 Thinkwell’s Microeconomics

Please note: the transcripts are unedited, uncorrected transcriptions of the video lectures.
THINKWELL’S MICROECONOMICS VIDEO TRANSCRIPTS

CHAPTER 1: INTRODUCTION TO ECONOMIC THINKING

Basic Economics Ideas
Defining Economics 1
Understanding the Concept of Value 3

Using Graphs
Using Graphs to Understand Direct Relationships 5
Plotting a Linear Relationship Between Two Variables 7
Changing the Intercept of a Linear Function 9
Understanding the Slope of a Linear Function 11

Advanced Graphical Concepts
Understanding Tangent Lines 13
Working with Three Variables on a Graph 15

Production Possibilities
Understanding the Concept of Production Possibilities Frontiers 17
Understanding How a Change in Technology or Resources Affects the PPF 21
Deriving an Algebraic Equation for the Production Possibilities Frontier 23

Comparative Advantage
Defining Comparative Advantage With the Production Possibilities Frontier 27
Understanding Why Specialization Increases Total Output 31
Analyzing an International Trade Using Comparative Advantage 33

CHAPTER 2: UNDERSTANDING MARKETS

Demand
Determining the Components of Demand 38
Understanding the Determinants of Demand 40
Understanding the Basics of Demand 43
Analyzing Shifts in the Demand Curve 45
Changing Other Demand Variables 47
Deriving a Market Demand Curve 49

Supply
Understanding the Determinants of Supply 51
Deriving a Supply Curve 53
Understanding a Change in Supply vs. a Change in Quantity Supplied 55
Analyzing Change in Other Supply Variables 57
Deriving a Market Supply Curve from Individual Supply Curves 59

Equilibrium
Determining a Competitive Equilibrium 61
Defining Comparative Statics 63
Classifying Comparative Statics 65

Elasticity
Defining Elasticity 68
Calculating Elasticity 69
Applying the Concept of Elasticity 71
Identifying the Determinants of Elasticity 73
Understanding the Relationship Between Total Revenue and Elasticity 75

Interfering With Markets
Understanding How Price Controls Damage Markets 77
Understanding the Problem of Minimum Wages in Labor Markets 79
Understanding How an Excise Tax Affects Equilibrium 82
Agriculture Economics
Examining Problems in Agricultural Economics  84

CHAPTER 3: CONSUMER CHOICE AND HOUSEHOLD BEHAVIOR

Utility Theory
Understanding Utility Theory  87
Finding Consumer Equilibrium  88

Budget Constraints and Indifference Curves
Constructing a Consumer's Budget Constraint  89
Understanding a Change in the Budget Constraint  91
Understanding Indifference Curves  92

Consumer Optimization
Locating the Consumer's Optimal Combination of Goods  95
Understanding the Effects of a Price Change on Consumer Choice  97
Deriving the Demand Curve  98

CHAPTER 4: PRODUCTION AND COSTS

The Basics of Production
Understanding Outputs, Inputs, and the Short Run  99
Explaining the Total Product Curve  101
Drawing Marginal Product Curves  104
Understanding Average Product  106
Relating Costs to Productivity  108

Variable Costs
Defining Variable Costs  109
Graphing Variable Costs  110
Graphing Variable Costs Using a Geometric Trick  111

Marginal Costs
Defining Marginal Costs  112
Deriving the Marginal Cost Curve  114
Understanding the Mathematical Relationship Between Marginal Cost and Marginal Product  116

Average Costs
Defining Average Variable Costs  118
Understanding the Relationship between Average Variable Cost and Average Product of Labor  119
Understanding the Relationship between Marginal Cost and Average Variable Cost  120

Total Costs
Defining and Graphing Average Fixed Cost and Average Total Cost  122
Calculating Average Total Cost  124
Putting the Cost Curves Together  126

Long-Run Production and Costs
Defining the Long Run  128
Determining a Firm's Return to Scale  129
Understanding Short-Run and Long-Run Average Cost Curves  131
Understanding the Difference Between Movement Along a Cost Curve and a Shift in a Cost Curve  134

Isocost/Isoquant Analysis
Constructing Iso-Cost Lines  135
Understanding Isoquants  137
Finding the Cost- Minimizing Combination of Capital and Labor  139

CHAPTER 5: PERFECT COMPETITION
The Basic Assumptions of Competitive Markets
Understanding the Role of Price 142
Understanding Market Structures 143
Finding Economic and Accounting Profit 146

Calculating Profit and Loss
Finding the Firm's Profit-Maximizing Output Level 149
Proving the Profit-Maximizing Rule 152
Calculating Profit 153
Calculating Loss 156
Finding the Firm's Shut-Down Point 158

Market Supply
Deriving the Short-Run Market Supply Curve 160
Relating the Individual Firm to the Market 164
Examining Shifts in the Short-Run Market Supply Curve 167
Deriving the Long-Run Market Supply Curve 169

Competitive Firms' Responses to Price Changes
Examining the Firm's Long-Run and Short-Run Adjustments to a Price Increase 171

CHAPTER 6: OTHER MARKET MODELS

Monopolies
Defining Monopoly Power 175
Defining Marginal Revenue for a Firm with Market Power 177
Determining the Monopolist's Profit-Maximizing Output and Price 179
Calculating a Monopolist's Profit and Loss 182
Graphing the Relationship between Marginal Revenue and Elasticity 184

The Social Cost of Monopoly
Determining the Social Cost of Monopoly 187
Calculating Deadweight Loss 189
Understanding Monopoly Regulation 192

Oligopoly
Introducing Oligopoly and the Prisoner's Dilemma 195
Understanding a Cartel as a Prisoner's Dilemma 198
Understanding the Kinked-Demand Curve Model 199

Monopolistic Competition
Defining Monopolistic Competition 200
Understanding Pricing and Output Under Monopolistic Competition 202
Understanding Monopolistic Competition as a Prisoner's Dilemma 204

CHAPTER 7: RESOURCE MARKETS

The Derived Demand for Labor
Deriving the Factor Demand Curve 206
Deriving the Least-Cost Rule 209
Analyzing the Labor Market 210

Monopsony
Understanding Labor Market Power and Marginal Factor Cost 213

Capital Markets
Analyzing Capital Markets 215

CHAPTER 8: MARKET FAILURES
### Overview of Market Failures
Understanding Market Failures 217

### Public Goods and Public Choice
Defining Public Goods 220
Analyzing the Tax System 223
Understanding Public Choice 225

### Uncertainty
Understanding Expected Value, Risk, and Uncertainty 229
Understanding Asymmetric Information as an Economic Problem 231
Understanding Moral Hazards in Markets 234

### Externalities
Defining Externalities 236
Explaining How to Internalize External Costs 237
Explaining How to Internalize External Benefits 239

### Solutions to Externalities
Finding a Market Solution to External Costs 240
Finding a Negotiated Settlement to an External Cost 242
Applying the Coase Theorem 244

### CHAPTER 9: INTERNATIONAL TRADE

#### The Basics of Open Economies
Determining the Difference Between a Closed Economy and an Open Economy 246
Understanding Exports in an Open Economy 248
Analyzing a Change in Equilibrium in an Open Economy 249

### CHAPTER 10: EVALUATING MARKET OUTCOMES

#### Normative Economics
Measuring the Benefits of Consumption 250
Using the Demand Curve as a Measure of Benefit 251

#### Calculating Total Economic Value
Quantifying Benefit 252
Quantifying Cost 253
Determining Total Social Cost 254
Understanding Economic Value 255

#### Consumer and Producer Surplus
Understanding Producer and Consumer Surplus 256
Calculating Total Economic Value 257

#### Market Interference and Economic Value
Understanding the Effects of Price Control 259
Understanding How Price Controls Destroy Economic Value 261
Evaluating the Effects of an Excise Tax 263
Assessing the Effect of an Excise Tax on Economic Value 265
Understanding How a Tax Can Create Deadweight Loss 266

#### International Trade and Economic Value
Evaluating the Gains from International Trade 267
Understanding the Effects of Tariffs on Consumer and Producer Surplus 269
Warning: What follows is an unedited, uncorrected transcript of the professor's lectures.
Introduction to Economic Thinking

Basic Economics Ideas

Defining Economics

Students sometimes begin the Economics course with a misperception. And that is that Economics is about money. Absolutely not. Economics is a much broader set of tools that can apply to all kinds of decisions that you make in daily life. The definition of Economics is this. Economics is the study of rational choice under conditions of scarcity.

That definition has two terms in it that are key. The first is scarcity. Scarcity means an imbalance between the amount of something that people want and the amount that's freely available. One of the best ways to come to grips with the concept of scarcity is to try to imagine something that isn't scarce. What would that be? Sometimes students suggest air isn't scarce. It's all around us. Yes, but the kind of air that people want, clean, breathable air is certainly scarce and especially scarce in cities with a lot of pollution. At one point in Tokyo there were vending machines on the street where people would insert a coin to buy a breath of breathable clean air. In that case air was scarce. There was an imbalance. People wanted more of it than there was freely available.

Well, what about space? There seems to be plenty of space. Yes, but there's not a lot of space in cities. There less space in your dorm room than you’d want. All this stuff has to be parcelled out. Because there isn't as much as people would like to have if it were freely available. What about garbage? There seems to be plenty of garbage. No scarcity of garbage. Aha! But see, that's where the definition comes into play. It's not that there's an infinite amount of garbage. There is a strictly limited amount of garbage. The reason garbage is not scarce is that nobody wants it. Scarcity is an imbalance between the amount of something that people want and the amount of that good that is freely available. Any time something is scarce we've got to figure out how to use it, how to share it, how to parcel it out among its competing uses. And that requires some kind of decision, some kind of choice. And that leads us to the second term, rational choice.

Rational choice or the word rationality, in Economics, refers to people making calculated, self-interested decisions. It requires that you be willing to consider cost and benefits, all of the factors that are involved in a decision and choose that course of action that is most satisfying to you. The one that maximizes your wealth, the one that maximizes your company's profits, the one that maximizes your satisfaction from the way you use your limited income or your limited amount of time.

We say that an agent is rational if that agent considers cause and effect. If that agent considers the consequences of his or her choices and chooses those courses of action that provide the most satisfaction. Rational choice is calculated self-interest. So if we have calculated self-interested people operating in a situation of scarcity, then we've got Economics.

One example of this is the concept of opportunity cost. We imagine that when people make a choice they consider the opportunity cost of that choice. The opportunity cost of a choice is the best alternative you give up when you make that choice. For instance, if you go to your Economics class one morning, the best alternative might have been an extra hour’s worth of sleep. Rather than staying in bed and sleeping, you chose to come to your Economics class instead. The sleep was the opportunity cost of your choice.

If you are enrolled in college this year, you're getting a good education which might be satisfying in it's own right. It may also be that that education is your key to a higher salary that leads to more goods and services in the future. But the opportunity costs of those goods and services in the future and the satisfaction of your expanded mind is the money that you're not making now and all of the toys that you could be buying with the income that you're giving up. Because you're in school, you aren't holding down a job with all of time that you are using studying and attending classes. The time away from work is lost income, lost goods and services and lost satisfaction.

So the opportunity cost of your investment in your future is the present satisfaction you can get from a higher paycheck. There's an opportunity cost. There's no such thing as free lunch. Anything that you enjoy, you enjoy at the cost of giving up something else. If you enjoy driving a pickup, the opportunity cost is you're not driving a Volkswagen that day. If you enjoy a vacation to the Bahamas, your opportunity cost might be a trip to California. Every choice you make involves an opportunity that you don't choose.

Well, what does that have to do with money? Not very much. See the point is, Economics is not about money. Economics is about analyzing the way people make choices in conditions of scarcity. We can make predictions about how you choose to use your time, between studying, working, playing, spending time with friends. We can come up
Introduction to Economic Thinking

Basic Economics Ideas

**Defining Economics**

with an economic model that predicts the way you'll respond, if we know enough about your preferences, your abilities and the constraints that are put on you from the outside. We could also come up with economic models of who people choose to marry, economic models of when countries go to war. There are even economic models of which religions people choose to affiliate with.

One economist is even suggesting that you can come up with an economic model of who chooses to commit suicide as an economic calculation. If the cost and the benefits are aligned properly, people just choose to check out. Now some people are offended by this. They think, "Well why should we have an economic model of these very personal things?" And, maybe they're right. But the point is Economics is a very flexible set of tools and it seeks to apply itself anywhere rational agents are operating in a situation of scarcity, anywhere that there are goods and services that are strictly limited that have to be shared in some way.

In the next lecture we'll look at another definition of Economics, one that has to do with value. And this might seem a little more applicable to business.
Introduction to Economic Thinking

Basic Economics Ideas

Understanding the Concept of Value

In the last lesson we defined Economics as the study of rational choice under conditions of scarcity. Now we will introduce a second definition of Economics that may seem to have a little more to do with business. Economics is the study of the creation and distribution of value.

To understand the importance of this definition we need to know what an economist means by the term value. And value sometimes becomes a bit of a squishy concept. We know what it is, but it's hard to measure in some cases. And therefore, seems difficult to define. Value is simply the difference between the benefit of an activity and the cost of that activity. It's a kind of measure of profit or extra satisfaction.

For instance, suppose I sell bread in my bakery. If I can sell my bread for $5.00 and the cost of producing the bread is $3.00, then I'm creating $2.00 worth of profit. In some sense the $2.00 worth of profit is the value of the bread. But it's a little more subtle than that. Suppose you have an apple and I come along and offer to trade you your apple for my peach. Would you do the trade? Sure you would, if you liked the peach more than the apple.

The trade has a benefit and a cost for you. The benefit is you get a peach. The cost is you give up an apple. The apple is the opportunity cost of the peach. The benefit from the peach is weighted against the cost of parting with your apple. Value is created if you like the peach more, because then the benefit is greater than the cost. But exactly how much value is created by this trade?

In order to know how much value is created we need some kind of yardstick, some way of measuring benefits and costs. And one way that we can do that is to use money as the yardstick. To assign dollar values to benefits and to costs. Suppose you think the peach is worth a dollar. That is, you be willing to spend a dollar on this peach. And suppose the apple is worth 40 cents to you. That is a measure of the satisfaction you would get from eating the apple. If I trade you your apple for my peach, you're getting a benefit that you value a $1.00. And the opportunity cost is the 40 cents worth of satisfaction that you would have received from the apple. One dollar’s worth of benefit minus 40 cents worth of cost means that we have created 60 cents worth of value for you when you trade the apple for the peach.

See, putting a dollar value on things allows us to measure economic value in a concrete way. But this sometimes we require us to put dollar values on things that we might resist. For instance, what is the benefit that you get from working at your job? Well, I get a salary and I can use that to buy toys. Well a good approximation of the benefit you get from your job is the dollar value that you’re paid when you work there. The paycheck that you receive might be a good approximation of the benefit you get from working at the job.

Now, what about the cost of working at the job? Well the cost is you’re not spending time with your family and friends. How would you put a dollar value on that opportunity cost? Well, you can imagine that we sit you down at a table and say, “Look, you can spend time with your family and friends or you can go to work.” And you say, “Well, I think I’d rather go to work if the salary is right.” And I’d say, “Well, what salary would make you not care whether you work or stay home with your friends?” And we’d find a number. And the number that may be the amount of money that you would be willing to accept in exchange for coming to work, rather than being with your friends.

If I’m careful about the way I structure this experiment I can find a dollar value that I consider your price, or your cost for giving up your time with your friends. I can put a dollar value on it by seeing how you behave. I ask you to go out with me for a pizza. And you say, “No, sorry I have to work this afternoon.” I found out that you’re only making $5.00 working this afternoon. So I know that you’d rather have $5.00 than spend the afternoon with me, with a pizza. Well, that tells me something about how you value this opportunity.

The point is, once we can assign dollar values to things it becomes very easy to calculate economic value. The value of the benefit or the economic number, the dollar value of the benefit, minus the dollar measure of the cost gives us some kind of profit left over. And that’s called economic value.

Let's consider another case. Suppose I'm baking bread and I can sell it for $5.00 a loaf.

Well, what is the opportunity cost of my baking the loaf of bread? I've got milk, eggs, and flour, all of these resources that I'm using. And they all have other uses, they have opportunity costs. If the price of eggs is a good measure of
Understanding the Concept of Value

the opportunity cost of eggs, that is, someone is willing to pay 50 cents for my egg that I’m about to use in a loaf of bread, that tells me that the egg is worth 50 cents to him. Rather than put the egg in my loaf of bread, I can give it to him to make an omelet. And the 50 cents that he’s willing to pay is a good measure of the opportunity cost of the egg. After all, I could have the 50 cents if he has the omelet. So I’m giving up 50 cents in order to make this loaf of bread using this egg.

What about my own labor? My own labor has an opportunity cost. I can be at home watching television or outside playing with my children. In that case, I would be enjoying another activity rather than using my labor to make bread. Well, what dollar value would I put on those foregone opportunities? That is a measure of the opportunity cost of my labor. If I’m making bread then there’s several factors of production that go into this business. There is labor, there is the land that my store sits on, there are all of the resources like milk, flour and eggs that I’m using to produce the bread. There is also the capital, the money that I borrowed to make my operation run, the money that I borrowed to build the store and to buy the resources before I’ve been able to sell the bread. There’s also what we call entrepreneurship. If I’m the owner of this business I’m taking a risk in bringing my talents in to help the business run well. There’s an opportunity cost to that also.

I could have peace of mind rather than being worried about whether my business is going to succeed. I can use my talent making art or writing stories instead of running a business. So entrepreneurship also has an opportunity cost. If you add up the opportunity costs of all the resources that are used in the production of bread, you have the opportunity costs of the production of bread. If you subtract the opportunity costs from the benefits of the loaf of bread that you’re creating, you have the economic value of that bread.

The economic value, the difference between the benefit and the cost of producing it. Well, who gets that economic value? That’s a good question. Who gets the bonus? Who gets the gravy? The gravy can be spread over all of the resources. I might decide that I’m going to pay myself as an entrepreneur a very, very large bonus, over and above the opportunity cost of my time and talent. That means the entrepreneur is getting that economic profit in the form of a bonus or extra profit. It could be that the person who owns the land that I’m sitting on has some power over me and charges me an extra high rent. And that way the land owner is getting a share of the economic profit by demanding a higher payment for the land than he could demand from someone else.

If I am using capital and people are putting their money at risk in my business, I may decide to give these capital owners some extra bonus from having taken the risk. If I pay them more than their opportunity costs, I’m sharing the economic profit with them. Economic profit is the difference between a benefit of an activity and its opportunity costs. And that economic profit will be created any time you create something of value. But, who gets the economic profit depends on the way the business is organized. It depends on how competitive the markets are for labor and capital. It depends on how specialized particular skills are that the entrepreneur brings into the business.

The important thing to know at this point is the concept of economic value, the difference between benefits and costs. Anytime economic value is created economist are usually pleased. This is the thing that they're trying to make more of, whether they're making it through trades, like the apple for the peach, or whether they're making it through the production of goods like the bread where the benefit is greater than the opportunity cost of production.

Economics is about figuring out full and fair values created and who gets the value in the form of benefits beyond their opportunity costs.
Economists use graphs to represent a lot of their analysis. So, we’ll take a while here to review how graphs work and see how we’ll be applying them throughout the lessons. You’re familiar with graphs like pie charts, which show you how one sum can be divided up into its parts. You’re also familiar with bar charts that show you how a sum changes over time or can be divided up also into its parts. In economics, we’re going to use graphs in two-dimensional space and these graphs usually represent relationships between two variables or in some case relationships among three variables. What we’ll do in this lesson is show you how those graphs work and prepare you for tools that we’ll be using in other lessons.

Let’s start by having a look at the two-dimensional graph space. We have a vertical axis, along which we measure the quantity of one variable and a horizontal axis, along which we measure the quantity of another variable. And, usually these two variables are related in some way that the economist finds interesting. Let’s look at an example of information that we represent in this two-dimensional space. Suppose we want to graph the relationship between two variables, consumption and household income. Consumption is the amount of money that a household spends on goods and services and income is their income, the money that they take in from work and other sources.

Now, how are we going to represent the relationship between consumption and income? Let’s start by graphing income on the horizontal axis. And, I will use this abbreviation, INC, to stand for a household’s income. And, we’ll measure this income in dollars, so I’ll put a dollar sign here to remind me that everything that’s measured on this axis is measured in dollars. Now, I can calibrate the axis by marking off these tick marks and putting numbers along them so that I have a scale along which to represent my information. Let me let each of these tick marks represent $10,000.00 worth of annual income. So, the first tick mark here I could label 10 for $10,000.00. The second tick mark I could label 20. The third I could label 30. And I might want to be careful here not to put too many numbers close together or the graph could start to get crowded and unwieldy. So what I’ll do is I’ll skip every other 10,000 and represent 10, 30, 50, 70 thousand. Here’s 80, I’ll skip it, 90, 100 thousand, 110, and 130. And the numbers of course continue as you go out. The numbers increase representing larger annual incomes as you move to the right.

On the vertical axis, now, I’ll represent the other variable. And this is going to be consumption spending, which I’ll represent with the letter C. So if I increase my vertical altitude here, I’m increasing the consumption spending of this household. I’ll but a dollar sign here to remind me that consumption is measured in dollars. And I’ll do the same thing that I did on the horizontal axis. I’ll calibrate using tens, and I’ll skip every other one so that I don’t get too crowded. Here's $50,000.00 in annual consumption spending. Here’s 70. Here’s 90. Here’s 110 and then I’ll just let the numbers go on up.

Now, what I do in this space—now that I’ve created it, is I can represent information about consumption spending. Suppose I have a table of numbers and each one of these numbers represents the income spending information for a particular household. Each household, each data point that I have represents the behavior of a household. So I’m going to put data points in this graph and each point will represent the behavior of one particular household.

So, suppose I have a household and I know that their annual income is $30,000.00 and their annual consumer spending is $40,000.00. Now you might ask, how can that happen? If their income is only 30,000 how can they spend 40,000. Well, perhaps they have other sources from which to get money. Perhaps they have savings that they can draw on. Or perhaps they receive payments from the government or other sources of income. So, I know that a household has a $30,000.00 annual income is spending $40,000.00 a year on consumption spending. I will go up to this point and put a dot where those two numbers come together. Now notice the horizontal coordinates of this point represent the income of the household. That is $30,000.00. The vertical coordinates represent the consumption spending.
Using Graphs to Understand Direct Relationships

So see, each point in this space represents a combination of two pieces of information about a single household. Their annual income and their consumption spending for that year. I can label this point with a pair of numbers that represent its coordinates. I'll write first the horizontal coordinates. That is 30 to represent their income, or 30,000. And second, I'll write the vertical coordinate of 40 to represent the spending data.

Now, sometimes you'll hear the information represented here called the X variable, or this number called the X-axis. The horizontal axis is sometimes referred to as X and the first number here represents the X coordinate. Sometimes this axis is called the Y-axis. And the number that's represented here is the Y coordinate. So the numbers in parenthesis are the X and the Y. The horizontal and the vertical numbers here, written side by side. And each point in this space will have two numbers, two coordinates associated with it.

Now let's suppose that I have data on a lot of households. And I know the income and the spending for several households. I can represent the points in this space and form what we call a scatter diagram. The scatter diagram just tosses the points out into space and looks at them. So why don't we do that? If I have information, say that a house that has an income of $10,000.00 was spending 30,000 that year. I can put a point here, with a horizontal coordinate of 10 and a vertical coordinate of 30. Suppose I know that a household that had $90,000.00 annual income was spending 70. So let's put a point up here with a horizontal coordinate of 90 and a vertical coordinate of 70. By the way, what was that household doing? If they were spending less than their income, then they were saving the difference.

And then, suppose I have a lot of other points. I'll just graph some samples here. Suppose I had a point like this one, and a point like this one, and a point like this one, and a point like this one, and so forth. Each point represents a combination of income and spending for one particular household. As I fill in the information that I have, in each case I'm representing the household's income by the horizontal and the household's consumption by the vertical. This is a scatter plot. It's information arrayed in a two dimensional space to represent two variables for each data point. That is, the income and the associated consumption spending for one particular household.

Now, once you've got a scatter plot like this the next thing that you're tempted to do is fit a line into that information to notice the general relationship between the two variables. Notice as we look at this information, in general as income increases consumption spending increases for these data points. We say that there is a direct relationship between income and consumption, a direct relationship or a positive relationship between these points.

If I wanted to notice the general relationship, I could fit a line to these points that would look something like this. And this line represents the general relationship between consumption and income. That is in general, for the points in this data set, as income increases consumption spending increases. The thing that's important to notice here is because of the direct relationship the line that passes through these points has a positive slope. Now what does that mean a positive slope? I'll get to that in just a moment. But I want to stop here and make the point that in Economics one of the first things that we do is we notice general relationships between data points. We notice, for example, that when income increases consumption increases. We observe a positive or a direct relationship among the points.
Introduction to Economic Thinking

Using Graphs

Plotting a Linear Relationship Between Two Variables

Now we'll look at another relationship between variables represented graphically. In this case we'll look at the behavior of the hamburger customer, Bob. And we'll look at what Bob does when the price of hamburgers changes. I'm going to now look at a data set that represents the price of hamburgers and the quantity of hamburgers that Bob consumes every week. Let's look at the relationship between these two variables.

When the price of hamburgers is $5.00 Bob didn't buy any, they're too expensive. Even if the price drops down to $4.50 per hamburger Bob still doesn't buy any hamburgers. If the price drops down to $4.00 per hamburger Bob decides to treat himself to one hamburger a week. If the price falls to $3.50 Bob buys two hamburgers each week, and so forth. Every time the price of hamburgers drops by 50 cents Bob decides to buy one more hamburger a week. And finally, if hamburgers drop down to a price of a dollar apiece, Bob buys one every day. That is, seven hamburgers per week.

Let's now represent this relationship between two variables, price and quantity, in a graph. The first thing that I'll do is I will draw two axes and calibrate them. Here are my axes for this particular problem. The two variables that I'm interested in analyzing are the price of hamburgers and the quantity of hamburgers consumed per week. So, let me go ahead then and calibrate my axes.

On the horizontal axis, I'm going to represent the quantity of hamburgers consumed by Bob every week. So if I increase my distance from the origin, that is, the place where the axes intercept, as I increase my distance from the origin horizontally Bob is buying more and more hamburgers every week. And those numbers are represented here, one, two, three, four and so forth. On the vertical axis I'll represent the price the hamburgers sell for. The sale price of hamburgers is represented on the axis here. Increasing $1.00, $2.00, $3.00, $4.00 and so forth, per hamburger. So, on the vertical axis we have price per hamburger. On the horizontal axis we have hamburgers consumed per week.

If you look in the box you'll see the information that we were just looking at about Bob's behavior. In the box you see the number of hamburgers that Bob buys at different prices. I'm going to now transfer that information that's in the tables into a picture here, a representation of the relationship between price and quantity.

Let's start then with a price of $5.00. Bob buys no hamburgers. At a price of $4.50 Bob also buys no hamburgers. So we have points here that are actually on the axis. That is, zero hamburgers are purchased, so Bob is behaving in a way that gives us points that are on the axis. So here's a price of $4.50 per hamburger and no hamburgers consumed in a week. Now, when the price drops down to $4.00 per hamburger, Bob actually buys a hamburger. So now we have a point in our space that's right here. One hamburger at a price of $4.00. So that gives us a dot right here to represent Bob's behavior. The coordinates of this point are 1 and $4.00. If the price drops down to $3.50, so we'll be looking at a price that's right here on the vertical axis, Bob buys two hamburgers per week. So now we have a point that looks like this. At a price of $3.00 Bob buys three hamburgers per week. So now we have a point that looks like this. At a price of $2.50 Bob buys four hamburgers. So that lets us drop down to another point like this, $2.50, four hamburgers. In each case I'm finding a pair of numbers, like $2.00 and five hamburgers and I'm putting a dot at the horizontal, vertical coordinates. Six hamburgers at $1.50. So find the $1.50 point and six hamburgers. Put a dot there. And finally, we get seven hamburgers when the price drops down to $1.00. And that gives us a point like this.

So, these dots represent Bob's behavior. That is, the quantity of hamburgers that he consumes as the price of hamburgers changes. Now, I can connect these dots and form what we call in economics a demands curve. This demand curve will be a relationship between the price of hamburgers and the quantity that Bob consumes in a given amount of time.

Now we can connect the dots in this diagram and form the demand curve for Bob. That is we can draw a line that represents the relationship between the price of hamburgers and the number of burgers that Bob consumes in a week. Let me go ahead and do that. I'll draw a line here that connects all of these dots, and it looks like this, and this is Bob's demand curve. We can even put a label on it. Say label it with the letter D to let us know that this is Bob's demand for hamburgers, the relationship between the price of burgers and the quantity that's consumed.

Now, if I want to I can write down the formula for this demand curve. Now I’m going to go ahead and do that. Now I knew the formula before we started because I used the formula to form the tables. The formula for Bob’s demand
Introduction to Economic Thinking

Using Graphs

**Plotting a Linear Relationship Between Two Variables**

curve is this. The price, that is, the vertical axis variable, is equal to $4.50, which is the price at which Bob will buy no hamburgers minus 50 cents times the quantity of hamburgers that Bob buys. That is, for every additional hamburger that we want Bob to buy in week we have to lower the price of the hamburgers by 50 cents.

This is the relationship between quantity and price for Bob and that is the formula for this line. Now if that takes you by surprise, remember from your algebra class that if we have a variable that's measured on the Y axis and another one that's measured on the X axis, that the formula for any straight line is equal to this. The Y axis variable is equal to A plus B times X. Where Y is the variable that we're measuring here, in this case, price. X is the variable that we're measuring here, in this case, quantity. A is the intercept or the height here on the vertical axis when the X variable has a value of zero. In this case, Bob is buying zero hamburgers when the price is $4.50 per hamburger. A represents the intercept and in our case the intercept is $4.50. B represents the slope of the line, the rise over the run. And in this case the slope is -50 cents. For each additional hamburger that we want Bob to buy we have to lower the price by 50 cents. So B, in this case, is -50 cents.

Let's take a moment and remind ourselves about the slope of a line. The slope of a line is equal to the rise over the run. And the rise over the run is the same thing as saying the change in the Y variable, that is the rise—the upward movement of the graph. Delta Y, which in Economics stands for the change in Y, divided by the run, which is the horizontal movement or the change in X. So if we want to know the slope of a line, we look at the change in the horizontal or the change in the vertical variable divided by the change in the horizontal variable, the rise over the run.

Let's look at an example here. If we increase the price of hamburgers by 50 cents, say from $2.00 up to $2.50, Bob's behavior changes in the following way. Instead of buying five hamburgers now, he reduces his hamburger consumption to four. This tells us the slope of the line, the rise over the run. The rise in this case is the change in the Y variable, which in this case is an increase in the price of the hamburger by 50 cents. And the run is the change in the X variable, how does Bob's behavior change. And in this case he reduces his hamburger consumption from five to four. That is a change of -1.

So if we wanted to calculate the slope of this line we would divide the rise by the run. That is, in this particular case, the slope is going to be equal to 50 cents divided –1 hamburgers or -50 cents. Negative 50 cents is the slope of the line and that's what you see right here in the formula.

So, for straight line the important terms are the vertical intercept, which shows up right here at this constant term and the slope of the line, the rise over the run, which shows up as the coefficient on this variable. In this particular case it's the coefficient on the quantity of hamburgers, -50 cents. Every straight line can be characterized by these two pieces of information, the intercept and the slope.
Let's look at what happens to the relationship between price and quantity when we change Bob's income. We see already that Bob has a relationship between the price of hamburgers and the quantity that he consumes per week. We looked at these table last time and we imagined, although I didn't say so, that Bob's income was $500.00 per week. Now if Bob gets a bigger income he's going to change his behavior.

