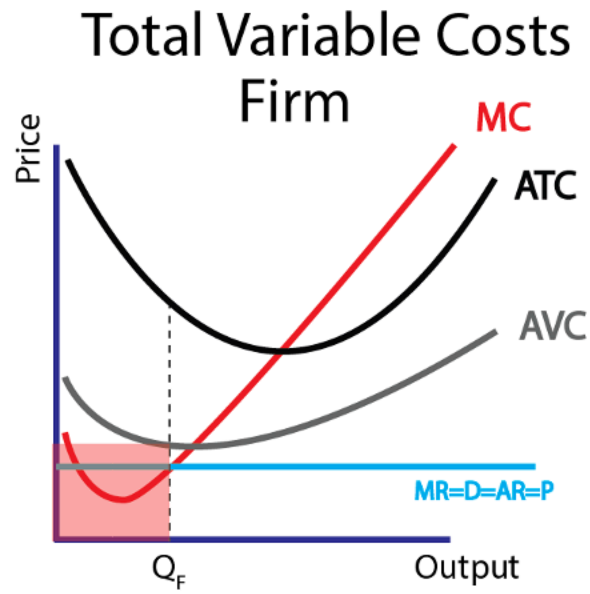
If we compare the total revenue to the total variable costs, we can see that TR>VC, so the firm should stay open in the short run. The firm still isn't making a profit, but it is at least covering its variable costs.



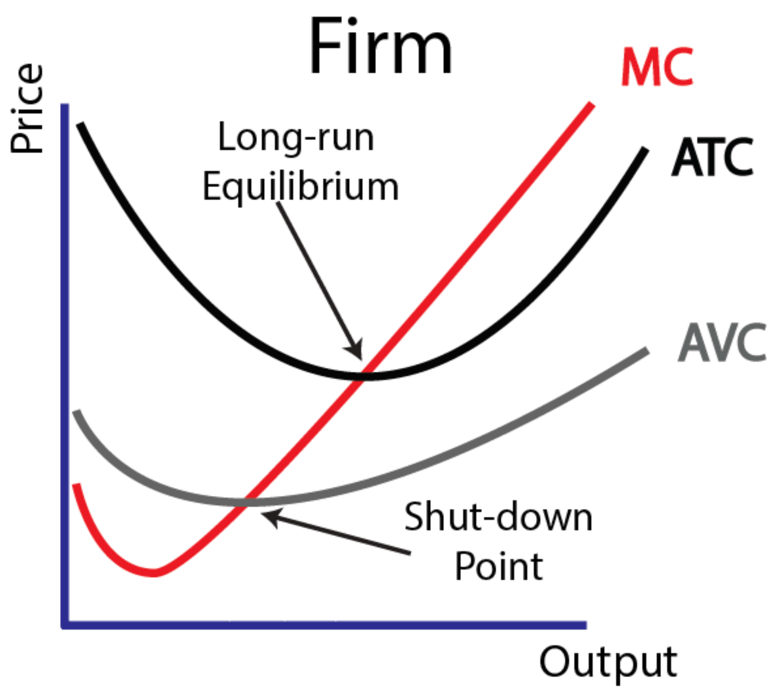


As you might imagine, the firm will shut down in the short run if AR falls below AVC, since this means revenue cannot cover the firm's variable costs. You can see an example of this below:





If we were to label this graph, then, we would label the minimum of ATC as the long-run equilibrium point (because below that the firm can't make enough money to cover all of its costs), while the minimum of the AVC curve would be the shut-down point (because below that the firm can't make enough money to cover its variable costs).



[Return to the top of the document]

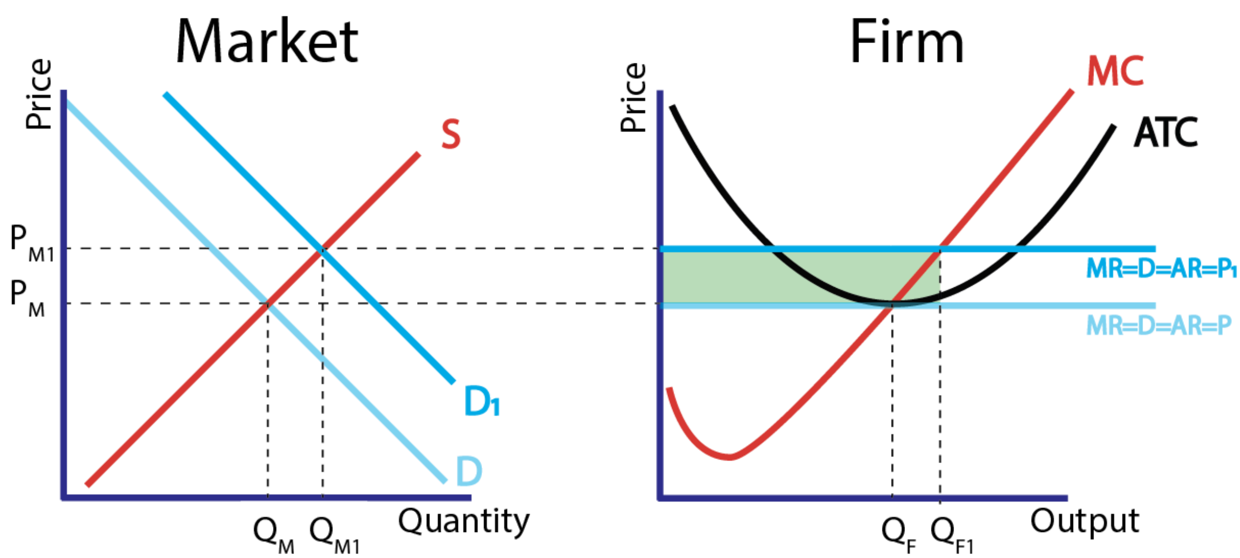
BONUS READING: The Most Common Mistake in Perfect Competition

I'm going to give you a sample free-response question and student response, and see if you can spot the mistake this student makes. I'll even give you a hint: the mistake is not in part a), nor is it on the Market side of the graph.

Question:

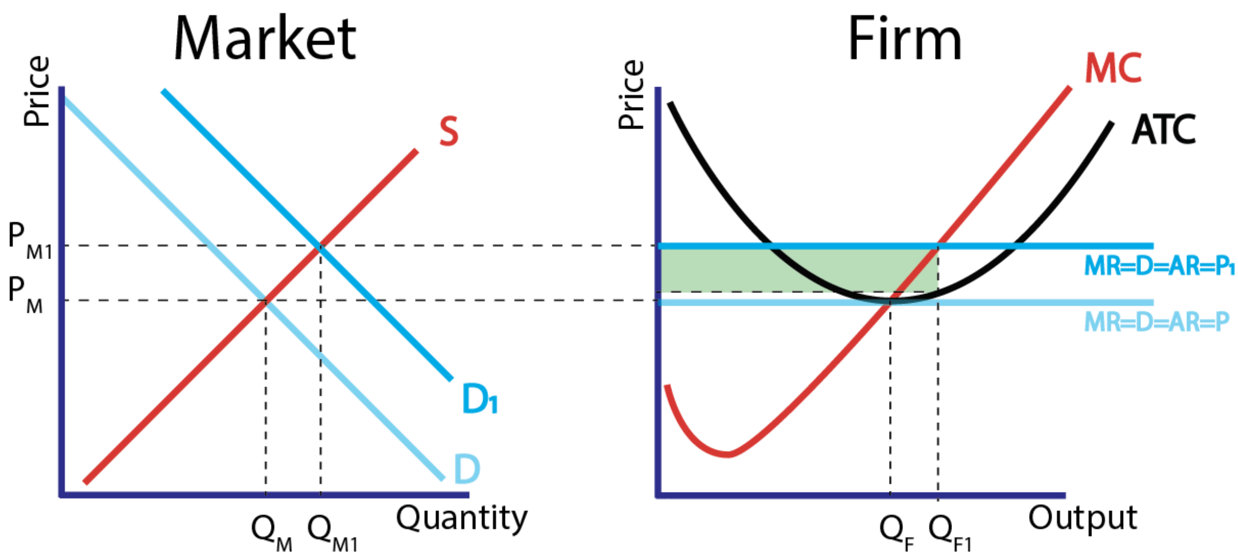
a) A firm in a perfectly competitive market is in long-run equilibrium. Draw the graphs for the firm and the market showing P_M, Q_M, and Q_F.
b) Assume that demand increases in the broader market. In the graphs from part a), show the impact of this increase in demand. Label the new price as P_{M1} and the new quantities as Q_{M1} and Q_{F1}. Shade in the area that represents the amount of profit or loss the typical firm is making.

Student Answer:



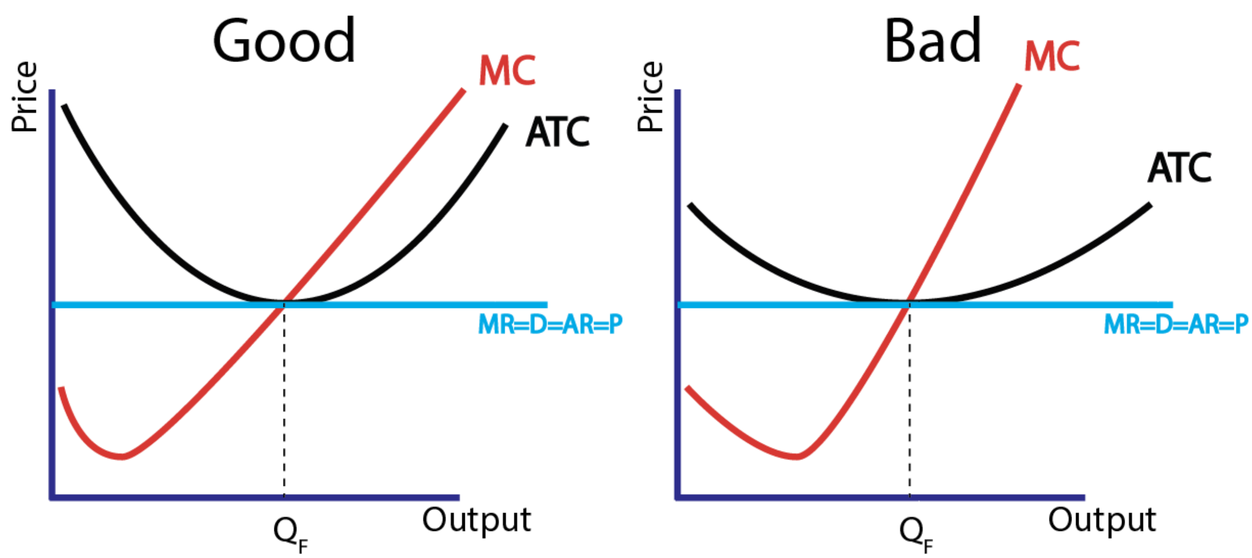
If you didn't catch the mistake here, don't worry: you're in the same boat as the vast majority of Econ students who see these questions.

The problem with this student's answer is that the rectangle representing profit goes all the way down to the original market price. But the height of the rectangle for profit should only extend down to ATC (remember, profit is the difference between Average Revenue and Average Total Cost). So the student should have drawn their graph like this:



I know, you're probably thinking that this difference is tiny, but it's the kind of difference that differentiates the 4's from the 5's on the AP exam.

To make sure you don't get caught in this trap, make sure you draw your graphs really large. You need to be able to show this tiny difference. Also, I strongly recommend you draw your MC curve at a 45° angle and make your ATC curve slope up sharply from the point where it meets MC. That way, any change in the broader market will make a large (and easy to draw) difference in quantity and profit levels.



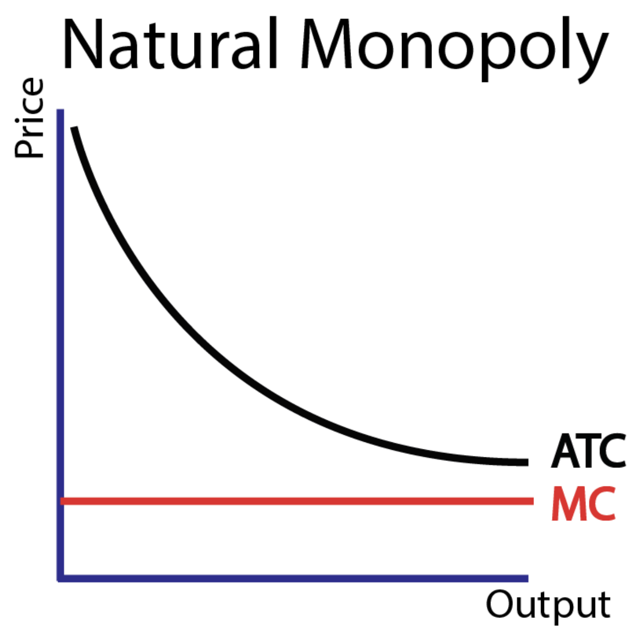
[Return to the top of the document](#)

4.1–4.2 Monopolies

In the previous unit, we drew the market and firm graphs side-by-side, since any one firm only represented a tiny fraction of the overall market. Today, we look at what happens when only one firm controls the entire market, a situation known as a monopoly. So we're going to be combining those two graphs from last unit into one.

There are two types of monopolies you need to know. The first is a regular, run-of-the-mill monopoly that can be caused by any barriers to entry. These can be legal, such as when the government gives a patent to a drug company, or they can arise from high start-up costs, like we saw in the previous unit.

The second type of monopoly is a **natural monopoly**, and this happens whenever the industry is structured such that a firm would always operate in the economies-of-scale portion of the long-run ATC curve, so its ATC curve is always downward sloping. This happens, for example, with social media companies. Meta (the parent company of Facebook and Instagram) holds a natural monopoly, because the marginal cost of adding another user to its platform is essentially zero, no matter how many users it adds. In fact, the more users it adds, the cheaper the entire system becomes to run per user, which means Meta will always be in the economies of scale portion of the LRATC. Its graph would look something like the one on the right.



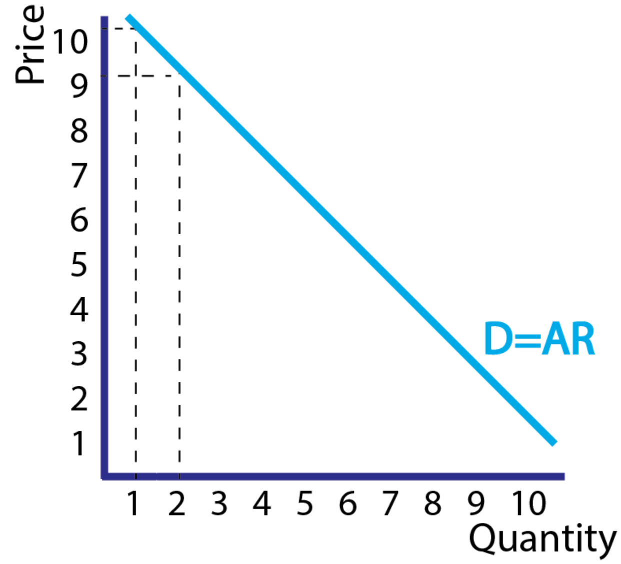
The result of a natural monopoly is that it will be impossible for a competitor to enter this market. Once Meta has become big enough, no firm can enter the market because their average cost will be so much higher than Meta's, so they can never compete on price. Incidentally, Meta was sued in 2020 by the FTC for abusing its monopoly power.

Marginal Revenue and the Demand Curve

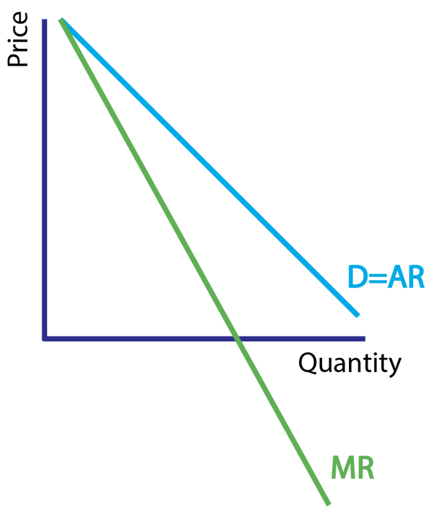
The biggest difference between monopolies and firms in perfect competition is that monopolies face a downward-sloping demand curve, and they are price makers, not price takers. That means our revenue calculations are going to be a little different.

Average revenue for monopolies is easy. Whatever the price is, that's the average revenue as well. That's because if I sell X number of goods for $5 each, the average revenue will always be $5. So I'm going to label my demand curve as Average Revenue as well.

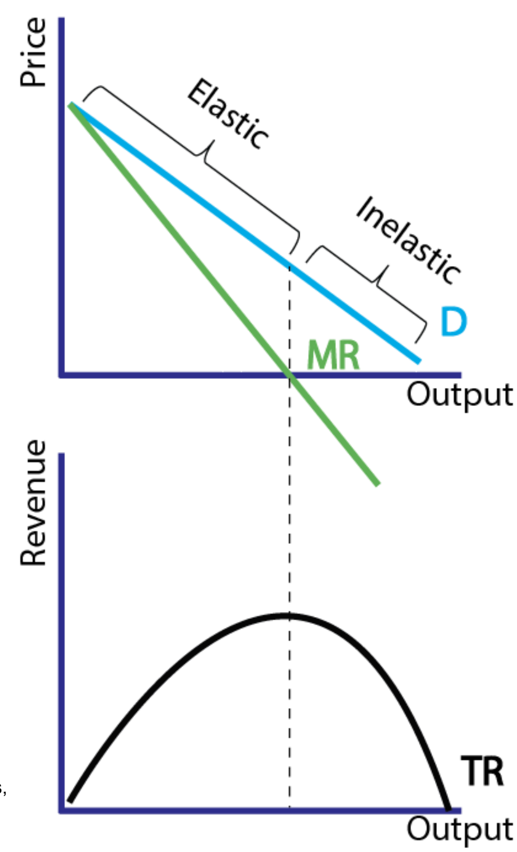
Marginal revenue is a bit weirder. To see how this works, look at the demand curve on the right. If I sell 1 good for $10 each, my total revenue is $10. If I want to sell more than one good, however, I have to lower my price, so to sell 2 goods I have to sell them for $9 each. At this price my total revenue goes from $10 to $18. So by selling one more item at $9, I'm only gaining an additional $8 of revenue (MR=8). Why? Because to sell the second item, I had to lower the price for the second item *and the first item*. So I'm gaining $9 in revenue from selling the second item, but I'm losing $1 of revenue from that first item, since I can't sell it at $10 anymore.