Let's suppose that Bob's income goes up to $600.00 a week. If it does, we might expect that Bob would spend more money on hamburgers than before. When we replace the old set of numbers with a new set, this is Bob's behavior when the income that he earns goes up to $600.00 per week. In this case we see that although he still buys no hamburgers at a price of $5.00 a piece, he'll now buy one when the price drops to $4.50. And then each time the price drops by 50 cents he'll add one more hamburger to his weekly consumption, so that by the time we get down to a dollar per hamburger Bob's buying eight hamburgers a week. That is, some days he is eating more than one.

Let's now represent this new set of data in the graph that we originally drew. I'll now back to my two-dimensional diagram with the quantity of hamburgers on the horizontal axis and the price at which hamburgers are selling on the vertical axis. Let's look at this new relationship between price and quantity, the numbers that are now in the table beside me. Let me take my pen and chart these numbers. We can see that now at a price of $5.00 Bob buys no hamburgers. So the intercept of our curve, that is, the lowest price at which Bob buys zero hamburgers has move—-the intercept has moved up to $5.00.

At a price of $4.50 Bob now actually buys a hamburger. So we have a point like this, $4.50 and one hamburger. At a price of $4.00 Bob is now buying two hamburgers. So we have a point like this, a combination of the price $4.00 and a quantity of two. At a price of $3.50 Bob is now buying three hamburgers. So I have the price $3.50 and the quantity of three, represented by this point, and so forth. Every time the price drops by 50 cents Bob buys one more hamburger than he bought before. As we keep moving down to lower prices, we get larger quantities. We still have the same negative relationship between price and quantity. That is, we have the negative slope, the same negative relationship, but the numbers are now different.

Here down at a price of a dollar Bob is buying eight hamburgers per week. I can connect the dots now and form Bob's new demand curve, and let me do that. With this new information, Bob's new demand curve is going to look like this. I'm going to use my straight edge here to draw things carefully. There's no point in being precise—being imprecise, if I have all the information that I need to be precise—and here's Bob new demand curve.

The new demand curve, which I might label D prime, and the prime or the apostrophe just means a new relationship to compare to the old relationship. The new relationship between price and quantity has the same slope as the old curve. That is, notice the curves are parallel. For every 50 cent increase in the price, Bob still reduces the quantity of hamburgers that he buys by one. So the rise over the run is still -50 cents. We've got the same slope. And sometimes we'll represent that by putting two marks on these curves and the two marks mean that they have the same slope. They are parallel.

The formula for this new curve is going to be given as follows. The price on a new curve is equal to the intercept, which is $5.00 minus the slope of 50 cents times the quantity of hamburgers that's consumed. The formula for the new line is going to be price is equal to $5.00, that's our intercept, minus 50 cents, the slope, times the quantity of hamburgers consumed. This is the relationship if Bob's income increases to $600.00. We might think of this as the curve shifting outwards. If we wanted to represent our original demand curve like this, then with the increase in income we get a new set of points that moves the demand curve outwards. You can think of this as a shift in the curve.

Let's look at another set of numbers. Now suppose that Bob's income drops from $600.00 a week back to a lower income of $400.00 a week. In that case, what's going to happen to Bob's consumption? Well here's a table of numbers that represents the quantity of hamburgers that Bob consumes if his income is $400.00. And let's look at these numbers together. What we see is at an income of $400.00 Bob is not going to buy a hamburger until the price drops to $3.50. That's the price at which he is first willing to purchase burgers. Then as the price falls by 50 cents, he'll buy another one, and so forth and so forth. And when the price gets down to $1.00 per hamburger Bob will buy six hamburgers per week.
Changing the Intercept of a Linear Function

So in this case we can see that at a lower income Bob will buy fewer hamburgers at every price. Let's represent this set of numbers, which I'll move over to the box, on the graph. Going back to our diagram with the price on the vertical axis and the quantity on the horizontal. If the number of burgers that Bob buys a week falls because of the decline in his income, we get a new set of points. Notice now, that Bob will not buy any hamburgers when the price is $4.00 a week. So we can put a dot here, at $4.00 and zero, to represent Bob's behavior with a lower income. This is going to be the new intercept. The lowest price at which Bob buys no hamburgers.

If the price drops down to $3.50 Bob will buy his first hamburger, so we get a dot like this. If the price drops down to $3.00 a hamburger Bob will buy two hamburgers. We get a dot like this and so forth. Just go through and fill on the numbers from the chart, $2.50 gets us three hamburgers. A price of $2.00 gets us four hamburgers. A price of $1.50 gets us five hamburgers. A price of $1.00 gets us six hamburgers per week, and so forth.

Now if I want to connect these numbers, I can form the new demand curve for Bob at the lower income. So let me connect these dots. And when I do I get a new demand curve. And this one we could represent by D double prime. Again the prime is just to remind us we have a new relationship or a new set of points. And again, it's going to have the same slope. All three of these lines are parallel. One of them represents Bob's behavior when income is high, one represents Bob's behavior when income is in the middle, $500.00, and finally the third represents Bob's behavior when income is low.

The slope in all three cases, that is, the rise over the run, is going to be the same. But the intercept is different. And that's what gives us these parallel shifts. This third relationship can be summarized by the equation the price is equal to $4.00, that is, the vertical intercept minus the slope of 50 cents times the quantity of hamburgers that Bob consumes each week. In this case we have a lower intercept. The curve is closer to the origin, but has the same slope as before. So, notice the difference between the three curves is their intercepts. The first curve is the one that's in the middle with an intercept of $4.50. The second curve is this one right here with the higher intercept and the third curve is the one with the lower intercept. Each of these three is a relationship between price and quantity for Bob. The difference is in the income and when the income changes we get a whole new relationship between price and quantity.
We've been looking at the relationship between two variables, price and quantity, for Bob who's making a decision about how many hamburgers to buy each week as the price changes. And we looked at how Bob's behavior changed when his income changed. I want to look at something else now, and that is the sensitivity of Bob's demand for hamburgers to changes in price. What happens to the number of hamburgers that Bob buys each week when the price changes? And I bring this up because I want to discuss how economists think about the slope of a line.

Let's look at the line that we were analyzing before, Bob's original demand curve for hamburgers. We saw that when the price of hamburgers fell by 50 cents, that is from $2.00 down to $1.50 per hamburger, that Bob's quantity demanded increased from five hamburgers per week up to six hamburgers per week, and this was represented by this slope of the line. Let's show how we calculate the slope. The slope, remember, is the rise over the run. The change in the vertical axis variable or the Y axis variable, in this case the change in price, divided by the change in quantity demanded, the change in the X axis variable. In this case the X axis variable was the quantity of hamburgers that Bob purchased each week as the price changed.

Now, let me plug in my numbers here. The change in price, in this case, is a movement from $2.00 per hamburger down to $1.50 per hamburger. So let's plug that in. The new price is $1.50 and the old price was $2.00, the price is actually falling. And the change in quantity is given here by the movement of Bob buying five hamburgers a week to Bob buying six hamburgers a week. So the new number of hamburgers is six minus the old number is five, and that gives us the change in quantity. Let's actually solve this equation and we get -50 cents divided by one hamburger, and the slope therefore is -50 cents.

So, Bob is relatively sensitive. That is when the price falls by 50 cents, he'll buy an extra hamburger each week. Let's suppose now we have a different relationship between price and quantity for Bob. And let's suppose that in this different relationship, when the price falls from $2.00 to $1.50 per hamburger Bob increases his quantity not just to six hamburgers per week, but all the way up to ten hamburgers per week. Now, this is an entirely different relationship for Bob. This is an entirely different relationship between price and quantity. So I would have to redraw the demand curve to represent the new combinations of price and quantity, and I'll do that using green. I'll show this new relationship with a green curve that has on it the two points that I've been talking about, $2.00 and five hamburgers and $1.50 and ten hamburgers.

In this new relationship Bob’s demand curve is flatter. That is the slope is smaller than the slope on the red curve. The curve is flatter. Let's actually calculate the slope of the green demand curve. If this were Bob's demand curve, and I've labeled this D prime to represent a new relationship between price and quantity for Bob. If D prime, the green curve, were Bob’s demand curve, what would the slope be? Well, I'll write the slope out here.

Now to use prime to remind us that we're talking about this new curve, the green curve. The slope on this curve is the rise over the run, and in this case we're dealing with the same rise, the same 50 cent change as before, only now we have a much bigger run. Bob increases his demand for hamburgers not just by one hamburger, but by five. So in this case my change in quantity was much bigger and I can call this delta X prime. That is a bigger change in the X axis variable, so the slope is going to be the change in price divided by this new change in quantity. The change is price is still going to be 50 cents, but the change in quantity is now much bigger.

Let's plug in the numbers and calculate. The change in price, as before, is -50 cents. The movement from $2.00 to $1.50. The change in quantity now is—the new quantity is ten minus the old quantity of five, gives us the change in quantity equal to five hamburgers, so the new slope is -10 cents. That is, when the price changes by 10 cents on average Bob adds an extra hamburger to his weekly consumption. Bob’s demand is more sensitive on the green line than it is on the red line, and that's why economists watch the slope of curves.

It is the slope of curves that tells us something about the sensitivity of one variable to changes in another. Now, One word of warning here. The slopes of curves depend entirely on how you are measuring the variables. Here we're measuring the price of hamburgers in dollars. But if we were measuring the price of hamburgers in cents instead, the
Introduction to Economic Thinking
Using Graphs

Understanding the Slope of a Linear Function

slope might change. That is if this were instead of $1.00, one penny, then Bob wouldn’t be buying any hamburgers until we were way, way up this axis. So if we change the way we measure the price, the slope is going to change.

And if we change the hamburger measure from burgers to boxes of burgers or bags of burgers or half hamburgers or parts of hamburgers, you can change the slope in all kinds of ways. The slope of a curve depends completely on the units in which you measure the variables. And that’s why economists like measures that are called elasticities. Elasticities don’t depend on the units in which you measure variables. Elasticities are based on percentage changes. And whether the price of a hamburger goes from $1.00 to $2.00, a doubling of the price, or whether it goes from 100 cents to 200 cents, it’s all the same. Economists are interested in measures that don’t depend on units, that depend instead on percentage changes. And when we get into our discussion of elasticity, in a later lecture, you will see how economists put this measure to work.
We've been looking at linear relationships between variables and how economists think about these. So far everything that we've looked at has had a constant slope. That is, we've been analyzing straight-line relationships, such as the relationship between price and quantity in Bob's demand for hamburgers. Now we're going to look at a relationship that is not linear. In fact, it's what we call curve linear. It is a line that's not straight.

We're going to see how we draw a relationship like that and how we calculate the slope of a relationship where the slope is not constant. In order to do that let's look at the relationship between labor and output for a firm that makes television sets. And I've sort of made up these numbers so that I can fit a curve linear line to that. Here is my relationship. If you have one worker working by himself, he can make a single television in a day. If you have two workers cooperating they can make four televisions in a day. If you have three workers, nine. Four workers, 16. Five workers can make 25 televisions, and so forth. Now notice this relationship is one of exponential increase. That is, the number of TVs are increasing much faster than the number of workers. So in this case our slope will not be constant.

Let's represent this relationship now in a picture. I'm going to take my axes here and use the axes to represent the amount of labor that the firm uses. So we're going to measure labor on the horizontal axis. And on the vertical axis we're going to measure output. In this case it's going to be output of television sets. So, each point in this space will represent a combination of labor and output. This is a production relationship.

So, for example, one worker is able to produce a single television if he's working by himself. So since I've calibrated these axes where each of my lines represents two TVs, my first dot will be half way up towards this point at one worker and one television set. So here's my first point on my product curve. The second point on my product curve will be two workers and four television sets, so that's going to be at this point right here. Three workers are able to produce nine television sets. So that gives me a point like this. Four workers are able to produce 16 television sets, so follow this line up here to sixteen. And five workers can produce 25 television sets, so follow this line all the way up to twenty-five, which will be right here.

Now connecting these dots is going to be a little bit more complicated then connecting the dots in our earlier story. In the earlier story all we had to do was pick any two points, draw a straight line between them and we had a demand curve. But the production curve is not linear. That is, the slope is not constant. So I have to draw a curve that connects these points together. That curve is going to look something like this. Let me take the curve here and figure out how to make it fit. I want it to go through this point and this point and this point and this point and this point. And so my curve is going to look something like this. It's going to be an exponential relationship.

So if I want to connect my dots, I'm going to have them lying along the blue line that looks like this. Let me go ahead and draw that blue line in. I could have kept that curve there and just drawn along it. I'm going to try—and it takes practice to do this well, to connect my blue dots along a single curved line. Like this. Well, that's not bad. So I'll go ahead and make this line a little bit darker. And there's my production function for a firm that hires workers to produce television sets. And we might call this curve the Total Product Curve. We'll be studying this later in these lectures.

So I'll go ahead and label the TP, Total Product. And this Total Product Curve depends on lots of things. Like the technology of the firm and the size of their factory and other things. What I'm mainly interested in, at this point though, is figuring out what the slope of this curve is at a particular point. Say for instance, we are looking at this point right here. The point where we're hiring two workers and making four television sets in a day. And I'd like to know what is the slope of the curve at this point, when I'm hiring two workers and making four television sets a day. What's the slope of the curve? That is, what's the rise over the run? How many extra television sets will I be able to produce if I hire an extra worker?

Well, to calculate the slope of the straight line is not quite as simple as looking at two points on the curve, because as you move from one point to another, what's happening is that the line itself is changing it's slope. Down here our blue curve is relatively flat. That is, if we look at a point down here we get a relatively flat line. But as we move up the curve, notice, the line is getting steeper and steeper and steeper. The slope is changing all along this blue curve. So, if we want to calculate the slope at a particular point, we can't just look at the slope between two points on the curve, because as you move from here to here, even that small movement represents a change in the slope or causes a change in the slope of my Total Product Curve. The blue curve slope is changing.
Introduction to Economic Thinking

Advanced Graphical Concepts

Understanding Tangent Lines

Well here's how we do this. We calculate the slope of a single point on this curve. We calculate the slope at a single point by taking the limit of slopes of lines that connect two points on the curve. So for instance, suppose we look at what happens as we move from two workers to four workers. As we move from two workers to four workers what happens is this. We're moving along a line that connects those two points, and the line that connects those two points is this line right here. This red line connects those two points. Now what's the slope of that red line that connects this dot with this dot? What's the slope of that line?

Well let's calculate it. The slope of my red line here is equal to the rise over the run, with the change in the vertical axis variable, that is, a change in the number of television sets divided by the change in the number of workers. Well the change in the number of television sets is going to be 16 minus four, which is going to be equal to 12. And divide that by the change in the number of workers, which is going to be four minus two, and that's going to give me a slope of 12 over two equals six. The slope of this red line, right here, that is its rise over its run, is going to be six. The slope of that line is six.

However, if we took a different red line. Say the red line that connects this dot to this one, its nearest neighbor, we're going to get a different slope. Let me go ahead and draw that. If I connect this dot with the nearest neighbor my red line is going to be a bit flatter. Now it's a little harder to see that line because it's almost right on top of the blue line. But what we've got here is a line that connects this dot with its nearest neighbor. I'd like to know what's its slope. What's the rise over the run as I move from this point to the next?

So let's look then at the slope as we move from two workers to three workers. In that case, the slope once again is rise over the run, but the slope along this point is going to be a little bit different. The slope in that case is going to be equal to—the output has changed to nine from four divided by the change in workers from two to three. Gives us a slope of nine minus four is five over one, which is going to be equal to five. So the slope of this point is five. Notice what's happening. As we move from this original red line that connected these two points to the red line that connects two closer points, the slope is getting flatter. It's gone from rise over run of six to rise over run of five.

Eventually, what happens is, if we let the point that we're looking at get closer and closer to our original point, if we let the point that we're looking at get closer and closer to our original point, we're eventually going to come up with a line that has a slope that is equal to the slope of the blue curve itself. Eventually, what we get is what we call a tangent line. The tangent line is the line that just touches the blue curve. If you have something straight like this ruler and you wanted to lay it against the blue curve, laying the ruler against the blue curve would give you the tangent line. Laying something straight against a curved line at a particular point gives you the tangent. Tangent means touching.

If you wanted to take the blue curve, like before, and lay it against the blue curve and have this blue curve be the straight line. The slope of this blue curve now is the slope of the tangent line, the line that just touches. So to calculate the slope of a curve at a particular point, you take the limit of the slope of the lines that connects that point to other points and that limit is going to be a tangent line, the slope that's tangent.

So, when we're looking for the slope of the curved line, we take the slopes of lines that connect points on the curve, let those slopes change as the point gets closer and closer to the point we're interested in, and finally we come up with a line that's tangent to the curve at the point we care about. In this particular case, from Calculus, I know that the tangent line that I've drawn here with the dash has a slope of four. I'm not going to do the Calculus right now, but what I'm interested in your understanding is that when you have a curved line, the slope of that line changes as you move along it. And the important thing for you to know, that the slope of that curve at a particular point is the slope of a straight line that touches the curve at that point.
We've seen how economists use graphs to represent relationships between two variables. Now we'll see how an economist can represent a relationship among three variables in a two-dimensional graph. The trick is you only change two variables at a time.

Suppose we're interested in a relationship among three variables, and a good example would be making a map of a terrain. In this case, you've got three variables that change as you move across the terrain. One is your east/west coordinate, called your longitude. The other is your north/south coordinate, called your latitude. Finally, there is the distance that the terrain lies above sea level or below sea level – the altitude or the height of the point. As you move through a terrain, all three variables are changing – east/west, north/south, and height above sea level.

So when we draw a map, how do we represent these three variables? Let's look at a set of data points, and then we'll show how we represent them in a graph. Suppose we have a set of combinations of latitude and longitude for which the altitude remains constant. That is, let's look at points that have a constant altitude of 1,000 feet above sea level. If we identify a set of these points, we'll find that the latitude and longitude that give us an altitude of 1,000 can be written down in a table. Suppose that at a latitude of 200 feet north and 600 feet east we have an altitude of 1,000 feet. Suppose that at another point, a latitude of 100 feet north and 100 feet east, we have an altitude also of 1,000, and finally, a third combination. Notice the trick here is I'm holding my altitude constant as I change latitude and longitude. All of these combinations of latitude and longitude have a constant altitude of 1,000.

So let's move these numbers over to the box at the side, and then show how to represent them in a graph. On the vertical axis we will measure the distance north or south – or the latitude. On the horizontal axis, we'll measure the distance east and west – or the longitude. Altitude, we'll represent in a special way in the graph, and I'll show you in just a moment.

So let's look again at those points that we started with. Suppose that we are considering the point 60 east/west, 200 north/south, and an altitude of 1,000 that will lie in this picture in this way. First, put the longitude on the horizontal axis. The longitude of 60 means that we are at a point like this one. The latitude of 200 means that we are up here at 200, so I'll put a dot here to represent that combination of latitude and longitude. With 100 and 100, we also get the same altitude. And finally with 280, we get the same altitude. Each of these three dots represents a latitude and longitude combination with a constant altitude of 1,000. So what I can do to make it easier is I can connect all of the points that have an altitude of 1,000 and label this curve “1000.” This collection of points is sometimes called an isoquant, from the word that means “same quantities.” That is, each of the points on this curve represents a combination of latitude and longitude with a constant altitude of 1,000.

Now, this is not the only isoquant that we can draw. We can now change the altitude and find another set of combinations of latitude and longitude that have a different constant altitude. Let’s consider another set of information here. Suppose we look at this table of numbers. Here are combinations of altitude and longitude that have a constant altitude of 2,000. Two hundred north and 200 east gives us an altitude of 2,000, the same with 140 and 300, and so forth. Each set of combinations here of latitude and longitude gives us a constant altitude of 2,000.

So now let’s represent this set of points with a different isoquant in the same graph. Let’s start with the combination of 200 and 200. With a longitude of 200 and a latitude of 200, we're going to get this point right here, and that has an altitude of 2,000. We also get an altitude of 2,000 with a longitude of 140 and a latitude of 300. That would be this point right here. Three hundred and 100 gives us another combination with an altitude of 2,000. Finally, one last point, here, at a latitude of 300 and a longitude of 380, we also have an altitude of 2,000.

Now, we can connect all of these dots to get another isoquant – that is, another collection of points where the altitude is constant, this time at an altitude of 2,000. However, notice this dot lies way up here; it doesn’t appear to be on the same curve as these three dots. In fact, if I had more points what I would see is that this isoquant is actually an oval shape that comes around and connects to itself. Unlike this isoquant, which doesn’t appear to bend back on itself, here’s a complete set of points where all of these dots form a closed oval with an altitude of 2,000.

If I had other altitudes, I could form other isoquants; that is, other points with constant altitude. For example, this oval might represent the set of points that have an altitude of 3,000. Finally, here in the middle there might be a point at the top of our hill that has an altitude of 3,500. Here, then, is our isoquant map. What I’ve done is I’ve collected points together that have constant altitude. You can imagine here a topographical map where the altitude is rising as we
move up to the northeast; and then after you go over the top of the hill, you begin to lose altitude. In fact, what you’re seeing here are cross-sections of a hillside. Imagine a plane coming through and lopping off a hillside at a certain altitude. Here’s what it would look like if it were lopped off at 1,000, here’s what it would look like if it were lopped off at 2,000, here’s the shape of all of those points with an altitude of 3,000, and so forth.

This is the way an economist represents three dimensions in a two-dimensional graph. We use the two axes to represent two of our dimensions – in this case east/west and north/south – and we use the third dimension in the graph, and we represent the third dimension in the graph as the numbers on the isoquant lines. Later, we’ll be using this tool to represent consumer preferences as well as the trade-off between factors of production in the making of goods and services.
Introduction to Economic Thinking

Production Possibilities

Understanding the Concept of Production Possibilities Frontiers

We've defined economics as the study of rational choice in situations of scarcity. In this lecture, we're going to take a closer look at the concept of scarcity and how economists describe scarcity. Scarcity means that you have a limited quantity of resources with which to satisfy potentially unlimited wants. But let's be more careful about that. What are those resources and what can we do with them? How do we use our resources to create the goods and services that satisfy our desires? In order to represent the possibilities of an economy, we're going to introduce a tool, and that tool is the production possibilities curve.

Now, we're going to be working towards that, but first I want to introduce a concept that will be very important in this lecture, and really, throughout the course, and that's the concept of efficiency. Efficiency refers to doing the best you can with what you have, or making the largest quantity of output possible with your given quantity of inputs. Efficiency refers to getting the most you can out of what you have. So for instance, if we have a cook who is talented with lots of good recipes, and he has a lot of food in the kitchen to work with, and a good set of knives and other tools and a stove. Efficient behavior would be using his talent and those resources to produce the most meal that he's capable of producing, getting the largest quantity of output from a given quantity of input. If he wastes some of the food, if he uses his time wastefully so that the quantity of output that goes out the door of his kitchen is less than is possible, then we say that his behavior is inefficient.

Suppose we have an economy that produces two goods, and for the sake of our example, we'll make this an agricultural example. Our two goods are wheat and rice. So let's suppose we have an economy that produces these two products. Efficient behavior would be for the economy to produce the most wheat that it can possibly produce for a given quantity of rice. That is, using all of the resources that are left over that aren't producing rice, and making the largest quantity of wheat possible. That's what we mean when we refer to efficiency—getting the most you can out of your limited quantity of resources.

Now, I want to describe the production possibilities of a particular economy. This hypothetical economy produces only two goods—wheat and rice. Before I describe the production possibilities, you might want a little additional information. The first thing you're going to want to know is, well, what resources are available in this economy with which to produce these two grains? How much arable land is in this economy? Is the land wet or is it dry? What kind of labor is there in this economy? What kind of tools? Are there tractors and combines or are there merely shovels and hoes?

The first thing you need to know when you want to describe the production possibilities of an economy is—what is the resource endowment of that economy? How much land, labor, and capital does that economy have to work with? The second thing that you're going to want to know is—what are the capabilities of this economy, given these resources? That is, what are the strategies this economy has for combining input to produce output? And we call that the economy's technology. When an economist refers to technology, the economist means a complete catalog of all the things this economy knows how to do, all the ways in which this economy can combine inputs to produce output. There are techniques that involve using shovels and hoes and low-skilled labor on particular kinds of land to produce rice or wheat. There are techniques that involve using tractors and combines and a different kind of land to produce rice and wheat. A technique is a particular combination of inputs, but a technology refers to all the possible combinations of resources that the economy can use to produce output. So when an economist refers to technology, the economist is referring to a catalog, a complete list of all the things the economy knows how to do. When your technology improves, it usually means that you're able to produce more output with less input, because you're using that input in a smarter way.

Now, given our knowledge of the economy's resources and what we know about the economy's technology, we are able to describe the production possibilities of the economy, and I'm going to do this first in this table. This table represents the production possibilities of our hypothetical economy. And every combination of numbers in this chart represents an efficient combination. That is, we're saying that if our economy uses its resources, all of them, in the production of wheat, it can produce 80 bushels of wheat a year. Well, since all the resources are devoted to wheat production, there isn't going to be any rice production in this case. If the economy moves some of its resources into the production of rice, and let's suppose that it moves those resources that are best suited for rice production, the wet land, the workers who know more about rice production, tools that are well suited for the production of rice, the economy can produce 20 bushels of rice and still produce 78 bushels of wheat. Notice that the wheat production is decreasing as resources are moved into the production of rice. But since we're moving first those resources that are
well suited for the production of rice, we can get a large increase in rice production for a small decrease in wheat production.

Next, we can get an additional 20 bushels of rice per year if we reduce wheat production by 8 more bushels of wheat a year. The combination 70 and 40 is another efficient combination. That is, the maximum amount of wheat that we can get out of our economy when we’re producing 40 bushels of rice, is 70 bushels of wheat if we are using all of the resources in the best uses possible and everyone is working as productively as they can. If we produce rice production to 60 bushels a year, wheat production falls to 55 bushels. And if we increase rice production to 80 bushels a year, then wheat production falls to 38. Finally, if we put all of the resources in our economy into the production of rice, we get 100 bushels of rice per year, and of course, no wheat, because all of the resources are devoted to rice production. So the production possibilities table represents those efficient combinations of rice and wheat production that are possible given the economy’s technology and given its resource endowment. That is, for each number in the right hand column, for each quantity of rice produced a year, the number in the left-hand column tells us the maximum quantity of wheat that this economy can produce using its resources with efficiency.

Now, I’ve got a production possibilities table here. I’ve got the combinations of wheat and rice that are possible to be produced in this economy. I’m going to now represent this information graphically. I’m going to put the production possibilities information into a diagram, and we’re going to call this diagram the production possibilities curve, or the production possibilities frontier. The first thing I need to do when I draw a diagram is make sure that I’ve carefully labeled the axes. So the first thing we’ll do is we’ll put wheat production on the vertical axis. That is, the number on the vertical axis refers to the quantity of wheat, in bushels, that’s produced in this economy in a year. The next thing will be to label the horizontal axis, and we’ll put rice production here. The number on the horizontal axis tells us the quantity of rice that’s produced in this economy in a year. Now the next thing I want to do, since I’m going to be using real numbers and I want to be careful about plotting this curve accurately, I want to calibrate the axes. I want to put some numbers in that tell me how things are measured on each axis. Since I’m going up to 80 bushels of wheat, let’s make each hash mark on this axis equal to 20. So here’s 20 bushels of wheat—here’s 40, here’s 60, and here’s 80. And I can keep on going, but I’m calibrating the axis here with each hash mark representing 20 additional bushels of wheat. The same thing with rice here. I’ll make this be 20, and this 40, and this 60, and here’s 80, and finally 100, and of course, it keeps going.

Now I’ve got my axes calibrated and labeled carefully, so I’m ready to draw my graph. And let’s now take the numbers, which are now over on the board, and put them as points on a curve in this diagram. So the first combination of wheat and rice that’s possible is all the resources devoted to wheat production. We produce 80 bushels of wheat and zero rice that gives us a point that’s right here on the vertical axis. The next combination is 20 bushels of rice and 78 bushels of wheat. So we go over here to 20 bushels of rice and up to 78 bushels of wheat and I put another point on the curve. If we produce 40 bushels of rice, then the maximum wheat production I can get out of my economy is 70, so there’s another point on the curve. With 60 bushels of rice I get 55 bushels of wheat. With 80 bushels of rice I get 38 bushels of wheat. And finally, with 100 bushels of rice, my wheat production has been pushed down to zero because all of the resources are devoted to the production of rice.

So what I did was take the numbers from the production possibilities schedule and put them in a diagram with wheat on the vertical axis, and rice on the horizontal axis. Now, these are not the only combinations of wheat and rice that we can imagine. In fact, you can imagine our economy moving resources continually or continuously out of the production of wheat into the production of rice, so that the quantity of rice produced increases gradually as the quantity of wheat produced falls gradually. If we plotted all the possible points, they would lie between these blue points, and if I connected the dots, I would get a nice, smooth, continuous curve, and this curve is called the production possibilities curve, or the production possibilities frontier.

The production possibilities frontier is defined thusly: the production possibilities frontier is a collection of points representing the maximum amount of wheat that can be produced for a given quantity of rice production. These points are all of the efficient combinations of wheat and rice that can be produced in the economy. For any point on this horizontal axis, you can go up to the blue curve and find the maximum amount of wheat that can be produced when the economy is producing this quantity of rice. That’s the production possibilities curve.

As we look at the production possibilities curve we can see that there are many things represented in this picture that are very basic concepts in economics. The first thing that we notice whenever we look at this picture is that there are
Understanding the Concept of Production Possibilities Frontiers

points that are outside the production possibilities frontier. There are points such as this one, where we have 80 bushels of wheat and 80 bushels of rice produced in a given year. This point would be a great point because it involves a lot of everything. However, this point is not attainable. This point reminds us that scarcity is a reality in this economy. Given the quantity of land, labor, and capital, given our resource endowment, and given the technology that's available to us in this economy, this point out here—we'll call this Point F, because it represents scarcity—this point is unattainable. We simply can't get to it. There's no way to make this much wheat whenever we're making this much rice. So points that lie outside the production possibilities frontier represent a larger quantity of goods than this economy can possibly produce. They remind us of the reality of scarcity.

Likewise, points that lie within the production possibilities frontier remind us of inefficiency. I'll call this point down here Point U. Point U is certainly possible. It's always possible for us to produce points that are within the production possibilities frontier. But why would we want to? We would prefer to be up on the production possibilities frontier with more of everything. If more is better, we would always prefer a point on the blue curve to a point inside it. But these are the points that we wind up with when the economy uses its resources inefficiently. If the economy is not getting the maximum possible productive output from its resources, we're going to be inside this curve. That happens when some resources are unemployed—that is, if some of the land or labor is not used. It also happens if some of the resources are underemployed—that is, if resources are used in a way that is not maximizing their output. For instance, if we take wetland and use it to cultivate wheat, and dry land and use it to cultivate rice, we're going to wind up with less grain than we can produce. Because wetland is better for rice and dry land is better for wheat, we're going to wind up inside the production possibilities frontier at a point of inefficiency—unemployment or underemployment of resources.