This same thing happens if I sell three items for $8 each. Now my total revenue is $24, which is only $6 higher than when I was selling 2 items. Why? Because now I've had to lower the price on all three of my items to be able to sell them, so I've lost that extra dollar on all three. In fact, as I get to the end of my demand curve, marginal revenue actually becomes negative. If I'm selling 9 items for $2 each, I get $18 in revenue, but if I sell 10 items for $1 each, I only get $10, so my marginal revenue on that 10th item is –$8. So if I were to graph the marginal revenue out, it would look like the graph on the left.



The marginal revenue curve also helps us analyze the elasticity of the different points on the demand curve. Whenever MR is positive, that means Total Revenue (TR) is increasing, and when MR is negative, TR is decreasing. So everywhere that MR is above the x-axis, a decrease in price will lead to an increase in revenue. If you remember your revenue test for elasticity, that means the demand curve is elastic at all those points. Where MR crosses the x-axis the demand curve will be unit elastic. And where MR is negative, a decrease in price will lead to a decrease in revenue, so the demand curve will be inelastic (see the graphs to the right).

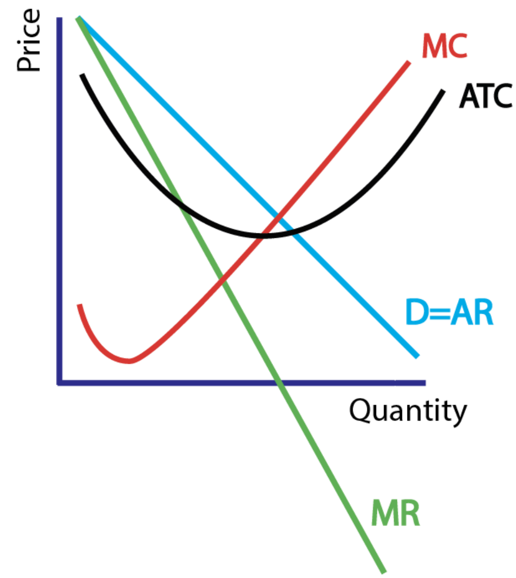


(reading continues on the next page)

Alex Douglas,

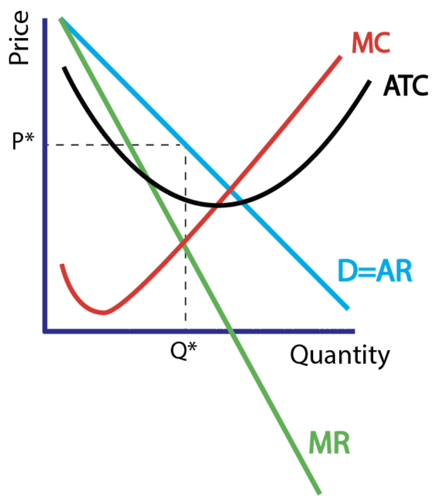
Monopoly Production Decisions

To analyze how monopolies function, we now need to add the firm's cost curves. In the graphs that follow, I leave out the AVC curve just to keep the graphs from getting too crowded, but the curve is still there, and it still works the same way: if price ever falls below AVC, the firm should shut down in the short run.

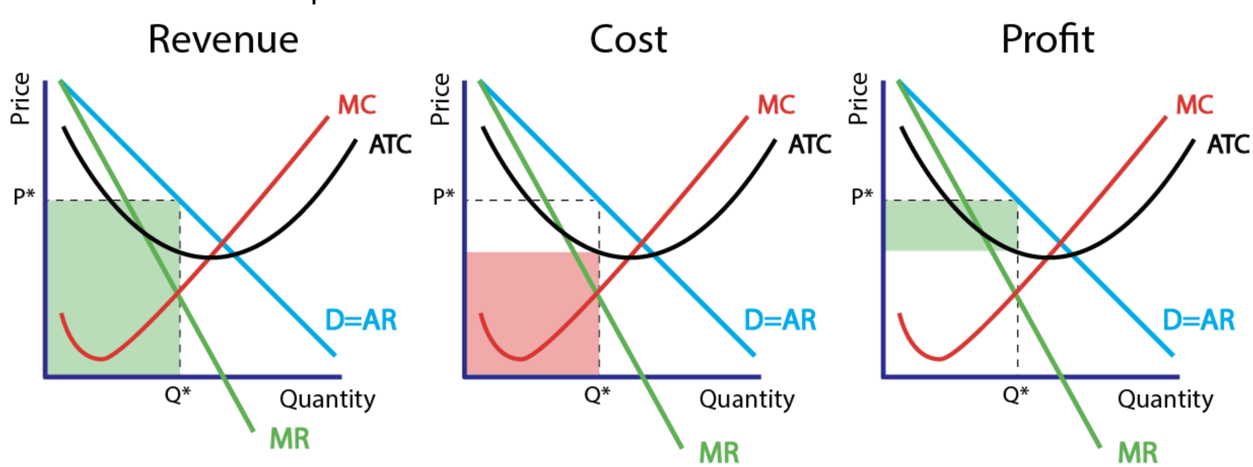


When I add the cost curves, we end up with a market that looks like the graph on the right. And just like we did in perfect competition, we're going to follow the exact same steps of using the marginal curves to figure out the profit-maximizing quantity, and then using the average curves to figure out profit.

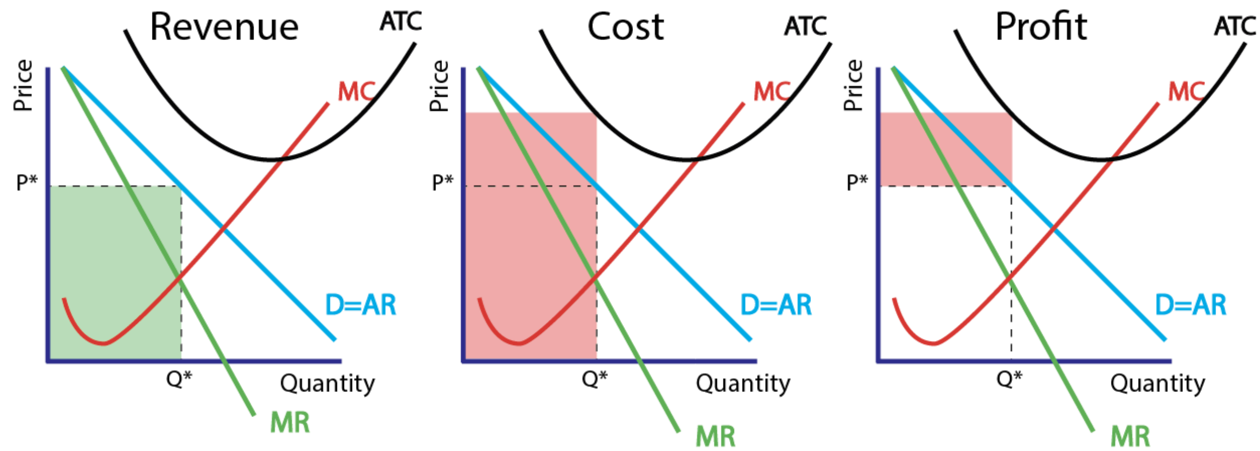
First, remember the ironclad law of economics: to maximize profit, you always produce at a level where MR=MC. Once you've set your quantity, the step that throws everyone off is that you don't use this intersection to set the price. Rather, a monopolist will set their price as high as the market will stand, i.e., they will go all the way up to the demand curve and set price there, as we see in the graph to the left.



To figure out profitability, we once again use the average curves at this quantity. Since Average Revenue is higher than ATC at this point, we know the firm is going to be profitable, but let's see by how much. Average Revenue x Quantity will give us total revenue, and Average Total Cost x Quantity will give us total cost, and we can see the two rectangles representing those areas below. The difference between them is going to be profit, and in this case, the firm is profitable.



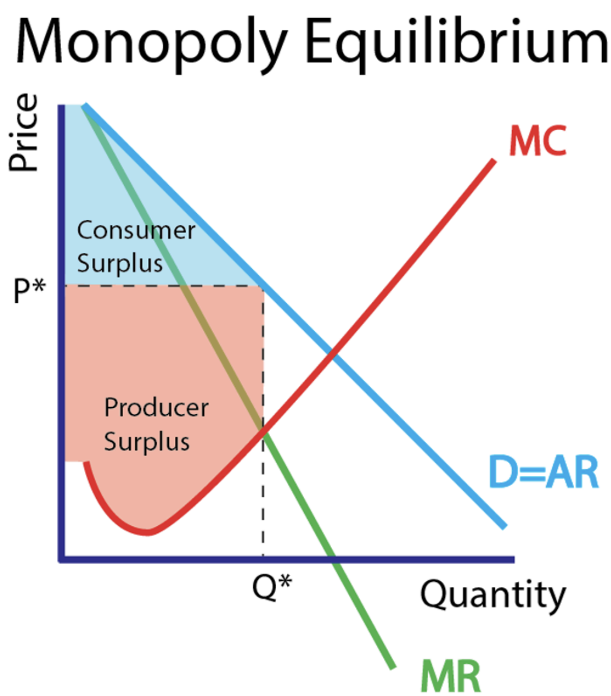
For the sake of completeness, let's look at a firm that is making a loss:

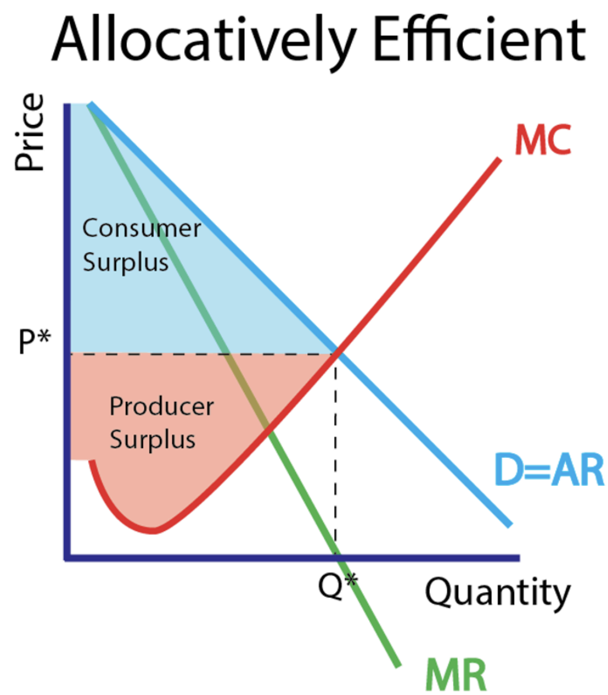


[Return to the top of the document](#)

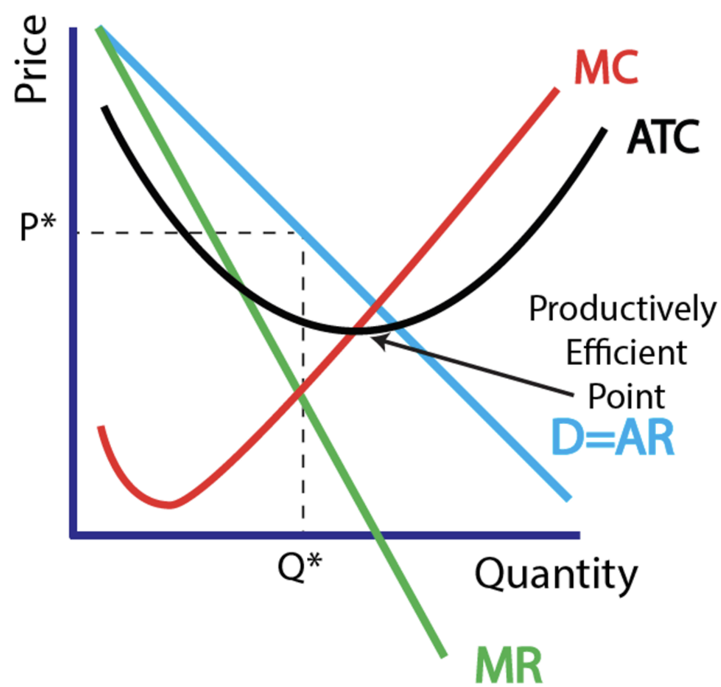
4.3 Price Discrimination

Having a monopoly may be good for the company that has it, but it is bad for society as a whole. This is because monopolies are not allocatively efficient, meaning they do not maximize the total amount of market welfare. Remember, allocative efficiency is achieved when the marginal cost of the last unit produced is equal to the marginal benefit of the last customer, i.e., when Q is set where supply meets demand. For a firm, the MC curve is their supply curve. You can compare below the actual surplus under monopolies vs. the surplus if monopolies produced at the allocatively efficient outcome.





Another shortcut to see if we are at the allocatively efficient outcome is if the price charged is equal to the marginal cost, which it is not in this case. Monopolies don't produce at this socially optimal quantity because doing so would hurt their profits. Monopolies also do not produce at the productively efficient point, at the minimum of ATC (as seen on the right).



Price Discrimination

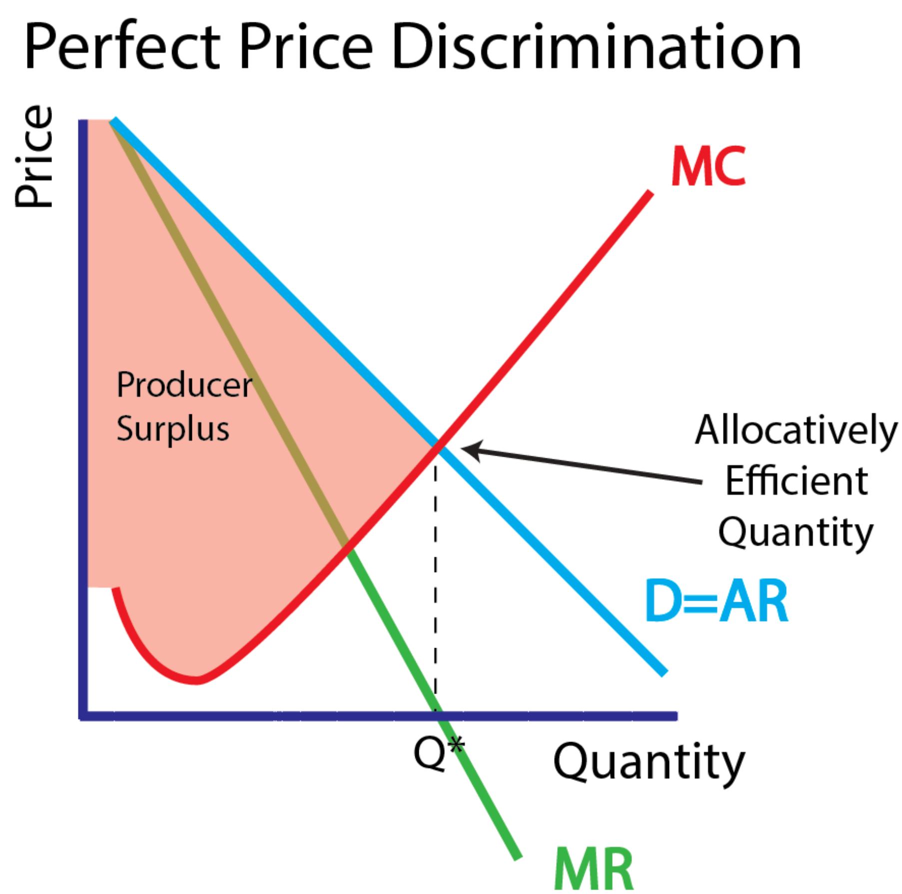
The reason that monopolies produce less than the socially optimal amount is because monopolies have that steep

Marginal Revenue curve, and they get the most profit if they produce where that curve meets Marginal Cost. And as we saw in the previous lesson, this MR curve is so steep because when monopolies increase production, they have to decrease the price for *all* goods in order to sell more, which cuts into revenue. But what if monopolies could find a way to sell extra quantity at a lower price while still selling some at a higher price? That is, what if they could figure out what kind of people are at the top of the demand curve—and thus are willing to pay a lot for the good—and sell the good at a higher price to them, while selling at a lower price to those who are lower down on the demand curve? In fact, most monopolies do just that. It's called **price discrimination**.

You've seen price discrimination all the time. For example, companies know that on average, college students are poorer (and thus are lower down on the demand curve) than other customers, so they usually offer a "student discount." They don't offer this discount because they feel bad for students; they offer it because it allows them to charge customers two different prices: those who are higher on the demand curve pay one price, and those who are lower pay another. It's not perfect—after all, there are plenty of rich students and poor non-students—but it works well enough to allow companies to increase their profits.

In fact, any time you see the same good being sold for different prices, this is the result of price discrimination. It happens with early-bird specials, coupons that you cut out of the newspaper, senior discounts, and many other places.