Here's another thing to notice. The production possibilities curve is downward sloping. That means that as we increase production of one, we are necessarily reducing the production of wheat. In order to produce additional bushels of rice, you have to give up wheat that you might otherwise have produced. That is, wheat is the opportunity cost of rice. Any time you want to increase rice production, you're going to have to move resources out of the production of wheat. Land, labor and capital that's used to produce wheat is transferred into the production of rice in order to increase rice production, and that means that wheat production will be decreasing. Wheat is the opportunity cost of rice.

We can actually calculate this opportunity cost. Look, whenever we increase rice production from zero to 20 bushels a year, wheat production is falling from 80 bushels to 78 bushels. The opportunity cost of the first 20 bushels of rice is 2 bushels of wheat. We can even calculate the opportunity cost per bushel of rice by doing a little calculation. First let's calculate the change in wheat production. The change in wheat production is the new quantity of 78 minus the old quantity of 80 for a change in wheat production of -2, a loss of 2 bushels. The rice production is 20 minus the old quantity of zero, so the change in rice production is 20 bushels—20 minus zero equals 20. Now, divide the change in wheat production by the change in rice production, and you'll get -2 over 20 or -1/10. That is, for each additional bushel of rice that we are producing between zero and 20, on average we're giving up 1/10th of a bushel of wheat. 1/10 a bushel wheat is the opportunity cost of each bushel of rice, on average, between zero and 20.

Notice something else—that if we draw a line connecting these two points, the original point of 80 and zero, and the new point of 78 bushels of wheat and 20 bushels of rice—if we draw a line connecting those two points, the slope of that line is going to be the change in wheat production divided by the change in rice production. Now remember, what's the formula for the slope of a line? The rise over the run. As we move between these two points, the rise is the change in wheat production. The rise is the change in the vertical coordinate. That is, we go from 80 to 78, -2, that's the rise. The run is the change in the horizontal coordinate. We go from zero to 20. So that change of 20 is the run. -2 divided by 20 is the rise over the run. The change in wheat production divided by the change of rice production. And in this case, that number is -1/10. So I'll go ahead and put a little mark here indicating that I'm calculating the slope between those two points and I'll make a note that it's -1/10. The slope of the production possibilities frontier represents the opportunity cost of rice production measured in terms of the wheat that we're giving up. As we move between these two points, each additional bushel of rice that we produce is costing us 1/10 of a bushel of wheat, on average. The slope of the production possibility frontier is the opportunity cost of rice production measured in terms of wheat given up.

We can calculate the slopes down here as we move from 80 to 100 bushels of rice and we'll find we get a different number. The change in wheat production is the movement from 38 bushels of wheat to zero bushels of wheat. Take
Introduction to Economic Thinking

Production Possibilities

Understanding the Concept of Production Possibilities Frontiers

the new quantity of wheat, which is zero, subtract the old quantity, which is 38, and you get -38, a loss of 38 bushels of wheat. The change in rice production is 100, the new quantity, minus 80, the old quantity. That gives us 20 bushels of rice added. -38 divided by 20 reduces to -19 over 10, or -1.9. If I connect these two dots, the slope of this dotted black line is the rise over the run, the change in wheat production divided by the change in rice production, and this slope in this case is -19 over 10, or -1.9. That is, for each additional bushel of rice that we're adding between 8 and 100, on average, we're giving up almost 2 bushels of wheat for each additional bushel of rice. The opportunity cost is 1.9 bushels of wheat lost for every bushel of rice added. Notice the opportunity cost down here is much larger. Each additional bushel of rice costs almost 2 bushels of wheat. Up here, each additional bushel of rice costs only 1/10 of a bushel of wheat. The opportunity cost is changing as we move along the production possibility frontier. And you can tell by looking at it. The slope of the curve is pretty flat up here near the vertical axis. As we move down the curve, the slope gets steeper and steeper and steeper and steeper as we approach the horizontal axis. Remember, the slope of the production possibility frontier is the opportunity cost of producing rice measured in terms of wheat that you give up. So down here, the opportunity cost is very high, whereas up here, the opportunity cost is relatively low.

What does that tell us? Well, this is the final thing to notice about the production possibility frontier, is that the production possibility frontier is concave. That is, it bows outwards. The slope of the production possibility frontier is pretty flat near the top and pretty steep near the bottom, and the slope is decreasing. That is, the slope is becoming a larger and larger negative number. The curve is getting steeper as we move down it. What does that tell us? It tells us that the opportunity cost of producing rice is increasing. The more rice we produce, the more expensive it gets to make another bushel. Why? That's because not all resources are equally well suited for the production of wheat and rice. Some resources are very well suited to the production of rice. These resources include wetland, workers who know a lot about rice cultivation, particular types of tools that are well suited for tending rice. Some resources are better suited for the production of wheat. Those include dry land. They also include workers who know a lot about wheat production, fertilizers that are good for wheat, and so forth. When an economy has a scarce, strictly limited quantity of resources, some of them will be very well suited to the production of one good, others will be better suited to the production of the other good.

In this case, when we first begin to produce rice, we’re going to be using the wettest land and the workers that know the most about rice production. That wetland probably wasn’t growing a lot of wheat. Therefore, when we first start to produce rice, we can do so with a small opportunity cost. Very little wheat is given up, because that land wasn’t very productive anyway for wheat. It was better suited for rice. But as we make more and more and more rice, we’re using up all the land that’s well suited for rice, and then eventually we have to draw on the land better that’s suited for wheat—dry land, and works who know more about wheat cultivation than they know about rice. That’s when the opportunity costs start to get very, very large. Down here, we have to move a lot of dry land out of the production of wheat just to get an additional 20 bushels of rice, and that’s very costly in terms of the wheat that we lose. That’s why the opportunity cost of rice production increases the more rice the economy chooses to produce.

So, there you have it—four things about economics that you can tell by looking at a production possibilities frontier. First of all, you see that scarcity is a reality. Some combinations of wheat and rice are simply unattainable given our technology and our resources. Next, notice that there are a lot of inefficient points. There are a lot of combinations of wheat and rice that aren’t taking full advantage of the resources and technology that we have. Third, notice that the curve is downward sloping. That is, there’s an opportunity cost. If you want to produce more rice, you have to give up wheat production. Once again, that’s because of scarcity. And finally, the curve is concave. That is, the slope is getting steeper as we move down it. The curve gets steeper because the opportunity costs of rice production increases the more rice we make. This increasing opportunity cost is due to the fact that some resources are well suited to rice production, and other resources are not.

So this simple tool captures in it most of the basic ideas of economics—scarcity, efficiency, opportunity cost, the suitability of resources. In the next lesson we’ll look at how changes in resources, and changes in technology affect the way this picture looks.
Understanding How a Change in Technology or Resources Affects the PPF

In the last lesson we introduced the production possibilities frontier, a tool that describes what an economy can produce given its resources and its technology. In this lesson we’re going to take a dynamic look at production possibilities. We’ll ask what happens to the productive capacity of an economy when there’s some kind of change in its environment. Perhaps its resources increase or decrease. Perhaps technology changes in some particular way. What then happens to the combination of goods and services that that economy can produce?

Let’s take another look at the production possibilities frontier. Remember, it’s a downward sloping line representing the fact that if you want to produce more rice, you’ve got to produce less wheat. The movement of resources from the production of one good to the other is necessary and creates a kind of opportunity cost, that is, the opportunity cost of increasing rice production is reduced wheat production. We draw this line when we know the economy’s resource endowment and its technological possibility. That is, what the economy is capable of doing with the resources that it has. Suppose now we change one of those constraints. Suppose we change the technological possibilities in the economy, or suppose we change the resource endowments. In that case, the production possibilities curve will shift. A movement along the production possibilities frontier represents a change in the economy’s choice of which combination of goods to produce. If an economy decides that it wants to produce a lot of wheat and a little rice, it will be at a point on the production possibilities frontier up here. If the economy wants to produce a lot of rice and a little wheat, they’ll be down at a point like this. If the economy wants a balanced output of wheat and rice, they’ll be at a point like this. Whenever the economy changes the combination of wheat and rice that it produces, it moves along the existing production possibilities curve.

Suppose now that we change one of those things that we hold constant when we draw the curve. Suppose we change resources or technology. Anytime you change something that was held constant when you drew the curve, you have to redraw the curve. Anytime you change a variable that’s not represented on one of the axes, well, you’ve got to redraw the curve to represent new production possibilities. Let’s start and suppose that we have general technological progress in the economy. Suppose there’s an agricultural revolution and this economy discovers new and more productive ways to make both rice and wheat with their given resources. In that case, the production possibilities curve will shift outward. The curve shifts outward, representing the fact that after technical progress, it’s now possible to produce more wheat for a given amount of rice than before. Before when we were producing 20 bushels of rice, we could only produce 78 bushels of wheat. Now, after technological progress, we’re able to produce maybe as many as 85 or 90 bushels of wheat when we produce 20 bushels of rice. For every quantity of rice produced, we can now make more wheat than before. For every quantity of wheat produced, we can now make more rice than before. Technical progress allows us to expand our production possibilities. We can now make more than before with our given quantity of scarce resources.

Suppose now that we have general migration of labor into the economy, and that this labor is well suited for either working in the production of rice or wheat. A migration into the economy of agricultural labor would also shift the production possibilities curve outwards. Now with more workers to employ in the production of wheat and rice, we can make more wheat and rice than before. For any quantity of rice we produce, we can now produce a larger quantity of wheat than before. For any quantity of wheat that we produce, we can now produce a larger quantity of rice than before. An increase in labor allows us to expand our production possibilities in this economy.

Suppose now that we have a different kind of change. Let’s suppose that a change occurs that increases this economy’s productivity in the wheat industry but not in the rice industry. Suppose a new fertilizer is discovered that increased the productivity of wheat farms, but doesn’t do anything for rice farms. How would we show that change in this picture? Production possibilities would change in such a way that there would be an increase in wheat production, but no change in potential rice production. That is, if the economy continued to devote all of its resources to the production of rice, it would only be able to produce 100 bushels of rice, just like before. But if the economy devoted all of its resources to the production of wheat, it might be able to produce 120 bushels of wheat with this new fertilizer. The new fertilizer increases the productivity of wheat without changing the productivity of rice. So we get a skewed shift in the production possibilities curve, one that is slanted in the direction of increases wheat production without changing the productivity of rice.

Let’s consider another possibility. Suppose the climate of our economy changes. With the change in climate of our economy, let’s suppose that the ground is wetter most of the year. It rains more often; the ground gets soaked and saturated. That’s that going to do to the productivity of our economy? What’s that going to do to our production possibilities? Well, let’s suppose that this extra wetness hurts the production of wheat and helps the production of...
Understanding How a Change in Technology or Resources Affects the PPF

We would then show the shift in the production possibilities curve looking something like this. The amount of rice that could be produced with this increased wetness will be larger, maybe at 420 bushels of rice a year. The amount of wheat that could be produced would shrink, maybe down to 60 bushels a year. With this change in the climate there is a reduction in the productivity of this economy for making wheat, and an increase in its productivity in making rice.

One final thing. Let’s suppose we have a large out-migration of labor; that a lot of people leave this economy to go to work somewhere else. Or suppose an earthquake occurs and wipes out a lot of the capital stock, a lot of the tools and equipment that are used. In that case, the production possibilities frontier would actually shift inwards. There would be a reduction in the productivity of this economy. It could produce less of everything. So that producing 20 bushels of rice might leave you only enough resources to make, say, 55 bushels of wheat instead of the previous 78. So it’s possible sometimes that the production possibilities curve would actually shift inward representing a reduction in the economy’s productivity.

So, a quick summary then. Things that increase the productivity of the economy: an increase in the stock of resources, an improvement in technology. Those things shift the production possibilities curve outwards. Those things that reduce productivity, such as a loss of labor or capital, or an unfavorable change in the climate, shift the production possibilities curve inwards. And sometimes changes in the economy will affect the two different sectors in different ways. They might actually cause one sector to become more productive, and the other sector to become less productive.

I’d like to end with a question, and this is a question that came up during your quiz in the last lecture. Look at this production possibilities curve. Of all of these choices, which is the one that our economy prefers? Which combination of wheat and rice is going to make consumers the happiest? The answer is—we have no idea. There’s no representation in this picture for consumer preferences. Unless you know something about consumer preferences, you’re not able to say which of these feasible points is the most desirable. If our economy likes a lot of wheat, this point up here might be the best one. If our economy likes a lot of rice, this point might be the best one. If our economy like a nice balanced mix, then this point might be the best one. The point I’m trying to make is this. The production possibilities curve tells you about the supply side of an economy. It tells you about what this economy is able to produce given its limited quantity of resources, given the technology know-how that the economy possesses, given the peculiarities of the resources—some being well suited for the production of wheat, some being well suited for the production of rice. But this diagram tells you nothing about which of the points is the most desirable. In order to tell which of these points is best we would have to bring in something from the demand side of the economy, knowledge about what pleases consumers. So the blue line here is purely a statement about engineering—what our economy is able to make. And the important concept from this lesson is the concept of efficiency—doing the best you can with the resources that you have and the know-how that you possess. That is, each point on the blue curve is efficient because it represents the maximum amount of wheat that our economy can produce when it produces a given quantity of rice. The blue line represents doing the best we can with what we have, and that’s the definition of efficiency.
Deriving an Algebraic Equation for the Production Possibilities Frontier

In these next two lessons, we're going to introduce and explore one of the economists' favorite ideas, comparative advantage. And we're going to show how two people with differing abilities can cooperate to increase their wealth. We're going to be using the tool that we developed in the last lecture, the production possibilities frontier. And we're going to first derive the production possibilities frontier for a cleaning service run by Bernie. In the next lesson, we'll be introducing a potential cleaning partner for Bernie and show how cooperation can increase the wealth of both cleaners. But first, let's look at Bernie.

Bernie has a service that provides two cleaning chores: Bernie scrubs rooms and Bernie sweeps rooms. The first thing that we want to do is explore Bernie's technology; that is, what is Bernie able to do? Well, let's look. Bernie takes 20 minutes to scrub one room, Bernie takes 10 minutes to sweep one room. In economics, we call these numbers the unit labor requirements, how much time it takes for Bernie to perform each of these tasks. And the unit labor requirements describe Bernie's productivity. Notice that Bernie can sweep faster than he can scrub. So 20 minutes to scrub or 10 minutes to sweep.

Now let's suppose that Bernie is going to work for 1 hour. How much can Bernie accomplish? Well, if he devotes an hour to scrubbing, then we can find out how many rooms he scrubs by dividing 1 hour, or 60 minutes, by the time that it takes Bernie to scrub one room. So 60 minutes divided by 20 minutes is equal to three rooms scrubbed in 1 hour. If Bernie devotes 1 hour to sweeping, on the other hand, divide 60 minutes by 10 minutes per room swept and you find that Bernie can sweep six rooms in one hour. So we'll use these numbers then to represent Bernie's productivity; that is, if Bernie spends all of his time scrubbing, in an hour he can scrub three rooms. If Bernie spends an hour sweeping, he can sweep six rooms.

Now, of course, if Bernie has an hour to work, he needn't do all sweeping or all scrubbing. In fact, he can do any combination of the two. So what we'll do now is derive a production possibilities schedule for Bernie by looking at different combinations of sweeping and scrubbing that are possible for Bernie in 1 hour. Remember Bernie can spend 10 minutes sweeping or 20 minutes scrubbing and divide up his hour any way he likes.

Well, let's look at some numbers that work for Bernie. And here we have a production possibility schedule, much like the one that we looked at last time for rice and wheat. Bernie can spend an hour scrubbing and scrub three rooms. That leaves no time for sweeping, so he sweeps zero rooms. If Bernie decides that he wants to sweep one room, that's going to take 10 minutes. Subtract 10 minutes from 60 minutes to leave 50 minutes. Bernie has 50 minutes now to spend scrubbing. And if you divide 50 minutes by 20 minutes per room scrubbed, you find that Bernie has time to scrub two and a half rooms. If Bernie wants to sweep two rooms, however, that's going to take 2 × 10 or 20, leaving only 40 minutes for scrubbing, so that he can scrub a total of two rooms, and so forth. If he wants to sweep three rooms, that leaves him time to scrub one and a half, if he wants to sweep five rooms, that leaves him time to scrub only half of a room, and if he spends all of his time sweeping, so that he spends no time scrubbing, he can sweep a total of six rooms. This is the production possibilities schedule for Bernie. Notice, as we look at these number, we are already getting a taste of Bernie's opportunity cost. Every room that Bernie sweeps causes him to give up one-half of a room scrubbed. The opportunity cost for Bernie of sweeping one room is one-half of a room that he doesn't scrub.

Well, the next logical step is to take the information from the production possibilities schedule and represent it in a graph. Here's how that would work. I'll move these numbers over to the board, so that you have them for reference, and I'll put up some axes that we can use to draw a production possibilities curve for Bernie's business. Let's put sweeping up here on the vertical axis and let's put scrubbing on the horizontal axis. So here's sweeping and here's scrubbing. Now, let's plot numbers from Bernie's production possibilities schedule into this diagram. And we see that, let's see here, 1, 2, 3, 4, 5, 6 rooms is the maximum that he can sweep and 1, 2, 3 is the maximum that he can scrub, so it looks like my chart here is going to work. I'm going to calibrate my axes a little bit differently, just for the sake of making things nice and neat. I'm going to let each of the hash marks represent two rooms swept, so here's one room, 2, 3, 4, 5, 6, and 1. And I'm going to let each of the hash marks mark down here on the bottom represent one room scrubbed, so here's 1, 2, 3, 4 and so forth. I could go ahead and put the numbers in to calibrate the axis, so let me do that, just to make things nice and neat. Here's 2, 4, 6, 8 and so forth, and down here we have 1, 2, 4, 5, 6, 8 and so on.

Now, let's put in the points from the schedule into the diagram so that we can plot the production possibilities curve. Let's start here with if we scrub three rooms, then we can sweep zero, so go to three rooms scrubbed and zero room swept and we get this dot right here on the production possibility frontier, three scrubbed, zero swept. If we sweep
Deriving an Algebraic Equation for the Production Possibilities Frontier

one room, then we are going to be only able to scrub two and a half rooms, so we’re going to get a point like this one right here. If we decide to sweep two rooms, then we’re going to be able to scrub two rooms, so that gives us a point like this one and so forth. Just keep plotting the points. If we are going to scrub one and a half rooms, that leaves us time to sweep three and if we scrub one, that leaves us time to sweep four. One-half leaves us time to sweep five. And if we don’t do any scrubbing at all, we can sweep six rooms.

Well, there you have it. There’s the production possibilities frontier for Bernie. Bernie can sweep anywhere from zero to six rooms. He can scrub anywhere from zero to three rooms and if he divides his time, he can achieve any combination that’s on this line. Let’s now imagine that Bernie can divide his time even more finely between sweeping and scrubbing and we’ll be able to connect those dots and get a straight line. This is what Bernie’s production possibilities frontier looks like.

Now let’s notice a few things about Bernie’s production possibility frontier. First, the vertical intercept represents the maximum number of rooms that Bernie can sweep if he spends a whole hour sweeping. The horizontal intercept represents the maximum number of rooms that Bernie can scrub if he spends all of his time scrubbing. The downward slope reminds us of scarcity. Bernie’s time is scarce. If he wants to scrub more rooms, he has to give up rooms swept. There are points outside the production possibilities frontier, representing Bernie’s limitations. He only has an hour, so he couldn’t, say, sweep six rooms and scrub six rooms. That would take too much time. There are also points inside the frontier if Bernie were using his time inefficiently.

So the intercepts tell us about the maximum amount of sweeping or scrubbing that Bernie could possibly get from an hour’s worth of time. What about the slope of the line? What does that tell us? The slope, as usual, tells us about the opportunity cost of scrubbing, measured in terms of sweeping; that is, if Bernie wants to scrub one additional room, if want to increase scrubbing from, say, two rooms to three rooms, that is a horizontal movement in this space. That means we would have to reduce the amount of sweeping that Bernie does. That would be a vertical movement in the space. And if we take the amount of sweeping that Bernie gives up to do an extra room’s worth of scrubbing, that gives us the opportunity cost of scrubbing the room. So the rise, which, in this case, would be a change in Bernie’s sweeping, a reduction in the amount of sweeping that Bernie is able to do, over the run, which would be the increase in scrubbing, is the slope of the production possibilities frontier. In this case, in order to have a change in scrubbing equal to one room, we get a change in sweeping equal to –2 rooms. For every room that Bernie scrubs, he gives up two rooms that he could have swept. The slope of this line is –2 and it’s the opportunity cost of scrubbing for Bernie, measured in terms of the sweeping that he can’t do. Of course, this is logical. It takes him 20 minutes to scrub a room, divide that by 10 minutes that it takes to sweep a room, and you see that he could be sweeping two rooms in the time that it takes him to scrub one. And that’s the slope of this line. The opportunity cost is the slope of the production possibilities frontier.

Notice that we have a line here, so we can describe it with an algebraic equation. And, in this case, the equation for this line is going to be, first of all, we have the variable on the vertical axis, so I can write this sweeping – that’s the thing we’re going to be describing with our line – sweeping is going to be equal to – the next thing you do is write down the vertical intercept. Six rooms swept, this is the maximum number of rooms that Bernie can sweep if he spends a full hour sweeping, minus the slope of the line, which, in this case, is 2, multiplied by the horizontal axis variable. The horizontal axis variable is scrubbing. So here’s the formula for the line: intercept – slope × x-axis variable. The amount of sweeping that Bernie does is equal to six rooms, the maximum that he can sweep, minus 2 times the number of rooms that he scrubs, because any time he scrubs a room, that takes 20 minutes, and that’s time enough to sweep two rooms. So the vertical intercept is the constant in this equation and the slope, -2, is the opportunity cost of scrubbing, measured in terms of sweeping that Bernie isn’t able to do. When he scrubs a room, he gives up two rooms’ worth of sweeping. There’s the equation for Bernie’s production possibilities frontier.

Well, it seems kind of arbitrary that we’ve put sweeping on the vertical axis and scrubbing on the horizontal axis. And it is, in fact, quite arbitrary. We could have just as easily put scrubbing on the vertical axis and sweeping on the horizontal. Well, I’d like to try to make that point clear now by taking the production possibilities frontier and showing you how to convert it into the other way of expressing things, with scrubbing up top and sweeping on the horizontal.

Here’s what I’m going to do, I’m going to put this line in for my production possibilities frontier and I’m going to secure it with tape. And what I’m about to do is simply transpose the axes. I’m going to just convert this diagram from one with scrubbing on the horizontal to one with scrubbing on the vertical. Bernie still has the same productivity, no matter
Deriving an Algebraic Equation for the Production Possibilities Frontier

how we express it. Bernie can still do the same combination of sweeping and scrubbing, but I'm just going to change the way I write it. And it's going to be as easy as picking this graph up and flipping it over, so that the scrubbing axis is vertical and the sweeping axis is horizontal. I'll have exactly the same expressions. So, let me do that then. I'll pick scrubbing up and put it on the vertical. I'll pick sweeping up and put in on the horizontal, and then I'll flip my diagram over, so that the scrubbing information is on the vertical and the sweeping information is on the horizontal. Now notice I've got a bit of a mess here, because I still have all the markings that I did for my previous graph. But if you'll bear with me and let me tape my production possibilities curve down to the axis firmly, and if you'll let me clear my pad here, so that I get rid of the old writing, voilà! We've got the new production possibilities frontier with sweeping on the horizontal axis and scrubbing on the vertical axis. The vertical intercept is now the maximum amount of scrubbing that Bernie can do if he scrubs for an hour. And, as we've seen, that's equal to three rooms. The horizontal intercept is the maximum amount of sweeping that Bernie can do if he sweeps for an hour and, as we've seen, that's six rooms. The formula for this new equation is going to be given by this expression. Bernie's production possibilities frontier, expressed this way with scrubbing on the vertical and sweeping on the horizontal, has exactly this mathematical expression. 3 is the vertical intercept or the maximum number of rooms he can scrub in an hour, and \(-\frac{1}{2}\) is the opportunity cost. Anytime Bernie sweeps an extra room that takes 10 minutes. 10 minutes divided by 20

minutes, the amount of time that it takes to scrub a room, gives us \(\frac{1}{2}\). Anytime Bernie sweeps a room, he's giving up one-half of a scrubbed room, because it takes half as long to sweep a room as it does to scrub the room.

So there you have it, Bernie's production possibilities frontier. Scrubbing is equal to three rooms max minus one-half the opportunity cost times each room that he sweeps. Now, let me show you something about the relationship between those two production possibilities frontier equations, and this is what's interesting.

Notice that the original equation was sweeping is equal to 6 – 2 times scrubbing. Well, let's rewrite this equation and move scrubbing over to the left-hand side, so that we have an equation that has scrubbing measured in terms of sweeping. Well, here's what I'll do, I'll move this \(-2\) scrubbing to the left-hand side, and that's equal to 6 minus sweeping. So all I did was move this expression to the right-hand side and this expression to the left-hand side. It looks like I need to make sure that I change the sign when I move it over. Then let's divide both sides by 2, and I get

scrubbing = 3 - \(\frac{1}{2}\) sweeping.

There you have it. The production possibilities frontier can be written either way, in terms of sweeping or in terms of scrubbing. This is the vertical intercept, the maximum amount of scrubbing and this, \(-\frac{1}{2}\), is the opportunity cost.

Here's one more interesting thing to note. Up here, this \(-2\) was the opportunity cost of scrubbing, measured in terms of sweeping. If Bernie wants to scrub a room, he gives up two rooms that he could have swept. Down here, this \(-\frac{1}{2}\) is the opportunity cost of sweeping, measured in terms of scrubbing. If Bernie wants to sweep a room, he gives up one-half of a room that he could have scrubbed. Now, this is the thing that's important to notice: the opportunity cost of scrubbing, measured in terms of sweeping, is the reciprocal of the opportunity cost of sweeping, measured in terms of scrubbing. If Bernie wants to sweep a room, he has to give up half of a room that he could scrub. If Bernie wants to scrub a room, he has to give up two rooms that he might otherwise sweep. See, the opportunity costs are reciprocals of each other, and that makes mathematical sense, because whenever you're moving and dividing, you're going to wind up with the reciprocals. But it even makes good intuitive sense. It takes 10 minutes to sweep, it takes 20 minutes to scrub. So if you scrub, you're giving up 20 minutes divided by 10 minutes, two rooms you could have swept. If you want to sweep, that takes 10 minutes, so that means you're giving up 10 minutes divided by 20 minutes, one-half of a room that you could have scrubbed. The opportunity cost of sweeping, measured in terms of scrubbing, is the reciprocal of the opportunity cost of scrubbing, measured in terms of sweeping. And that's really what it means to pick this diagram up and flip it around. You wind up with exactly the same line, only now you're measuring the axes differently. You're measuring the different variables on the different axis, so that the intercepts are switched around,
Introducing Economic Thinking

Production Possibilities

Deriving an Algebraic Equation for the Production Possibilities Frontier

and the slope of the new line is the reciprocal of the slope of the original line. The original line, with sweeping on the vertical, had a slope of $-2$. The new line, with scrubbing on the vertical, has a slope of $\frac{1}{2}$.

Now, there’s one more thing that we ought to notice about this picture before we begin to apply it to the bigger question of comparative advantage and gains from trade. That’s this: Bernie’s production possibilities curve is a straight line. What does that tell us about Bernie’s technology, his abilities? Remember the production possibilities frontier from our previous example with rice and wheat was bowed outwards. The outward-bowed production possibilities frontier reminded us that not all resources are equally well suited to all uses. Some land is wet, it’s good for rice, some land is dry, and it’s good for wheat. Bernie, on the other hand, apparently has time and talent that’s equally well suited to sweeping or scrubbing. There is no outward bow to Bernie’s production possibilities frontier, because there is increasing opportunity cost. Anytime Bernie wants to sweep an extra room, it’s going to take him 10 minutes. Anytime Bernie wants to scrub an extra room, it’s going to take him 20 minutes. These are the assumptions that we’ve used to describe Bernie’s technology. Well, because Bernie’s costs are constant in this way, because all of his time and talent is equally well suited for sweeping and scrubbing, Bernie’s opportunity cost is constant. The constant opportunity cost of one-half of a scrubbed room for every room that you want to sweep, or two swept rooms for every room that you want to scrub, gives us a straight line production possibilities frontier.

So, a quick summary: a bowed-outward production possibilities frontier, a concave production possibilities frontier is for a case where the technology has increasing opportunity cost, and that’s because not all resources are equally well suited to all uses. But a straight line production possibilities frontier is for a case where the opportunity cost is constant. And that will be the case when all resources are equally well suited to the production of either of the goods or services.

So now we’ve given a thorough description of Bernie’s technology. We started with his unit labor requirements, then we looked at what he was able to do with 1 hour of his time. Then we looked at a schedule of all the possibilities of sweeping and scrubbing that Bernie could turn out. Then we drew a picture to represent his production possibilities, and we saw that it was a straight line, representing constant opportunity costs. Then we showed that we could represent the production possibilities with either good on either axis, and what we found was that the opportunity cost of one, in terms of the other, was the reciprocal of the opportunity cost of the other, in terms of the original good, $-2$ versus $\frac{1}{2}$.

Now, these are pretty technical points. In the next lesson, we’re going to apply these technical points as we look at potential cooperation between Bernie and a business partner, Anne, as they seek to increase the wealth of their cleaning businesses.
Introduction to Economic Thinking

Comparative Advantage

Defining Comparative Advantage With the Production Possibilities Frontier

In this lesson, we’re going to introduce the concept of comparative advantage and show how two people with different abilities can cooperate to increase their wealth. In the last lesson, we introduced Bernie and his cleaning service and we discussed his productivity and his production possibilities. In this lesson, we’re going to introduce a partner for Bernie. Her name is Ann and she has her own cleaning service. We’re going to find out that because of their different abilities, Bernie and Ann can form a joint venture and in fact do a lot more cleaning together than they can do separately.

Let’s begin by looking at Ann’s productivity. Let’s suppose that Ann has the following unit labor requirements. Ann is able to scrub a room in five minutes and Ann can sweep a room in five minutes. The first thing you notice, looking at this chart, is that Ann is able to sweep and scrub faster than Bernie. We say that Ann has an absolute advantage in the provision of both services. Whenever you can do something with less effort or fewer resources than another person, you have an absolute advantage in the provision of that service. Because Ann can scrub faster, she has an absolute advantage in scrubbing. Because Ann can sweep faster, Ann has an absolute advantage in sweeping.

Now, let’s suppose Ann spends an hour doing sweeping or scrubbing. What is the maximum output that she could expect? Well, divide 60 minutes, one hour, by five minutes that it takes to scrub a room and you get that Ann’s maximum number of rooms scrubbed in one hour is 60 divided by five or 12. Likewise, if Ann wants to sweep for an hour, she can sweep a total of 60 minutes divided by five minutes per room swept to get a total of twelve rooms swept. So, with Ann’s unit labor requirements here and one hour’s worth of time to work, Ann’s maximum output would be given by this table. Just as Bernie can scrub a maximum of three rooms or sweep a maximum of six rooms, Ann can scrub a maximum of 12 rooms in an hour or sweep a maximum of 12 rooms in one hour.

We can now take this information and look at Ann’s production possibilities schedule. With one hour’s worth of time, if Ann wants to divide her time between sweeping and scrubbing, what possible combinations of output can she get? Well, let’s have a look. Here’s a table then that represents Ann’s possible outputs, sweeping and scrubbing. If Ann spends all of her time scrubbing, she can scrub 12 rooms. That means she does no sweeping. If she wants to sweep two rooms, that’s going to take her two times five minutes per room for a total of 10 minutes. Subtract 10 minutes from 60 minutes and that leaves 50 minutes. 50 minutes divided by five is 10 rooms worth of scrubbing that Ann can do. 50 minutes is long enough for Ann to scrub 10 rooms at five minutes per room, and so forth.

The opportunity cost for Ann of sweeping another two rooms to bring her to a total of four would be a loss of two rooms worth of scrubbing. She’d be down to eight rooms scrubbed. Six minutes would be or six rooms of sweeping would be 6 times five minutes or 30 minutes sweeping. Leaving her a half an hour for scrubbing, divided by five minutes is a total of six rooms scrubbed. Notice that all of these numbers in this column, 12 and zero, 10 and two, eight and four and so forth, all of the combinations add to a total of 12. That’s because Ann’s opportunity costs are one room swept for one room scrubbed. How did I calculate that? Well, the cost of sweeping a room is five minutes of Ann’s time, divided by five minutes that it takes to scrub a room. That means every room that Ann sweeps is a room she’s not scrubbing.

Her opportunity cost is constant. The opportunity cost of sweeping one room for Ann, is one room worth of lost scrubbing. Likewise, because of this reciprocal principle, if Ann scrubs a room that’s a room that she’s not sweeping. She spends five minutes scrubbing, that’s the same five minutes she could be spending sweeping. So the opportunity cost of one room worth of scrubbing is one room worth of lost sweeping. Here’s Ann’s production possibilities schedule. Now, let’s convert this production possibilities schedule into a production possibilities curve. Let’s draw a diagram like the diagram that we drew for Bernie. I’ll clear my pad here and I’ll move these numbers over onto the board, so that you can use them for reference.

Well, over here, I’ve got Bernie’s production possibilities frontier that we drew in the last set of lectures. And to remind us about Bernie’s situation, we’ll go ahead and put up his equation. Bernie’s production possibilities frontier is described by the following expression. The total number of rooms that Bernie can scrub is equal to three. That is, if he spends a whole hour scrubbing, minus one-half the opportunity cost time the number of rooms that he sweeps. Each room that he sweeps is 10 minutes and that means that he’s giving up 20 minutes divided into 10, or 10 divided by 20 or one-half of a room worth of scrubbing. We saw this last time.

Every room that he sweeps is 10 minutes worth of time and that’s half of a 20 minute job worth of scrubbing that he’s not able to do. The intercept tells you the maximum amount of scrubbing that Bernie can do and the slope of the line
Introduction to Economic Thinking

Comparative Advantage

Defining Comparative Advantage With the Production Possibilities Frontier

tells you the opportunity cost of a room swept measured in terms of lost scrubbing. So, here’s Bernie’s equation and here’s his production possibilities curve. Now we can draw the curve for Ann. Well, let’s start. If Ann does no sweeping and all scrubbing, then she can scrub 12 rooms. So let’s put a dot here in her curve that represents 12 rooms worth of scrubbing and zero rooms worth of sweeping.

If she decides that she wants to do a little bit of sweeping, every time she wants to sweep an additional room she has to give up a room worth of scrubbing. So that gives us combinations like 10 and two and eight and four and so forth. And we get these points all the way down the curve. Every time Ann decides that she wants to sweep an extra room, she losses a room worth of scrubbing until finally we get down to this horizontal axis, where she’s sweeping 12 rooms and scrubbing zero rooms. Notice we can connect the dots here to find Ann’s production possibilities frontier. So, let’s connect the dots and just like Bernie, Ann’s production possibilities frontier is a straight line. What does that tell you?

What does it tell you that Ann has a straight-line production possibilities frontier? It tells you that her opportunity cost is constant. Any time she wants to sweep an extra room, she gives up one room worth of scrubbing. Why is her opportunity cost constant? It’s constant because her time and resources are equally well suited to providing each of the two services. Ann would just as soon scrub a room as sweep a room. She can just as easily, with her time and talent, do either. There is no change in opportunity cost. If Ann does more sweeping, she always does so by giving up one scrubbed room for each additional room she sweeps. Her opportunity cost is constant.

Let’s look at the curve here and see what we find. The vertical intercept of Ann’s curve is going to be 12. That’s the maximum amount of scrubbing that Ann can do. Down here, the horizontal intercept is also going to be 12. That’s the maximum amount of sweeping that Ann can do with an hour’s worth of time. The slope of the production possibilities curve is rise over run or negative one. That is, every additional room that Ann sweeps is one room worth of scrubbing that she no longer has time to do. You find the opportunity cost by dividing the number of minutes that it takes to sweep by the number of minutes that it takes to scrub. Five minutes worth of sweeping divided by five minutes worth of scrubbing gives you the opportunity cost of sweeping, measured in terms of scrubbing that Ann’s not able to do. Every time she sweeps a room, that’s a room worth of scrubbing that she gives up. The slope of the curve is minus one. Now, one more thing to do before we start talking about comparative advantage and that is suppose Bernie and Ann were working independently of one another. What kind of outcome could they expect? Well, here’s Ann’s production possibilities frontier. The formula is the amount of scrubbing she can do is equal to 12 rooms worth of scrubbing minus the number of rooms that she sweeps. The coefficient on sweeping is one, negative one. That’s the cost to Ann to sweep a room. It’s one room worth of lost scrubbing.

Here’s Bernie over here and we know that his production possibilities are represented by this equation. Let’s suppose that Bernie and Ann are both running cleaning businesses and that they are working independently of each other. If Bernie and Ann worked independently of each other and don’t cooperate, then let’s suppose that each of them have to sweep the same number of rooms that they scrub. You go to a house and it’s got two rooms in it. You’ve got to sweep each room and you have to scrub each room in order to clean that house.

So the question is, if Bernie and Ann have to sweep and scrub the same number of rooms each, that is, Bernie has to sweep say two rooms and scrub two rooms and Ann has to say sweep six rooms and scrub six rooms. Well, what numbers are going to work out? What’s the maximum number of rooms that Bernie can clean by himself? What’s the maximum number of rooms that Ann can clean by herself? Remember, the numbers have to be equal for sweeping and scrubbing for Bernie and the numbers have to be equal for sweeping and scrubbing for Ann. Well, the numbers that work for Bernie, as it turns out, are two rooms. Bernie can sweep two rooms, and that takes him two times 10 minutes each for a total of 20 minutes worth of sweeping. That leaves him 40 minutes for scrubbing and since it takes him 20 minutes to scrub a room, 40 minutes divided by 20 means that he has time left to scrub the two rooms.

So, if Bernie works by himself, then he is able to sweep and scrub two rooms each. So, it looks like that point would be right here on the diagram. Bernie working by himself wins up scrubbing two rooms and sweeping two rooms. If Ann works by herself and she sweeps six rooms, that’s going to take her six times five minutes each, is 30 minutes. Leaving her half an hour for scrubbing. Since it takes her five minutes to scrub a room, she can also scrub the six rooms. So, Ann working by herself is going to be able to sweep six rooms and scrub six rooms.
Now, if you'll look over on the board, I'm going to do a summary of this outcome. When they work independently of each other and each one of them has to sweep and scrub the same number of rooms, then, Bernie is able to sweep and scrub two rooms in an hour. Ann is able to sweep and scrub six rooms in an hour. Here's what you'll notice over on the board. I've done a table of what Bernie and Ann are able to produce when they work independently. When they work independently, Bernie can sweep two rooms, Ann can sweep six rooms for a total of eight rooms of sweeping. Eight total rooms get clean when Bernie and Ann work independently of each other.

Next, I'm going to show how Bernie and Ann can increase the profits of their business. That is, how Bernie and Ann can cooperate and clean more rooms together than they can when they work independently. And this is where we come to the concept of comparative advantage. Here we have Bernie and here we have Ann and next to each of our players we have their present production patterns. Bernie is sweeping two rooms and scrubbing two rooms. Ann is sweeping six rooms and scrubbing six rooms. Now, we're ready to introduce the concept of comparative advantage. Comparative advantage refers to more opportunity cost relative the opportunity cost of someone with whom you might trade. We say that you have a comparative advantage in the provision of a particular good or service if your opportunity cost of providing that good or service is lower than the opportunity cost of someone with whom you might trade.

Let's use an example now to try to make this concept very clear. We'll start by looking at the opportunity cost of our two players and look at who has a lower opportunity cost for providing he service of scrubbing. Let's start with Bernie. If Bernie scrubs a room, it takes him 20 minutes. Now, since it only takes him 10 minutes to sweep a room, Bernie's opportunity cost for scrubbing one room is two rooms worth of sweeping that he's not doing. So, that's his opportunity cost. Let's look now at Ann's opportunity cost for scrubbing. Ann takes five minutes to sweep a room and also five minutes to scrub a room. So the opportunity cost for Ann of scrubbing a room is five minutes worth of time, which means sweeping one room. So, if Ann scrubs a room it's because she's not sweeping a room. Her opportunity cost of scrubbing one room is one room that goes unswept. Given this information, we know who has a comparative advantage in scrubbing. For Bernie scrubbing costs two rooms worth of sweeping for Ann scrubbing costs only one room worth of sweeping. Ann has a lower opportunity cost for scrubbing and therefore Ann has a comparative advantage in providing scrubbing services.

Well, why should you care? Why should you care that Ann has a comparative advantage? Because comparative advantage is a guide to how Bernie and Ann can specialize, and therefore increase their wealth. By specializing and trading, Bernie and Ann can clean more rooms than the can working independently. Let me show you how this works. Let's suppose Bernie decides to reallocate his time and specialize in sweeping. That is, to do more sweeping and less scrubbing. Why am I having Bernie do more sweeping and less scrubbing? Because scrubbing is relatively expensive for Bernie to do. He doesn't have a comparative advantage in scrubbing. He would like to work out a deal where Ann does more scrubbing because her cost is lower. Bernie will do more sweeping because his cost for sweeping is lower.

Remember, when Bernie sweeps a room, he only gives up half a room worth of scrubbing, Ann gives up a whole room worth of scrubbing. So, Bernie has a comparative advantage in sweeping just as Ann has a comparative advantage in scrubbing. So let's have Bernie specialize according to his comparative advantage and see what happens. Well, first thing that happens is, Bernie reallocates 20 minutes of time away from scrubbing and instead winds up sweeping two additional rooms. So, with his newly allocated time, Bernie adds two rooms worth of sweeping and loses one room worth of scrubbing. Well now he's sweeping four rooms and scrubbing one, so he needs a little bit of help here from Ann in order to make sure that he's getting a full house clean.

What he does then, is he offers to trade with Ann two rooms worth of sweeping, those newly created two rooms and see if Ann will give him some scrubbing in exchange. So what Bernie does then is give some sweeping services to Ann. He sweeps two rooms for her. Well this now frees Ann not to have to sweep those two rooms herself and with those rooms not having to be swept by Ann, she has five minutes here and five minutes here that she can now reallocate towards scrubbing. So, Ann sweeps less and scrubs more. She moves this scrubbing, this time into
scrubbing and increases her scrubbing activity as she cuts back on her sweeping. So now Ann is sweeping four rooms, plus she’s getting two rooms worth of sweeping services from Bernie and that allows her time to scrub a total of eight rooms.

She can now give one of those scrubbed rooms to Bernie, that is, scrub one of his rooms and he’s just as well off as he was before. Now with this help from Ann, Bernie has two rooms scrubbed and two rooms swept. He’s also sweeping two rooms from Ann, so she can scrub two additional rooms. And if you’ll look, our total output now is eight rooms worth of sweeping and nine rooms worth of scrubbing. Where did this extra room of scrubbing come from? It came from specialization. Because Bernie specialized according to his comparative advantage these two working together were able to scrub an additional room. They were able to turn out a little bit of extra labor, relative to what they did before.

So, now Bernie and Ann have eight rooms swept and nine rooms scrubbed. Can they do any better? Well, probably so. Right now we’ve got an unequal number of sweeping and scrubbing. So we’ve got this scrubbed room that’s unswept. Let’s see if we can do any better. Well, of course we can. And what should we do? We should have Bernie do even less scrubbing. That is, here’s the room that Bernie’s still scrubbing, let’s have him take those 20 minutes and sweep instead. That way he can turn out two more rooms worth of sweeping services. This promises to be a good deal. Now Bernie is doing no scrubbing at all. He’s doing all sweeping. He’s got one room worth of scrubbing services from Ann that he’s acquired and now he is ready to give Ann two more rooms worth of sweeping.

Once he gives her those two more rooms, Ann is now freed to scrub some more. She converts her time from sweeping away to scrubbing and now we’ve got all these scrubbed rooms that Ann has done and she can send one of them over to Bernie. Bernie’s just as happy as he was before. He’s got two rooms swept, two rooms scrubbed. Ann is scrubbing both of these rooms for him and he is sweeping these two rooms, plus for more rooms from Ann. But the total now that we’ve got here is one, two, three, four, five, six, seven, eight rooms swept and one, two, three, four, five, six, seven, eight, nine, ten rooms scrubbed. Well, we’ve got a little bit of an imbalance here and the imbalance is that we’ve got more rooms scrubbed than we have swept.

If Ann wants to be able to sweep and scrub equal numbers of rooms then she might want to convert one of these scrubbings back into a sweeping, so that she now has equal numbers of rooms swept and scrubbed. Well, look at this. Ann has now got a total of one two, three, four, five, six, seven rooms swept and one two, three, four, five, six, seven rooms scrubbed. Bernie has two rooms swept and two rooms scrubbed. By cooperating and specializing according to comparative advantage, by Ann doing more scrubbing and Bernie doing more sweeping and by trading with each other, Bernie is just as well off as he was before and Ann is better off than she was before.

Now, is that necessarily going to be the case? Well, not necessarily. I mean, certainly since we have nine rooms clean now, instead of eight before, the two of them combined are able to make more money with their cleaning business. Ann could stay where she was before and leave Bernie with three rooms clean and Ann would still have the same six rooms swept and scrubbed as before. It doesn’t really matter who gets the benefit. That would be determined by which of these two was a more persuasive bargainer, but the point that I want to make is this. The point that I want to make is this. By Bernie specializing according to his comparative advantage, and by Ann specializing according to her comparative advantage, the two of them working together specializing and trading are able to produce nine clean rooms. Nine swept, nine scrubbed rooms, rather than simply eight as before.

Now, let’s take a look at the production possibilities frontier and show how this story translates into a graph.
Comparative Advantage

Understanding Why Specialization Increases Total Output

Remember these production possibilities diagram. Here’s Bernie and here’s Ann. The slope of Bernie’s production possibilities frontier is minus one-half. That’s the opportunity cost of sweeping measured in terms of lost scrubbing. Ann’s opportunity cost is negative one. That’s her opportunity cost of sweeping in terms of lost scrubbing. Because sweeping is less expensive for Bernie, we had Bernie specialize in sweeping services and we showed that we could get a large increase in total output if we had Bernie completely specialize in sweeping.

So here’s what we’re going to have Bernie do. Rather than staying at sweeping two rooms and scrubbing two rooms, let’s have Bernie completely specialize in sweeping, the task at which Bernie has the comparative advantage. So, Bernie’s production point now moves all the way to the end of the production possibilities frontier. Bernie completely specializes in sweeping. So, here’s Bernie’s new output point right here. It’s at the end of the production possibilities frontier, where Bernie is now sweeping six rooms and scrubbing zero. We had Ann then specialize by moving up in this direction. Ann reduces her sweeping, taking advantage of Bernie’s low cost sweeping and Ann scrubs instead.

Ann moves her production point up to this dot on her production possibilities frontier. Ann winds up scrubbing nine rooms and sweeping only three rooms. When she does that, she’s able to take advantage of Bernie’s low cost sweeping and increase the total number of rooms that she cleans. Now, if you’ll have a look over on the board, you’ll see our totals with trade. With trade Bernie sweeps a total of six rooms and scrubs zero. Ann sweeps a total of three rooms and scrubs nine. The total number of scrubbed rooms is nine plus zero or nine. The total number of swept rooms is six plus three or nine. Working together, Bernie and Ann are able to sweep and scrub a total of nine rooms.

Looking at this picture, how could we have known that Bernie was going to wind up specializing in sweeping services? Well, we would know because Bernie’s production possibilities frontier is flatter than Ann’s. His slope is minus one-half, Ann’s is minus one. Because Bernie’s production possibilities curve has a flatter slope, we know that Bernie’s opportunity cost is smaller for sweeping than Ann’s is. Bernie has an opportunity cost of one-half rooms scrubbed. Ann has an opportunity cost of one whole room scrubbed. Therefore, Bernie will specialize in sweeping because he has a comparative advantage in sweeping. Ann will then specialize in scrubbing.

Now, finding the exact numbers where this story worked out requires a little bit of puzzle solving. I had to figure out, well, if Bernie specializes, that’s six rooms worth of sweeping. So Ann has to do at least six rooms worth of scrubbing. Which leaves her, six time five is 30 minutes from an hour, that leaves Ann 30 minutes of half an hour for additional work. And since six and six is already taken care of then she would have to split that remaining time between sweeping and scrubbing. And that’s how I wound up with her sweeping three rooms and scrubbing three additional rooms, for a total of nine.

The things you should learn from looking at the diagrams are as follows. First, the trader that has a flatter production possibilities frontier will have a comparative advantage in providing the good or service that is written on the horizontal axis. Here, Bernie has the comparative advantage in providing sweeping services, because his opportunity cost is lower in terms of sweeping than Ann’s opportunity cost. The next thing to notice is that if you have a comparative advantage in one good, you have a comparative disadvantage in the other good. Bernie has an opportunity cost of two rooms worth of sweeping for every room of scrubbing he does. Ann can scrub a room by only giving up one room worth of sweeping.

Because these slopes are reciprocal, if you have a comparative advantage in one good, you have a comparative disadvantage in providing the other good. Just as Bernie has a relatively low opportunity cost for sweeping, he has a higher opportunity cost than Ann for scrubbing. All right. So you’ve always got a comparative advantage in something. Even though Ann can sweep and scrub faster than Bernie, Bernie is relatively better at sweeping than he is at scrubbing and therefore his opportunity cost is lower and that’s why he and Ann can get together and make a deal that benefits both of them. It’s not driven by absolute advantage. The story is driven by comparative advantage and no matter how unproductive you are you always have a comparative advantage in something. Even though Bernie has higher cost for providing all services and Ann can do everything faster, Bernie is still relatively less unproductive at sweeping than he is at scrubbing. And that’s how he and Ann can get together and make a profit.

The last thing to notice is that by getting together and trading, Bernie is able to make a deal that’s satisfying to him. Before he was sweeping two rooms and scrubbing two rooms. Now that he’s completely specialized in sweeping services, he would be willing to trade as many as four rooms worth of sweeping to get two rooms worth of scrubbing in exchange. That would put him right back where he was before and Ann is happy to give him two rooms worth of
Introduction to Economic Thinking

Comparative Advantage

Understanding Why Specialization Increases Total Output

scrubbing, because that would leave her ahead of where she was before. So, once Bernie specializes, according to his comparative advantage, there’s room for him and Ann to work out a deal that makes one or both of them better off. That’s the magic of comparative advantage. By people specializing according to comparative advantage, everyone who’s trading can gain.

In the next example, we’ll use a story that sounds a little bit more like the international trade stories you read in the newspaper. We’ll look at two countries that are producing two agricultural products and show that by specializing and trading, they can both be made better off.
Analyzing an International Trade Using Comparative Advantage

In the last lesson, we introduced the concept of comparative advantage and showed how two players with different opportunity costs could cooperate, specialize, and trade and both be better off. In this lesson, we’ll give another illustration of the concept of comparative advantage. This one is a story about international trade and we’ll show how two countries with different opportunity costs for producing two products can specialize and trade and both be better off.

Let’s look first at the country of Pakistan, and we’ll suppose that Pakistan uses one input, labor, to produce two outputs, two agricultural products, wheat and rice. We’ll begin with a description of Pakistan’s technology. Suppose that the unit labor requirement for producing one good, wheat, is two workers; that is, in Pakistan, it takes two workers to produce one bushel of wheat. Suppose also that the unit labor requirement for rice is three workers. It takes three workers to produce one bushel of rice. Now, if this is the case, the opportunity costs can be calculated from the unit labor requirements; that is, if you want to produce a bushel of rice, it’s going to take you three workers. Those three workers divided by the two workers that are needed to produce a bushel of wheat let you know that anytime you produce one bushel of rice, you are giving up one and a half bushels of wheat. The opportunity cost in Pakistan to produce one bushel of rice is one and a half bushels of wheat.

Now, let’s introduce the constraint on Pakistan’s production. Suppose that there are 60 workers in Pakistan. These 60 workers can be used to produce wheat or they can produce rice, but the amount of wheat and rice that can be produced will be limited by the technology. In particular, suppose we take those 60 workers and divide them by the two workers that are needed to produce one bushel of wheat. The total amount of wheat that can be produced in Pakistan in a given period of time is 30 bushels of wheat. 60 workers divided by 2 workers per bushel of wheat equals a total of 30 bushels of wheat, the maximum wheat output possible in Pakistan.

Suppose instead that all 60 workers are used to produce rice. Divide 60 workers by three workers per bushel of rice and you get a maximum output of 20 bushels of rice. Therefore, we’ve got the constraints on production possibilities in Pakistan, 60 workers, two workers per bushel of wheat, three workers per bushel of rice. We’re ready now to draw a production possibilities frontier for Pakistan’s economy.

But before I do that, let me go ahead and introduce a potential trading partner for Pakistan, and that is Malaysia. Malaysia is going to have a different technology, and this different technology will become the basis for trade between Pakistan and Malaysia. Let’s suppose that in Malaysia it takes only one worker to produce a bushel of wheat. Notice that Malaysia has an absolute advantage in the production of wheat; that is, Malaysia can produce wheat with fewer workers than it takes to produce wheat in Pakistan. Malaysia is therefore better at producing wheat. They can make more with less and therefore we say Malaysia has an absolute advantage in the production of wheat.

Likewise, it takes two workers to produce one bushel of rice in Malaysia; therefore we say that Malaysia also has an absolute advantage in the production of rice. It only takes two workers to produce a bushel of rice in Malaysia. It takes three workers to produce a bushel of rice in Pakistan. Because Malaysia can do more with less, we say that Malaysia has an absolute advantage.

Notice the way I’ve set up this example. I’ve cooked up a case where Malaysia has an absolute advantage in the production of both goods. Economists like examples that work like this. We like to show you that even though Malaysia is better at doing everything; that is, even though Malaysia can produce more wheat and more rice with a given amount of labor than Pakistan, in the end we will find that it is still advantageous for Malaysia and Pakistan to cooperate and trade. Even though Malaysia has an absolute advantage in the production of each good, it will still turn out that Pakistan has a comparative advantage in something.

Now, let’s suppose then that these numbers represent technology in Malaysia. What is the opportunity cost for Malaysians of producing a bushel of rice? Well, if they produce a bushel of rice, they are going to need two workers to do so. And those two workers, divided by one worker needed per bushel of wheat, means that in Malaysia anytime you’re producing one bushel of rice, you are giving up two bushels of wheat. The opportunity cost in Malaysia of producing one bushel of rice is two bushels of what. Now, in a minute, we’re going to compare Malaysia’s opportunity cost with Pakistan’s opportunity cost to calculate comparative advantage.

But first, let’s look at the constraints on the production possibilities in Malaysia. Suppose there are 60 Malaysian workers and they can spend their time either producing wheat or rice. What’s possible in Malaysia? If you take the 60
Analyzing an International Trade Using Comparative Advantage

workers and divide by one worker needed per bushel of wheat, you get the total maximum wheat production possible in Malaysia is 60 bushels of wheat. 60 workers at one worker per bushel gives you a maximum output of 60 bushels of wheat. Likewise, if those 60 workers go into the production of rice instead, 60 workers divided by two workers per bushel of rice gives you a maximum rice output possible of 30 bushels. Notice that because Malaysia has lower unit labor requirements, Malaysia can make more output with its 60 workers than Pakistan can.

Now, given these numbers, we can write out a table of the production possibilities for Pakistan and for Malaysia. Let's look at those tables now. First, we'll look at the production possibilities for Pakistan. If Pakistan has 60 workers, then it's possible for Pakistan to produce any combination of wheat and rice given in this table. Pakistan can use all of its workers in wheat and produce no rice. Pakistan can produce 20 units of wheat, moving some workers into the production of rice. It can produce 15 and 10, 10 and 13, and so forth. Any of these combinations is possible.

Remember anytime Pakistan wants to produce an extra unit of rice, it has to move three workers into the production of rice. Those are three workers that are not available for wheat production and three divided by two workers needed per bushel of wheat means that anytime Pakistan produces one extra bushel of rice, it's giving up one and a half bushels of wheat that it might otherwise produce. So we will use these numbers in a moment, when we draw the production possibilities frontier for Pakistan. Notice as you look at these numbers, these are the production possibilities. We'll move this number over to the board and we'll look at the production possibilities now for Malaysia.

Notice the numbers are bigger, because Malaysia has an absolute advantage. It is more productive than Pakistan with its labor. Here are all the possibilities, 60 and 0, 40 and 10, 20 and 20 – I say all, these are some. What I mean is all of these combinations are possible with 60 workers and Malaysia's technology. Notice that the opportunity cost of producing 10 bushels of rice is 60 – 40, or 20 bushels of wheat that Malaysia has to give up. As we saw before, anytime Malaysia wants to produce an extra bushel of rise, Malaysia has to give up two bushels of wheat. We'll move these numbers now over to the board and use them whenever we plot the production possibilities frontier for Malaysia.

Let's go ahead now and plot that production possibilities frontier. The first diagram we'll let represent the production possibilities of Malaysia. The second set of axes we'll let represent the production possibilities of Pakistan. Now, we've got to label the axes carefully or the diagram doesn't mean anything, so let's put wheat on the vertical axis in each case and we will put rice on the horizontal axis in each diagram. Now we are ready to draw the production possibilities frontiers and we have two options. The first option is to plot points directly from the schedules that are over on the board. That's easy enough to do, so we can start with Malaysia's. Malaysia, if they produce only wheat, can produce up to 60 bushels of wheat, so I'll go up here to the point 60. That leaves them no labor for producing rice, so we have this point, 60 and 0. If they use all of their labor to produce rice, they can produce 30 bushels of rice and no wheat, so that would be this point, 30 and 0, on the horizontal axis. And then we could just go through and plot other combinations, like 40 and 10, 20 and 20, 10 and 25 and so forth on down. Once we have plotted all of these points, we can connect them in the diagram and draw the production possibilities frontier.

Now, I didn't go through and plot all of those points, because I don't want to make my diagram too cluttered, but it's clear to you that you can plot all of these points and then just connect them with a straight line. Notice the production possibilities frontier will be a straight line in this case. You'd find that out if you'd plotted all of the points. What does it mean? What does it mean that the production possibilities for Malaysia is a straight line? It means that the opportunity cost is constant. Anytime Malaysia wants to produce an extra bushel of rice, it has to give up two bushels of what. If Malaysia wants to move one more bushel of rice outwards in the horizontal axis, then it has to move down two bushels of wheat. And that doesn't matter whether it's producing its first bushel of rice, its tenth or its thirtieth bushel of rise. Every bushel of rice in Malaysia has the same opportunity cost of two bushels of wheat. That's why the slope of Malaysia's production possibilities frontier is constant at –2. There is no increasing opportunity cost. Opportunity cost is constant for Malaysia, and that means that all resources that are used to produce wheat and rice are equally well suited to the production of either good. There is no kind of specialization or special characteristics of resources. All resources are equally well suited and that's why the opportunity cost is constant and the production possibilities frontier is a straight line.

Alternatively, if we hadn't wanted to use the numbers and plot the points from over in the diagram, we could have used the formula for Malaysia's production possibilities frontier. This is the formula that I used to get those numbers to begin with. Before I started this example, I wrote down this formula, because it's what I wanted my production
Analyzing an International Trade Using Comparative Advantage

possibilities frontier to look like. If I’d given you this formula to being with, you could have simply plotted the equation. Here’s the formula for Malaysia’s production possibilities frontier: wheat = 60 – 2 times the amount of rice that Malaysia produces. I’ll move this equation now over to the board and you can see that the equation is what I’ve drawn here in this diagram, that the vertical intercept is at the point of 60, the maximum wheat production that Malaysia can turn out if it uses all 60 of its workers to produce wheat. The horizontal intercept is going to be 30; that is, if you plug in 30 for R, the rice production, you get zero wheat left over to be produced. If all workers are in rice, you get 30 units of rice produced. And the slope of this line is –2. That’s the coefficient on rice in the formula. Anytime Malaysia wants to produce one more bushel of rice, it gives up two units of wheat. So the slope of this line is –2, and I’ll go ahead and write that in. –2 is the slope of Malaysia’s production possibilities frontier.

I can use the same technique now to derive the production possibilities frontier for Pakistan. I know what the end points are. The end points are going to be 30 bushels of wheat if Pakistan uses all of its labor to produce wheat. I got this number for over in the table. Also, if it uses its labor to produce only rice, then we get 20 bushels of rice, so there’s the other end point. You could plot some of the intermediate points and you’d see that they lie on a straight line, or you can simply connect the two dots at the ends and get the production possibilities frontier for Pakistan. So let me draw this line in carefully. There’s the production possibilities frontier for Pakistan. The end points tell us about the maximum production that’s possible if Pakistan completely specializes in one good or the other; that is, the maximum wheat production Pakistan can get is 30 bushels of wheat and the maximum rice production that Pakistan can get is 20 bushels of rice. I could get the same information if I use the formula for Pakistan’s production possibilities frontier. Wheat = 30 bushels, that’s the maximum production, minus \(\frac{3}{2}\) times the amount of rice production. Remember - \(\frac{3}{2}\) or negative \(\frac{1}{2}\), is the opportunity cost in Pakistan of producing another bushel of rice. Anytime Pakistan adds one unit of rice to its output, it does so by giving up one and a half bushels of wheat. So the slope of the curve here is going to be \(-\frac{3}{2}\). That’s the opportunity cost of rice, measured in terms of wheat, that Pakistan would have to give up. Again, notice in Pakistan’s case that we have the production possibilities frontier being a straight line, indicating that all resources being used in Pakistan are equally well suited to the production of rice or wheat.

So here we have our situation. Here are Malaysia’s production possibilities and here are Pakistan’s production possibilities. If now these two countries wanted to produce independently of each other, that is, if we had them separated so that they can’t trade, then we could pick a point on each production possibilities frontier and call that the consumption point in the separated economies. Let’s suppose that Malaysia cannot trade with Pakistan. If Malaysia cannot trade with Pakistan, then all of the consumption possibilities lie on this curve, too. If Malaysia is going to eat some combination of wheat and rice, then it has to make that combination of wheat and rice for itself. And let’s suppose that the mix that they choose is a mix like 20 and 20; that is, Malaysia chooses to produce at this point right here with equal quantities of wheat and rice produced. So here we have 20 bushels of rice produced in Malaysia and 20 bushels of wheat produced in Malaysia. This combination, 20 and 20, is one of Malaysia’s production possibilities. And it’s the production possibility that we are going to imagine that Malaysia chooses if they have to produce on their own.

We can say the same then about Pakistan. Pakistan has this table of production possibilities and if it wants to operate by itself without trade, it’s got to choose some point on this line and that be its consumption. So let’s suppose that Pakistan chooses this point right here, which would be at 15 and 10. I’m going to imagine that Pakistan chooses to produce and consume 10 bushels of rice and to produce and consume 15 bushels of wheat.