If companies are even more savvy, they will differentiate between customers to an even greater degree. When you look at airline tickets, everyone is buying the same service, but almost everyone is paying a different price based on what should be fairly meaningless variables, such as the time of day the ticket is bought, how far in advance you buy it, whether you are traveling on a weekday or a weekend, whether you sit by a window, whether you've searched for that itinerary before on your browser, or even what zipcode you are in when you log on to buy the ticket. Airlines track all of this data because it helps them identify how much people are willing to pay on average, and they can change prices accordingly.



If a monopoly had perfect information, they could engage in

perfect price discrimination, where they charge each person exactly how much they would be willing to pay. In a situation like this, every customer would pay a different price, and since this price would be the max they're willing to pay, there would be no more consumer surplus. And since they don't have to lower prices to sell more quantity, the MR curve wouldn't matter anymore, so they could increase quantity all the way up until MC=Demand. Everything then would be producer surplus, as in the graph above.

From an economist's perspective, this is actually a good outcome: total market welfare is as high as it could possibly be, and there is no longer any deadweight loss. Of course, consumers generally aren't happy in this scenario, as they have lost all their surplus.

Return to the top of the document

4.4 Monopolistic Competition

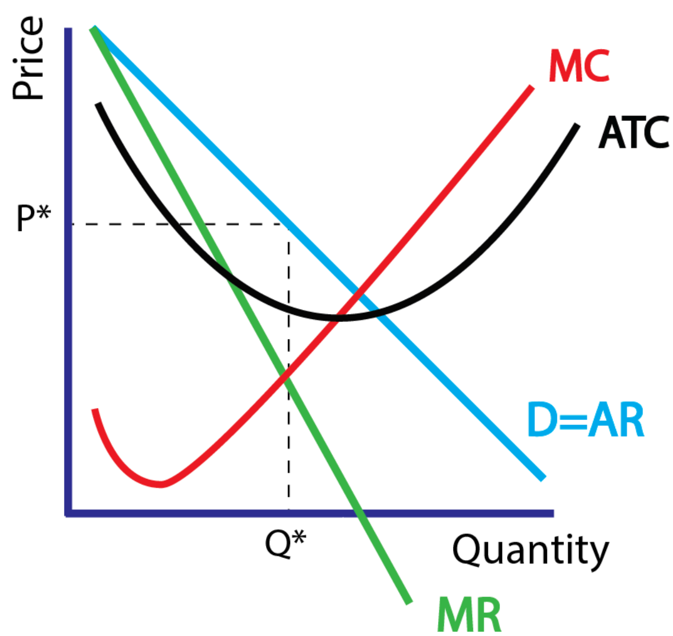
So far we've covered two types of markets: monopolies exist when there is only one firm in a market, and perfect competition exists when there are many firms selling identical goods.

The vast majority of goods, however, do not meet either of these criteria. Instead, most goods are sold under brands. Shoes are differentiated between Nikes and Reeboks, clothing comes from Abercrombie or Gap, iPhones are not the same as Androids, cars come in different makes and models, and the list goes on. We call these kinds of markets **monopolistic competition**, because each company has a monopoly on their particular brand (after all, Abercrombie would be sued if they tried to sell Gap clothing), but they exist in competition with other firms in their same industry. The two most important characteristics of monopolistic competition are:

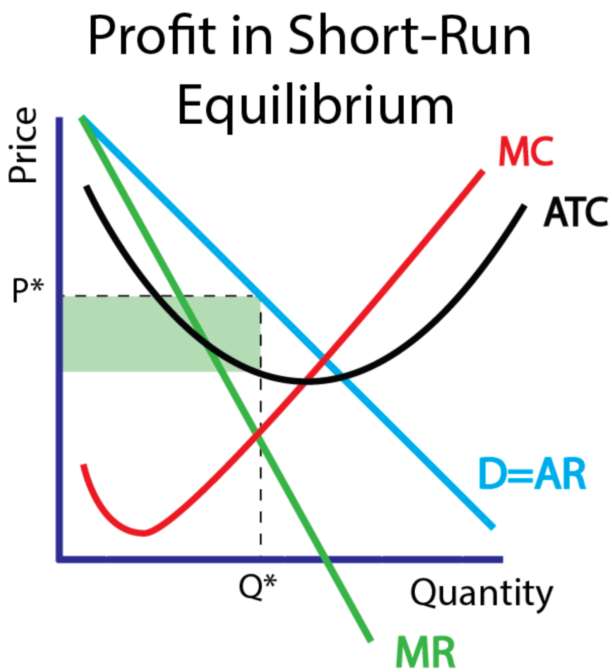
- *Goods are differentiated (usually by brand)*, giving each company a monopoly on their particular corner of the market.
- *There are no barriers to entry or exit*. If I see profits being made in the phone market, nothing will stop me from starting my own phone company and launching my own brand. Remember what we said before: *if there are no barriers to entry and exit, then we know that in the long run, economic profits will equal zero*.

Monopolist Competition Market Structure

Since each firm in this kind of industry has a monopoly on its brand, each firm will get its own downward-sloping demand curve. After all, while the demand for Nikes is clearly related to the demand for Reeboks, they are two different products that sell for different prices. And since the firm has a downward-sloping demand, that means it will also have a steeper MR curve, since it will have to lower prices if it wants to increase Q. Each firm will also have all the usual cost curves (once again, I am not including AVC here just for the sake of keeping the graphs simple, but remember that it's there). So the market for Nikes, for example, will look like the graph on the right:



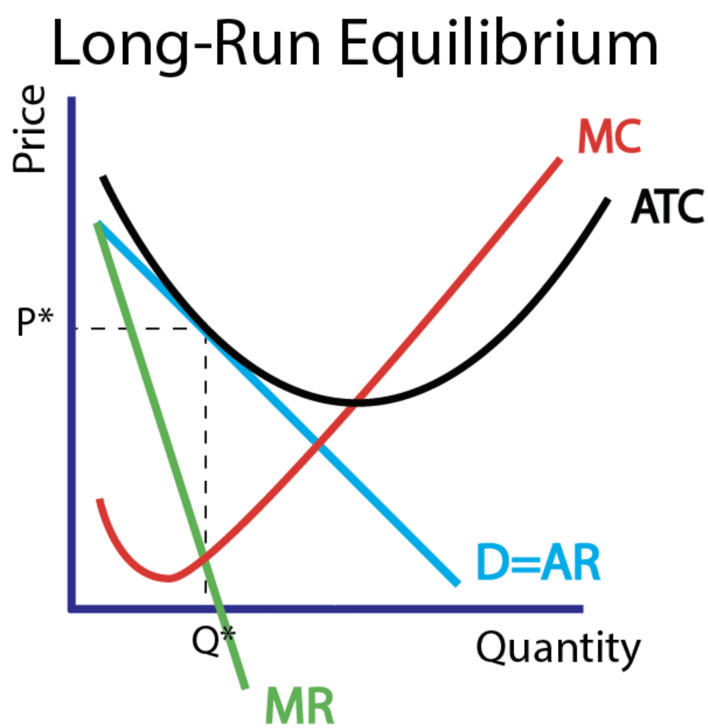
As you can see, our firm will follow the profit-maximizing strategy of setting quantity where MR=MC, and it will set prices as high as the market will bear, just like with a regular monopoly. In the example here, the firm is making a profit, since at



this quantity AR is higher than ATC (see graph on the left).

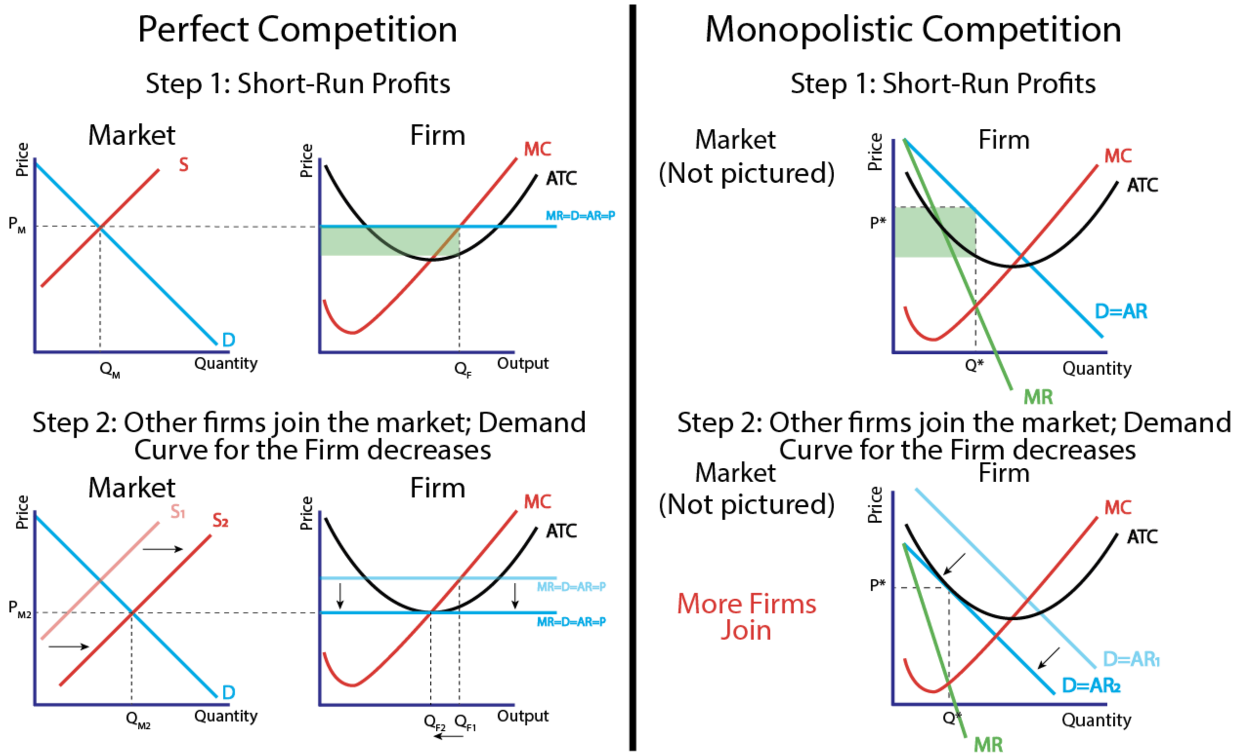
But here's where things are different from a regular monopoly: remember, there are no barriers to entry and exit here, which means that other firms will see that Nike is making a profit selling shoes, and they're going to start creating and selling their own shoe brands: Reebok, Sketchers, Converse, Vans, Adidas, and the list is practically endless. As long as there are profits being made, firms will continue to join the market.

As all these firms join, some people are going to buy these new brands rather than Nikes, so with each new firm, the demand for Nikes will decrease. Firms will continue to join (and demand will continue to decrease) until all profits in the shoe market are gone, and our long-run equilibrium point will be at a quantity where AR=ATC. But since monopolistically competitive firms still set MR=MC, this long-run equilibrium point looks a little weird. AR and ATC will be the same when the ATC curve is tangent to the AR (Demand) curve (see graph on the right).

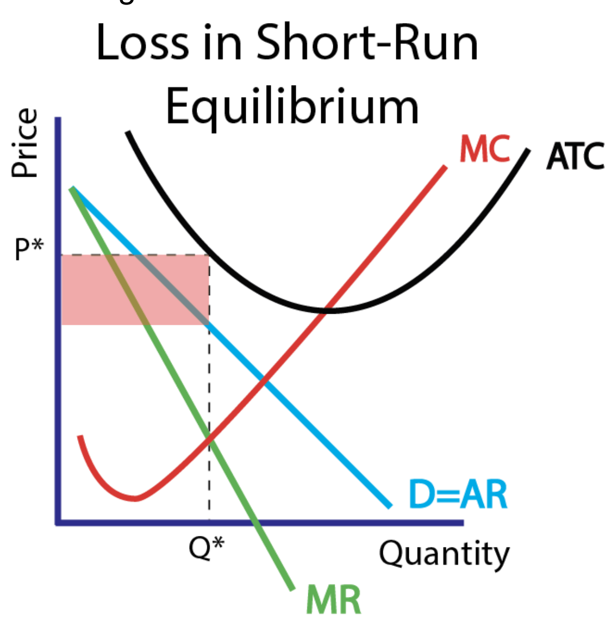


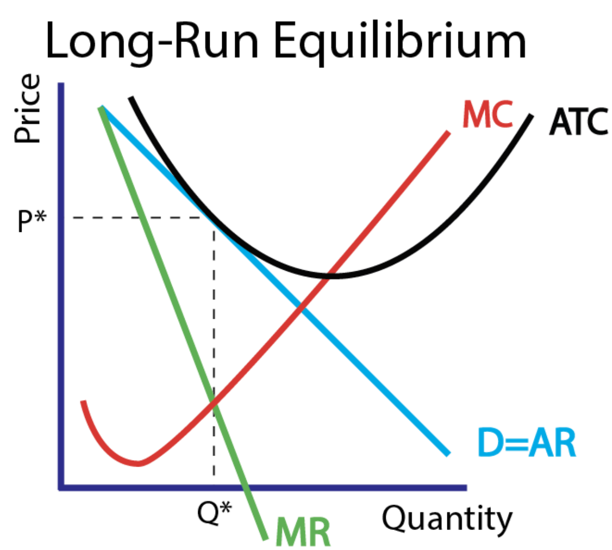
Notice how similar this is to monopolies: the firm does not produce at the allocatively efficient point (where MC meets the D curve), and it has excess capacity (i.e., it does not produce at the productively efficient point, at the minimum of ATC). There is deadweight loss. The main difference, however, is that there are no profits in monopolistic competition in the long run.

And notice how similar this is to perfect competition as well. In both cases, profits in the short-run cause more firms to enter the market, which causes the demand curve faced by the firm to decrease, thus eliminating profits.



If the firm were initially making a loss, the exact same process would play out, only in reverse. Each firm in the monopolistically competitive market would be making a loss. This would drive some firms to exit the market, causing the demand curve for all the remaining firms to increase until we are back at long-run equilibrium.

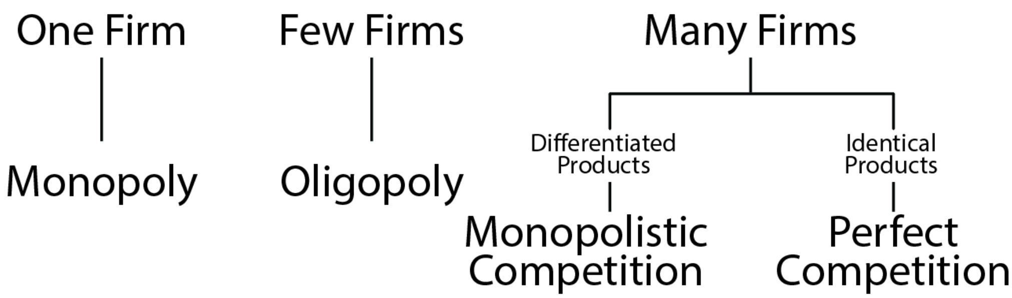




[Return to the top of the document]

4.5 Oligopoly and Game Theory

So far we've seen three different market structures: perfect competition is when there are many firms and goods are identical. If there is only one firm, that is a monopoly. And if there are many firms that sell differentiated goods, that is monopolistic competition. The final market for this unit is when there are only a few firms that control the entire market. This is called an **oligopoly** (*oligo-* means "few"). This happens when there are high barriers to entry in the market. Some examples of oligopolies would be the airline industry, cell phone carriers, or home internet providers. On the right, you can see major airlines by market share. Notice how only four airline companies dominate almost the entire airline market.



Up until now, we've been able to assume that all firms act the same—they maximize profit by setting MC=MR, they exit the market when they are making losses, etc. Even in the monopolistic competition graphs, it didn't matter whether we were looking at Nike or at Reebok, since all companies face the same market conditions and respond in the same way. This has let us largely ignore what other firms are doing. But with oligopolies, that's no longer true. If you only have two firms that dominate a market, then the price that one firm sets will have a major impact on the profit and revenue of the other.

To analyze oligopolies, we don't use a graph; rather, we use something called **game theory**, which is a way to analyze decisions when multiple parties make choices that impact each other. Each interaction between firms can be thought of as a "game," with each firm trying to maximize its own payout based on what the other firm does.

To set up a game, you put Player 1 on the left and Player 2 on top, and you write down each player's options. In this sample game, let's assume that Player 1 can either charge full price or run a sale, and Player 2 can either send out coupons or not send out coupons. So our game would look like this:

		Player 2	
		No Coupons	Coupons
Player 1	Full Price		
	Sale		

Now we write down what each firm would get in each circumstance. In the top left box, if Player 1 charges full price and Player 2 does not send coupons, then they would each get $250. In the top right box, if Player 1 charges full price and Player 2 sends coupons, then everyone is going to go to Player 2's store, so he will get $300 while Player 1 only gets $100. Notice that Player 1's payout goes on the left (since he is further left) and Player 2's goes on the right (since he is further right). I'm color coding it here, but they won't do this for you on the exam.

In the bottom left box, if Player 1 runs a sale and Player 2 does not send coupons, then Player 1 will get all the customers, and he'll get $300 while Player 2 only gets $100. In the final box, if Player 1 runs a sale and Player 2 sends coupons, then they each only get $150.

		Player 2	
		No Coupons	Coupons
Player 1	Full Price	$250, $250	$100, $300
	Sale	$300, $100	$150, $150

The question is: if the two players aren't able to coordinate between themselves, what will they do? To figure this out, you start with Player 1 and figure out what is in their best interest *depending on what the other player does*. Let's look at the left column, which gives the possible outcomes if Player 2 does not send coupons. In this case, Player 1 has the choice between the two black payouts in this column, receiving either $250 (if he goes full price) or $300 (if he runs a sale). He would obviously prefer $300, so we're going to circle $300 to show his preference.

Scenario 1: Player 2 does not send coupons

		Player 2	
		No Coupons	Coupons
Player 1	Full Price	$250, $250	$100, $300
	Sale	$300, $100	$150, $150

Alternatively, Player 2 could decide to send coupons. If he does, Player 1 has the choice between receiving $100 (by going full price) or $150 (by running a sale). Again, he will prefer to run a sale since the payout is higher, so we are going to circle $150.

Scenario 2: Player 2 does send coupons

		Player 2	
		No Coupons	Coupons
Player 1	Full Price	$250, $250	$100, $300
	Sale	$300, $100	$150, $150

If you notice, Player 1 is always better off running a sale, no matter whether Player 2 sends coupons or not. We call this a **dominant strategy**. Since Player 1 is always better off running a sale, we know that's what he's going to do.

Now we figure out what Player 2 is going to do, and we follow the exact same process. If Player 1 charges full price, then Player 2 has the option of either receiving $250 (if he does not send coupons) or $300 (if he does send them). He's going to prefer getting $300, so we circle it.

Scenario 3: Player 1 charges full price

		Player 2	
		No Coupons	Coupons
Player 1	Full Price	$250, $250	$100, $300
	Sale	$300, $100	$150, $150

Alternatively, if Player 1 runs a sale, then Player 2 has the option of receiving either $100 (if he does not send coupons) or $150 (if he does send them). He's going to prefer getting $150, so we circle it.

Scenario 4: Player 1 runs a sale

		Player 2	
		No Coupons	Coupons
Player 1	Full Price	$250, $250	$100, $300
	Sale	$300, $100	$150, $150

In this case, Player 2 also has a dominant strategy: he is always better off sending coupons, no matter what Player 1 does.

The solution to this problem, then, is that Player 1 will run a sale, and Player 2 will send coupons. We can tell because this is the only box where there are two circles inside it. Whenever there are two circles in a box, we call this **Nash Equilibrium**, named after the famous economist John Nash who invented game theory. A game with a Nash Equilibrium (i.e., two circles in a box) will always end up at that point and will stay there, no matter how many times the players play the game. The reason they get stuck here is because neither player can improve their outcome by unilaterally changing strategy (i.e., if only one changes strategy and the other does not).

Notice, by the way, that neither player actually wants to end up at that point. If they could coordinate among themselves, they would choose for P1 to go full price and P2 to not send coupons. That way they would end up at the top left cell and get a combined $500 to split among themselves, the highest combined payout of all the cells in the table. But because each player is acting in their own best interest, they end up at Nash Equilibrium, at a sub-optimal point.

This is why oligopolies often try to collude and form **cartels**. They want everyone in the market to charge higher prices, thus increasing profits for everyone, and if a cartel is successful, it will act like a pure monopoly. Cartels are illegal in the US for this very reason. If they can't form a cartel, firms often try to make informal agreements among themselves to keep prices high. But without any binding way to enforce these agreements, they usually fall apart, since everyone in the market has an incentive to undercut the competition's price and get a higher profit for themselves. Due to these competing forces (wanting to get higher profits but also trying to undercut the competition), oligopolies usually tend to have higher prices and lower output than we see in perfect competition, but they usually aren't able to get prices to be as high as they would be if there were a monopoly.

And that's it. When analyzing an oligopoly, all you need to do is set up the payout matrix, then analyze it column by column and row by row, circling each player's preferred payout in each scenario. The question will then usually ask whether each player has a dominant strategy or not. Finally, whenever you get a box with two circles in it, that is Nash Equilibrium, and that's where you know the game will end up.

Be careful: sometimes a game will not have any Nash Equilibria, in which case we can't predict how the game will turn out. Or sometimes there can be more than one Nash Equilibrium. In that case, we can predict that the game will end in one of two spots, and once it gets there, neither player will change strategies (since they can't be made better off by unilaterally changing strategy). See the following scenarios for some of the more common examples you'll see on the test.

Player 2 does not have a dominant strategy;
there is a Nash Equilibrium

Player 1 \ Player 2	No Coupons	Coupons
Full Price	1, 5	3, 4
Sale	2, 4	4, 5

Two Nash Equilibria

Player 1 \ Player 2	No Coupons	Coupons
Full Price	1, 1	2, 2
Sale	2, 2	1, 1

No Nash Equilibrium

Player 1 \ Player 2	No Coupons	Coupons
Full Price	1, 5	3, 4
Sale	2, 4	2, 5

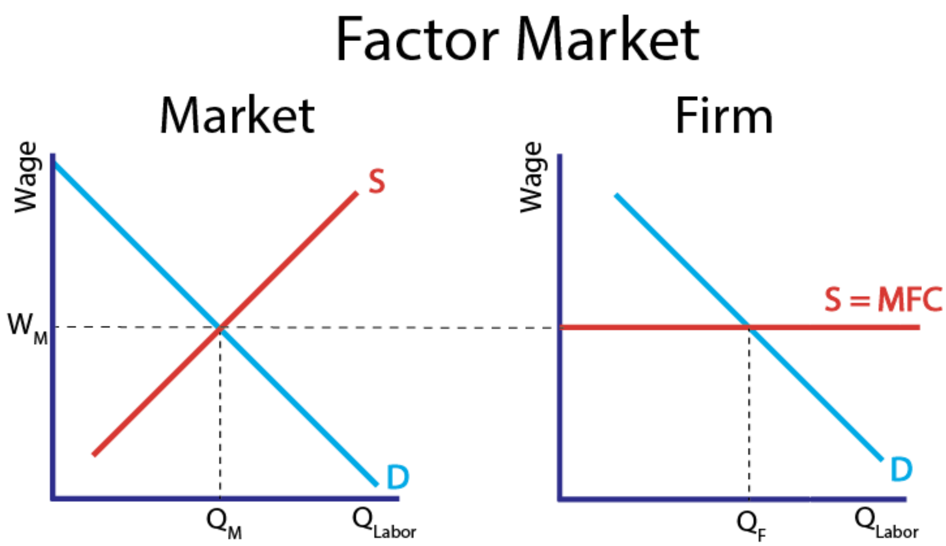
Return to the top of the document

5.1 Factor Markets and Labor Supply

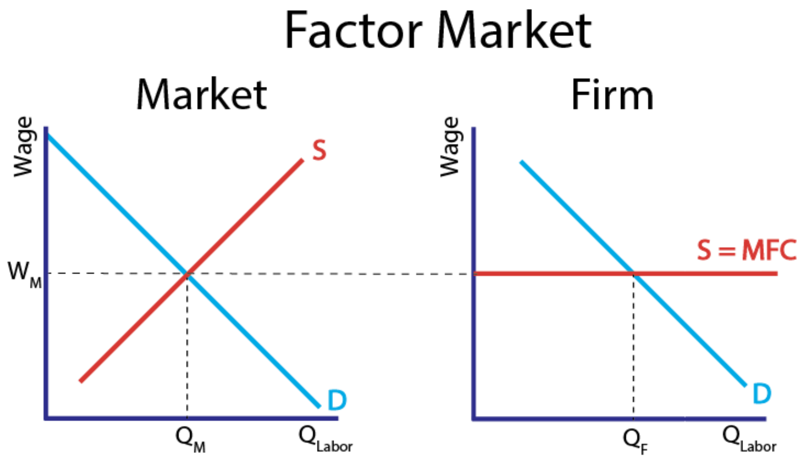
The past few units have focused on the product market and how firms produce and set prices for goods. In this unit, we'll learn about factor markets: how firms decide how much land, labor, and capital to purchase, as well as how wages and prices are determined. Over the next few lessons, we're going to look at the labor market, but everything we do here could be done in the exact same way for labor or capital.

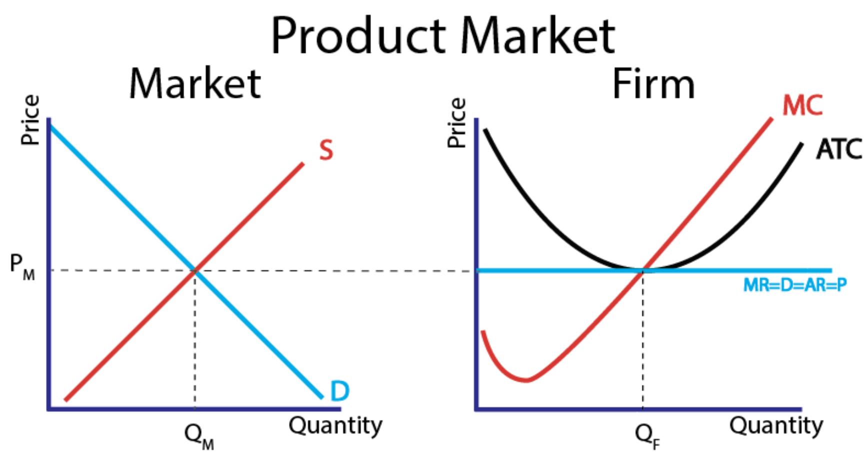
Factor markets are set up in much the same way as product markets, only now the suppliers and demanders are reversed. Firms are the ones who demand laborers, and the population is the supplier. On the x-axis we'll put Q_{Labor}, and on the y-axis we put "wage," since this is the price of labor. As you might expect, the demand for laborers is downward sloping (since at a high wage, firms can't afford to hire many workers), while the supply of labor is upward sloping (since as wages rise, people want to work more). That gives us the Market side of the graphs.

On the Firm side of the graphs, we can assume that in a perfectly competitive labor market, any individual firm is a price taker: no matter how many workers they hire, they can't impact market wages. The supply of workers they face is also perfectly horizontal, since they can hire as many or as few workers they want at the going market wage. Since firms always analyze their decisions in marginal terms, we're going to call this the **Marginal Factor Cost (MFC)**.



Notice how this is reversed from the perfectly competitive product market. In that market, firms face a perfectly horizontal *demand* curve, but in the factor market they face a perfectly horizontal *supply* curve. That's because firms are suppliers in the product market but demanders in the factor market.



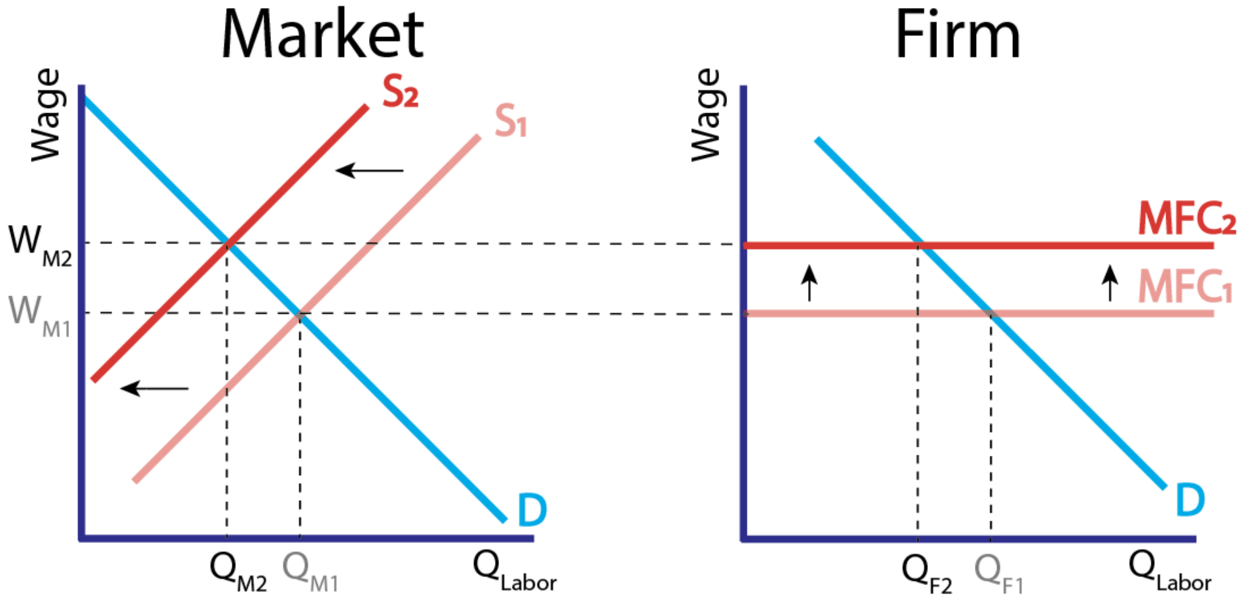


Changes in the Supply of Labor

In the broader market, there are a number of influences that can shift the market supply of labor. You need to know these. These influences are:

- *Changes in population*. This is impacted by birthrates and immigration, so an increase in immigration would lead to an increase in the market labor supply curve.
- *Wages in other markets*. If I'm trying to hire janitors, it matters a lot what wage people can get by waiting tables. If the wage of waiters goes up, the supply of labor in my market will go down.
- *Tastes and preferences*. Sometimes people decide they don't want to work as much. In Europe, for example, people tend to value leisure time more highly than Americans do. If we started to mirror European attitudes, our supply curve would shift left.

As the market supply shifts, this causes wages to rise or fall, which impacts the Marginal Factor Cost that the firm faces. Below you can see an example of what would happen if wages rose in another market:



Wages would rise in the market, and fewer workers would be hired in the market. The firm would also hire fewer workers.

[Return to the top of the document]

5.2 Labor Demand

Firms demand labor, but they do not have full control over how much labor they need. That is because the demand for labor is known as a **derived demand**, meaning that this curve is derived from the demand for something else (namely, the final product). I don't demand auto workers; rather, I demand cars, and the amount of my demand for cars will determine how many auto workers each car manufacturer will need.

When it comes to hiring workers, firms have to balance two things: 1) how much will each worker cost me? And 2) how much revenue will each worker bring in for the firm? If hiring a 10th worker would bring in $100 in revenue and only cost me $90, then I should hire him. If he would only bring in $80 in revenue, then I shouldn't hire him.

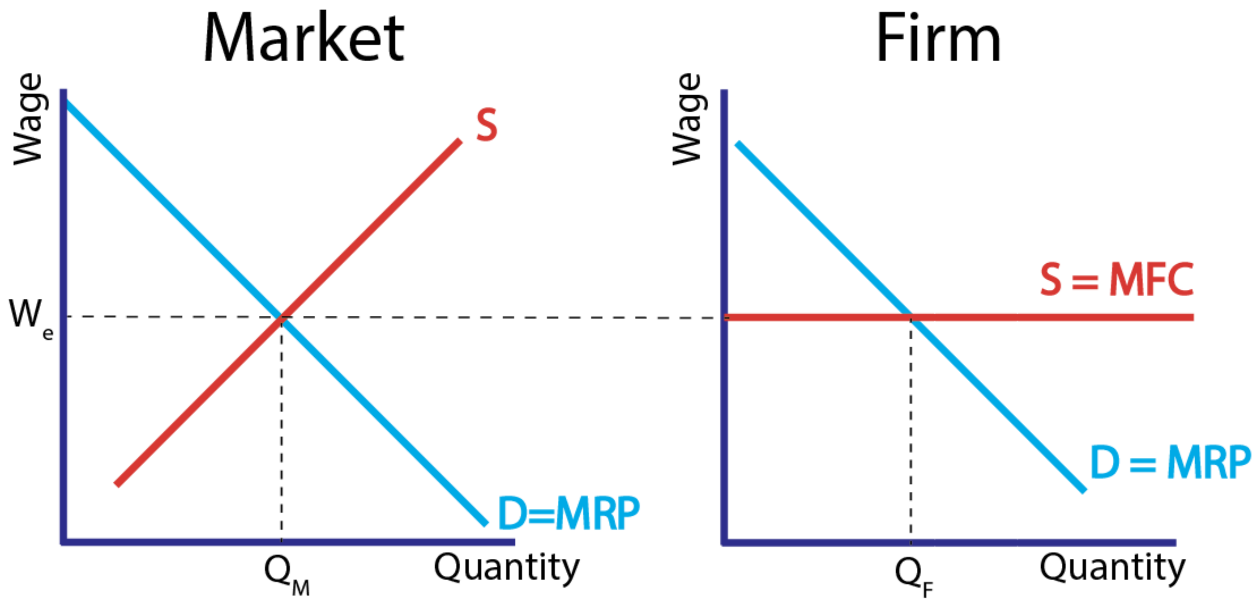
The first concern—how much each worker costs—is determined by supply and demand in the broader labor market, as we saw last lesson. The second concern—how much revenue each worker brings in—is determined by two things:

1. **How much does the final product sell for?** If cars sell for $100,000, then each worker will bring in a lot of revenue for the firm, so I can afford to hire a lot of workers. But if cars only sell for $10,000, then my workers won't bring in much revenue at all. Hence why the demand for labor is a derived demand.
2. **How productive is each worker?** If you recall the production function from previous lessons, the first few workers that your firm hires will be very productive—they might make 7 cars each. But as you hire more and more workers, they become less and less productive. By the 100th worker you hire, each additional worker might be able to make only 2 cars each.

If you combine these two things, you can figure out how much additional revenue each worker brings in for the firm. Let's say that cars sell for $10,000 each, and my first worker hired can make 7 cars. That means he is bringing in 7 x $10,000 = $70,000 in revenue for the firm. If market wages are $50,000, then I should definitely hire this worker, since he brings in more money than he costs me to hire.

For the 100th worker, however, he can only make 2 cars, so he is only bringing in 2 x $10,000 = $20,000. If market wages are $50,000, then I should not hire this 100th worker.