So the black dots in each of these diagrams represent the pattern of production and consumption when the two economies operate independently of each other. Now, let’s show how, by cooperating, Pakistan and Malaysia can improve their situation, how, by cooperating, Pakistan and Malaysia can produce more wheat and more rice than they are producing operating independently.

First, let’s notice how much that they are producing together when they’re operating independently. Malaysia’s producing 20 bushels of wheat, Pakistan is producing 15, for a total wheat production of 35 bushels. Malaysia is
Introduction to Economic Thinking

Comparative Advantage

Analyzing an International Trade Using Comparative Advantage

producing 20 bushels of rice, Pakistan 10, for a total rice production of 30 bushels. Let’s see now how, by cooperating, specializing and trading, we can increase the quantity of agricultural goods that the two countries have available. Let’s look at an example.

In Malaysia, the opportunity cost of one bushel of rice is two bushels of wheat. The same amount of labor that will produce these two bushels of wheat would otherwise produce this one-bushel of rice. So the opportunity cost can be represented by this ratio: one bushel of rice always costs two bushels of wheat in Malaysia. In Pakistan, on the other hand, one bushel of rice costs only one and one-half bushel of wheat. Or we could write this a different way: the opportunity costs of two bushels of rice in Pakistan is three bushels of wheat. The same labor that could produce three bushels of wheat would otherwise be producing two bushels of rice. Now, we can show here that there is a possibility of gain from trade. Notice that the opportunity cost of producing two bushels of rice, for instance, is lower in Pakistan than it is in Malaysia. If Malaysia wants to produce two bushels of rice, it’s going to be giving up four bushels of wheat. The opportunity cost in Malaysia is two bushels of wheat for one bushel of rice. In Pakistan, on the other hand, if Pakistan wants to produce two bushels of rice, it’s only giving up three bushels of wheat. The opportunity cost of rice in Pakistan is lower than it is in Malaysia; therefore we know by definition that Pakistan has a comparative advantage in the production of rice, because it can produce rice at a lower opportunity cost than can Malaysia. These are the numbers that are in the production possibilities frontier equation. Malaysia’s slope is \(-2\); that is, two bushels of wheat for every bushel of rice. Pakistan’s slope is \(-\frac{3}{2}\); that is, one and a half bushels of wheat for every bushel of rice.

Now that we know that Pakistan has a comparative advantage in the production of rice, we can tell a story whereby Malaysia and Pakistan cooperate to increase their wealth. Here’s the way the story works. Since Pakistan has a comparative advantage in the production of rice, Malaysia is going to try to find some way to get Pakistan to want to produce rice and send it to Malaysia instead. Malaysia wants to specialize in the direction of wheat; Pakistan would specialize in the direction of rice. Here’s a story that I think will make this comparative advantage comparison clear.

Let’s suppose that Malaysia wants to try to make things good for itself. What it will do then is it will cut back its production of rice by two units. So get these two bushels of rice off of the board and replace it with four bushels of wheat. That’s what Malaysia can produce instead. Now, Pakistan down here is producing three bushels of wheat, but remember Pakistan has a comparative advantage in rice, so anything that allows Pakistan to reduce its wheat production is going to increase wealth for these two economies operating together.

So let’s have Malaysia send these three bushels of wheat down to Pakistan, which then allows Pakistan not to have to produce these three bushels of wheat. Instead, Pakistan can devote its labor to rice production and increase its rice output by two units. So now that Pakistan has someone else producing wheat for it, Pakistan can produce two extra bushels of rice, which it then sends to Malaysia. Notice now we’re back where we started. Pakistan has three bushels of wheat, which now Malaysia is producing and sending to it. Malaysia has two bushels of rice, which it got from Pakistan, but we now have one extra bushel of wheat. This extra bushel of wheat was created by specialization and trade. This bushel of wheat did not exist before. We’re back where we started with one extra bushel of wheat. That’s what happens when two countries specialize in trade.

Let’s suppose now that, since Pakistan has a comparative advantage in the production of rice, we have Pakistan produce all of the rice for the two economies. And let’s suppose that, in addition, that we have Malaysia produce 12 bushels of rice and spend its remaining labor on wheat. With this particular production of specialization and trade, we can now see that both countries can be better off. Let’s look back at our diagram.

Here we have Pakistan, specialized according to its comparative advantage, producing 20 bushels of rice. Now, these 20 bushels of rice are available for trade, so Malaysia can now change its pattern of production and specialize in the direction of wheat. We’re going to have Malaysia cut back its production to 12 bushels of rice and that leaves \(12 \times 2 = 24\), subtracted from 60, is 36 bushels of wheat that Malaysia can produce. So here’s a new point on Malaysia’s production possibilities frontier. If Malaysia produces here at this orange point and Pakistan produces here at this orange point, we can see that they have created more wheat and more rice for their combined economy. Let’s add up the totals.
Analyzing an International Trade Using Comparative Advantage

Pakistan is producing 20 bushels of rice; Malaysia is producing 12 bushels of rice, for a total of 32 bushels of rice. They were only producing 30 bushels before, that’s two extra bushels of rice created by trade. Let’s look at wheat production. Well, Pakistan is not producing any wheat and Malaysia is producing 36 bushels of wheat. 36 is one more bushel than we were producing before, when the economies were operating independently and producing only 35 bushels of wheat. Aha! By specialization according to comparative advantage, we have created an additional two bushels of rice and an additional one-bushel of wheat. These two economies now can agree to trade in such a way that both of them are better off. They can find some ratio of trading rice for wheat that makes them both better off than they were before. Those extra two bushels of rice and extra one bushel of wheat are available to be divided between Malaysia and Pakistan. And the rate of trade that they will agree on is somewhere between the slopes of their production possibilities frontier. Pakistan is quite happy to trade with any terms of trade that give it more than one and a half bushels of wheat per bushel of rice. Malaysia is happy to trade with any terms of trade that give it less than two bushels of wheat per bushel of rice. So the two economies can now agree upon a particular price of wheat, in terms of rice, that will allow them both to be better off. By specializing, they have created more wheat and rice together than they could have created separately. And this is the basis of their gains from trade. Notice that each country produced more of the good in which they had comparative advantage, more of the product for which its opportunity cost was relatively low.
What determines the price at which a good trades in the market? What determines the quantity of the good that's traded? To answer these questions, we're going to need a model of how the market works, and this is one of the economist's favorite models. In the coming lectures, we will be developing this model of the market in stages. We'll begin by discussing the behavior of consumers or households. We call this demand. Then we'll be looking at the behavior of producers – those who make and sell goods. We call this supply. Then we will put demand and supply together to see how they interact to determine prices and quantities in the market.

Let's begin, then, with the discussion with the behavior of consumers and households, the determinants of demand. Let's pick for the sake of a concrete example a product that we can all relate to, bread. And let's consider the factors that influence the quantity of bread that a household is able and willing to purchase in a given period of time. How much bread do they buy in a week? We call this the household's demand for bread. What are the factors, then, that influence the amount of bread that a household is willing and able to purchase in a week? Let's make a list of those factors.

The first factor may be the most obvious. It's the price at which bread trades. This price could be 25¢ a loaf, $1 a loaf, $2 a loaf, $5 a loaf. As the price of bread rises, households are usually willing and able to purchase less bread. When the price of bread falls, households are willing and able to purchase more.

Another factor that influences the quantity of bread that's traded is the price of substitute goods. In my picture I have a bagel which might be regarded as a substitute for bread. Also potatoes might be a substitute for bread, or pasta. Anything that households would buy instead of bread could be considered a substitute. When the price of substitutes rises, people will avoid the substitutes and buy bread instead. When the price of substitutes falls, substitutes look like a good alternative, and people will buy bagels or pasta instead and avoid the bread.

Another factor that influences the quantity of bread that a household purchases is the price of complementary goods. In my picture I have cheese. Cheese is something that people enjoy consuming along with bread. We might also have butter, jam, peanut butter, or any other products that people enjoy consuming along with bread – products that are complementary, products that go with bread. When the price of a complement increases, then we'll find that people are consuming less bread. Here's the logic: If peanut butter gets more expensive and you enjoy peanut butter with bread, say in peanut butter sandwiches, higher prices for peanut butter raise the overall price of your peanut butter sandwich. That means that you want fewer peanut butter sandwiches, and thus less bread. When complements get less expensive, then your peanut butter sandwich becomes a better bargain. You buy more peanut butter and more bread. So the price of complementary goods also influences the amount of bread that a household buys.

Another factor that influences household behavior is the price of input goods. For instance, in my picture here I have two bottles of milk and some eggs. If the price of inputs is low, you may make your own bread. If the price of inputs goes up, then you'd prefer to buy your bread at the store. So the price of input goods can influence the amount of bread that a household chooses to buy.

Another factor that influences the amount of bread that you buy in a week is your income. Income will definitely play a role here. The more income you have, the more bread you can afford; and thus, you may be likely to buy more bread in a given period of time. There are some goods, of course, for which more income would lead you to buy less of those goods, and we'll be discussing those later. But let's say for now that changes in your income are likely to influence your buying behavior – your demand for bread.

Finally, one more factor that we want to take into account is expectations of future prices. If you think that bread prices are going to go up next week, then you may go out and hoard bread today, buying it all off the shelves and putting it in your freezer. If you think bread prices are going to fall next week, you may put off your purchases of bread, choosing instead to wait until bread becomes less expensive. So if you have expectations of future prices, those can influence your behavior today.

Now all of these factors can play a role in determining the quantity of bread that a household is willing and able to purchase in a given period of time. We'd like now to write down a mathematical relationship that describes household behavior, and this mathematical relationship will be called a demand function. The demand function will show us how the quantity of bread purchased is related to all of these other factors. Since we want to do this mathematically, we're
Understanding Markets

Demand

Understanding the Determinants of Demand (1)

going to need some symbols. Unfortunately, we won’t be able to carry the pictures around with us when we’re doing careful, logical reasoning, so we’re going to use some mathematical symbols instead.

The quantity of bread demanded in a given period of time we’ll represent with this symbol: \( q_d \) – quantity demanded. The price at which a loaf of bread sells, we’ll represent with the letter \( p \) – the price of bread. The price of substitute goods we’ll represent with this symbol – \( p_s \). The price of complementary goods we’ll represent with \( p_c \). And the price of input goods will logically be \( p_i \). Income, we’re going to use the letter \( m \) simply because that’s a symbol that economists usually use to represent income; and finally, we’ll use \( ex \) to represent expectations of future prices. If we know all of these variables and if there is a well-established relationship between consumer behavior and these factors, we can form what we call a demand function. A demand function is a mathematical relationship that predicts the quantity of bread demanded as a function of each of these factors that influences consumer behavior.

Here’s how we write down a demand function. The variable that we want to explain is the quantity of bread demanded – the quantity of bread that a household is willing and able to purchase in a week. We’re going to explain this as a function of several other variables. So to represent that we have a function, we’ll use this letter \( d \). It’s going to be called our demand function – I’ll use the letter \( d \) to stand for demand – and the color red because I’m going to use red for demand throughout these lessons. Now I’m going to say that the quantity of bread demanded is equal to this function of several variables. That means that if you know what the variables are, you can say exactly what the quantity of bread demanded is. So I put a parenthesis after \( d \), and then I make a list of all of the factors that influence the quantity of bread demanded, and we know what those factors are. First, the price of bread itself, and then I put a comma to indicate that there’s something else in the list, and that’s going to be the price of substitute goods, like bagels. Then I put another comma to indicate that there’s something else on the list, the price of complementary goods, like cheese or jam. Then comes the price of input goods, like milk and eggs, and then the income – the consumer’s income – and finally expectations of future prices. Then I close my parentheses to indicate that these are all of the variables that I’m going to be considering in this story.

So what we have is a demand function. If you know the values of all of these variables, the function tells you the quantity of bread demanded in a week as a function of the variables. Well, this is a lot to consider, and any of these things could be changing at any time, changing the quantity of bread demanded. In order to build a model of the market and in order to predict how the price of bread is determined by the actions of buyers and sellers, we want to focus on one variable that becomes especially important, and that is the price of bread itself. We’re now going to hold constant all of these other prices. We’re going to make an assumption that economists call ceteris paribus. The assumption ceteris paribus means “all other things being equal.”

We want to look now at how the quantity of bread demanded in a week varies as the price of bread changes, and we’re going to do that by focusing on changes in the price of bread holding constant all of the other factors that influence consumer behavior. We’re going to assume that the price of bagels, the price of cheese, the price of milk and eggs, the consumer’s income, and the expectations of future prices all remain constant. When we do that, we will be able to draw a graph representing the relationship between the price of bread and the quantity of bread demanded, and we call this relationship a demand curve.

So, in the next lesson we will be developing a demand curve, which is the relationship between the price of bread and the quantity of bread demanded in a week, holding constant all of the other factors that influence consumer behavior.
Understanding Markets

Understanding the Determinants of Demand

With this lesson, we begin building the most important model in economics – the model of the market. A market is a place where buyers and sellers trade some good or service, and their interaction will determine the price at which that good trades, and the quantity of the good that's actually traded. In order to build this model, we’re going to break it into pieces. First, we’re going to examine the motivations of buyers. What determines the quantity of a good or service that people want to purchase in a given period of time? This is called “demand.” Then we'll look at the motivation of sellers. What determines the quantity of a good or service that businesses are willing to sell in a given period of time? This is called “supply.” Then we’ll put the buyers and the sellers together, and look at how their interactions establish a stable price and quantity traded. This is called equilibrium. So – demand, supply, and equilibrium.

Let’s start with demand. Demand describes the behavior of households, and answers the question, How much of a particular good or service do households purchase in a given period of time? Let's take some explicit example. Suppose we’re talking about the good “bread” and suppose we want to answer the question: How much bread does a typical household purchase in a week? Well, what we want to do in order to answer this question is make a list of all of the factors that influence the quantity of bread that a household would be willing and able to buy. And probably chief among the factors that influence people’s quantity of bread purchased is the price of bread.

We can also consider the price of other products like cheese or bologna that people enjoy along with bread. We can also consider the price of substitute goods like bagels that provide the same satisfaction in a different way. We can also consider the purchasing power, or income, of the household. Tastes and preferences probably play a role also, as well as expectations of what’s going to happen to the price of bread in the future.

If we want to see how any of these variables influences the quantity of bread that a household purchases in a week, we’ve got to examine it in some orderly fashion. That is, if we want to know how changes in the price of bread influence the quantity of bread demanded, we’ve got to isolate the effect of the price of bread by holding all of the other variables constant.

This “holding constant” is accomplished by an assumption that economists call “cederis paribus,” from the Latin phrase that means “everything else held the same.” So if we want to know how a change in the price of bread influences the behavior of the household, we’ve got to say “cederis paribus” all of the other factors that might influence the quantity demanded. If we hold these other things constant then, what do we expect will happen when the price of bread rises? Likely what will happen is the quantity of bread demanded will fall. This is what economists call the law of demand – that an increase in the price of a good typically leads to a reduction in the quantity of that good demanded.

And, if you think about it, it does kind of make sense. After all, if the price of bread goes up, then people’s income now can buy less bread; so the reduction in their purchasing power is going to leave them able to purchase less. Also, when the price of bread goes up, the opportunity cost of bread, measured against other goods, has increased as well. This is what economists call the substitution effect – people buy less bread because they’re willing to buy less. They’d rather buy other goods that provide the same satisfaction at lower opportunity cost. So the law of demand says that there is an inverse or opposite relationship between the price of bread and the quantity of bread demanded.

What about another factor, like the price of a complementary good? A complementary good is a good that is enjoyed along with the good in question. So a complementary good for bread might be butter or cheese or bologna or peanut butter, or anything else that you enjoy along with bread. See, you’re buying bread not for its own sake, because bread is one element in creating something that you really want, like a peanut butter sandwich. And if the price of peanut butter goes up, or if you’re after a grilled cheese sandwich and the price of cheese goes up, then, overall, the price of the good that you really desire – the sandwich – has risen. And, by the law of demand, you’re going to then want fewer of these sandwiches, or you’ll be able to afford fewer of these sandwiches. And since you want fewer sandwiches, that means less bread. So we have the rule that an increase in the price of a complementary good leads to a reduction in the quantity of this good demanded. A decrease in the price of that complementary good, on the other hand, will lead you to buy more bread, because now cheese sandwiches are a bargain, you’re going to want more of them, and that means buying more bread as well as more cheese.

Let’s consider now the effect of the price of a substitute good, holding constant everything else. A change in the price of a substitute good – cederis paribus – is going to lead to a change in the demand of our original good in the same direction. So if the price of bagels goes up, all of a sudden bagels don’t look like such a good way to get your carbohydrate satisfaction, and you switch to bread instead. On the other hand, if the price of the substitute good falls –
Understanding Markets

Understanding the Determinants of Demand

If bagels are now less expensive – go buy bagels instead, and the quantity of bread demanded will be reduced. So ceders paribus – an increase in the price of a substitute good increases the quantity demanded of the original good; and a decrease in the price of the substitute good causes a decrease in the quantity of the original good demanded.

What about income? And here things are a little bit more complicated, because an increase in your income can have either effect, really, on the amount of a good you purchase, depending on what kind of good it is. Most goods – what we call “normal goods” – see that when your income increases, you’ll actually purchase a larger quantity; your purchasing power is increased, you feel wealthier, and you buy more bread.

On the other hand, there’s a certain category of goods called “inferior goods” for which the opposite is true. When your income goes up, you actually buy less of these goods – some goods like public transportation, potted meat products – things that people can’t wait to have enough money so that they can buy something else instead. For example, beans – poor people tend to spend more money on beans than people at higher income levels, and when your income level rises, you may typically buy less beans and more meat instead. In that case, beans are acting as an inferior good; an increase in your income leads to a reduction in the quantity of beans demanded. So the effect of income – ceders paribus – depends on whether the good is a normal good or an inferior good. A normal good, an increase in income leads to an increase in the quantity of the good demanded; an inferior good has exactly the opposite effect – an increase in income leads to a decrease in the quantity of the good demanded.

What about tastes and preferences? And here the relationship is pretty straightforward. We don’t really have a number to put on tastes and preferences, so we describe this qualitatively. We say that if people just suddenly decide they like bread better than other things, they’ll buy more bread – ceders paribus. On the other hand, if people decide that they don’t like bread, or they’re concerned about the effect of bread on their health, or something like that, than tastes and preferences shift away from bread, and people will buy less bread – ceders paribus.

Finally, there’s the effect of the expectation of future prices. What are households going to do if they think the price of bread is going to be rising in the future? Well, ceders paribus – the expectation of a higher future price for bread leads people to buy more bread now. Either they’re going to stock up, or they’re going to enjoy bread before it moves out of their price range. On the other hand, if people expect that the price of bread is going to fall in the future, they’re going to reduce the quantity of bread that they purchase today, and hang back and wait for the lower price.

So we have these factors, then, that influence people’s purchasing behavior. And we can now collect these factors in a mathematical way, and build what we call a demand function. The demand function describes the quantity of bread that the household is willing and able to purchase in a given period of time as a function of all of these variables that influence that decision. It will look something like this: The quantity of bread demanded is the variable that we want to explain. So we’re going to say that that is determined by, or equal to, a function of all of those variables that we just listed. I’m going to use here the letter “D” to represent the function, kind of like sometimes we say “F” of “X” is a function. This time it’s going to be ”D” – or demand – as a function of the variables that influence people’s buying behavior. So here’s the price of the good in question, and remember ceders paribus – a higher price means a smaller quantity demanded; a lower price means an increase in quantity demanded.

The next variable that influences this decision is the price of complementary goods, and remember the higher price of complements – ceders paribus – the lower the quantity demanded. And the lower the price of complements, the higher the quantity demanded.

The next factor is the price of substitute goods. Remember the price of substitute goods moves in the same direction as the quantity of the good demanded. So a higher price for substitutes – ceders paribus – leads households to buy more of the good in question. A lower price of substitutes leads households to buy less of the good.

The next factor is income, and the effect of income depends on whether the good is a normal good or an inferior good. For a normal good, more income leads to more quantity demanded – ceders paribus. For an inferior, more income leads to a lower quantity demanded – ceders paribus.

The other variables that influence people’s behavior are tastes, and preferences, and the expectation of what the price is going to be in the future. At this point I can put a parenthesis on the other side, and indicate that I’ve collected all of the variables that I’m going to be concerned about.
Understanding Markets

Demand

Understanding the Determinants of Demand

There may be other things that influence people’s behavior, but we can probably lump those things under tastes and preferences – the little idiosyncratic things that may affect one household but not affect another. So here we’ve got a range of variables – a collection of factors that influence people’s willingness and ability to purchase bread.

So what do we do next? Well, what we’re going to do next is we’re going to focus our attention on the variable that’s most important for our building the model of the market. Remember the market is going to explain the quantity of the good traded and the price at which those trades occur. So we want to focus our attention now on how the price of the good itself influences the quantity of the good that people are willing and able to buy. And to focus our attention on the effect of the price, we’re going to hold constant – and that’s what I mean when I draw this horizontal bar above the name of the variable – we’re going to hold constant everything else that influences people’s behavior – the price of complements, the price of substitutes, income, tastes, and expectations. *Ceteris paribus* all of these variables. We’re now going to focus our attention on the way in which changes in the price of bread influence the quantity of bread demanded. And this is how we begin to construct what economists call “the demand curve.”
In the last lesson, we introduced the concept of the demand function. The demand function shows the relationship between the quantity of bread that a household is willing and able to purchase in a week and all of the factors that influence household behavior. We showed that the quantity of bread demanded depends on the price of bread, the price of substitute goods, the price of complement goods, the price of inputs, household income, and the expectations of future prices. In this lesson we want to focus our attention on a particular variable, and that is the price of bread itself and how the price of bread influences the quantity of bread demanded.

In order to focus on the relationship between the price of bread and the quantity of bread demanded, we’ll want to hold constant all of the other variables that influence household behavior. There’s an assumption that means “holding all things constant.” Do you remember the name of that assumption? The assumption is called ceteris paribus, and in this lesson we’ll be deriving a demand curve. The demand curve shows the relationship between the price of bread and the quantity of bread demanded. Ceteris paribus, or holding constant all other factors that influence consumer behavior.

Let’s begin, then, by looking at a demand schedule. A demand schedule shows a relationship between price and quantity demanded, holding constant all of the other variables that influence household behavior. Let’s start with a list of bread prices. Here are a list of bread prices: $5 per loaf, $4 per loaf, $3 per loaf, and so forth, on down to 50¢ per loaf. Now for each one of these prices, holding constant everything else, we’ll find that household behavior changes and a different quantity of bread is purchased with each different price for bread.

Let’s look now at a table that shows the quantities associated with each of these prices. Now remember, behind this table of numbers there is the assumption that we are holding constant bagels at a particular price, we’re holding constant cheese at a particular price, we’re holding the price of milk and eggs constant, we’re holding consumer income constant, we’re holding expectations constant. The only things we are allowing to vary are the price of bread and the quantity of bread that the household will buy.

So when the price of bread is $5 per loaf, this particular household will buy one loaf of bread per week. If the price falls to $4 per loaf, holding constant everything else, the household will now increase the quantity demanded to two loaves of bread per week. As the price falls to $3, this household will increase its quantity demanded to two loaves of bread per week. If the price falls to $2, this household will increase its quantity demanded to two loaves of bread per week. If the price falls to $1, this household will increase its quantity demanded to two loaves of bread per week. This is called a demand schedule – a relationship between the price of bread and the quantity of bread that the household will purchase.

In order to turn this table of numbers into a powerful, logical tool, we’re going to introduce a graph with price and quantity and turn these numbers into a demand curve, the demand curve we’ll later use to describe how price and quantity are set in the market when buyers and sellers are interacting.
Understanding Markets

Understanding the Basics of Demand

Seven loaves if the price is $1.50 per loaf, so that’s going to be this point here. And we get eight loaves if the price is $1 per loaf, so let’s see if I can find that point. And nine loaves if the price is 50¢ per loaf, so that’s this point here. What I did was take each combination of price and quantity from my table and represent it as a red dot in this diagram. Red is going to be my official color for demand, so I’ve represented all of these combinations with red dots.

Here’s a question: Are these the only imaginable quantities of bread? Are these the only imaginable prices? Absolutely not. We can imagine first of all that the price can vary continuously from zero – that is, free bread – all the way up to $10 per loaf and anything in between…any price imaginable. The price could be $1.01. It could be $1.05. It could be $1.74. All of the prices on the vertical axis are possible. Also, all the quantities are possible. Not just discrete quantities like one loaf of bread per week or five loaves of bread per week, but we could also imagine two and one-half loaves of bread per week. What would that mean? That would mean that you’d buy five loaves of bread every two weeks so that on average each week you’re buying two and one-half loaves. So we can imagine that bread could be divided up into very, very small fractions, and the price also varies continuously. If that’s the case, we can connect the dots and represent the consumer’s behavior with a smooth curve that shows a continuous relationship between price and quantity. So let me connect those dots. This gives us what we call in economics a demand curve. I’m going to label this with a d to make clear that I know that it’s a demand curve. Here’s the definition of a demand curve: The demand curve shows the relationship between the price of bread and the quantity that the consumer is willing and able to purchase in a given period of time holding constant all other factors that influence consumer behavior, such as the price of substitutes, the price of complements, income, expectations, and so forth.

The first thing you want to notice about the demand curve is the demand curve is downward sloping. Now what does that tell us? What does it tell us that economists almost always draw the demand curve downward sloping? Well, the downward-sloping demand curve is a way of representing that if the price of the good rises, the quantity demanded falls. Economists call this relationship the Law of Demand. The Law of Demand says that when the price of a product increases, consumers are willing and able to purchase less of it. Now it’s not true necessarily that this is always the case. Some goods, people will actually buy more of them as they get more expensive. But in general we observe that when the price of goods and services increase, people are willing and able, and in fact do, buy less of them.

Does this make sense? Well, in fact, it does make sense for two reasons. The first reason is what we call the substitution effect. When the price of a good rises, consumers, seeking to get the most satisfaction from their limited income, will stop buying this more expensive good and switch to other substitutes instead. For example, when bread gets more expensive, consumers try to get their carbohydrate satisfaction by buying bagels or potatoes instead. The higher price for bread leads consumers to substitute away from bread and seek less expensive substitutes.

The second reason behind the Law of Demand is what we call the income effect. When the price of bread rises, then a consumer is able to purchase less. That is, if bread goes from $1 per loaf to $3 per loaf to $6 per loaf, meanwhile I’m living on a limited income that’s not changing – a constant income – now my purchasing power – the purchasing power of my income measured in bread – is shrinking. If I have $10 to spend on groceries each week and the price of bread is $1 per loaf, I can afford ten loaves of bread; but if the price goes up to $5 per loaf, I can only afford to purchase two loaves of bread. This is the income effect. I buy less bread because my purchasing power is shrinking measured in terms of bread. So the substitution effect and the income effect are two rationales behind the Law of Demand. When the price of a good rises, we expect people to purchase less of it. Because they are willing to purchase less, they’re going to look for less expensive substitutes, and they are able to purchase less because the purchasing power of their income is shrinking.

In the next lesson, I’m going to look at what happens to the demand curve when we allow those other factors, those factors that we held constant, to change. Suppose now we allowed the price of bagels to change or consumer income to change, how would that influence the relationship between the price of bread and they quantity that this particular household demands in a week?
Analyzing Shifts in the Demand Curve

In the previous lesson, we looked at the demand curve, the relationship between the price of bread and the quantity of bread that a household is willing and able to purchase in a week. When you draw a demand curve, you hold constant all of the other factors that influence household behavior, factors like the price of substitutes, the price of complements, household income, expectations and so forth. You might ask yourself, “How do we represent a change in one of those other factors?” After all, it’s very easy to show how the household responds to a change in the price of bread, you just move up or down the demand curve. But what if the price of bagels changes, or what if consumer income changes? In that case, it turns out that we get an entirely new relationship between the price of bread and its quantity demanded. And once we have a new relationship, a new demand schedule, we have to draw a new demand curve. In this lesson, we will show how the change in one of the factors that we usually hold constant will cause us to have to shift the demand curve; that is, to redraw it in a new position.

Let’s go back to the original demand schedule we used in our last story. Here’s a particular household and this household has behavior that’s represented by these numbers. At a price of $5.00 per loaf, they demand one loaf of bread per week, $4.00 a loaf means two loaves of bread per week for them and so forth, these numbers representing the quantity of bread that they are willing and able to purchase at these different prices for bread.

But suppose now we make a change in one of the factors that we held constant when we drew this schedule; that is, suppose now we change household income and make this household richer. Suppose their household income goes from $500.00 per week to $700.00 per week. The household is now wealthier and one of the things that this household might choose to do with its newfound income is purchase more bread. This would be the case if bread were what economists call a normal good. A normal good is a good that a household buys more of when the household income increases. Let’s suppose that bread is a normal good for this household and, in that case, we will get a new higher quantity of bread at every price after income is increased. I’ll write these numbers in and, to show that this is a new demand schedule, I’ll use the symbol prime, or apostrophe, by the $D$ to represent that this is a new demand relationship. Prime usually means new. We’ve got something new, so we’re indicating it with this little mark.

Well, here’s our new demand schedule after the increase in household income. Now, at a price of $5.00 per loaf, this household wants two loaves of bread per week instead of one. At a price of $3.00, this household used to want three loaves of bread, now they want four, and so forth, all the way down the schedule. You will see that, after an increase in income, this household wants to buy a larger quantity of bread at every price. And that’s going to be important language for us as we look at how changes in variables influence the demand curve.

So, let’s take this demand schedule and move it over to the board and draw the new demand curve in the axes that we used to draw the old demand curve. So what I’ll do now is move all this stuff over to the board, so you’ll have it for reference. And I’ll go back to the diagram, in which we graphed the demand curve.

After the increase in income, we’ll see that there’s a change then in the demand curve. For example, now at a price of $5.00 per loaf, the household wants two loaves of bread per week instead of one. At a price of $4.00 per loaf, we have a new quantity of three loaves demanded per week instead of two. At a price of $3.00 per loaf, the quantity demanded after the increase in income is going to be $4.00 per loaf. Notice the quantity demanded is larger at every price after income increases. This household now wants five loaves of bread when the price of bread is $2.50 per loaf. This household wants seven loaves of bread when the price is $2.10. At a price of $1.80, the household wants eight loaves of bread. At a price of $1.50, the household wants nine loaves of bread. At a price of $1.00 per loaf, the household wants 10 loaves of bread, and finally the household wants 12 loaves of bread if the price of bread is down to 0.50 per loaf. Now, are these the only imaginable prices and quantities for this household? Absolutely not. We can imagine price changing continuously, we can imagine quantity demanded changing continuously, and we can connect these dots to get a continuous downward sloping demand curve, just like before.

Here’s the demand curve that represents the relationship between price and quantity demanded after income is increased. Notice that the new demand curve lies further away from the axes. We might use the language that the demand curve has shifted outwards; that is, after the increase in income, since bread is a normal good, the household wants to buy more bread at every price than it was buying before. The outward shift in the demand curve came about because we changed one of the variables that we had previously held constant. When we drew the old demand curve, we allowed price to change to see how the household behavior would change, how quantity would change. But we held income and the price of other goods constant. Whenever we changed income, we get an entirely new relationship between price and quantity. At every price, this wealthier household now wants to buy a larger quantity of
Analyzing Shifts in the Demand Curve

bread than they were buying before. The increase in consumer income leads the demand curve to shift outward. It causes an outward shift in the demand curve; that is, the household buys a larger quantity of bread at every price.