We call this calculation the **Marginal Revenue Product (MRP)**, which is honestly a terrible name. I prefer to think of it as the "Marginal Revenue Produced (by each worker)", and because workers become less productive as you hire more and more, this curve will be downward sloping. Each firm will have its own MRP curve, which is its demand for labor, and every firm's demand combined will give us a market MRP curve, or a market demand curve. That means our labor market graphs will look like this:



Remember, MRP is determined by two things: how much the final good sells for, and how productive each worker is. That means the MRP curve can shift if:

- **The price of the final good changes**. If the price goes up, that means each worker can now bring in more revenue, so the MRP curve will increase (shift to the right).
- **Each worker becomes more productive**. That can happen if either:
    - **Technology changes** (e.g., if new technology is introduced, the MRP curve will shift to the right). OR
    - **The amount of other factors of production changes**. If I get more capital, then my laborers will be more productive, thus increasing my MRP curve.

Keep in mind, however, that if the price of the good changes or if technology changes, the MRP curve for the market *and for the firm* will both increase. The only time the MRP curve for the firm would increase without the market changing is if the firm discovered some new technology that was not available to other firms in the industry, or if the firm increased or decreased the amount of other factors of production it is using.

[Return to the top of the document]

5.3 Profit-Maximizing Behavior in Factor Markets

On the exam, there are three kinds of problems you'll see dealing with factor markets. These three types are:

1. *The question tells you that at a specific quantity (i.e., a specific number of workers hired), the MRP is greater or less than MFC.* You have to know that the firm should hire more or fewer workers, respectively, to maximize profits.

2. *The question gives you a chart, usually listing the number of workers and total product* (see the chart on the right). You have to be able to figure out the MRP from this chart, which means you have to find the Marginal Product first, then multiply it by the price of the final good. In this case, let's say the good sells for $20 and wages are $250. How many workers should the firm hire? You need to create the chart below:

Number of Workers	Total Product per Hour
1	8
2	20
3	40
4	55
5	65
6	70
7	65
8	55

Number of Workers	Total Product per Hour	Marginal Product	Marginal Rev. Product
1	8	8	$160
2	20	12	$240
3	40	20	$400
4	55	15	$300
5	65	10	$200
6	70	5	$100
7	65	-5	-$100
8	55	-10	-$200

Since MRP falls below wage after the 4th worker, this firm needs to hire 4 workers. When you're working with numbers like this, remember your formula:

- MRP = MP x MR (Marginal Revenue Product = Marginal Product x Marginal Revenue). Of course, if the firm is in a perfectly-competitive product market, MR is just equal to Price, so to find MRP, just multiply MP by the Price.

3. *The question is in consumer-choice format*. Remember when we did consumer-choice problems, where people had to maximize their marginal utility per dollar spent? The formula was

$$\frac{MU_X}{P_X} = \frac{MU_Y}{P_Y}$$

Now we're going to do that same thing, only now firms are the consumers, and they are consuming labor and capital. They are going to consume these two things until the marginal product of labor per dollar spent is equal to the marginal product of capital (which we abbreviate as K) per dollar spent. If a firm is getting more productive laborers than capital, they need to increase the amount of labor and decrease the amount of capital they are buying until the two are equal.

$$\frac{MP_L}{P_L} = \frac{MP_K}{P_K}$$

NOTE: The following topic is listed by the College Board as something you need to know for this class, but I have never seen any questions about it, so you're probably safe to ignore it. I'm including it here for the sake of completeness, but if you want to stop reading here, that is totally fine.

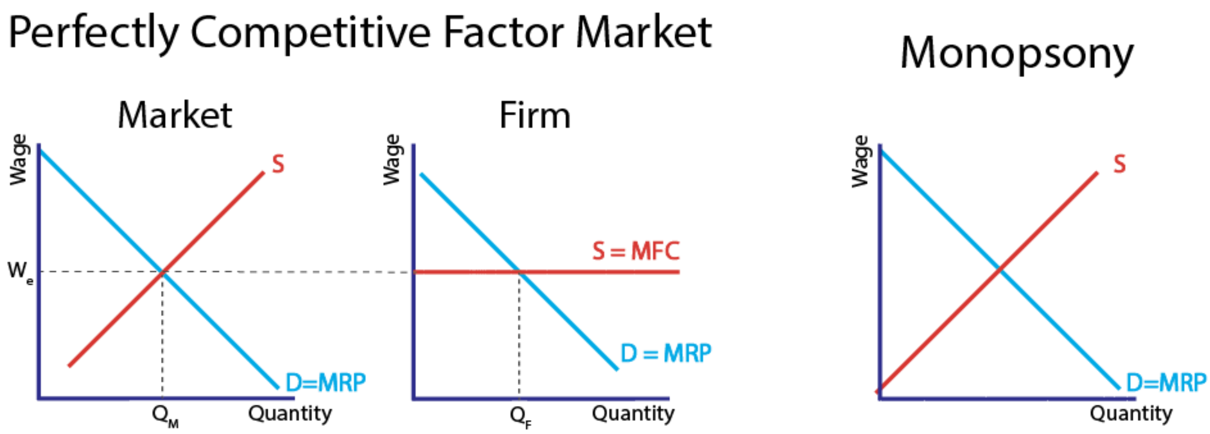
Another way to think about how much labor firms hire is by measuring the Value of the Marginal Product of Labor, abbreviated VMPL. The formula for VMPL is MPL x P (Marginal Product of Labor x Price of the final good). This looks almost exactly like MRP, which is MP x MR. In fact, they look so similar because if the firm is operating in a perfectly competitive output market, MR is equal to P (think of MR=D=AR=P). That means that for these kinds of firms, VMPL and MRP will be equal.

Of course, if the firm is in a monopoly or monopolistic competition, then MR will no longer be equal to P, so MRP will not equal VMPL. But that's a level of complication I've never seen on the AP exam. If you just memorize MRP, you'll be fine.

Return to the top of the document

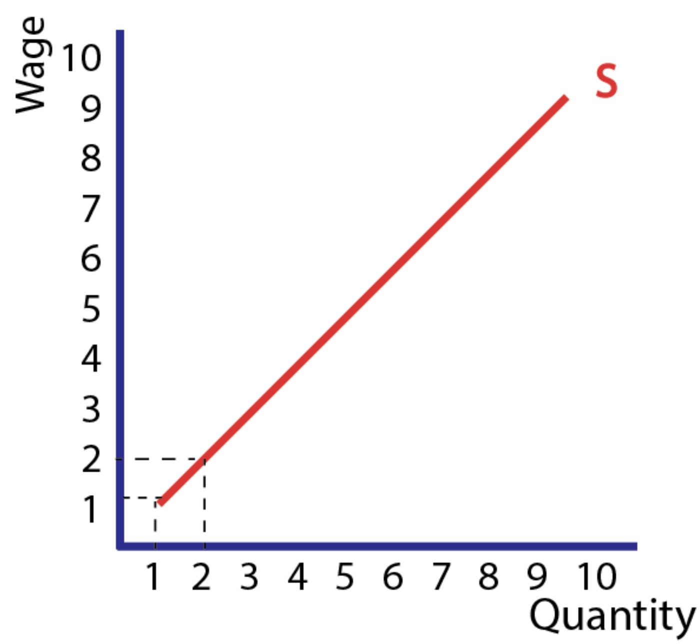
5.4 Monopsonistic Markets

A monopoly is when a firm is the only seller of a good in the product market. But sometimes one firm is the only buyer in a factor market. For example, in a small mining town, that one mining company will be the only demander of labor, which means we will no longer break our graphs up into “Market” and “Firm”. In this case, one firm will face the entire market, and it will no longer be a price-taker when it comes to wages. We call this a **monopsony**.



When we studied monopolies, we found that there was a separate MR curve because if the firm wanted to sell two goods, it had to lower the price of the second good *and the first good*, which led to a larger decrease in revenue. That same principle is going to hold true here, with a monopsony. But instead of splitting the Demand curve into two separate curves, now we’re going to be splitting the Supply curve.

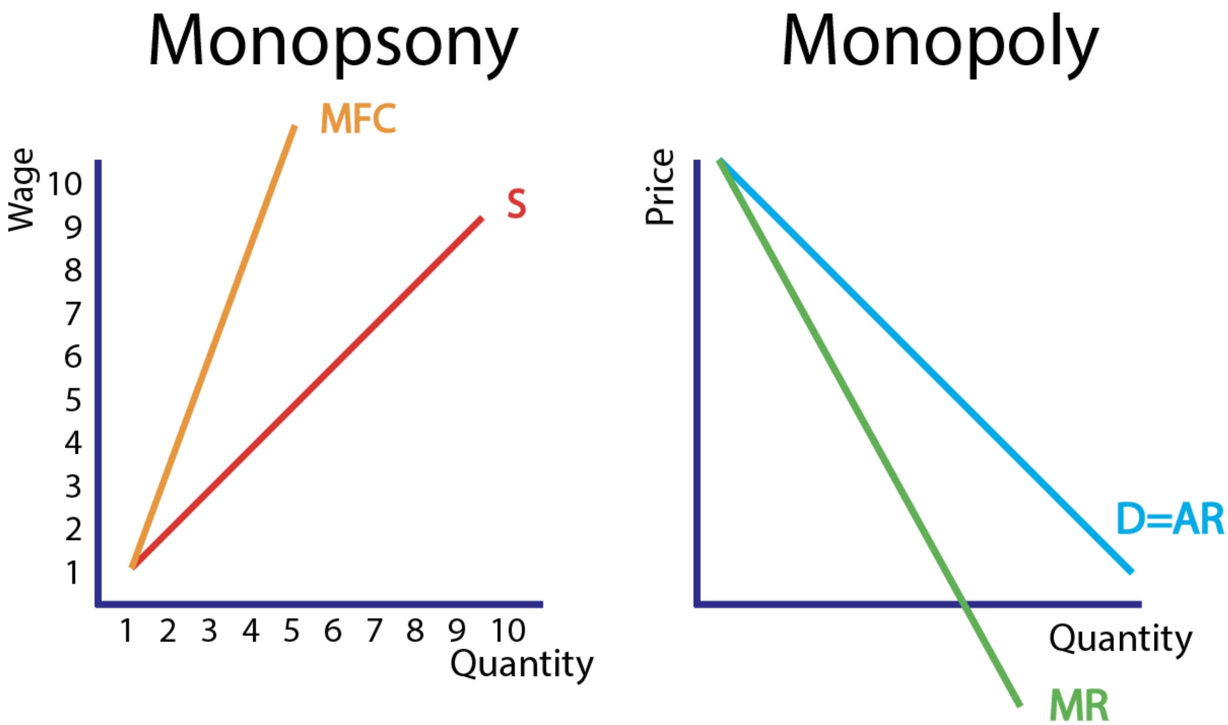
To see why, look at what happens if this firm tries to hire one worker. They can pay $1, which means the Total Cost of hiring one worker is $1.



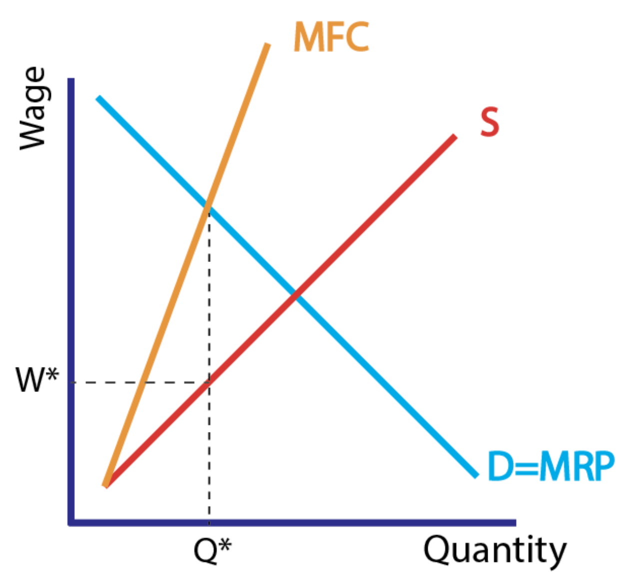
If they try to hire a second worker, they have to raise wages to $2, but the marginal cost of adding that second worker is more than $2. In adding that second worker, their total cost increases by $3 because they have to pay their second worker $2 *and their first worker an additional $1*. So the Marginal Factor Cost of hiring that second worker is $3.

This same principle holds true as you go up the supply curve, which gives us a Marginal Factor Cost curve that increases at a steeper rate than the Supply curve. It will look like the graph below.

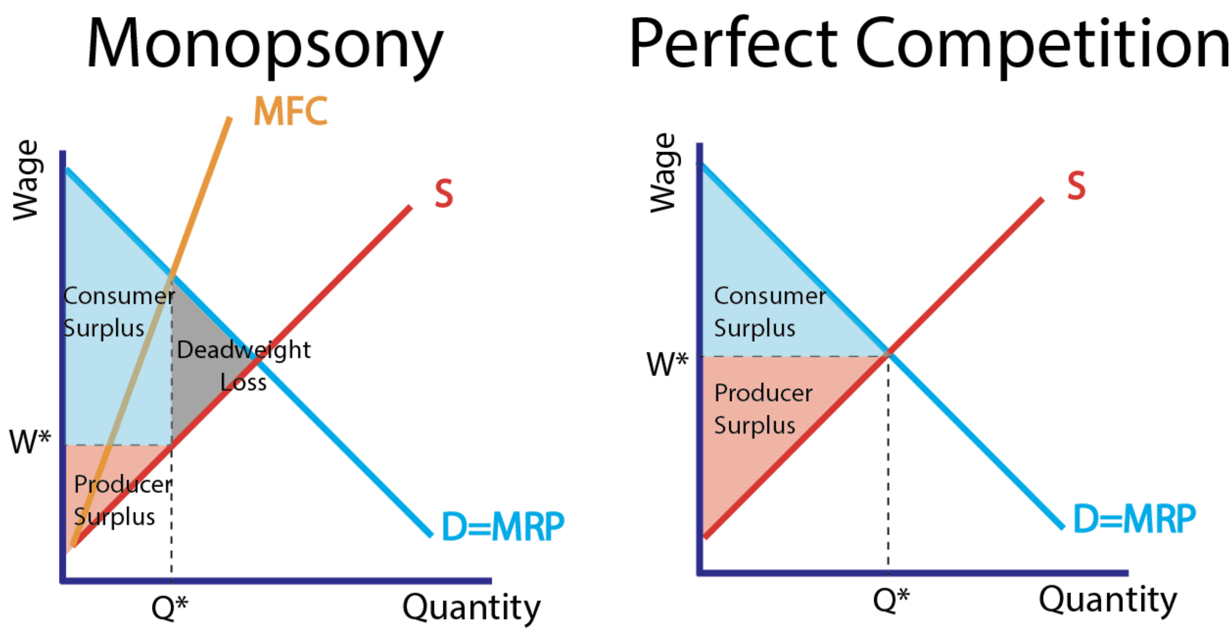
Once again, notice how similar this is to the monopoly graph. The only difference is that we are splitting the Supply curve here vs. the Demand curve in the monopoly graph, since in the factor market, Supply represents the “not-the-firm” part of the market.



To see how a firm will act in a monopsony, we’re going to simply follow the normal rules of profit maximization: determine your quantity by setting marginal cost (in this case, Marginal Factor Cost) equal to marginal revenue (in this case, Marginal Revenue Product). Once you’ve done that, we’re going to pay these workers the absolute minimum they will accept, i.e., at the supply line. That gives us a final graph that looks like this:



When compared to perfect competition, you need to know that monopsonies pay lower wages and hire fewer workers, and they create deadweight loss (see the next page for a comparison)



[Return to the top of the document]

6.1 Socially Efficient and Inefficient Market Outcomes

In normal markets, every good that is bought and sold has a certain cost to produce (marginal cost), as well as a certain benefit to the consumer (marginal benefit). When the marginal cost of the last unit produced is equal to the marginal benefit of consuming that last unit, we have maximized total market welfare, and we have achieved complete allocative efficiency. But sometimes markets fail, leading to inefficient outcomes. In these cases, economists recognize that some outside force (usually government) needs to intervene to set markets right and maximize efficiency. In this unit, we're going to be talking about these market failures and how to fix them.

We've already seen many market failures. Any time there is DWL, the market on its own has failed to maximize total surplus, and we see this with monopolies, monopolistic competition, oligopolies, and monopsonies.

Markets can also fail when buyers and sellers have **asymmetric information**, meaning one party knows more about the good than another. When selling a used car, for example, the seller knows much more than the buyer about what problems the car has, and since they are incentivized to hide these problems, people are wary about buying used cars. Even if two people got together who could make a mutually beneficial transaction (i.e., the buyer is willing to pay what the car is worth, and the seller is willing to sell it for that much), the fact that the buyer suffers from asymmetric information means that he or she is likely to not buy the car at all, due to a fear of being ripped off. This is a market failure.

Another failure happens when **property rights are not well defined**. Factories can spew out pollution, and this polluted air harms the people around the factory. But who owns the air? If the neighbors around the factory owned it, they could charge the factory for the right to pollute their air, and the factory and the neighbors could come to a mutually beneficial agreement, say with the factory paying the neighbors a certain amount for the right to pollute. Yes, the neighbors would be living with pollution, but they would be compensated enough so that they would prefer the compensation rather than living in a pollution-free area. As it stands, however, since no one really owns the air, factories spew out way more pollution than they should, and the neighbors have no recourse but to live with it or leave. Since a mutually beneficial transaction could be happening but is not, this is also a market failure.

Markets also fail when there are **high transaction costs**. I would love to buy a new house, and I have no doubt that there are many houses on the market that would fit my budget and my needs really well. But searching for a house costs a lot of time that I don't really have to spare, and even if I found the right house, buying it comes with other transaction costs that make it even worse: I have to pay for inspectors, the mortgage origination fee, the real estate agents, and a whole host of other fees. As a result, even though you might have a home that I'd be willing to buy, these transaction costs make it less likely that this transaction will take place.

To solve these problems, governments have various solutions. For ill-defined property rights, governments will occasionally create such rights and then give them to individuals (or auction them off). So for pollution, governments might create a permit to pollute a certain amount and then auction those permits off. Or if someone is playing loud music, they might make loud music illegal (i.e., give people the right to call the police on the person playing).

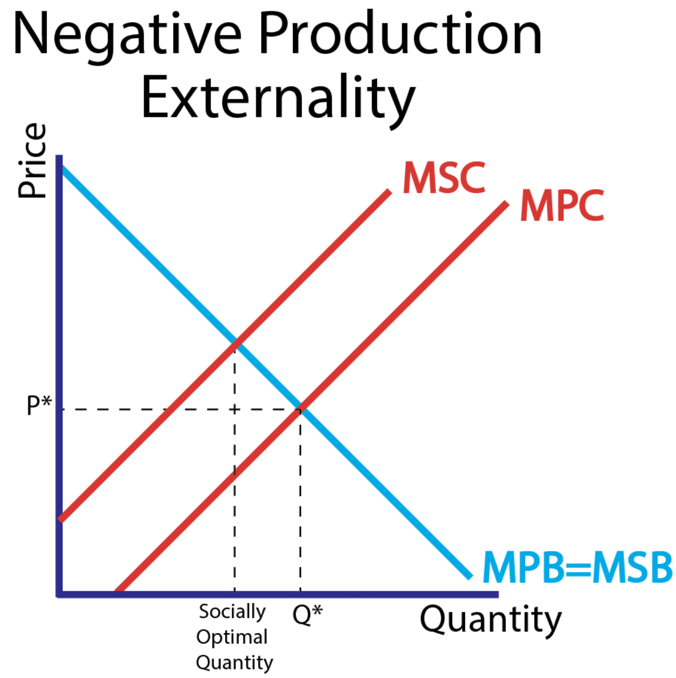
For markets with high transaction costs, governments will occasionally step in to create a marketplace that functions more efficiently. This is what the government did with healthcare under the Affordable Care Act, for example.

Return to the top of the document

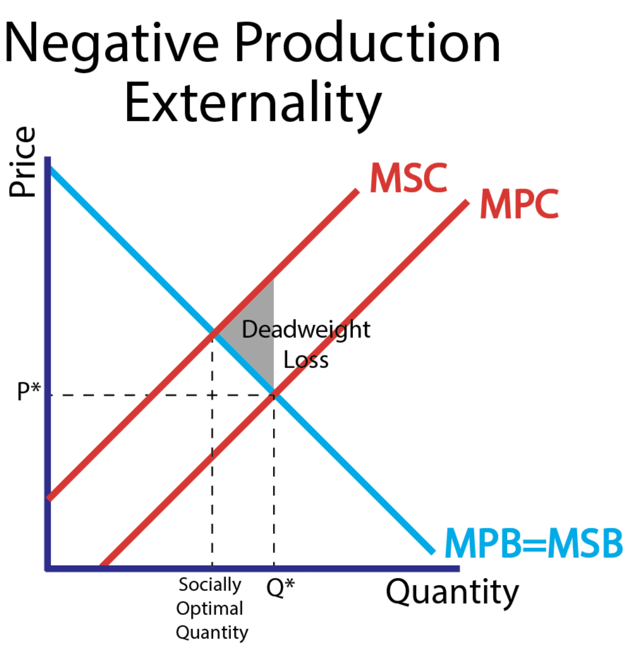
6.2 Externalities

Another type of market failure happens when the production or consumption of a good affects people beyond just the buyer or seller. Consider, for example, the production of plastic, which produces a significant amount of pollution in the manufacturing process. No matter what you think about the environment, from a purely economic standpoint pollution has some serious economic costs: it leads to increased medical problems (and thus higher healthcare costs and days of missed work), it reduces property values, and the process of cleaning it up is very expensive. But when firms are figuring out the marginal cost of producing plastic, they don't count these costs, because they aren't the ones who have to pay them. These costs are borne by society as a whole. As a result, we have a mismatch between the **Marginal Private Costs (MPC)** that firms pay and the **Marginal Social Costs (MSC)** that society as a whole pays.

In any scenario where the production or consumption of a good impacts other people, either in positive or in negative ways, we call this an **externality**. In the case of pollution, this is a **negative production externality**, because the production of the good has a negative impact on society beyond the costs that firms pay. We would graph this externality out as seen on the right, with a higher social cost than private cost (in this case, the consumption of the good does not carry an externality, so marginal private benefit is equal to marginal social benefit). In this graph, we end up at P* and Q*, because firms only respond to private costs, not to social costs. The point where social costs = social benefits is the socially optimal quantity, so if we are at Q*, then our society is producing too much of this good.



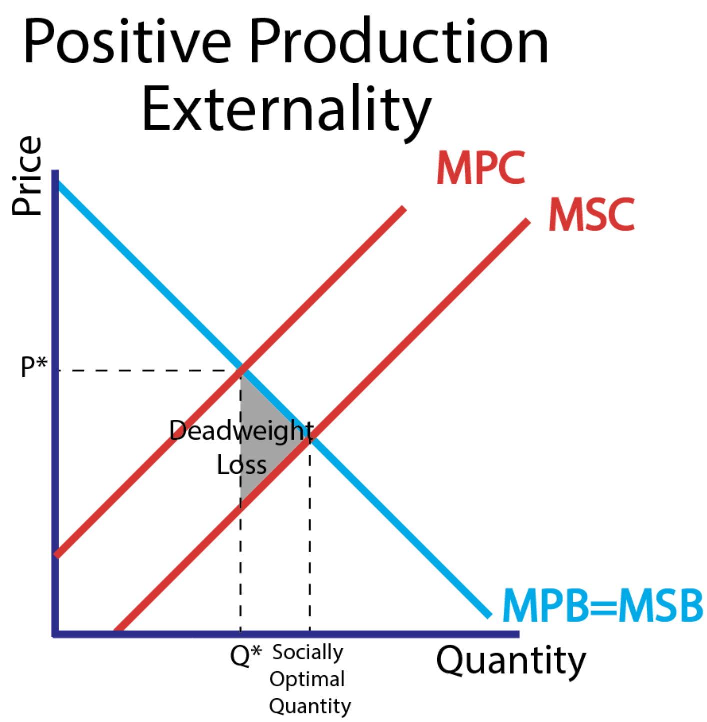
This isn't to say that no one should produce plastic. Rather, we're saying that at the current level of production, the costs to society outweigh the benefits to society, so our current production levels are inefficient. How inefficient are we being; how much deadweight loss is there? If we're examining market welfare, then the *social* cost and *social* benefit curves are the only ones we care about. These are the "real" supply and demand curves, and deadweight loss always falls



between supply and demand. So to find DWL, we just need to find where equilibrium should be, then put the triangle of DWL between Supply and Demand, with one edge touching our current production levels and one edge touching how much we *should* be producing. These are all transactions that, from a social standpoint, should not take place.

If we want to fix this, we need to find a way to make the firm internalize these social costs and decrease the levels of production. The easiest way to do this is with a tax. If we create a tax for the exact amount of the difference between MSC and MPC, this will shift MPC up/back so that these two curves will now be in line. Importantly, *imposing a tax does not create any deadweight loss*. Normally taxes do lead to DWL, but that's only because taxes usually push you away from the socially optimal production level. In this case, taxes are pushing us toward that level, so there is no DWL.

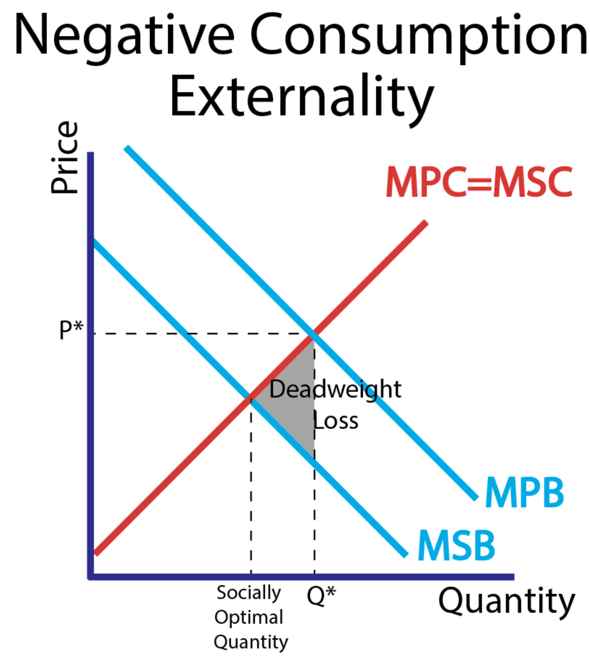
We could run the same analysis for a **positive production externality**, such as having a sports team in a city. If you have a sports team in Atlanta, this team attracts all kinds of additional benefits for the city that the team itself does not receive, such as increased spending at hotels, restaurants, and other shopping done by fans going to games. Since this is a production externality, we're going to split the supply curve into MSC and MPC, only now the Marginal Social Cost is going to be lower than Marginal Private Cost. When compared to the socially optimal quantity, we are now under-producing, hence why DWL is to the left of the socially optimal equilibrium (between the "true" supply and demand curves).



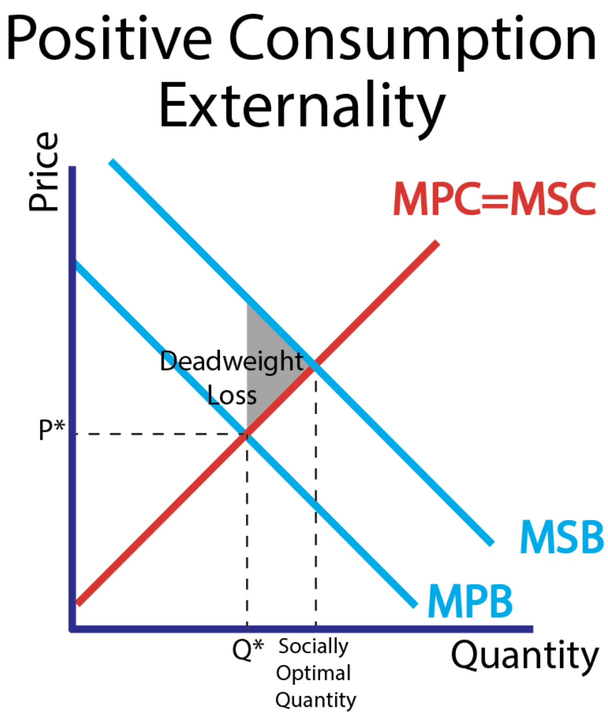
To fix a positive production externality, we need to subsidize this sports team. If we create a subsidy in the exact amount of the difference between MPC and MSC, MPC will shift right to the point where we will start producing the socially optimal quantity, and we will eliminate all DWL.

We also have consumption externalities, where the consumption of a good or service impacts other people. For a consumption externality, we're going to split the demand curve into **Marginal Private Benefit (MPB)** and **Marginal Social Benefit (MSB)**.

Smoking is the usual example given of a **negative consumption externality**. Smoking brings some benefits to smokers, but the benefits to society are considerably lower: health problems from second-hand smoke, bad smells, and increased healthcare premiums for everyone as smokers die prematurely (dying of cancer, it turns out, is insanely expensive, which drives up healthcare costs generally). Our graph will look like the one on the right. Once again, we are over-consuming, so the best solution is to tax cigarettes. As consumers pay a higher price, their personal benefits from smoking will decrease, leading us to the socially optimal outcome and eliminating DWL.



Alternatively, we could examine a **positive consumption externality**, such as schooling. Education clearly has benefits for students, but it also creates benefits for society as a whole: with more education, more jobs are created, crime decreases, cities become safer, people become healthier, and democracy becomes stronger. In other words, you will reap some private benefits from being educated, but our society will get all of those benefits plus more, so if we leave the market on its own, we are under-consuming education. As a result, we should subsidize it—which, in fact, we do. That's the whole idea behind public school.



The main takeaways from today:
- With a production externality, split the supply curve into MPC and MSC.
- With a consumption externality, split the demand curve into MPB and MSB.
- When you have a positive externality (like schooling), we produce too little of the good. Fix it with a subsidy.
- When you have a negative externality, we produce too much of the good. Fix it with a tax.
- These subsidies and taxes do not create any DWL—in fact, they eliminate it.

Return to the top of the document

6.3 Public and Private Goods

Another kind of market failure happens whenever we have non-private goods. Economists classify goods according to two metrics:

- A **rival good** is any good where the consumption of it leaves less for everyone else to consume. Almost every good we use every day is rival, from food to cars to ukeleles. Alternatively, you can consume a **non-rival good** without it impacting anyone else. This is what happens when you listen to the radio, watch Netflix, or enjoy the national defense that our country provides.
- An **excludable good** is one that suppliers can restrict to only paying customers. Radio is a **non-excludable good**, since no one can prevent you from tuning in to a station, regardless of whether you contribute to that station or not. Netflix, on the other hand, is excludable (at least in theory), since they can crack down on password sharing and make sure that you can't watch it unless you pay for an account.

Using these two metrics, we can classify goods as follows:

	Rival	**Non-rival**
Excludable	Private Goods	—
Non-Excludable	"Tragedy of the Commons"	Public Goods

Private goods are the goods we've been studying all year—suppliers can charge customers for them, and they are scarce. And you don't need to worry about the top-right of the chart (excludable, non-rival goods). The bottom row, however, presents a number of problems when it comes to free markets.

Let's start with **public goods** (non-rival, non-excludable). These include things like national defense, radio, roads, public parks, and even knowledge. The problem with these goods, and indeed with all non-excludable goods, is that every individual has an incentive *not* to pay for their production. We all know that we need roads and national defense, but if I know these things are going to happen anyways, then why should I personally pay for them? I should just wait for someone else to pay, and they won't be able to stop me from enjoying the good. After all, you can't provide national defense to all of the US *except for those who don't file a tax return this year*. This is known as the **free-rider problem**. And whenever there is an incentive to free-ride, the market is not going to produce the socially optimal amount of the good. Even though we all value national security, if the government didn't provide it for us and force us to pay for it, the free-rider problem would lead to us not having a large enough army to protect ourselves. The same goes for roads, public radio, lighthouses, and knowledge—hence why the government pays for all of these things (the "knowledge" piece is paid for through free public schools, state-run universities, and federal research grants).

The **"tragedy of the commons"** describes when a good is rival but non-excludable. The term comes from back in the day, when towns used to be built around a grassy field (called a “commons”) where everyone could let their horses and cattle graze. The problem with this setup is that the grass is non-excludable, so there’s nothing to stop me from letting my animals eat all they want to. But if everyone lets their animals eat all they want, then the grass will quickly die out, and no one will have any. The same thing happens in the ocean with commercial fishing. Since fishing in international waters is non-excludable, companies tend to fish until the oceans are depleted and entire species die out.

There’s not really a good solution to the tragedy of the commons. Private incentives will always lead to overconsumption in these situations, so unless people band together in collective action (by passing laws limiting fishing or grazing, for example), we are essentially guaranteed to run out of the resource eventually.

[Return to the top of the document]

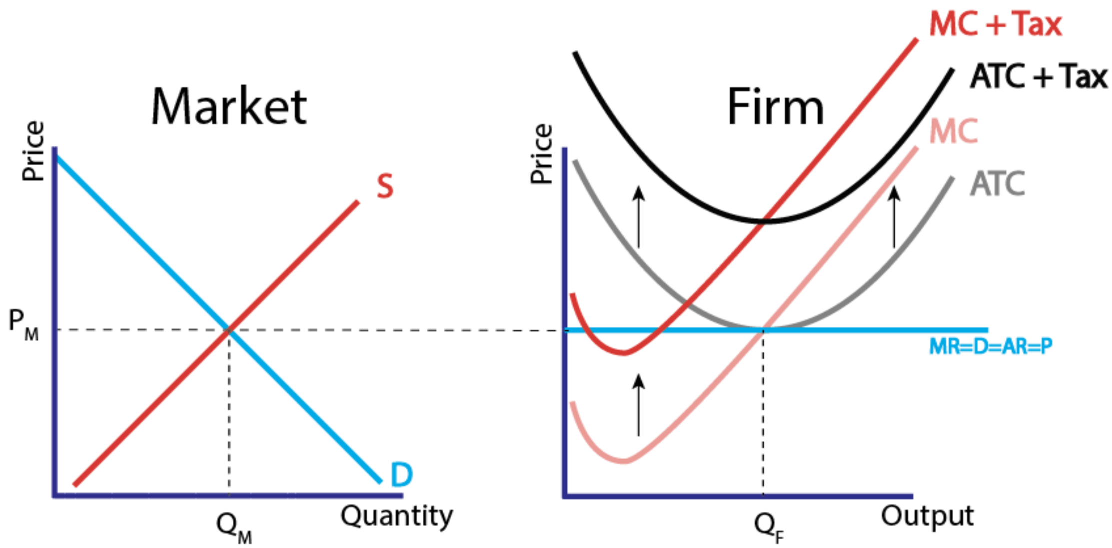
6.4 Government Intervention in Different Market Structures, Part 1

We've studied four markets so far that create DWL: monopoly, monopolistic competition, oligopoly, and monopsony. Governments often intervene in these markets to try and eliminate DWL. They also intervene in perfectly competitive markets in order to help producers and consumers (though this intervention *creates* DWL rather than getting rid of it). You have to know how government intervention operates in each of these markets.