Now, here’s some important language for you to know. Whenever there’s a change in the price of bread, the household changes its behavior by moving along the existing demand curve. When a household moves along an existing demand curve; that is, when the household responds to a change in the price of bread, we call this a change in quantity demanded. However, if we change one of the variables that we previously held constant, income, the price of other goods, expectations, this causes us to have to redraw the entire demand curve. It leads to an entirely new relationship between the price of bread and the quantity demanded. We call this shift in the demand curve a change in demand. The entire demand schedule has changed. We have a whole new relationship between price and quantity, and that’s what’s meant by a shift in the demand curve or the language a change in demand.

So, a movement along the demand curve is a response to a change in the price of bread itself, and that’s called a change in quantity demanded. A shift in the demand curve is caused by a change in one of the variables that we usually hold constant when we draw a demand curve, and this shift is called a change in demand. In the next lesson, we’ll do a complete inventory of all the things that can shift the demand curve, and we’ll discuss how those various factors that we usually hold constant influence consumer behavior.
In this lesson, we’re going to do a summary of all of the factors that influence consumer demand, and we’re going to show how each of those factors appears in the demand curve diagram.

Let’s start with the easiest case. How do we show a change in consumer behavior in response to a change in the price of bread itself? This is easy. If the price of bread changes, the consumer moves along the existing demand curve. For instance, if the price of bread goes up, the consumer will be buying less bread than before. We see a decrease in the quantity of bread demanded. If the price of bread, on the other hand, falls, we’ll see the consumer buying a larger quantity. Here we have the law of demand at work. The consumer is willing and able to buy more bread when the price is lower, and that’s a movement down the demand curve. A lower price for bread corresponds to a larger quantity demanded. Anytime you change the price of bread itself, you are moving along the demand curve. The demand curve never shifts in response to a change in the price of bread. You can never shift a curve when you’re changing one of the things that’s measured on the axes. You just move along the curve.

Suppose now we change another one of the factors that influences consumer demand, one of the factors that we hold constant when we draw the original demand curve. Let’s consider first the case of the price of substitute goods. What happens if the price of bagels increases? Well, let’s think about the household’s decision. With bagels now being less attractive, this household will be more inclined to get their carbohydrates from bread instead. Now, at any given price of bread, the household is going to want to buy more bread than before and fewer bagels. Bagels are relatively unattractive, bread has become more attractive. So the household wants to buy more bread now at every price, and this shifts the demand curve outward. It changes the relationship between the price of bread and the quantity demanded and it gives us this entirely new demand curve. With the price of substitutes going up, bread is more attractive and consumers want a larger quantity of bread at every price.

Suppose now that the price of bagels falls. If the price of bagels falls, bagels are now relatively attractive. Consumers get more of their carbohydrates from bagels instead and the quantity of bread demanded at every price shrinks. We represent this change by shifting the demand curve inwards. Now, at every price for bread, consumers demand a smaller quantity of bread than before. The demand curve shifts inwards. We call this a decrease in demand.

So, in summary, when the price of substitute goods goes up, there is an increase in demand, a larger quantity demanded at every price. If the price of substitute goods falls, we have a decrease in demand, a smaller quantity demanded at every price.

Let’s consider another factor that influences consumer behavior, the price of complementary goods. Suppose the price of cheese, a complementary good, falls. In that case, consumers say, “Well, we like to eat bread with cheese. The price of a cheese sandwich has just decreased. That means we want more cheese sandwiches and that means buying more bread.” When the price of a complement good falls, consumers are going to want more of the original good at every price. Bread is now more attractive and, after the price of the complement falls; households will want to buy more bread at every price than they were buying before. So here’s the new demand curve. A decrease in the price of a complement good shifts the demand curve outward. A larger quantity of bread purchased at every price.

If, on the other hand, the price of cheese should increase, the demand curve would shift inwards. Now with cheese more expensive, consumers say, “We don’t want cheese sandwiches, we’ll eat soup for lunch instead. And if we’re not buying cheese for cheese sandwiches, then we’re not going to buy the bread to go with it,” and that means a smaller quantity of bread purchased at every price, an inward shift of the demand curve. When the price of complementary goods increases, the demand curve shifts inward, a smaller quantity of bread demanded at every price.

Here’s another thing that can change: consumer income. Now, I’ve saved this case for later in this lesson, because it’s relatively subtle. There are two different cases that we want to consider with income. The first is the case of what we call a normal good. A normal good is defined as any good for which the demand increases when income increases. This would be the case with luxury goods, like jewelry or vacations. When you get wealthier, when your income increases, you’re inclined to want to buy more of these goods at every price. So let’s consider first the case of normal goods. With normal goods when your income increases, you want to buy more of the good at every price. This would be the case for the demand curve for jewelry or vacations. On the other hand, if we have a normal good and your income decreases, you’re going to buy less of that good at every price. The demand curve shifts inwards, because now, with less purchasing power, you will spend less on those goods than before. So for a normal good,
Understanding Markets

Demand

Changing Other Demand Variables

higher income means an increase in demand and a smaller income means a reduction in demand. The demand curve shifts outwards when income rises and inwards when it falls, if the good is a normal good.

Now, the other case with income is what we call the case of inferior goods. Inferior goods are those odd goods, for which your demand actually decreases when your income increases. One example of inferior goods that's often given is the case of beans for poor families. Poor families eat beans probably because they can't afford higher protein alternatives, like meat. If their income increases, we might see that these families buy less beans and more meat instead. So here's an example of an inferior good. As income increases, the demand for that good actually decreases. A family that gets a higher income is inclined to buy less beans and more meat. So we have the demand curve shifting inward, representing a smaller quantity of beans demanded at every price when the income increases. However, if the good is inferior and your income decreases, then you're inclined to actually buy more of the good. When you get poorer, you're inclined to buy more beans at every price and less meat instead. So with an inferior good, you get exactly the opposite relationship between demand and income than you get with a normal good. With a normal good, higher income shifts the demand curve outwards, lower income shifts it inwards. With an inferior good, higher income shifts the demand curve inwards and lower income shifts the demand curve outwards.

Finally, one more thing we can consider is the expectation of future prices. If you think that the price of bread is going to be rising in the future, then you may go out and buy more bread today, stocking up, making sure your freezer is full before the price increase. A higher expected price in the future leads households to buy more bread today at every price. However, if you think the price of bread is going to go down in the future, you may delay your purchases and wait for a bargain. In that case, your demand curve actually shifts inwards; that is, at any given price for bread, you now want a smaller quantity than before. You're saving up, waiting for the day when the price actually falls. So higher expected prices in the future lead to an increase in demand. Lower expected prices in the future lead to a decrease in current demand.

So there you have it, a catalog of things that can shift the demand curve. All of those factors that we hold constant when we draw the demand curve can change. And if they change, we have to redraw the demand curve. We get an entirely new relationship between price and quantity, and we represent that entirely new relationship by a shift in the demand curve.

Here's the important language for you to remember. I've said it before, but I'll summarize it, because it helps you to be very clear about these important concepts. When you have a change in the variable that's measured on the axis, a change in the price of bread itself, you represent the change in the price of bread as a movement along the demand curve. We talk about a change in the price of bread, a movement along the demand curve, and a change in quantity demanded. They all mean the same thing.

If you change one of the factors, on the other hand, that we usually hold constant, we represent that change as a shift in the demand curve. We talk about an increase in income for a normal good, an increase in the price of substitutes or a decrease in the price of complements, or an increase in expected future prices, we talk about these things leading to an increase in demand; that is, an outward shift in the demand curve. Shifting the curve is the same thing as a change in demand, an increase in the quantity demanded at every price.

So all of these phrases that I've used, all of these expressions are important for you to memorize so that you can use them carefully when you're trying to describe the response of consumers to changes in their environment.

In the next lesson, we're going to take this logic that we've developed for a particular household and turn it into a market demand curve by adding up the demand curves of the individual households to get what we call a market demand curve.
Understanding Markets

Demand

Deriving a Market Demand Curve

In this lesson we’ll move from the individual _______ [tape skips] supply curve to the supply curve for the market. And we’ll use the same procedure we used when we derived the market demand curve. We’ll be adding up the supply quantities of the individual firms at any given price to find the total market quantity supply.

Let’s look first at a particular bakery. And now we’ll call this Jim’s Bakery. Jim is willing and able to supply one loaf of bread per week when the price of bread is 40 cents per loaf. If the price is 60 cents per loaf, Jim is willing and able to supply two loaves per week, and so forth. The information in this table represents the quantity of bread that Jim is willing and able to supply at different prices, holding constant the price of inputs, technology, and expectations about future prices.

Now let’s look at the supply behavior of another firm. This is Dan’s firm. Dan is willing and able to supply two loaves of bread per week when the price of bread is 40 cents per loaf. He’s willing and able to supply three loaves of bread at a price of 60 cents per loaf, and so forth. As the price of bread increases, the quantity supplied from Dan’s bakery increases. Suppose that Dan and Jim have the only bakeries in town. What does the market supply curve for bread look like? The market supply curve for bread will be the sum of the individual supply curves of Dan’s firm and Jim’s firm.

At any given price, to find the market quantity supplied, simply add the quantity supplied by Dan’s firm to the quantity supplied by Jim’s firm. For instance, at a price of $1.50 per loaf, Dan’s firm supplies five loaves of bread per week. Jim’s firm supplies four loaves of bread. The market quantity supplied in that case will be five plus four or nine loaves of bread each week. You can do the same for any given price. Find the quantity that Dan’s firm is willing and able to sell, add it to the quantity that Jim’s firm is willing and able to sell at the same price, and that gives you the market quantity supply. Now, we can put Dan and Jim’s information together and wind up with the market supply schedule for bread at any given price. This table tells you the quantity supplied by Dan plus Jim’s bakery.

The next step in this lecture is to move this information over to the board and represent it in diagrams. Here is the supply curve for Dan’s firm. I took the numbers from Dan’s table and I put dots in this diagram to represent the price/quantity combinations that describe the relationship for Dan’s bakery. At a price of 40 cents per loaf we get two loaves a week from Dan. At a price of 60 cents per loaf we get three loaves, and so forth. Now we know that these are not the only price/quantity combinations. Price and quantity can both vary continuously. So I can connect the dots and get Dan’s supply curve for bread. I’ll call that Sd. The same thing with Jim’s information. I’ll move it from the table into the diagram by plotting the combinations of price and quantity that describe the behavior of Jim’s bakery. And if I connect the dots, I get Jim’s supply curve for bread.

If I want the market supply curve I have to add together these two individual supply curves. And just like I did before, I’ll use horizontal summation. I will pick a particular price, take the quantity that I get from Dan, add it to the quantity that I get from Jim, and that will give me the market quantity supply at that price. Let’s start by choosing the price of 40 cents per loaf. At a price of 40 cents per loaf we get two loaves from Dan and one loaf from Jim. Add Dan’s two loaves to Jim’s one loaf, and that gives us a total market quantity supplied of three loaves at a price of 40 cents per loaf. So I put a dot here in the market supply curve with a quantity of three and a price of 40 cents.

That’s the same thing if I took a ruler and I added two loaves from Dan’s bakery to one loaf from Jim’s bakery, and that gave me a total of three loaves, and I graphed it right here in this diagram. Three loaves in the market at a price of 40 cents per loaf. I can do the same for any price on the axis. Let’s say a price of 60 cents per loaf. In that case, I get three loaves from Dan, add that to two loaves from Jim, for a total of five loaves supplied when the price is 60 cents per loaf. Keep picking any price you want and add the quantity from Dan’s firm to the quantity from Jim’s firm to get a total quantity supplied in the market.

Alternatively, you can just go back to the table representing quantity supplied—that table that we got when we added Dan’s output to Jim’s output and you can graph the numbers directly. At a price of $1.00 per loaf, we get a quantity supplied of seven. At a price of $1.50 per loaf, we get a quantity supplied of nine. At a price of $2.10 a loaf, we get a total of 11 loaves supplied. At a price of $2.50 per loaf, the quantity supplied is up to 13 loaves per week. There’s the market supply curve. I can connect the dots and label this with an S to indicate that this is the supply curve for the whole market. The supply curve is the horizontal summation of the supply curves of the individual firms in the market. To get this curve, pick a price and add the quantity from Dan’s bakery to the quantity from Jim’s bakery at that price to get the quantity for the market as a whole.
Understanding Markets

Demand

**Deriving a Market Demand Curve**

Now, what’s going to shift the market supply curve? The answer is anything that shifts the individual supply curves will shift the market supply curve. After all, the market supply curve is simply the sum of the individual curves, so anything that moves Dan’s curve or Jim’s curve will shift the market supply curve as well. And what are those factors? They’re all of the factors that held constant when we originally drew the individual firm supply curves. The price of inputs, technology, and expectations about future prices. If any of those factors changes, we have to redraw the individual firm’s curves and consequently we have to shift the market supply curve as well.

Now, this wraps up our introduction to the behavior of suppliers and the behavior of buyers. Now we’re ready to put the two sides of the market together and find the conditions under which we have a stable situation. In the next lecture we’ll be putting supply and demand together in one diagram to find the equilibrium price and quantity traded.
Understanding Markets

Supply

Understanding the Determinants of Supply

In the last lectures, we've been considering the consumer side of the economy, those factors that influence the demand for particular goods and services. Now we'll shift our attention to the supply side of the economy. We'll look at the determinants of the supply of a particular good or service. Let's think now about the agents we're studying, people who offer bread for sale. These are people, like bakers and other stores that are trying to make a profit by selling bread.

Profit is the difference between the revenue that a company earns and its costs of providing its good or service. So any factors that influence profit; that is, any factors that influence revenue or costs, are going to have an effect on the supply decisions of firms. Let's look then at some of those factors, because they turn out to be the determinants of supply.

Think about the market for bread. What are the determinants then of the profitability of a company that might want to sell bread? What are the factors that are going to influence how much bread they are profitably able to offer for sale? Well, one important component of their decision will be the price at which they can offer bread for sale. If the price of bread is high, the prospects for profits are good and it's more like that there will be lots of companies offering lots of bread for sale. If prices of bread are low, on the other hand, it's unlikely that a lot of bakers will be able to make a profit and the quantity of bread offered for sale will be lower. Well, that works on the revenue side of things, what about on the cost side of things?

One important factor will be the cost of inputs, like milk and eggs, things that are used to make bread. So the input prices are important. The higher the input prices, the higher the company's costs and the less profitable it will be for companies to offer bread for sale. When input prices are lower, it's going to be more like that a company can make a profit selling bread and the amount of bread offered for sale is likely to be higher. However, input prices are just one part of the cost side of a firm. Another side of this equation is going to be technology. How is this firm going to be able to combine inputs to produce bread? If you have good technology, you can get by with very little milk and eggs to make your bread. If you have bad technology or are wasteful or unproductive, you're going to have to buy a lot more inputs, and that raises your costs of production. So technology is important; the better technology is for producing bread, the lower the costs of the company and the more like it is that that company will offer lots of bread for sale. If technology is not as good, costs will be higher and the bread supplied will probably be a smaller quantity.

One more thing that influences the supply decision will be expectations. Just like in the case of households, producers are looking forward to the future, trying to figure out whether the prices of bread will go up or down, what will happen to the prices of their inputs. And as they change their expectations, they may be changing the timing of production. If they think bread prices are going to rise next week, they may save bread on their shelves this week and try to sell it later. They may freeze bread in anticipation of being able to charge a higher price in the future.

So we have then four factors that influence the supply decisions of firms: the price of bread, the price of inputs, technology for production and expectations of future prices. Now, of course, I'm going to want to turn these into a supply function, so I need mathematical symbols for all of these ideas. We'll let the quantity of bread supplied be represented by $Q_S$, and I'm using the color blue here, because blue will be my official color for supply. The price of bread we'll represent, as before, with a $P$, the price of inputs with $P_I$. The technology of the firm we'll represent with the letter $T$ and the expectations of future prices with the letter $EX$.

Now, we're ready to write down a supply function. A supply function shows the amount of bread that a firm is willing and able to supply profitably as a function of all of the variables that influence its decision, the price of bread, the price of milk and eggs, the technology of production and the expectation of future prices. So let me take my icons off the pad and I'll turn these symbols into a function. We want to now describe the quantity of bread supplied as a function and I'm going to use, for my function, the letter $S$, my blue $S$, to be supply, so I'm going to have a function called the supply function. And the quantity of bread supplied depends on all of these variables: the price that we can charge for a loaf of bread, the price that we have to pay for milk and eggs and other inputs, the technology; that is, how effectively we can turn inputs into output, and finally, the expectation of future prices.

There you have it, the supply function for a particular firm. If you know the value of prices, input prices, the state of technology and the expectations for the future, you'll be able to say how much bread a particular firm will offer for sale. And there you have it, the quantity of bread supplied, expressed as a function of all of the variables that influence the
Understanding Markets

Supply

**Understanding the Determinants of Supply**

supplier's behavior. Now what we'll do is look at a supply schedule, the behavior of a particular firm in response to different possible prices for a loaf of bread. Then we'll take the supply schedule and turn it into a supply curve.
Understanding Markets

Supply

Deriving a Supply Curve

In the last lecture, we made a list of those factors that influence the behavior or suppliers. In this lecture, we’re going to look at a particular factor, that is, the price of bread, and how the behavior of suppliers is related to the price that they can charge for this product. Remember that we said that the quantity of bread that was supplied was influenced by the price of bread, the technology of production, the price of inputs and the producer’s expectations about future prices. What we’ll do now is look at the relationship between the quantity of bread supplied and the price of bread, holding constant all of these other factors. Once again, we’ll be invoking the assumption ceteris paribus, to look at how the quantity of bread supplied in the market is influenced by changes in the market price of bread, holding constant expectations, technology and the price of inputs. This will give us a supply schedule, and we’ll begin by looking at the supply schedule for an individual firm.

Suppose we have a baker, whose behavior can be summarized in this table; that is, the quantity of bread that our baker supplies is related to the price of bread. When I draw the supply schedule, I am invoking ceteris paribus. I am holding constant everything else that this baker cares about, technology, input prices, and expectations about the future. I’m looking only at how his behavior would change in response to a change in the market price of bread.

So, for instance, our particular baker here will supply one loaf of bread per week; that is, he is willing and able to offer one loaf of bread for sale is week when the price of bread is 0.40 per loaf. If the price of bread goes up to 0.60 per loaf, he is willing and able to offer two loaves of bread for sale a week, and so forth. At a price of $1.00, the quantity supplied is three loaves of bread. At a price of $1.50, the quantity supplied is four loaves, and so forth, all the way to the end of our chart. Again, this supply schedule represents a relationship between the market price of bread and the quantity of bread that our seller is willing and able to offer, holding constant all of the other factors that influence the supplier’s behavior.

The next step in this analysis is to take the information from this supply schedule and convert it into a supply curve; that is, a graph that represents the relationship between price and quantity supplied for our particular baker. So that’s where we’ll go next. I’ll move this information over onto the board and I’ll put up axes here that I can use to graph the relationship between price and quantity for this baker.

So on the horizontal axis, we will be measuring the quantity of bread; that is, supply, how much bread our seller is willing and able to offer for sale. And on the vertical axis, we will represent the price of bread, which again will be measured in dollars. Now, I’m going to take the information from the table and represent it in the graph. So at a price of 0.40 per loaf, we get a quantity of bread supplied of one loaf per week. So here’s a blue dot that represents that combination, 0.40 from the vertical axis and one loaf of bread from the horizontal axis. We get two loaves of bread when the price goes up to 0.60 per week, so let me go on into my diagram. Two loaves of bread at a price of 0.60 give me a point like this one. I get three loaves when the price is up to $1.00. And this particular baker will offer four loaves for sale if the price of bread rises to $1.50. We get another loaf of bread at the price of $2.10 per week; so let me put that point on the diagram. And we get six loaves of bread when the price is all the way up to $2.50 per week. In each case, I’m measuring price in the vertical axis and quantity on the horizontal axis. I get seven loaves of bread at a price of $3.00 per loaf. At a price of $4.00 per loaf, this particular baker is willing to offer eight loaves of bread for sale each week, so that gives me this point. And at a price of $5.00, this baker offers nine loaves of bread for sale each week. So that gives me this point.

Now, are these the only imaginable quantities of bread? No, you can imagine fractional quantities. The baker could offer two loaves for sale over the course of three weeks, which would average out to be \( \frac{2}{3} \) of a loaf a week. Or he could offer three loaves for sale every two weeks, which would average out to be \( \frac{1}{2} \) loaves per week. So it’s possible that there would be fractional loaves of bread sold. He could even bake a half-loaf, if he was so inclined. And it’s also possible to imagine prices varying continually from zero all the way up to $6.00, $7.00, $10.00. So the continuous variation of prices and the continuous divisibility of quantities of bread allows us to connect these dots and make a smooth supply curve, representing this baker’s behavior. So let me connect the dots, and when I connect the dots, I have created a supply curve.
Understanding Markets

Supply

Deriving a Supply Curve

The supply curve is a collection of points, representing the quantity of bread that a baker is willing and able to offer for sale in a given period of time as a function of the price of bread per loaf. So it’s a relationship between price and quantity supplied for this particular baker. I’m going to label this curve with an S to indicate that it’s the supply curve.

Now, there is something you need to notice immediately about the supply curve, and that is that the supply curve slopes upwards. This is a representation of what economists call the law of supply. The law of supply is a regularity that we notice in the economy; that is, as the price of a good or service increases, the quantity that is offered for sale generally increases. Now, does this make sense? What’s the intuition behind the law of supply? The intuition behind the law of supply is a concept that we’ve already considered in this series, and that’s the concept of increasing opportunity cost; that is, the more bread a particular baker offers for sale, the higher the cost of each additional loaf. The cost of each additional loaf tends to rise. Here’s an example that might make that clearer: think about what the baker puts into this operation. Every time this baker bakes a loaf of bread, that’s taking extra time and extra resources. To bake one loaf of bread might take time away from watching television, time that’s not especially valuable. His opportunity costs for that hour is small. But if he spends more and more time in the kitchen baking bread, eventually he’s taking time away from fishing, he’s taking time away from spending with his family, he’s taking time away from pleasurable leisure activities, like reading. He may also be taking time away from other opportunities to make money. And the more bread that he produces, the more he’s having to go to hours of time that are more and more valuable to him. Giving up his TV time is no big deal. Giving up his fishing time is a little more serious. Giving up his family time is more serious still. So we have to pay him larger and larger amounts of money to get him to produce larger and larger quantities of bread. Getting him to produce a small quantity of bread doesn’t take much of his time, so he’s willing to do that for a small amount of money. But getting him to produce a large quantity of bread begins to take away from time that’s more valuable. His opportunity cost of producing additional loaves of bread continuously rises as we eat into more and more valuable time, as we take him away from more and more pleasurable activities. That’s why it takes higher and higher prices to get this baker to make more and more bread. The supply curve is upward sloping because of increasing opportunity costs.

We’ll see later in this series how the upward sloping supply curve is related to increasing marginal costs associated with diminishing productivity of resources. But for now, a good answer to the question, “Why does the supply curve slope upwards,” is this. The supply curve slopes upwards because the opportunity costs of production rise as the baker makes more and more bread.

In the next lesson, we’ll look at those factors that shift the supply curve; that is, what if we allow those things to vary that we previously held constant? What if we relax the assumption of ceteris paribus and allow the price of inputs to change or technology to improve? How would that influence the relationship between the price of bread and the quantity that our baker is willing and able to offer for sale?
Understanding Markets

Supply

Understanding a Change in Supply vs. a Change in Quantity Supplied

The supply curve shows the relationship between the price of bread and the quantity of bread that a seller is willing and able to offer for sale, holding constant other factors that influence supply, such as the price of inputs, technology and expectations about future prices. What happens to the supply curve if we change some of those other variables? Of course, the supply curve shifts. In this lesson, we'll be looking at a particular case of a shift in the supply curve. Then, in the next lesson, we'll be looking at a catalog of all of the factors that can shift the supply curve. So let's start with a specific example.

Remember we’re talking about a particular baker, who has this relationship between price and quantity. This is the supply schedule that we used in the last lesson. Suppose now that we allow a change in one of the factors that we held constant when we wrote down this schedule. Suppose, in particular, that we allow the price of inputs to increase. An increase in the price of milk and eggs and other inputs used to produce bread is going to influence the profitability of this bakery. In particular, with higher input prices, it's going to be less profitable to make bread and the response of the baker will be to scale back his operation. With the higher prices for inputs, we will see a smaller quantity of bread supplied at every price. So we can put these numbers in the table to represent the quantity supplied after the change in input prices. And, as before, I’m putting a little apostrophe after the S, a symbol we usually call “prime” to indicate S’ is a new supply schedule, after the change in input prices.

So with the higher prices of inputs, this baker is unwilling to offer any bread for sale at a price of 0.40 per loaf. We get the first loaf of bread on sale at a price of 0.60 per loaf, two loaves offered at $1.00 per loaf, three loaves at $1.50 per loaf, three and a half loaves per week sold at a price of $2.10 per loaf, and so forth. Notice that we have a smaller quantity supplied than before at every price. A change in one of the variables that we previously held constant required us to redraw the entire supply schedule.

Now, I’ll move this information over to the board and we'll represent this change in supply as a shift in the supply curve that we drew. Here I go, moving information to the board, and I will go back to the diagram that we were looking at before. With a change in input prices with an increase in the price of milk and eggs and other factors that go into producing bread, we're going to get a change in the supplier's behavior. Let me simply graph the numbers from the table over on the board. At a price of 0.40, we have zero loaves of bread offered for sale, so the new dot for the new supply curve will be over at a quantity of zero and a price of 0.40. So it shifts over to the axis itself. At a price of 0.60 per loaf, we now get one loaf of bread offered for sale per week, so here's a point on the new supply curve at 0.60 per loaf and one loaf per week. At a price of $1.00 per loaf, we get two loaves of bread for sale, so here's another point on the new supply curve. The price is $1.00 and the quantity that the baker is willing and able to sell is two loaves.

Another point is three loaves, offered for sale at a price of $1.50 per loaf, so here’s a point on the new supply curve. We get three and a half loaves per week sold at a price of $2.10 per loaf; so here’s another point, price $2.10, quantity three and a half loaves. We get four loaves for this particular baker when the price is up to $2.50 per week, so that's going to be this point, four loaves, $2.50. We get five loaves at a price of $3.00 per loaf, that's going to be this point in the diagram, and six loaves at a price of $4.00 per loaf. That's be this point in the diagram. And finally, we get six and a half loaves at a price of $5.00 per loaf, so that's going to be this point up here on the diagram, $5.00 on the vertical axis and six and a half loaves on the horizontal axis. I'm trying to be careful here and make sure that my numbers line up.

Now, I can connect these dots, because all the prices and quantities in between are imaginable, and I'll label my new curve S’. It's a new relationship between price and quantity for this baker after the price of inputs increases.

So the supply curve has shifted inwards. The increase in the price of inputs made bread a less profitable business and resulted in our supplier offering a smaller quantity of bread for sale at every price.

Now, let's go back and review something that we talked about when we introduced the concept of shifts, back in our study of the demand curve. If the price of bread changes, you move along your given supply curve. A change in the price of bread leads to a change in the quantity of bread supplied, and that's represented as a movement along a given supply curve. But if you change one of those variables that we hold constant when we draw the supply curve, then you have to shift the supply curve. The entire relationship between price and quantity changes for that baker, because input prices have increased. We call this change in one of those variables and the shift in the supply curve a change in supply. A change in quantity supplied is a response to the price of bread changing, and that's a movement along the supply curve. But if you change one of those other factors, like the price of inputs, technology and so forth, then you have to redraw the entire supply curve, and we call that a change in supply.
In the next lesson, we'll go through an entire catalog of factors that can shift the supply curve. We will look at all of those things that can cause a change in supply, and I'll give you a chance to make some predictions about which way you think the supply curve will move in each case.
Understanding Markets

Supply

Analyzing Change in Other Supply Variables

In this lesson, we will show how the determinants of supply show up in the supply curve diagram.

Let’s start with a list of the factors that influence the amount of bread that the baker will offer for sale. They include expectations about future prices, technology, the price of inputs and finally the price of bread itself.

Let’s start here, with the price of bread. How does a change in the price of bread show up in this diagram? The answer is a change in the price of bread leads to a movement along the supply curve, a change in the quantity supplied. When the price of bread increases, the baker is able to cover the rising opportunity costs associated with expanding his output. Therefore, an increase in the price of bread leads to an increase in the quantity of bread supplied. We call this movement along the supply curve in response to a change in the price of bread a change in quantity supplied. When the price of bread increases, we move up the supply curve. The quantity increases. When the price of bread falls, we move down the supply curve, the quantity supplied decreases.

Let’s now look at another factor, and this is a factor that we usually hold constant when we draw the supply curve; that is, the price of inputs. What happens to the supply curve if the price of inputs, like milk and eggs, increases? What happens to the relationship between price and quantity? You answer. What happens to the supply curve when the price of inputs increases?

The answer is with higher input prices, the profitability of bread decreases; that is, the baker can make less profit by baking bread. So the quantity supplied reduces or is reduced at every price. There is a smaller quantity of bread supplied at every price when the price of inputs increases. We call this a decrease in the supply of bread.

What would happen, on the other hand, if the price of inputs were to decrease? Let’s suppose that milk and eggs go on sale and the baker is able to buy them at a lower price. In that case, what would happen to the supply curve for bread? The answer is with lower input prices, the supply of bread increases; that is, a larger quantity of bread is supplied at every price. Again, we have a change in supply, because we’ve changed one of the variables that we hold constant when we draw the supply curve. An increase in input prices leads to a decrease in supply, and decrease in input prices leads to an increase in supply.

Let’s consider another factor that influences the supply of bread. This time we’ll consider technology. Now, it’s hard to talk about technology increasing or decreasing. Instead, we talk about technology improving or worsening. Technology improves, meaning that the firm is now able to make more output with a given amount of input. You can stretch your inputs further, you can use them more productively, and thereby increasing the amount of output you get from the given quantity of eggs and milk and flour. If technology improves, what will happen to the supply of bread? You decide.

If technology improves, the cost of production will fall. Making bread becomes more profitable and the supply of bread increases. The supply curve shifts outward, indicating that the baker is now willing and able to offer a larger quantity of bread for sale at every price. There is an increase in supply in response to an improvement in technology.

Now, we don’t usually talk about a worsening of technology. I mean, that would require that you forget how to do things that you used to know how to do. But sometimes it becomes more difficult to organize your activities or something like that. There may be changes that influence your ability to use a particular technology. If technology should actually worsen due to some other kinds of administrative problems or perhaps you lose access to a technology that you used to be able to use because you had a license for a particular process, or something like that, what would happen to the supply curve for bread in that case? The answer is if technology worsens, the cost of production increases. You’re not able to do things that used to lower your costs, and therefore the cost of production being higher makes the production of bread less profitable, so bakers will offer less bread for sale at every price. A worsening of technology leads to a decrease in supply, which is an inward shift of the supply curve.

Let’s now consider one further factor. Suppose the expectation is that bread prices are going to be higher in the future? Maybe bakers believe that the price of bread is going to rise next week. How would that influence the amount of bread that they offer for sale today at any given price? With the expectation of higher future bread prices, bakers will take the bread off the shelves today, put it in the freezer and wait and sell it next week; that is, the supply of bread will shrink. Bakers will offer less bread for sale than they would if they expected prices to remain constant. Bakers
Understanding Markets

Supply

**Analyzing Change in Other Supply Variables**

will hold off on producing bread or they will hold their existing inventory and wait to sell it after the price has increased. So the expectation of future prices being higher leads to a decrease in the supply of bread today.

What would happen, on the other hand, if the sellers expect that the price of bread is going to be lower next week? Would that cause an increase or a decrease in supply? The answer is supply would increase. Bakers are trying to sell their bread before the price falls, and that would lead to a larger quantity of bread supplied at every price. The supply increases when the expectation is for lower prices in the future.

So there you have it, a catalog of things that can shift the supply curve. We formed this catalog by starting with a list of all of the factors that influence supply. Once we had the list, the considered how each factor in the list would change the seller’s behavior. We showed, first of all, that a change in price causes a movement along the supply curve, a change in quantity supplied. Remember the rule: you never shift a curve when you're changing one of the variables that is on the axis. If you're changing price and quantity, you’re moving along the curve. However, if we change some of the other factors that are held constant when we draw the curve, we have to redraw the curve or shift it. In the case of input prices, higher input prices reduce the profits of the baker and lead to a decrease in supply. Lower input prices increase the profitability of making bread and lead to a higher quantity supplied at every price.

Another factor that that influences supply is technology. If technology improves, the cost of production will be lower. You can get more output from a give amount of input. That's going to lead to a larger quantity supplied at every price. We call that an increase in supply. If technology worsens, the supply curve shifts inwards; that is, firms will supply less bread at every price, a decrease in supply.

Finally, expectations of higher future prices lead to withholding bread today, a reduction in supply. And expectations of lower future bread prices lead to a big increase in supply today, as sellers try to sell their bread before the price falls.

So there you have it. Any of the factors that change the behavior of sellers are going to be reflected either as a movement along the supply curve or a shift in the supply curve. If it’s a change in price, we move along the curve, a change in quantity supplied. If it’s a change in any of the other factors that we usually hold constant, we reflect that by a shift in the supply curve, and we call that a change in supply.

Now we'll look at how we move from the individual firm supply curve to the supply curve for the whole market.
Understanding Markets

Supply

Deriving a Market Supply Curve from Individual Supply Curves

In this lesson we’ll move from the individual _______ [tape skips] supply curve to the supply curve for the market. And we’ll use the same procedure we used when we derived the market demand curve. We’ll be adding up the supply quantities of the individual firms at any given price to find the total market quantity supply.

Let’s look first at a particular bakery. And now we’ll call this Jim’s Bakery. Jim is willing and able to supply one loaf of bread per week when the price of bread is 40 cents per loaf. If the price is 60 cents per loaf, Jim is willing and able to supply two loaves per week, and so forth. The information in this table represents the quantity of bread that Jim is willing and able to supply at different prices, holding constant the price of inputs, technology, and expectations about future prices.

Now let’s look at the supply behavior of another firm. This is Dan’s firm. Dan is willing and able to supply two loaves of bread per week when the price of bread is 40 cents per loaf. He’s willing and able to supply three loaves of bread at a price of 60 cents per loaf, and so forth. As the price of bread increases, the quantity supplied from Dan’s bakery increases. Suppose that Dan and Jim have the only bakeries in town. What does the market supply curve for bread look like? The market supply curve for bread will be the sum of the individual supply curves of Dan’s firm and Jim’s firm.

At any given price, to find the market quantity supplied, simply add the quantity supplied by Dan’s firm to the quantity supplied by Jim’s firm. For instance, at a price of $1.50 per loaf, Dan’s firm supplies five loaves of bread per week. Jim’s firm supplies four loaves of bread. The market quantity supplied in that case will be five plus four or nine loaves of bread each week. You can do the same for any given price. Find the quantity that Dan’s firm is willing and able to sell, add it to the quantity that Jim’s firm is willing and able to sell at the same price, and that gives you the market quantity supply. Now, we can put Dan and Jim’s information together and wind up with the market supply schedule for bread at any given price. This table tells you the quantity supplied by Dan plus Jim’s bakery.

The next step in this lecture is to move this information over to the board and represent it in diagrams. Here is the supply curve for Dan’s firm. I took the numbers from Dan’s table and I put dots in this diagram to represent the price/quantity combinations that describe the relationship for Dan’s bakery. At a price of 40 cents per loaf we get two loaves a week from Dan. At a price of 60 cents per loaf we get three loaves, and so forth. Now we know that these are not the only price/quantity combinations. Price and quantity can both vary continuously. So I can connect the dots and get Dan’s supply curve for bread. I’ll call that $S_d$. The same thing with Jim’s information. I’ll move it from the table into the diagram by plotting the combinations of price and quantity that describe the behavior of Jim’s bakery. And if I connect the dots, I get Jim’s supply curve for bread.

If I want the market supply curve I have to add together these two individual supply curves. And just like I did before, I’ll be using horizontal summation. I will pick a particular price, take the quantity that I get from Dan, add it to the quantity that I get from Jim, and that will give me the market quantity supply at that price. Let’s start by choosing the price of 40 cents per loaf. At a price of 40 cents per loaf we get two loaves from Dan and one loaf from Jim. Add Dan’s two loaves to Jim’s one loaf, and that gives us a total market quantity supplied of three loaves at a price of 40 cents per loaf. So I put a dot here in the market supply curve with a quantity of three and a price of 40 cents.

That’s the same thing if I took a ruler and I added two loaves from Dan’s bakery to one loaf from Jim’s bakery, and that gave me a total of three loaves, and I graphed it right here in this diagram. Three loaves in the market at a price of 40 cents per loaf. I can do the same for any price on the axis. Let’s say a price of 60 cents per loaf. In that case, I get three loaves from Dan, add that to two loaves from Jim, for a total of five loaves supplied when the price is 60 cents per loaf. Keep picking any price you want and add the quantity from Dan’s firm to the quantity from Jim’s firm to get a total quantity supplied in the market.

Alternatively, you can just go back to the table representing quantity supplied—that table that we got when we added Dan’s output to Jim’s output and you can graph the numbers directly. At a price of $1.00 per loaf, we get a quantity supplied of seven. At a price of $1.50 per loaf, we get a quantity supplied of nine. At a price of $2.10 a loaf, we get a total of 11 loaves supplied. At a price of $2.50 per loaf, the quantity supplied is up to 13 loaves per week. There’s the market supply curve. I can connect the dots and label this with an S to indicate that this is the supply curve for the whole market. The supply curve is the horizontal summation of the supply curves of the individual firms in the market. To get this curve, pick a price and add the quantity from Dan’s bakery to the quantity from Jim’s bakery at that price to get the quantity for the market as a whole.
Supply

Deriving a Market Supply Curve from Individual Supply Curves

Now, what’s going to shift the market supply curve? The answer is anything that shifts the individual supply curves will shift the market supply curve. After all, the market supply curve is simply the sum of the individual curves, so anything that moves Dan’s curve or Jim’s curve will shift the market supply curve as well. And what are those factors? They’re all of the factors that held constant when we originally drew the individual firm supply curves. The price of inputs, technology, and expectations about future prices. If any of those factors changes, we have to redraw the individual firm’s curves and consequently we have to shift the market supply curve as well.

Now, this wraps up our introduction to the behavior of suppliers and the behavior of buyers. Now we’re ready to put the two sides of the market together and find the conditions under which we have a stable situation. In the next lecture we’ll be putting supply and demand together in one diagram to find the equilibrium price and quantity traded.
In this lecture we put together our previous analysis about supply and demand to come up with a model of the market. And the payoff is we have a way of predicting how factors in the environment will influence the price of a good and the quantity of that good trade. We’re going to be developing a tool that economists use called the competitive equilibrium.

The competitive equilibrium is what happens when you get together a bunch of buyers and sellers, all of whom believe that they have no influence over the price. That is, all of the agents in our market are price takers. They accept the market price as given and do the best they can subject to that price or the constraint. So the buyers take the price as given and choose their quantity demanded. The sellers take the price as given and choose their quantity supplied. When the price is such that the quantity supplied is equal to the quantity demanded we say that we have a competitive equilibrium.

Now first we should define the concept of equilibrium in Economics. An equilibrium is a situation from which there is no tendency to change. And as we will see in our model of supply and demand, there is only one price at which there is no pressure for something to change. Let’s go back to the analysis that we developed earlier. The red dots in this diagram represents quantities demanded at different prices. This information comes from the table that’s over on the board. At a price of $5.00 the quantity demanded is one loaf of bread per week. At a price of $4.00 per loaf the quantity demanded is two loaves per week, and so forth, giving us this collection of red dots.

We can collect these red dots and form the demand curve. But we’re going to wait to do that for just a moment, because I think that leaving this in the form of dots will help me make my argument more clearly. The blue dots represent the supply curve. That is, at a price of 40 cents per loaf we get one loaf of bread supplied per week, two loaves at 60 cents, three loaves at $1.00, and so forth. The blue dots represent the supply curve.

Now, we’ve got these price taking buyers and price taking sellers. All of them are doing the best they can subject to the price that they are given by the market. Let’s find now the price at which an equilibrium obtains. Suppose, first of all, that we consider a price of $1.00 a loaf for bread. What would happen in that case? Well, with a price of $1.00 per loaf there are three loaves the bakers are willing and able to supply each week. The quantity supplied is three at a price of $1.00. At a price of $1.00 the quantity demanded, however, is eight loaves per week. That is, buyers are willing and able to purchase eight loaves of bread at a price of $1.00.

The fact that there are more buyers than sellers, or the quantity demanded is greater than the quantity supplied at a price of $1.00, means that we have excess demand. We have excess demand, which is the difference between the quantity demanded and the quantity supplied at the price of $1.00. I'll label this with a little sign that says “excess demand.” Now when there is excess demand the bidding mechanism takes over. The bidding mechanism is the process by which unsatisfied buyers try to bid up the price of bread to guarantee that they get some. After all, we’ve got eight loaves of bread demanded here and only three loaves of bread supplied. We’ve got five loaves of bread excess demand. That could be five people who aren’t able to buy their loaf of bread at a price of $1.00 per loaf.

Those five people are now going to offer sellers a higher price. They’re going to bid the price of bread up to $1.10 or $1.20 or $1.50 or $2.00 a loaf. They’ll continue bidding until either it’s no longer worth it to bid higher prices or until enough bread becomes available that the shortage of bread is eliminated. Look, if the price of bread goes up to a $1.50 per loaf the quantity supplied increases to four loaves. Some baker has a reservation price of $1.50. That is, he offers his loaf of bread for sale as soon as the price goes up to $1.50. At the same time, when the price goes up to $1.50 we have one of the buyers drop out. That is, there was one buyer who was willing to pay $1.00 but not $1.50.

So the excess demand shrinks as the price rises, because additional sellers enter the market and some of the buyers drop out of the market. Finally, when the price of bread rises all the way up to $2.10 a loaf, the quantity supplied is equal to the quantity demanded. Enough buyers have left the market and enough sellers have entered the market to give us an equality of quantity demanded and quantity supplied. At a price of $2.10 per loaf the quantity demanded equals five loaves of bread per week. And the quantity supplied equals five loaves of bread per week. This means that we have a competitive equilibrium. Taking prices as given the quantity demanded by buyers and the quantity supplied by sellers are equal at a price of $2.10 per loaf. And we get there by the bidding mechanism.

Buyers who can’t get bread at the lower price will bid up the price and as buyers and sellers respond to the rise in price, an equilibrium is established. The excess demand is eliminated as the quantity demanded falls and the quantity
Understanding Markets

Equilibrium

Determining a Competitive Equilibrium

supplied increases with the rise in price. Now, you can see the same analysis working from the other direction. Suppose we start with a high price for bread. Suppose $4.00 a loaf. What would happen in that case?

At a price of $4.00 per loaf the quantity demanded is going to be smaller, that is, two loaves of bread per week. Because the price of bread is high the quantity demand would be much smaller. However bakers are very excited about this high price and the quantity supplied is quite high because a lot of bakers can cover their opportunity cost at that higher price and there’s a larger quantity supplied at $4.00 per loaf. That quantity in our example is eight loaves of bread per week. Well, there you’ve got a problem, because we’ve got an excess supply. That is, the quantity supplied is eight loaves at $4.00 per loaf and the quantity demanded is two loaves at $4.00 per loaf.

We have an excess supply of six loaves of bread every week. Bread piles up un-purchased on the shelves. What’s going to happen in that case? The bidding mechanism will operate from the other direction. In this case sellers who are unable to move bread from their shelves will begin to put the bread on sale. They will lower their price. Some sellers will offer their bread for $3.50 a loaf or $3.00 or $2.75, and the price will begin to fall. As competition from the bidding mechanism pushes down the price of bread two things happen.

First, additional buyers will enter the market. Notice when the price drops to $3.00 a loaf you can then sell three loaves of bread because three people are willing and able to purchase bread at that price. As the price drops down to $2.50 a loaf the quantity demanded increases to four loaves of bread and so forth. So additional buyers enter the market as the price falls. As the price falls from the supplier’s perspective things are getting less attractive. So at a price of $3.00 per loaf you find that a seller or two may have left the market and your quantity supplied is down to seven loaves of bread. And at a price of $2.50 per loaf the quantity supplied drops even further to six loaves of bread.

Finally, when the price is $2.10 per loaf, the quantity demanded is equal to the quantity supplied and the excess supply has been eliminated. Partly by bringing additional buyers into the market and partly by reducing some of the over-supply by pushing some of the sellers out of the market, sellers who can’t cover their opportunity costs as the price of bread falls. Well, what’s the outcome?

The outcome is we wind up with a stable situation. The price of bread is set at $2.10 per loaf. We call that the equilibrium price. Since the quantity supplied and the quantity demanded are equal and they are both equal to five loaves at that price of $2.10, the bidding mechanism is going to just sit there and do nothing. There is no tendency for prices to change. There’s no excess demand, so prices don’t rise. There’s no excess supply, so price won’t fall. When quantity supplied and quantity demanded are equal we’re in a stable situation, a competitive equilibrium.

So when the price is $2.10 the quantity of bread traded will be equal to five loaves of bread per week. We can label this P*, $2.10, that’s the price at which supply and demand are equal and we can label this Q*, the quantity of bread that’s traded at a price of $2.10. So there you have it, the stable outcome, the competitive equilibrium. The bidding mechanism takes us to the point where the quantity supplied and the quantity demanded are equal. And once we reach that point, there is no further tendency to change.

In the next lecture we’ll look at factors that can shift the supply and demand curve, upset this equilibrium and lead to a new equilibrium. But first let’s connect the dots and make this picture look like something that we might be more familiar with. Connect these dots and they become the demand curve. Connect the blue dots, that is, allowing prices and quantity to vary continuously and we get a supply curve. The competitive equilibrium is found at the place where the supply curve and the demand curve intercept. As in most economic graphs, the most interesting point is where the curves cross and it’s where the curves cross that we find the price and quantity where supply and demand are equal. And when quantity supplied and quantity demanded are equal at a particular price there is no further tendency to change. This is what a competitive equilibrium looks like. And usually we will mark the price and quantity with lines down to the axis to show us the price and quantity at which the market has a stable outcome.
What happens to the price of a good or service when something changes in the market? Can you predict how a change in the environment will influence the price of a good and the quantity traded? Well, you can using the tool that we developed in the last lecture. Here’s what a competitive equilibrium looks like. The competitive equilibrium is the price and quantity where supply and demand are equal. That is, it’s the price at which the quantity demanded on the red curve is equal to the quantity supplied on the blue curve.

Now, anytime we change anything that is held constant when we draw this picture, we have to redraw the picture. That means if we change consumer income, the price of substitute goods, the price of compliment goods, expectations, or if we change input prices or technology, we have to shift the curves. And when the curves shift, we have to find a new equilibrium price and quantity. Let’s take an example. Suppose we consider a case where in the market for bread there’s a change in the price of a substitute good. Let’s suppose that bagels become more expensive. If bagels become more expensive, what will happen to the price of bread and the quantity of bread traded?

Well, there is a series of steps that you should go through to answer a question like this, and this series of steps will help you anytime you try to solve one of these problems. The first step is, identify which side of the market is affected. If the price of bagels goes up, who cares, buyers or sellers? The answer is, buyers care because bagels are a substitute for bread for consumers. That’s the first thing you need to resolve. Who cares, buyers or sellers, because that tells you which curve you’re going to be shifting.

The second question—how will the change affect the curve? So if the price of a substitute good increases, we know we’re going to be dealing with the demand curve, but will it shift in or will it shift out? The answer in this case is, the curve shifts outwards. When the price of a substitute good increases, the demand for bread increases as well. That is, customers are going to want to buy more bread at every price, so the demand curve for bread shifts outwards.

The third question we want to answer is, what happens to the equilibrium? That is, after you shift the curve, which direction will the bidding mechanism take the price of the good? Will it take it up, or will it take it down? Let’s look at this example carefully. If the price of a substitute good increases, we know that the demand for bread is going to increase, so the demand curve is going to shift outwards, in this direction, and we’ll have a new demand curve which we could label D prime. Here it is. And this is the demand curve after the change in the price of bagels.

The next thing you want to notice is, at the original price for bread, say $2.10 per loaf, the quantity of bread demanded is now larger than the quantity supplied. The quantity supplied hasn’t changed, because at a price of $2.10, we would still be at this particular point on the blue curve, still with, say, five loaves of bread supplied each week. However, with the increase in the price of bagels, the quantity demanded has increased to a new point over here on this new red curve. The old red curve doesn’t apply anymore because of a change in the environment. With an increase in the price of a substitute good, the quantity demanded is now over here on this new red curve, even though the supply is still over here on the old blue curve. We’ve now got a gap between quantity demanded and quantity supplied at the original price of $2.10.

Well, you know what’s going to happen next. We have an excess demand for bread, and that’s going to cause the bidding mechanism to kick in. The price of bread is going to be bid up, and as the price of bread is bid up, some buyers will leave the market and new sellers will enter the market. As the quantity demanded shrinks with the rising price, and the quantity supplied increases, we will approach a new competitive equilibrium. We will approach a new equilibrium with a higher price and a larger quantity of bread traded.

Let me go ahead and draw this curve in and show you this adjustment process very carefully. We have a new demand curve. The new demand curve results from the increase in the price of a substitute good. With consumers wanting to buy a larger quantity of bread at every price, the demand curve is now shifted out here. Well, that gives us a problem, because at the original price, the quantity supplied, the point on the blue curve is now less than the quantity demanded over here on the new red curve. So the distance between these two points is an excess demand at the original price. That causes the bidding mechanism to kick in, and as the price rises, buyers begin to buy less bread than they were buying before. We have a movement along our new demand curve. We have a decrease in quantity demanded as the price is bid up. Over here on the supply curve, the rising price leads new suppliers to enter the market, or existing suppliers to increase the quantity they supply. As the price rises, suppliers can cover rising opportunity costs associated with providing more bread to the market.
Eventually, we reach this point, at which the quantity demanded and quantity supplied are equal with the new demand curve and the original supply curve. This new price and new quantity are our new competitive equilibriums. So if I draw a dashed line over to the price axis, I can call this $P^*$prime, the new higher price. And if I draw a dashed line down to the quantity axis, I could label this $Q^*$prime, an increase in the quantity traded. So the market for bread has gotten a little bit bigger, because bagels, a substitute, are now less attractive. More consumers buying more bread at every price leads to an increase in the price of bread and an increase in the quantity of bread traded.

We got rid of our excess demand through the bidding mechanism, which pushed up the price. And as buyers and sellers adjusted to the new higher price, the quantity traded changed. This is just one example of how a change in the environment will shift one of the curves and create a new competitive equilibrium. In the next lesson we’ll look at a whole catalog of factors that can shift the demand curve or shift the supply curve, thereby changing the competitive equilibrium price and quantity traded.
In this lesson we present a catalog of comparative statics. Comparative statics are static experiments where you change something and see what happens. We call them statics because it refers to comparing one state with another state, the state before the change with the state after the change. In our last lesson we did an example of comparative statics. We looked at what happened with competitive equilibrium when we changed the price of a substitute good. We saw that the demand curve shifted and the price and quantity traded at competitive equilibrium changed.

Well, that was only one example. Now we’re going to present all the examples that I can think of. The thing to remember is there are a limited number of comparative statics that you can do in a supply and demand diagram. That’s because there are only two curves you can shift and only two directions to shift each curve. So a complete catalog of comparative statics would look like the following.

We would have one case where the demand curve shifts outwards. One case where the demand curve shifts inwards. One case where the supply curve shifts outwards and one case where the supply curve shifts inwards. Each of those cases is going to lead to a different combination in changes of price and quantity. And each of those cases will correspond to a different set of stories about the real world. So what I’ll do in this lesson is go through each of those cases one at a time until I’ve done all four.

Let’s start with case number one, the case we considered in the last lesson. Here’s the case where the demand curve shifts outwards. If the demand curve shifts outwards, then if you hold the price steady at the original level, you will have an excess demand. The new quantity demanded would be larger than the old quantity demanded. The quantity supplied would be the same as before at that original price, and you’ve got excess demand or a shortage. The bidding mechanism would push up the price of the good. Some buyers will drop out and some new sellers will enter the market. The end result when the demand curve shifts outward is the equilibrium price rises and the equilibrium quantity traded increases.

Now what are the stories that could cause that to happen? The first story would be an increase in the price of a substitute good. When the price of the substitute good increases the demand for bread itself will increase, as consumers consume bread instead of bagels. The second case that would shift outwards the demand curve would be a decrease of the price of a complimentary good. If the price of cheese falls, people want more bread and cheese sandwiches. So the demand for bread increases, pushing up the price and increasing the quantity traded in equilibrium.

The third case would be if there were a change in consumer income. And here we have two possibilities. If bread is a normal good then an increase in income will lead to an increase in demand for bread and a higher price and larger quantity in equilibrium. If bread is an inferior good, on the other hand, then a decrease in consumer income would cause the demand curve to shift outward. If this were beans instead of bread and beans were an inferior good, then lower consumer income would lead people to buy more beans at every price and that would lead to a new higher price in equilibrium and a larger quantity of beans traded.

Another thing that would shift outward the demand curve would be expectations changing. If people expect the price of bread to be higher in the future they’re going to buy more bread today. And quite ironically that would push up the price of bread. It becomes a self-fulfilling prophecy. If people are hoarding bread today for fear of higher prices in the future, their actions have the result of fulfilling the prophecy. The price of bread indeed rises.

Finally, since this is the market demand curve for bread, an increase in the number of buyers in the market will shift the demand curve outwards. If the number of consumers grows in this market or more people decide they like bread and enter the market for bread that would push up the price and increase the quantity traded. So, that’s case one, an outward shift in the demand curve. And we looked at what happens and the stories that could lead to it.

Let’s consider now case number two. Case number two is a case of an inward shift in the demand curve. If the demand curve shifts inward then at the original price we have now an excess supply of bread. With the demand curve shifting inward consumers want less bread at every price. So at the old price the quantity of bread demanded is less than the quantity of bread supplied. The point on the blue curve. This causes the bidding mechanism to kick in. Sellers put their bread on sale and the price of bread falls. As the price of bread falls some buyers come back into the market for bread and some sellers drop out of the market.
Understanding Markets

Equilibrium

Classifying Comparative Statics

The bidding mechanism comes to rest when the quantity demanded is equal to the quantity supplied at the new equilibrium price. The new equilibrium price will be lower. It had to fall to get rid of that excess supply, and the quantity of bread traded would shrink. These people want less bread now; the quantity of bread actually traded in the market would be smaller. Let’s look at the stories that could result in an inward shift in the demand curve.

The first thing would be if the price of substitute goods falls. If bagels become less expensive then people would buy less bread and more bagels instead. If the price of a complimentary good increases, like cheese, if cheese gets more expensive, people want fewer cheese sandwiches. And that means they would buy less bread at every price. The demand curve shifts inward because of the higher price for a compliment.

The third case would be a change in income. And once again we have two cases. If bread is an inferior good then the demand curve shifts inwards when income increases. Consumers want less bread because now they’re better off and they buy other things instead. But if bread is a normal good, then the inward shift corresponds to a decrease of income. When your income falls you buy less bread at every price.

Another thing that would shift inward the demand curve would be the expectation that the price of bread will fall. People delay their purchases of bread and the demand for bread shifts inwards. It is reduced. However, you can see that the expectation that future prices will be lower turns out to be a self-fulfilling prophecy. By reducing their demand for bread people lead to a lowering of the price. The bidding mechanism kicks in, lowers the price of bread, and the expectation becomes self-fulfilling.

Finally, the last thing would be if there are fewer traders in this market. With fewer traders in the market, the demands will be less and therefore the equilibrium price will fall. All of these are the same factors that shift the demand curve outwards, only now we’re considering the opposite cases. So there’s case two, an inward shift in the demand curve. This will occur anytime something causes consumers to want to buy less bread at every price. And the result is a lower equilibrium price and a smaller quantity traded in equilibrium.

Let’s look now at case number three, an inward shift in the supply curve. Now we have a different set of factors that lead to an inward shift in the supply curve, because we’re no longer talking about the behavior of consumers, now we’re talking about the behavior of sellers. In this case, when the supply curve shifts inwards the immediate results is that we have an excess demand for the product. That is, at the original price, the original equilibrium, where the original demand curve and the original supply curve cross we have P* as our equilibrium price and Q* as our equilibrium quantity. If we keep the price constant and the supply curve shifts inwards, that is, if we have a decrease in supply, we’re now going to have a small quantity supplied and a larger quantity demanded. So we have an excess demand for the product given the new quantity supplied and the original quantity demanded. Well, this is going to cause the bidding mechanism to kick in and as usual the bidding mechanism deals with an excess demand by causing the price to rise. The price is bid up. And as the price is bid up some buyers leave the market and some of the sellers who thought they were going to leave the market come back in, because the rising price makes it worth their while to continue to offer bread for sale.

The result is that the price of the good rise with the equilibrium. The new equilibrium price is higher and the equilibrium quantity traded is smaller. This will be the case any time the supply curve shifts inwards. Any time there is a decrease in supply. Well what are the factors that lead to a decrease in supply?

The first would be an increase in the price of input. If milk, eggs or flour become more expensive, then our baker finds it less profitable to make bread, and bakers will bake less bread and sell less bread at every price. A second factor would be a change in technology. If technology worsens or our baker looses access to a low cost technology then the cost of production will increase. It’s less profitable to make bread and the quantity of bread offered for sale will be lower at every price.

Finally, there could be a decrease in the number of sellers in this market, for any number of reasons. Remember this is the market supply curve and the market supply curve is the sum of the supply curves of the individual sellers. If we have fewer sellers in the market the curve will shift inward because fewer curves are being added together. Any of
Understanding Markets

Equilibrium

Classifying Comparative Statics

these factors will lead to an inward shift in the supply curve. And with that inward shift comes a higher equilibrium price and a smaller quantity traded. That’s case number three.

Let’s go now to the final case. And that is an outward shift in the supply curve. The outward shift in the supply curve occurs if sellers are offering more bread for sale at every price than they were before. Originally the equilibrium price was $P^*$ and the equilibrium quantity was $Q^*$. However, something occurred that causes sellers to want to sell more bread at every price. At the original price therefore, we now have excess supply of bread. The quantity supplied is greater than the quantity demanded at the original price. The bidding mechanism once again takes over and leads sellers to cut the price of their bread, to make sure they can sell their product, they lower the price. And as the price falls, some sellers decide that they’re going to drop out of the market, they can no longer cover their opportunity costs. Meanwhile, the falling price draws some buyers into the market. These buyers are now willing and able to purchase more bread than they were buying before. The result is a new equilibrium price that’s lower than the original. The equilibrium price falls in this market and the equilibrium quantity traded increases.

What are the real world events that could lead to this outcome? The first would be if the price of input goods falls, making the production of bread more profitable. Sellers would make and sell more bread at every price than before. The second possibility would be an improvement in technology. With an improvement in technology, once again production costs will fall, making it more attractive to offer bread for sale. Producers respond by offering more bread for sale at every price, an outward shift in the supply curve.

And finally, if the number of sellers in this market increases, since it is the market supply curve, the supply curve shifts outwards. All of these factors shift the supply curve outwards, leading to a lower equilibrium price and a larger equilibrium quantity traded.

So, we have four cases, and each of these cases results in a different combination of changes of price and quantity. If the demand curve shifts outwards, the price increases, as does the quantity traded. If the demand curve shifts inwards, we have a decrease in the price, as well as a decrease in the quantity traded. If the supply curve shifts outwards, the price will fall, but the quantity traded will increase. And if the supply curve shifts inwards, then the price will increase and the quantity traded will fall. Those are all the possibilities we have for comparative statics. So now that you’ve got the catalog, you’re ready to try some experiments of your own.
Understanding Markets

Elasticity

Defining Elasticity

Suppose you run an ice cream store. Right now you’re selling your ice cream cones for $2.00 apiece and you’re selling 20 cones a day. Should you try to improve your business by putting your ice cream cones on sale for $1.00 apiece? In order to answer this question you’d need to know how your customers would respond to a change in your price, and this brings up the question of the elasticity of demand.

Elasticity means responsiveness. Suppose you have a piece of elastic. If you stretch it, it responds and changes shape. If it didn’t change shape, that is, if it were unresponsive to your pulling on it, it wouldn’t be elastic, would it? Elasticity means the extent to which the quantity demanded changes when there’s a change in the price of a good. We define elasticity as the percentage change in quantity demanded that results from a percentage change in the price of a good.

Let’s look at those ice cream stores now. Suppose Angie runs an ice cream store where she charges $2.00 apiece for her ice cream cones, and right now she’s selling 20 cones a day. If Angie puts her cones on sale at $1.00 apiece, she discovers that her customers want to buy more cones, that is, as the price falls, the quantity demanded increases. But they only want to buy 10 additional cones daily for a total of 30 cones per day at a price of $1.00 apiece. Angie’s demand curve for ice cream cones looks something like this: her customers want to buy a few extra cones when the price falls, but the demand curve is relatively steep. We’ll label this $D_a$ for the demand on the part of Angie’s customers.

Across town Barney has an ice cream store. Barney starts, like Angie did, with a price of $2.00 apiece for his ice cream cones. He’s also selling 20 cones per day. Now, when Barney puts his cones on sale at $1.00 apiece, his customers get very excited and come in and buy 50 cones a day. That is, Barney’s customers are much more responsive to a change in the price of cones than were Angie’s customers. If we connect these two dots for Barney, we see that Barney’s customers demand curve is flatter. That is, for a given change in the price of ice cream, Barney’s customers will buy many more additional cones than will Angie’s customers.

When the price goes from $2.00 to $1.00 at the two stands, Angie’s customers respond a little, but Barney’s customers respond a lot. We say that Barney’s customers demand for ice cream is more elastic than Angie’s customers demand for ice cream. Now, why should you care about this? See if you can answer this question. What happens to the total revenue that Angie raises when she changes the price of her ice cream cones from $2.00 apiece to $1.00 a piece. Does her total revenue increase, or does it decrease?

If you’ll look over on the white board, you can see my calculations. When the price of ice cream cones is $2.00, Angie sells 20 cones for a total revenue of 20 cones times $2.00 apiece, or $40.00 in total sales. When she drops her price to $1.00 per cone, she sells 30 cones. That means each day she earns 30 cones times $1.00 apiece, or $30.00 in total revenue. When Angie drops the price of her ice cream cones from $2.00 apiece to $1.00 apiece, Angie’s total revenue actually falls.

Now, what happens to Barney’s total revenue whenever he drops his price? You can see my calculations. Whenever Barney is charging $2.00 apiece for ice cream cones, he’s also selling 20 cones. His total revenue is 20 cones times $2.00 apiece or $40.00. When Barney drops the price of his cones to $1.00 apiece, he sells a total of 50 ice cream cones. Fifty times $1.00 apiece gives a total revenue of $50.00 a day. Barney’s total revenue increases when he drops his price, that is, when he has a sale on ice cream cones.