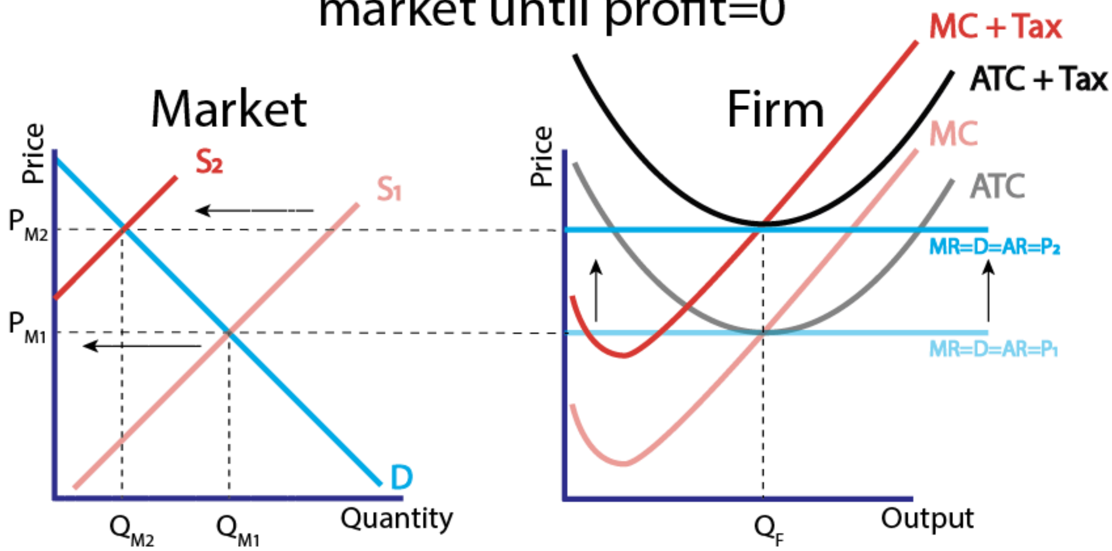
You already know about lump-sum taxes, per-unit taxes, lump-sum subsidies, and per-unit subsidies (see lesson 2.8). The most important thing to know here is that per-unit taxes and subsidies move both the Marginal Cost curve and the Average Total Cost curve, while lump-sum taxes and subsidies only move the Average Total Cost curve. I'm going to show you how the taxes work out below, but you can figure out subsidies if you just do the opposite. DON'T MEMORIZE WHAT FOLLOWS. As long as you know the rules about lump-sum and per-unit taxes, you'll be able to figure it out on the test. That's much easier than memorizing every change in every market for every kind of tax or subsidy.

Per-Unit Tax, Perfect Competition

Step 1: Tax leads to increase in MC and ATC



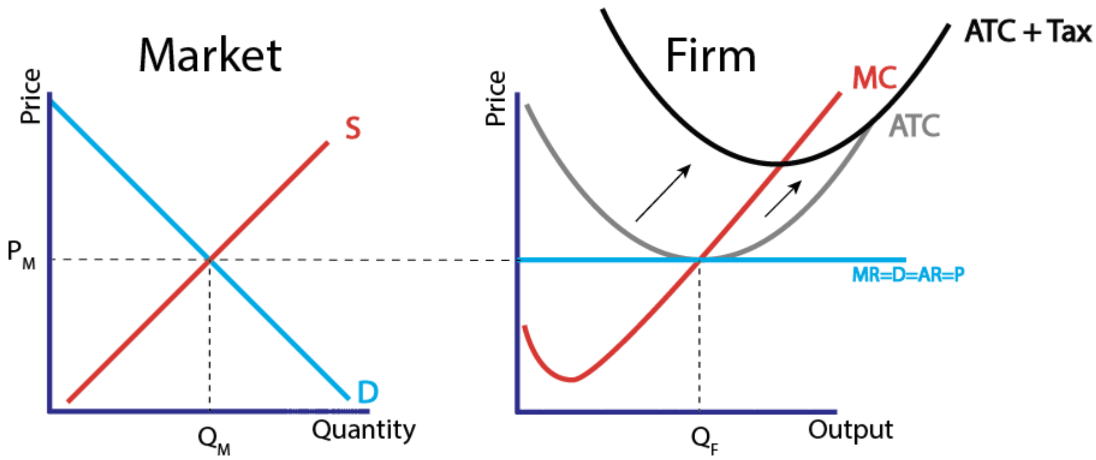
Step 2: Firms make losses, so they exit the market until profit=0



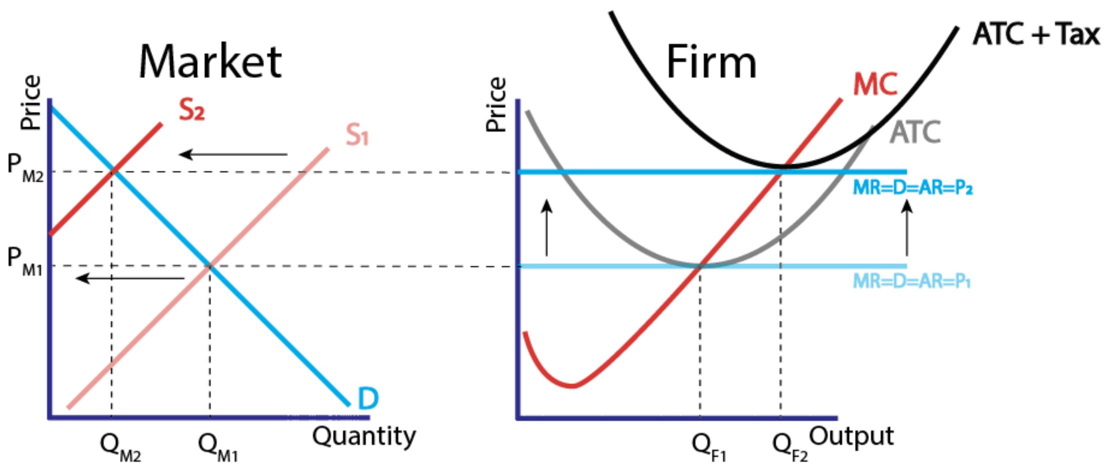
Notice that Q_F is the same as before, but Q_M↓

Lump-Sum Tax, Perfect Competition

Step 1: Tax leads to increase in ATC



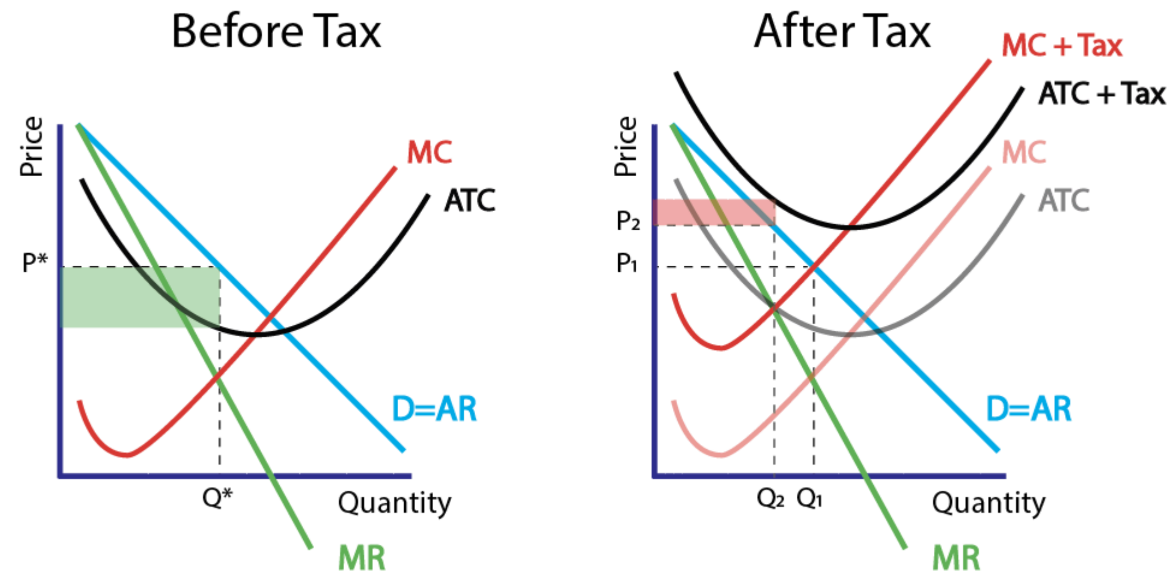
Step 2: Firms make losses, so they exit the market until profit=0



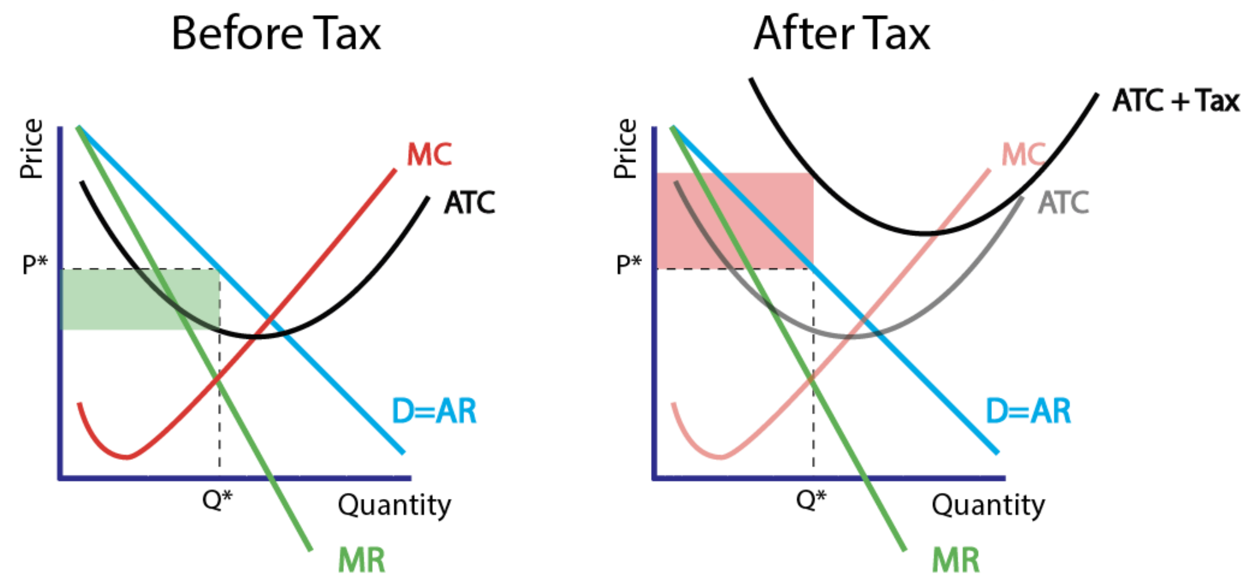
Notice that $Q_F\uparrow$, but $Q_M\downarrow$

Notice that in both cases, market supply decreases. This is why in Unit 1 we said that a tax will decrease the supply curve.

Per-Unit Tax, Monopoly



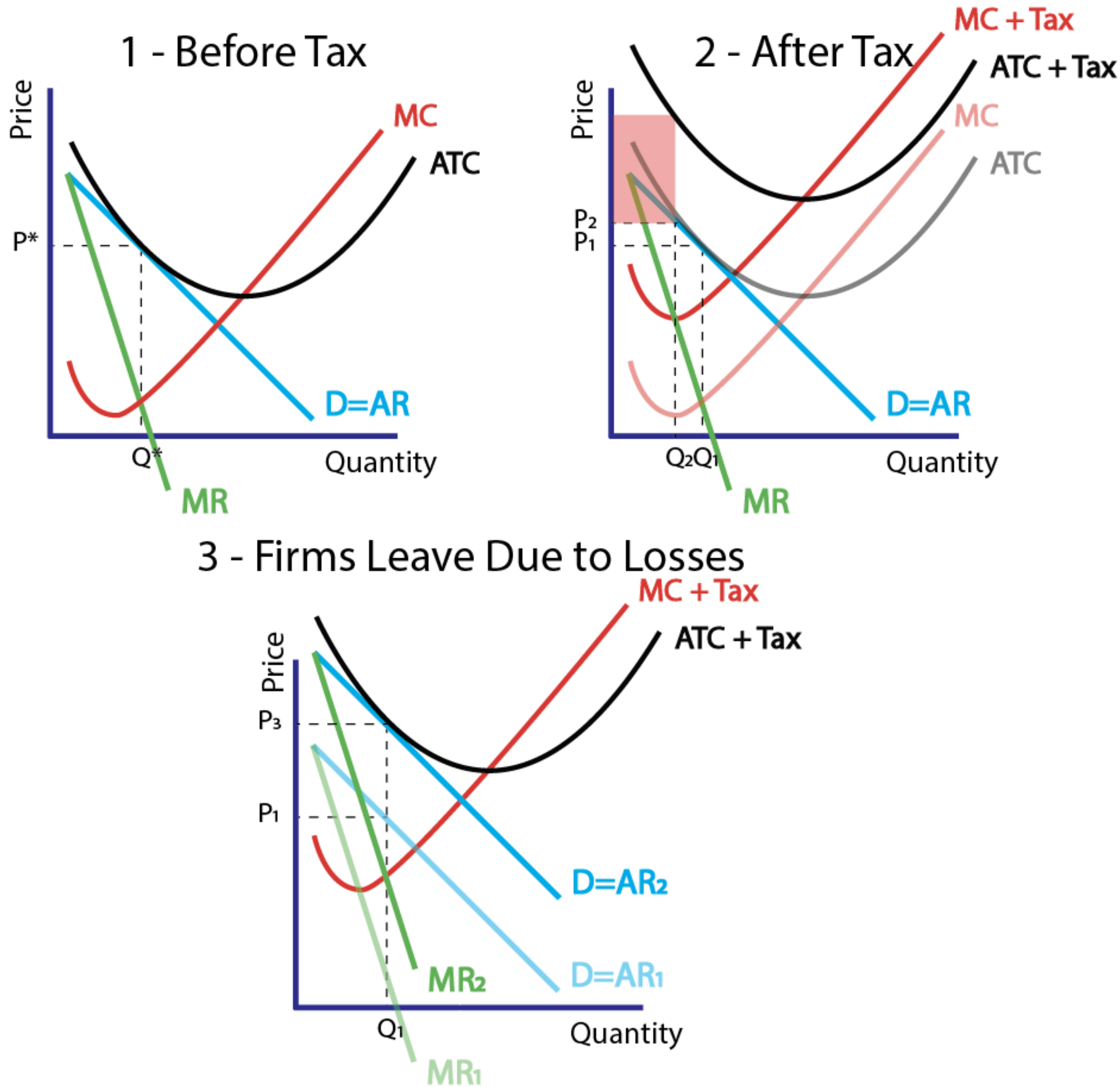
Lump-Sum Tax, Monopoly



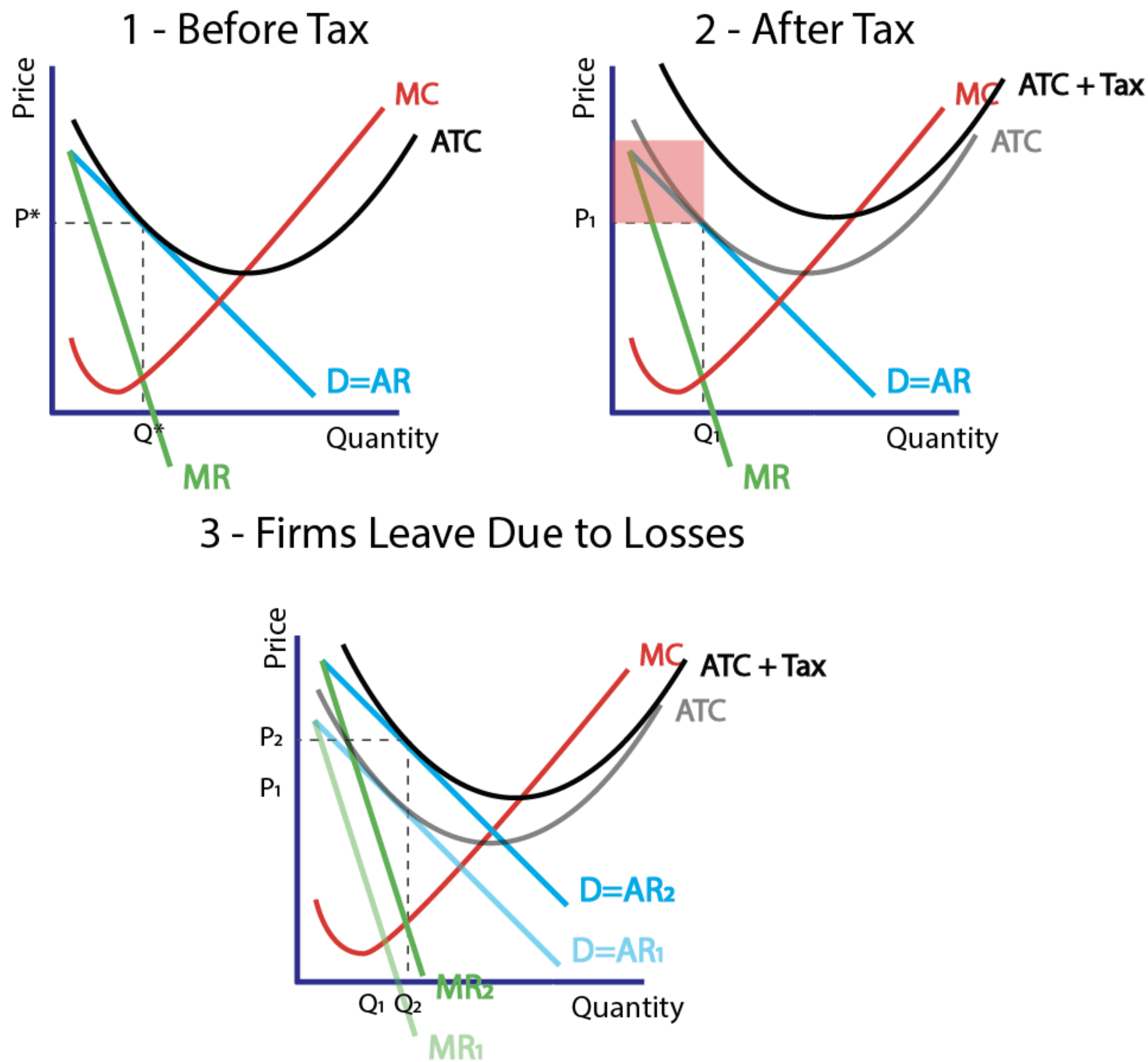
With monopolies, a lump-sum tax does not change P or Q because the MC curve is unaffected. So the only impact of a lump-sum tax will be on the firm's profits. Notice also that putting a per-unit tax on a monopoly is a terrible idea. It leads to less output and higher prices, so everyone is mad.

For the next two sets of graphs dealing with Monopolistic Competition, don't sweat over the details. I've never in my life seen a question about taxes and monopolistic competition, but it's helpful to see the pattern, since it is always the same.

Per-Unit Tax, Monopolistic Competition



Lump-Sum Tax, Monopolistic Competition

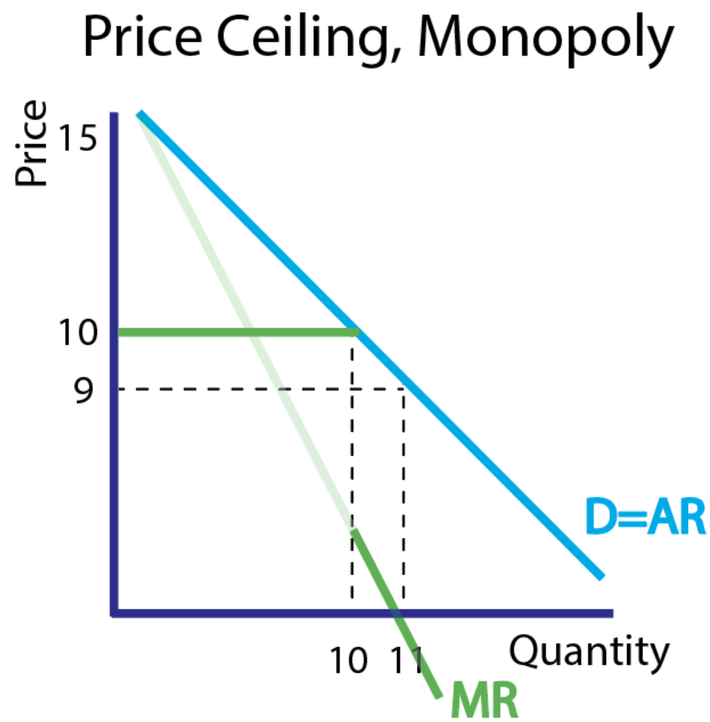


You do not need to know taxes or subsidies dealing with oligopolies. You'll notice a general pattern here: if governments want to reduce DWL, or even if they just want to help consumers, taxing monopolies and monopolistically competitive firms is a terrible idea. It always either doesn't solve the problem or makes the problem worse.

Return to the top of the document

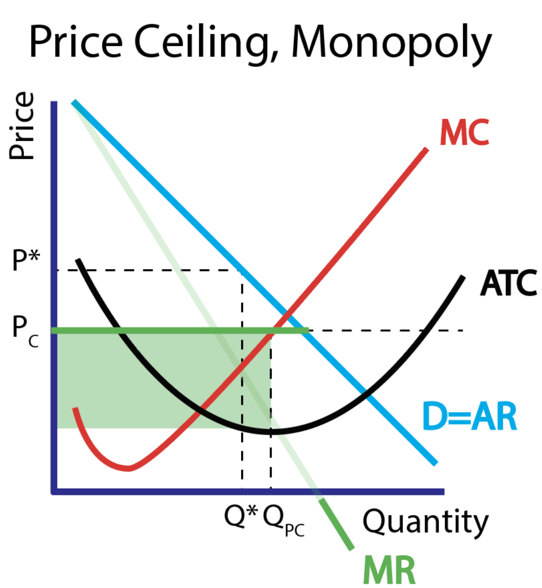
6.4 Government Intervention in Different Market Structures, Part 2

Since taxes are so ineffective with monopolies, one common intervention governments use is price ceilings. Unfortunately, this is really going to mess with our graphs, so brace yourself.



The main idea is this: with a monopoly, Marginal Revenue slopes downward so steeply because when the firm increases production (from 1 to 2, for example), they have to lower the price of all goods sold, so that lower price causes them to lose revenue twice (on the first good, and on the second good). But if the government sets a price ceiling, this whole idea of decreasing Marginal Revenue no longer applies. If they can't charge more than $10 for a good, then if they sell one good they get $10 in revenue, and if they sell 2 goods they get $10 more, so Marginal Revenue is flat. But that only applies until the price ceiling reaches the demand curve - in this case, selling 10 goods. If the firm wants to sell more than 10 goods, they have to lower prices, but now they are back in the situation they were in before the price ceiling was imposed: if they lower the price, they have to lower the price for *all* goods sold, so they are no longer protected from that decreasing MR. In other words, our MR curve is going to pick back up from where it was before the price ceiling was imposed.

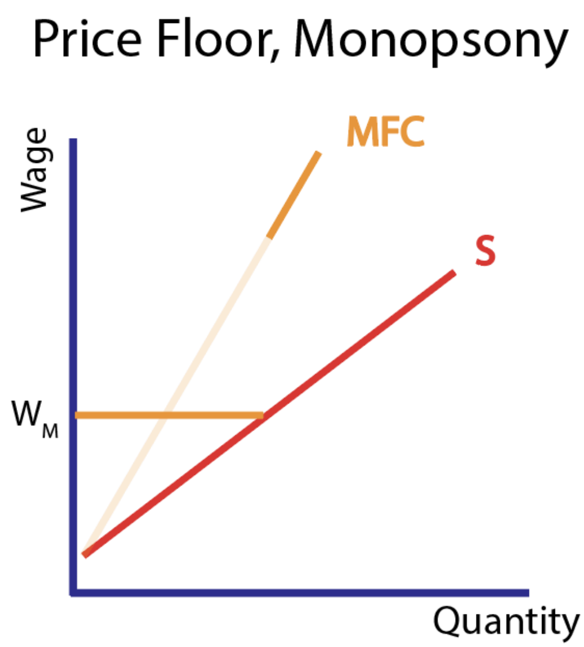
Before you panic, know that you don't need to be able to draw this curve on a free-response question. But you do need to know how it works, since you'll have to answer questions about equilibrium price and quantity when a price ceiling is imposed. The main thing to remember here is that <u>marginal revenue is derived from the downward-sloping demand curve, so the flat MR in a price ceiling extends only to the demand curve</u>.

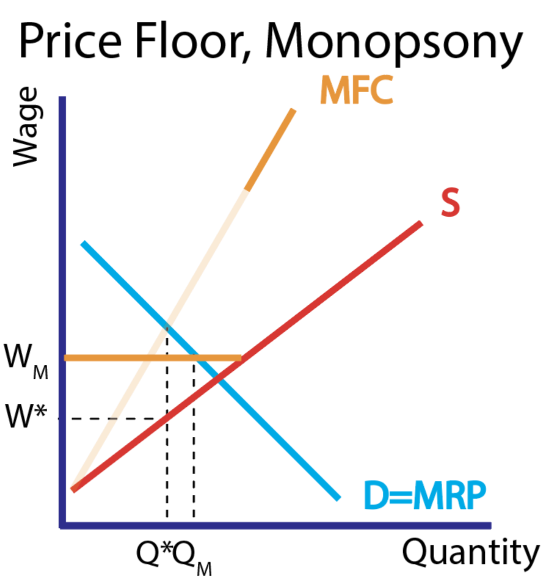


Now let's put our cost curve in and see how a monopoly will act. As always, firms will set quantity where MC=MR, so in the case to the left, the monopoly is now going to produce more, at Q_{PC}. Normally they'd set price as high as they can, up where the demand curve is, but now they can't set it higher than the price ceiling. Notice, by the way, how we are now producing closer to the allocatively efficient quantity, and consumers are better off with the lower price (hence why governments like to use price ceilings with monopolies).

What makes these questions tricky is that they're not going to give you the flat MR curve. You have to know that MR is going to be flat until it hits the demand curve, and then figure out the optimal quantity and price from there.

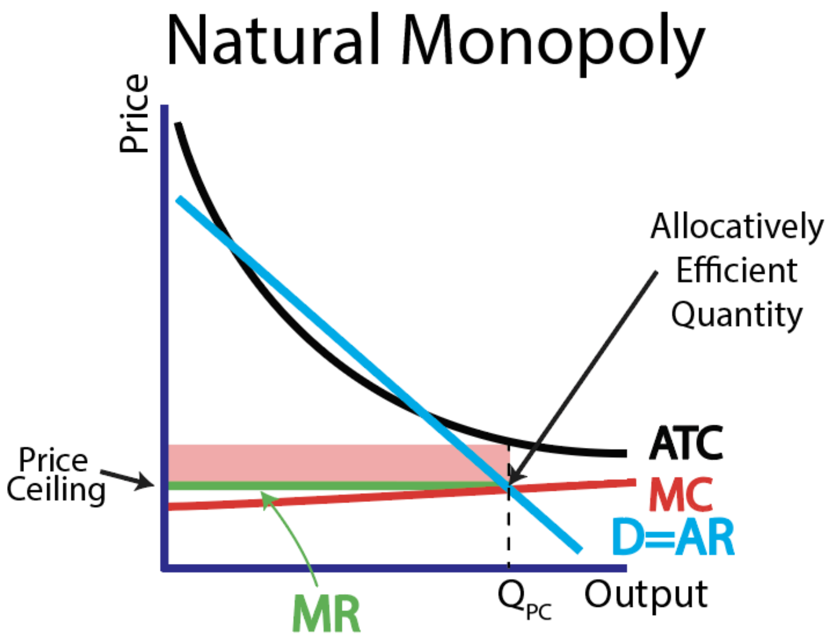
The same idea is going to hold true with a monopsony, only in reverse. The idea of a steep-upward-sloping MFC curve only holds true because when you hire more people, you have to raise wages for everyone. But if the government imposes a minimum wage, then no matter how many people you hire, your marginal cost is the same—all the way until you run into the supply curve. Past that point, if you want to hire more people than the market can supply at a minimum wage, you have to raise wages for everyone, so you jump right back up to where the MFC curve was originally. Once again, the thing to remember here is that <u>marginal factor cost is derived from the upward-sloping supply curve, so the flat MFC in a price floor extends only to the supply curve</u>.





On the left you can see how this works once we add the firm's D=MRP curve back in. As always, the firm will set quantity where MC=MR (or MFC=MRP, in this case). Normally they would pay the lowest wages that workers (the supply curve) could bear, but now they can't due to the minimum wage. So we end up at the rather surprising outcome that minimum wages actually lead to the firm hiring *more* people than they otherwise would, and everyone is making more money. This is why governments like minimum wages to fix monopsonies.

The final thing to know about intervention in various markets is that when the government tries to regulate a natural monopoly, things go a little haywire. If you recall, in a natural monopoly the firm is always facing a decreasing long-run ATC curve, which means the MC curve will never cross the ATC. As a result, if the government tries to make this monopoly produce the allocatively



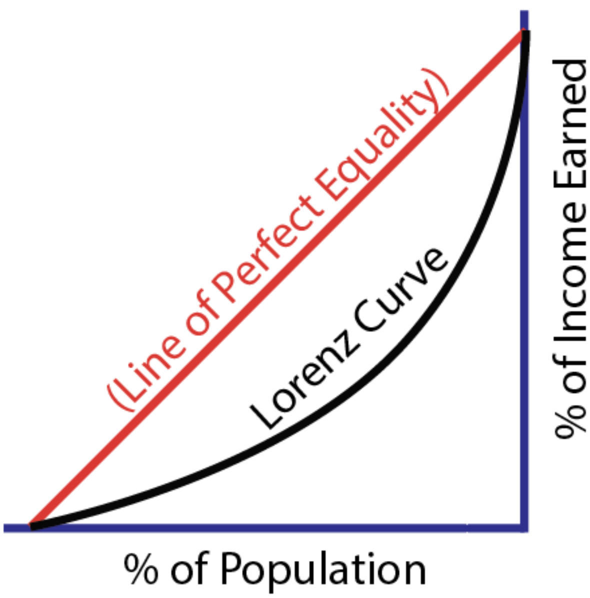
efficient quantity with a price ceiling, **it will also have to provide a lump-sum subsidy** to the monopoly in order to allow it to stay in business. To see why, look at the graph to the left. Since MC never crosses the ATC, at the allocatively efficient quantity ATC will always be higher than AR, which means the monopoly would be making a loss. To offset this, the government has to artificially lower the ATC curve by providing a lump-sum subsidy.

[Return to the top of the document]

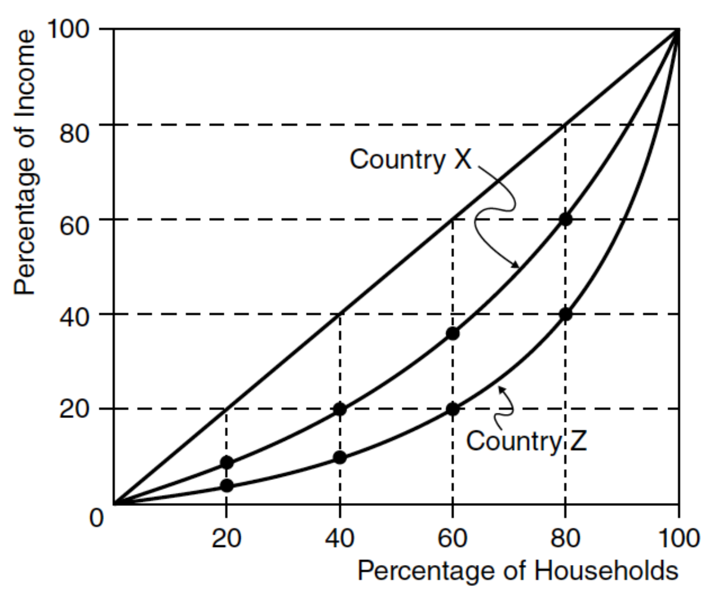
6.5 Inequality

The final kind of market failure happens in any economy where markets are left to run on their own: some people end up with a lot of money, and some people end up with almost nothing. The reasons why inequality happens are complicated, and the field of economics has nothing to say about whether inequality itself is morally right or morally wrong. But when levels of inequality get really high, it leads to serious economic problems (decreased lifespan and other health outcomes, increased crime, lower social mobility, decreased trust in institutions, etc.). These problems occur even in rich countries when inequality reaches high levels.

Economists measure inequality on a scale from 0 to 1, called the **Gini coefficient**. In a society of perfect equality, where everyone has the exact same amount of money, the Gini coefficient would be 0 (as in, "there is *zero* chance that this kind of society would ever exist"). In a society where one person owns everything and everyone else has nothing, the Gini coefficient would be 1 (as in, "*one* person has everything"). All societies exist somewhere in the middle of this spectrum. As of 2024, the US is .4, France is .32, and South Africa is .63. You don't need to be able to calculate Gini coefficients for the exam, but you do need to know that higher means more unequal.



We can also show inequality graphically through a **Lorenz curve**. A Lorenz curve puts the percent of the population on the x-axis and the percent of income earned on the y-axis. If a society were perfectly equal, the relationship between these two should be 1-to-1, or a straight diagonal line—that is, 1% of the population should earn 1% of the income, 50% should earn 50% of the income, 75% should earn 75% of the income, etc. The more the Lorenz curve bends away from that line of perfect equality, the more unequal a society is. Look for example at the Lorenz curves below. In Country X, the bottom 60% of the population only gets ~38% of the income, while in Country Z, the bottom 60% of the population only gets 20% of the income, so it is more unequal. You don't need to be able to draw a Lorenz curve on the exam, but you have to be able to interpret one.



One of the major drivers of inequality is the fact that each factor of production will only receive the value of its marginal product (as we saw last unit in our study of factor markets). That means in a place like rural Appalachia—where infrastructure is bad, capital is limited, access to financial markets is poor, and education is generally low—labor can't be very productive, which means wages will necessarily be depressed. In addition, workers have almost no bargaining power in such areas due to the fact that unemployment there is so high, leading to even lower wages.

Tax Systems

The way governments tend to fix inequality is through taxation and redistribution. There are two types of taxation systems you need to know: a **progressive tax system** is set up such that as your income rises, you pay a higher percentage of your income in taxes. This is the kind of system we use in the US, and you can see an example to the right.

TABLE 1

2017 Tax Schedule and Rates

Single filers		
Taxable income ($)		Current marginal rate (%)
Over	But not over	
0	9,325	10
9,325	37,950	15
37,950	91,900	25
91,900	191,650	28
191,650	416,700	33
416,700	418,400	35
418,400	and over	39.6

Source: IRS tax brackets.

Alternatively, in a **regressive tax system**, as your income rises, you pay a lower percentage of your income in taxes. These kinds of taxes tend to make inequality worse. Such a system might sound crazy (after all, who would want to design a system that makes inequality worse?), but we actually have some regressive taxes in the US as well. Sales tax, for example, disproportionately impacts the poor. To see how, consider Bobby and James. Bobby makes $100 a month, and he spends it all on groceries. If there's a 6% sales tax, that means he spends 6% of his income on sales tax. James, on the other hand, makes $10,000 a month. He buys $1,000 of groceries every month, which means he pays a total of $60 in sales tax. In absolute terms, that's more money than Bobby pays in sales tax, but this $60 represents only 0.6% of James's income. So this tax is a major burden for Bobby, but it barely even registers for James.

Return to the top of the document