You can see it’s very important to know whether the demand for your product is elastic or inelastic, that is, by how much the quantity will respond to a change in the price of a good. If the quantity demanded is very responsive, your total revenue will increase when you reduce the price of your product. If, however, the quantity demanded is unresponsive, or relatively inelastic, then when you reduce the price of the good, your total revenue will actually fall.
Elasticity

Calculating Elasticity

We’re looking for a measure of the elasticity of demand that would be useful to a firm that’s trying to decide whether to raise its price, lower its price or keep the price of its product the same. We’re going to use Barney’s demand curve for ice cream as an example. Recall that when the price of ice cream is $2.00 at Barney’s store he sells 20 cones a day. If he lowers his price to $1.00 per cone, he sells 50 cones a day. We would like now to calculate a measure of the elasticity of demand that can be reached with Barney.

First of all, we want a measure of elasticity that doesn’t depend on the way in which we measure the units, or the price or the quantity. That is, it shouldn’t matter whether we measure the price of Barney’s ice cream in dollars or pennies or nickels or French francs. It also shouldn’t matter whether we measure the quantity in terms of cones or dozens of cones or cases of ice cream. We’re looking for a unit-less measure of the responsiveness of the quantity demanded to a change in price. And this is where we get our elasticity measure.

An economist defines elasticity as “the percentage change in quantity demanded that results from the percentage change in price.” Let’s say a couple of things about this measure. The first thing is you’ll notice that we write it as a fraction. Anytime you have a fraction in Economics we can read the fraction with these words, “that results from”. The percentage change in quantity demanded that results from a percentage change in price. The next thing to notice about the way we’ve written the elasticity measure is we’ve put it between two bars to indicate absolute value. Remember, price and quantity vary inversely. As the price increases, the quantity demanded falls. That is, price and quantity are inversely related. This means that the fraction that compares price and quantity will always have a negative value. So we put the absolute value bars around this fraction to get rid of the negative sign. That way the elasticity measure is always a positive number.

Finally, the third thing to notice about this measure is that it is a unit-less measure. When you calculate percentage changes, you get rid of the units. That is, it doesn’t matter whether we’re measuring the price in dollars or pennies. The percentage change will be the same in either case. Elasticity is a unit-less measure of the responsiveness of quantity demanded to a change in price. Now, let’s do a quick reminder of how you calculate a percentage change.

To calculate a percentage change, take, for example, the new quantity demanded, subtract the old quantity demanded and that gives you the absolute change in quantity. To turn this into a percentage change you divide it by some base quantity, usually the original quantity or the old quantity.

So, for example, in the case of Barney’s ice cream store, the percentage change in quantity demanded as we move from $2.00 to $1.00 per cone will be—the new quantity is 50 cones, subtract the old quantity of 20 cones and divide by the original quantity of 20 cones, to get a percentage change of 30 divided by 20 or 1.5. So there’s a 150 percentage change in the quantity demanded. Divide that now by the percentage change in price. To calculate percentage change in price take the original price of $2.00 per cone and subtract that from the new price of $1.00 per cone. $1.00 minus $2.00 is negative $1.00. Divide that again by the original price of $2.00 per cone, and you get a percentage change of negative one over two, or negative one-half, or negative 50 percent.

Now, to calculate the elasticity of demand as we move from point A, the original price and quantity, to point B, the newer price and quantity, we take the percentage change in quantity demanded, which is 1.5 or 150 percentage, and divide that by the percentage change in price, which is negative 0.5 or negative 50 percent. That gives us negative three for our fraction. Now apply the absolute value bars and you get the elasticity of demand calculated for Barney’s ice cream as we move from point A to point B is three. That is, for a one percent change in the price of ice cream we’re getting a three percent change in the quantity demanded.

Now, let’s do as a calculation the same exercise going from point B to point A. That is, let’s suppose Barney starts with a price of $1.00 and a quantity demanded of 50 cones and he moves to a new price of $2.00 and a quantity demanded of 20 cones per day. See if you can calculate the elasticity demanded as we move from point B to point A. The elasticity of demand is calculated as follows. Take the new quantity of 20, subtract the old quantity of 50 and divide by the base quantity of 50, the original quantity demanded. That gives you 20 minus 50 or negative 30, divided by 50 or negative 0.6 or negative 60 percent. The percentage change of price can be calculated as follows. The new price is $2.00. Subtract the original price of $1.00 and divide by that original price of $1.00 for a fraction of one over one, which gives us a percentage change of one or 100 percent. The elasticity in this case is negative 60 percent divided by 100 percent or negative 0.6. Apply the absolute value sign and the elasticity measure as we move from point B to point A is 0.6.
Understanding Markets

Elasticity

Calculating Elasticity

Now, wait a minute. Something’s wrong here isn’t it? We’re getting one elasticity measure as we move from point A to B and we’re getting a different elasticity measure as we move back from B to A. Somehow, that seems confusing. We’d like to have a measure of responsiveness that doesn’t depend on whether we start at A and move to B, or start at B and move back to A. That is, we’d like to have one number that represents the responsiveness of quantity to a change in price that would be constant whether we’re going this direction or back in the other direction.

Well, the problem here stems from using different bases to calculated percentage. When you move in this direction, this price and quantity are the base of our fractions. And when we move back from B to A, then this price and quantity are the base of our fractions. That’s what gives us the inconsistency. So in order to have a consistent measure of elasticity over this range of prices and quantities we’re going to use as the base of our percentage calculation the price and quantity that’s halfway between A and B.

We’ll call this point M to represent the midpoint. When we calculate the elasticity moving along a demand curve, we view the midpoint as the basis of our percentage calculation and that gives us a consistent measure. Why don’t you try now, for Barney’s ice cream store, to calculate the elasticity of demand this time using as the basis of your percentage calculations the midpoint? That is, the point halfway between the original price and the new price, or the point halfway between the original quantity and the new quantity. That is, calculate those elasticity numbers again, and this time use $1.50 for the basis of your percentage calculations. Then whenever you go to calculate the change quantity demanded us 35 as the basis of your quantity calculations. See what you get.

So, what do we have? Negative $1.00 divided by $1.50 is negative two-thirds. Now, to calculate the elasticity of demand, take the percentage change in quantity demanded, which is six-sevenths, and divide that by the percentage change in price, which is negative two-thirds. And what do you get? Let’s solve this problem together. The percentage change in quantity demanded is six-sevenths. The percentage change in the price of the product is negative two-thirds. So the elasticity will be six-sevenths divided by negative two-thirds, absolute value. Well, let’s calculate that. Six-sevenhs, and remember in order to divide by a fraction you invert the fraction and multiply, multiply by negative three-halves, and we get, the six will cancel with the two to turn that into a three and we get as our fraction negative nine over seven. The absolute value sign gets rid of the minus sign and nine-sevenths is one and two-sevenths. The elasticity of demand, moving from point A to point B for Barney, is one and two-sevenths.

Now, recalculate the elasticity going in the other direction. Go from point B back to point A, again, using the midpoint as the bases for your percentages. In this case, the original quantity would be 50 and the new quantity would be 20. So if you take 20 minus 50 and divided by 35, this time you get negative six-sevenths. In this case, the new price is $2.00 a cone and the old price is $1.00 a cone. Two minus one is one, divided by 1.5 gives us two-thirds. This time we’ve got six-sevenths, negative, divided by two-thirds, the absolute value sign gets rid of the minus sign and once again we have one and two-sevenths. The great thing about using the midpoint formula for calculating elasticity of demand is you get the same answer whether you’re going from point A to B or back from B to A. This is how we calculate elasticity of demand in Economics.
We’re still studying calculation of elasticity. For more practice now, let’s consider the case of Angie’s ice cream store. Remember, when Angie lowers her price from $2.00 per cone down to $1.00 per cone, the quantity demanded at Angie’s store increases from 20 cones a day to 30 cones a day. What I want you to do first is forget about the midpoint formula. Forget about what I told you was the right way to solve the problem. Why don’t you go back to that old way and find Angie’s elasticity of demand when she lowers the price of her product from $2.00 per cone to $1.00 per cone, but start by using point A as the base point for calculating percentages. That is, use $2.00 as your base price, and use 20 cones as your base quantity. See what you get when you calculate elasticity of demand, moving from this point to this point.

Now let’s do those calculations together. The new quantity is 30, subtract from that the old quantity of 20. That gives you a change in quantity of 10 cones. You want to divide now by the original quantity, which was 20, and that gives you a total change of 10 over 20, or one-half, a 50 percent change in quantity demanded. Now, divide by the percentage change in price. The price goes from $2.00 to $1.00. Take the new price of one, subtract the old price of two, that gives you negative one, and divide by the old price of two. Now we have negative one-half or negative 50 percent. If we take one-half, the percentage change in quantity, and divide by negative one-half, the percentage change in price, we get negative one. Then, apply the absolute value sign to get rid of the minus, and the elasticity measure, calculated, moving from point A to point B, using point A as the base point, is equal to one. We call this unitary elasticity, or unit elastic.

Now, go from point B back to point A. That is, calculate the elasticity of demand as we move from a price of $1.00 up to a price of $2.00, and the quantity falls from 30 to 20, only this time use point B as the base points for your percentage calculations. That is, use $1.00 as your base price, and use 30 cones as your base quantity, see what you get.

All right, the new quantity is 20 cones. Subtract the old quantity of 30 cones and divide by 30 to get negative one-third as the percentage change in quantity demanded. The percentage change in price will be the new price of $2.00 minus the old price of $1.00 divided by that old price of $1.00, which gives us one over one. So we have negative one-third over one, which is negative one-third. Apply the absolute value sign, and you get the elasticity measure moving from point B to point A, using B as the basis for your percentage, is equal to one-third. Aha! We’ve recreated that problem that we had with the demand curve originally, with Barney’s demand curve. We’ve recreated the problem, but we get a different elasticity measure moving from A to B than we get when we move back from B to A. And that problem is all related to the way you measured percentages.

Now you know what the right way to solve the problem is. You know how to get a consistent measure moving from A to B or B to A. That is, you want to use as the basis for your percentages the point halfway between A and B. You want to use this midpoint in order to calculate the price and quantity percentage changes. The midpoint will be $2.00 plus $1.00 divided by two, or $1.50. $1.50 will be the midpoint price. The midpoint quantity will be 20 plus 30, which is equal to 50, divided by two, equal 25.

Now calculate the elasticity of demand for Angie’s ice cream store. As you move from a price of $2.00 down to a price of $1.00, only this time use the midpoint formula. That is, use $1.50 as your base price, and use 25 as your base quantity. See what you get. The new quantity is 30, the old quantity is 20, 30 minus 20 is 10, divided by the midpoint quantity with is 25 gives you 10 over 25 or two-fifths. Two-fifths or 0.4 is the percentage change in quantity demanded. The percentage change in price now will give you the new price of $1.00, minus the old price of $2.00, that is, negative one divided by the midpoint price which is $1.50, and that gives you negative one over $1.50 or negative two-thirds. So what we have here is two-fifths divided by negative two-thirds. Apply the absolute value sign and multiply out and you get three-fifths. Three-fifths is Angie’s elasticity of demand moving from A to B using the midpoint as the basis of your percentages.

Now, just as a check, find the elasticity of demand as you move from B back to A, again using the midpoint as the basis for your percentages. In this case, the new quantity is 20 minus the old quantity of 30 divided by 25 is negative two-fifths. In this case, the percentage change in price will be the new price of $2.00 minus the old price of $1.00 divided by $1.50 for a percentage change of two-thirds. Once again, applying the absolute value sign, and multiplying out those fractions, we get the elasticity of demand for Angie’s ice creams is three-fifths, or 0.6. Notice then it doesn’t matter whether you’re going from A to B or B to A. Once you’re using the midpoint formula, you get a consistent measure of the elasticity of demand in either direction.
Understanding Markets

Elasticity

Applying the Concept of Elasticity

Notice that in Angie’s case, the measure of the elasticity of demand is 0.6. That is, it is less than one. If the elasticity of demand for your product has an elasticity number of less than one, we say that the demand for your product is inelastic. That is, a given change in the price of your product results in a relatively small change in the quantity demanded. You can lower your price by a fair bit and not get a significant increase in the sales of your product.

In this case, what happens is, the smaller price has a big impact on your sales revenue. You’re making less money on each unit of the good that you sell. And since you’re not bringing in a lot of extra customers, your total revenue suffers. Anytime the demand for your product is inelastic, anytime your customers are relatively unresponsive to a change in the price of your product, total revenue will fall when the price of your good is decreased.

If we go back to Barney’s store, we'll see that his measure of elasticity was greater than one. In fact, it was one and two-sevenths. What that means is, a given change in the price of Barney’s product, or in Barney’s market, results in a relatively large response of customers. When Barney cuts his price by $1.00, he brings in a lot of new customers, and even though he’s making less money on each ice cream cone than he was before, he’s bringing in so many additional customers that his total revenue actually increases. Anytime the demand for your product has an elasticity measure greater than one, we say that the demand for your good is elastic. And in that case, a reduction in the price of your good generates a big response from your customers and total revenue increases.

So to summarize, if the demand for your product is inelastic, a reduction in the price of your good results in less total revenue. If the demand for your product is elastic, then reducing your price will increase your total revenue. If elasticity of demand for your product is exactly equal to one, then the change in price and the change in quantity demanded exactly offset each other, and there is no change in total revenue.
The elasticity of demand is the responsiveness of the quantity demanded to changes in the price of the good. We’ve discussed how to calculate the elasticity of demand, and we’ve also discussed the relationship between elasticity of demand and total revenue. One question remains. What determines the elasticity of demand for a particular product? What determines whether the demand for a product is elastic, that is, very responsive to changes in the price of the product or whether it’s relatively inelastic? That is, a large change in the price of the product may not produce a sizeable change in the quantity demanded. What’s the difference between goods with elastic demand and goods with inelastic demand?

But first, and probably most important determinant of the elasticity of demand is the available of close substitutes, products that have a lot of close substitutes are likely to have more elastic demand. For example, if one airline decides to raise the price of its tickets, customers can switch to another airline and fly to perhaps the same destination. The available of substitutes makes the demand for a particular airline ticket relatively elastic. The demand for a particular kind of soda might also be elastic. If one company raises the price of its soft drinks, customers may switch to another comparable drink.

Elasticity of demand often follows from the available of close substitutes. The more substitutes are available, the more elastic the demand for the product is likely to be. The more close substitutes, the more price responsive. Products that are inelastic in demand are likely to be those that have no close substitutes available, or few good close substitutes. For example, the demand for insulin to treat diabetes is usually viewed as inelastic. Whatever the price of insulin is, a diabetic is likely to pay it rather than do without because there are no good substitutes. However, even insulin is not a perfectly inelastic good. As the price of diabetic treatment goes higher and higher, some people who suffer from diabetes may look to alternative ways of managing their condition. They may choose to exercise more rigorously or watch their diet more closely as opposed to paying higher and higher prices for the medicine.

It’s difficult to come up with any product that’s going to be completely unresponsive, any product that has no substitutes at all available. Even though you need gasoline to make your car go, it may be that if the price of gasoline gets high enough, you will stop driving your car altogether and use mass transportation or walk or simply make fewer trips. The point I’m trying to make is this: if there are few close substitutes available for a product, the demand for that product is likely to be unresponsive to changes in the price of the good. When there are few close substitutes, people have to buy that product, and their demand for the product is unresponsive or inelastic. On the other hand, the more close substitutes there are available, the more likely you are to find an acceptable substitute, and therefore, the more price-responsive your demand will be.

The second determinant of the elasticity of demand is time. The more time you have during which to search for substitutes, the more likely your demand will be elastic. Suppose I’m building a house and I think I need lumber for framing my house and finishing it. If the price of lumber goes way up and I have to make a decision now, I’m likely to just swallow the extra cost and finish my house with lumber. But if I have two or three months to make a decision, two or three months to search for a substitute, then I’m more likely to hold out and look for a possible alternative. During those months producers may convince me that steel framing is a very good substitute for lumber. As lumber prices rise, producers will begin to offer customers alternatives.

Alternatives don’t come on line quickly, so the more time you have to get an alternative, the more likely you are to find an acceptable substitute, and the more price-responsive your demand for lumber will be. That is, the more time you have for making a decision, the more likely you are to find a good substitute, and therefore, the more price-responsive your demand will be.

The third determinant of the elasticity of demand is the percentage of your budget that is spent on the good in question. If you consider your elasticity of demand for bubble gum, your demand for bubble gum is probably relatively inelastic. That is, when the price of bubble gum goes up from two cents apiece to a nickel apiece, you probably don’t change the quantity that you consume during a week. Your demand for bubble gum is inelastic because bubble gum is not a big percentage of your budget. You may not even look at the price tag when you purchase gumballs.

On the other hand, if you consider your demand for housing, your housing cost is such a large percentage of your budget that if your apartment rent were to go up by 20 percent, you may have to look for another place to live or take in a roommate. Because housing occupies a large chunk of your budget, your quantity demanded will be responsive to a given percentage change in price.
Identification of Elasticity

Well, as a quick summary then, your demand for a product is likely to be more elastic if there are close substitutes available, if you have lots of time over which to search for substitutes, and if the products in question occupies a big chunk of your budget, so that price changes really require you to make some kind of adjustment. On the other hand, your demand for a particular product is likely to be inelastic or non-price-responsive if there are not good substitutes available, if you have to act quickly and don’t have time to search for substitutes, and finally, if the good in question occupies a small chunk of your budget so that you don’t have to change your behavior when the price changes.

These three factors are likely to influence whether the demand for product will be inelastic or elastic, and as you know, once you know which it is, you can say how a change in price will affect total revenue raised in that market. Inelastic demand means that lower prices shrink total revenue. Elastic demand means that lower prices increase total revenue.
Understanding Elasticity

Understanding the Relationship Between Total Revenue and Elasticity

Why should you spend all of this time and energy studying elasticity of demand? It’s mathematical, it’s not immediately intuitive, it’s difficult. What’s the payoff to you? The payoff is, once you understand the concept of elasticity of demand and how to calculate it, you have a useful tool for predicting the relationship between the price of a product and the total revenue that you earn when you sell your product at that price. Let me give you an example. Have you ever wondered by airline tickets go on sale all the time and appendectomies never do? The answer is this—if airline company ABC puts their airline tickets on sale, they will attract a lot of customers. They’ll attract customers from their competitors, and they will attract customers from people who didn’t even plan to travel. A small change in the price of tickets brings in a lot of new customers for ABC Airline. Their total revenue increases because the demand for their product is very price-sensitive.

On the other hand, the hospital doesn’t put its appendectomies on sale because if they cut the price they wouldn’t bring in a lot of customers. People don’t buy appendectomies because of the price they buy them because they need them. And if the price changes, the number of customers that come into the hospital may not change at all. You lower the price of an appendectomy and all you’re going to do is reduce the total revenue that you earn from doing appendectomies.

This goes back to an idea that we discussed. That is, the price sensitivity of demand. If demand is price-sensitive, a small change in price brings in a lot of new customers and total revenue increases. On the other hand, if demand is not price-sensitive, a big increase in the price of the product may not change your customer base at all. You may still sell just as many appendectomies and total revenue goes up when price goes up. The relationship between elasticity and total revenue is a good reason to learn about elasticity of demand. It’s a payoff for learning that difficult concept.

What I want to do now is show you in mathematical terms how elasticity of demand and total revenue are related. In order to do that I’m going to start with a mathematics trick and you can see this in the box next door. The mathematics trick is this—if you have the product of two numbers, A times B, and you want to know how they change, you can approximate the percentage change in the product, A times B, as equal to the percentage change in the first factor, A, plus the percentage change in the second factor, B. This is a good and very useful approximation that’s used all the time in economics. Now, I can’t prove this without doing some kind of heavy duty calculus with logarithms, and you don’t want to see me do that. So instead what I’m going to do is show you in a picture how the proof might look.

Suppose we have a diagram with A on the vertical axis and B on the horizontal axis. If you pick a combination of A and B and represent that with a point, the area that is seen in the diagram is the product of A times B. Now suppose you change that area by moving A and moving B. You can add to A by adding height to the box. When you add to B you add width to the box. And if you look at those two areas, you can see the percentage change in A and the percentage change in B. All that’s left out is the little piece up in the corner. But this lets you know that you can take on faith my assertion that the percentage change in the product is equal to the sum of the percentage changes in the components.

Now, let’s put that to work to understand how elasticity and total revenue are related. What is the formula for total revenue? Well, it’s the price of the product times the quantity that a firm sells, P times Q. What's the percentage change in total revenue going to be? Well, the percentage change in total revenue is going to be equal to the percentage change in the price plus the percentage change in the quantity demanded. How do I know that? Well, I just used my math trick. I used P as one variable and Q as the other variable.

Now, here’s a question I want to ask. If I change the price of the product—we know that quantity demanded changes in the opposite direction. That is, a percentage change in price has the opposite sign from the percentage change in quantity demanded. We know that because the demand curve is downward sloping. When the price goes up, people are willing and able to buy less of the product. What I want to know is, which of these two effects dominates total revenue. Does the price effect dominate? In that case, total revenue and price will move in the same direction. Or does the quantity effect dominate? If it does, then price and total revenue will move in opposite directions. Let's do a little bit of math here and we'll come up with a familiar expression. Let’s suppose then that the percentage change in price plus the percentage change in quantity demanded adds up to a number that’s greater than zero. And we’re going to suppose that that happens when the price of the good goes up.
Understanding the Relationship Between Total Revenue and Elasticity

Now, let’s take this equation and mess around with it a little bit. Let’s start by taking this term—percentage change in quantity, and moving it over to the other side of the inequality. Now, I’ve got to be real careful with my signs here. I can get confused. I decided that the percentage change in price was positive, and the percentage change in quantity demanded was negative, so since I’m moving it over to the other side of the inequality, the percentage change in quantity is now going to have a positive number. The numbers on both sides of the inequality are now positive numbers. And the inequality tells me that the price effect, the percentage change in price, is bigger than the percentage change in quantity demanded.

Now, divide both sides by the percentage change in the price of the product, and that gives you this fraction being less than one. Now, does that fraction look familiar to you? Have you see it before? You have. You have seen it before. That’s the formula for the price elasticity of demand. The percentage change in quantity that results from a percentage change in price. That negative sign that we wound up with outside of that fraction, that does the same job that the absolute value sign did before. It just made sure that the top and the bottom of that fraction have the same sign that makes this a positive number.

That’s the formula for the elasticity of demand. And if it’s less than one, we say that the demand for this product is inelastic, like the demand for appendectomies. If that’s the case, then price and total revenue move in the same direction. The company that has a product that has inelastic demand, when they raise the price of their product, what’s going to happen to total revenue? It’s going to go up. And why is it going to go up? Because the company is not going to lose customers. The customers are price-insensitive. They stay with this firm even though the firm is charging a higher price. And because the price effect dominates the quantity effect, total revenue increases.

Now the opposite is going to be the case if we decide that the quantity effect is greater. If the quantity effect is greater, then total revenue is going to move in the opposite direction of the change in price. Again, the intuition is this—the firm may try to raise their revenue by charging a higher price. Well, what happens? What happens is all those price-sensitive customers defect. They go and buy some other product, some substitute product, or they just stay home. Because the quantity effect is so big, because this firm loses so many customers at the higher price, total revenue actually shrinks. And if you manipulate my equations like I did before, you’ll see that the measure of elasticity in that case is greater than one. That is, our definition of a good whose demand is elastic.
Understanding Markets

Interfering With Markets

Understanding How Price Controls Damage Markets

Let's look again at the effect of a price control on a market. We'll start with the classic example, the market for rental housing in a city where there's a rent control. If we had a free market for apartment space, we would have people supplying apartments, more apartments being supplied as the price rises, and people demanding apartment space. The demand curve sloping downward, meaning fewer people are willing and able to buy apartments or rent them as the rental price of apartments goes higher and higher.

Suppose that a city votes to try to protect its low-income members by imposing a ceiling of the rental price of an apartment. We call this a rent control. So we put in a rent control at this level. This is a restriction, and I'll put a bar over the top of it to represent the price control. It can go no higher than this regulated level. If the equilibrium price of an apartment is $1,000.00 a month, the city may decide that that's too high and control the price to be $500.00 a month and no more. If that's the case, what would happen?

Well, one thing that could happen is that landlords and tenants simply avoid the rent control and find sneaky ways to get around it. The landlord might charge you $500.00 to rent an apartment, but also charge you another $500.00 to rent the key or the air-conditioner or the rugs. In that case you're really paying $1,000.00 a month for the apartment and we go right back to our free market equilibrium. That's the case where the rent control is ineffective and avoided. The second case would be where the rent control is actually strictly enforced and the landlord cannot get more than $500.00 for each apartment from a tenant. If that's the case, the situation actually becomes very costly.

The first thing that happens, as usual, with a price restriction is that all these landlords whose opportunity cost is greater than $500.00 a month will simply withdraw their apartments from the market and the total quantity of apartments rented will drop, from this amount right here with a free market, to this smaller quantity of apartment space. And we might call this Q bar. The restrictive quantity of apartment space that's available in the presence of the rent control. Well, what that tells you right off the bat is there's going to be some kind deadweight loss. You've got a restriction in trade and since the original situation was good in some sense, this restriction of trade cannot be an improvement.

The second problem is of course the problem of the rent seeking. Who all wants a apartment at a price of $500.00 a month? Well, look, all of these people, all of them react to the demand curve here. We've got a huge excess demand for apartment space. What usually happens when we have an excess demand? Usually what happens is the bidding mechanism pushes up the price of apartments until we return to equilibrium. In this case, however, the bidding mechanism has been blocked by the price control. Therefore, all of these people who want to rent an apartment for $500.00 a month will find themselves engaging in non-price competition. Now what does non-price competition mean in this market? It means trying to get to apartment ahead of other people. It means reading the obituaries in the morning to find out what apartments may have opened up overnight. It also means hiring kamikaze apartment locators. It means locators with very sharp elbows who have a talent for getting to the front of the line for you. That is, taking apartments away from other people to make sure that you get them.

Who can afford a kamikaze locator? Well, as these people, all of the these potential tenants begin to compete with each other and hire kamikaze locators, the price of locators will be bid up until the true cost of getting an apartment is equal to the P prime. P prime is a higher price, perhaps $1,300.00 a month. $1,300.00 a month includes the $500.00 rent plus an additional $800.00 that you have to pay to a locator to guarantee that you get the apartment. This is $800.00 a month, which you might think of as a payment that you owe your locator for having done this service for you. Well, here's the new equilibrium. Tenants pay $1,300 a month, $500.00 of it to landlords, and $800.00 of it to apartment locators, and we have a huge deadweight loss. All of this area here represents economic value that's been destroyed by the price control.

The price control restricts trade as apartments leave the market because landlords can't cover their opportunity cost. And all of these tenants who can't afford to keep up with the competition are priced out of the market. All of these trades are blocked by the price control. On top of that, you have all of these tenants who are paying money to these kamikaze locators. Now these are not your usual locators who just happen to have good information about where apartments are. These are people that you hire to go out and wrestle for you, people that you hire to go out and take apartments away from other competitors. They're not really adding any value to society. They're your own hired army that you hire to try to get a bigger share of the benefit for yourself. They're not creating value for society; in fact, they cost society value. Because instead of going out and wrestling for apartments, they could be hooking rugs, writing children's books, assembling computer boards, or doing other things that create value.
As they take their time away from value-creating activities and direct it instead into rent seeking activities, it becomes costly for society. And the amount of the cost is approximated by this area here, that is, the amount of money they get paid to attract them away from other valuable activities. If you add up the amount of rent seeking activity and the amount of deadweight loss, you can see that the price control is very costly to society. All that’s left in this market, instead of the big consumer surplus and producer surplus that we had before the price control, is a small amount of consumer surplus up here and a small amount of producer surplus down below. The price control has messed up this market for a deadweight loss and the cost of rent seeking activities.

That’s the story about a price control in the market for rental housing. The same story applies if you think about the market for tickets to concerts. You have a demand curve for concert tickets a supply curve for concerts and you have a price control often that’s imposed by the act itself, a band that wants to make sure that people get to come and watch it in concert. If there’s a price control on the tickets, we wind up with a similar situation. Tickets are sold at a price below the true equilibrium price. As a result, it may be that fewer concerts are produced. Producers of this act or of similar acts may decide that they can’t cover their opportunity cost of bringing more art for people to see.

Meanwhile, there are a lot of people who would like to see the concert but can’t afford it because the price of tickets has been bid up. Once we have a restrictive quantity to tickets people will resort to non-price competition. And you know what that means, standing in line or buying tickets from scalpers. This box right here represents the rate seeking activity, the cost to society of having people who could be writing children’s books standing on a street corner scalping tickets. Or the wasted time of people standing in line when they could be doing something productive. Now you might argue, what if I like to stand in line? Well, if you like to stand in line, you can stand in line, but the point is you shouldn’t be forced to stand in line to allocate tickets. It’s cheaper for society if people allocate tickets through the price mechanism.

If we want to help people who can’t afford tickets get tickets, perhaps we should give them a subsidy or increase their income. But using a price control does not help people in this model who have low incomes. It doesn’t help the people who can’t afford to pay as much for tickets. In fact, the people who wind up getting the tickets are the people who can afford not only the old equilibrium price but the higher price that scalpers will charge in this event.

The economist ends this story by saying that price controls rarely achieve their goals. Because price controls often create a situation of shortage, and on top of that, the pressure on non-price competition and the waste associated with it, price controls usually create a bigger mess than the problem they were intended to solve.
Understanding Markets

Interfering With Markets

Understanding the Problem of Minimum Wages in Labor Markets

Suppose you are concerned about the plight of low wage workers in the economy. You want to make sure that people who are working hard make enough money to take care of their basic needs or feed their family and therefore, you are lead to support a minimum wage. Now, frequently in Economics we study unanticipated consequences of economic policy. That is, we enact the policy for one reason, but the policy has unanticipated consequences that may actually work against the policy's intention. Economists like to make this argument about the minimum wage. That in fact minimum wage laws may in many cases hurt the very people that they are designed to help.

Now, there’s a lot of controversy these days about the minimum wage, whether it actually helps or hurts low wage workers. But what we’ll do in this lecture is look at the standard economic argument using supply and demand analysis for why the minimum wage has unanticipated consequences that may in fact harm low wage workers. We’re going to look at this story in our supply and demand diagram that we’re so familiar with. I want to make sure that we're clear on what all of the curves represent. Let’s suppose now, that we have price and quantity, and since this is the market for labor, the price will be the wage. This will be the hourly wage or the amount of money that must be paid per hour for labor services.

On the horizontal axis we’ll have the quantity of labor and this will be the quantity supplied by workers or demanded by employers. Now, putting these two together, we create a graph in which we can represent the market for labor. Let’s begin with the demand curve for labor. The demand curve for labor is downward sloping representing the fact that as labor gets less expensive, firms will hire more workers. Firms will hire more workers for two reasons. The first is, as the wage falls, the cost of doing business in declining and therefore the business may be inclined to expand its operations. Remember, when costs fall at the margin, businesses find it profitable to produce more of their output for sale on the market. So, as the wage falls, the business may actually be expanding and hiring more workers.

A second reason why lower wages lead to an increased quantity of labor demanded is the substitution between labor and other factors of production. That is, when labor gets less expensive, employers are inclined to use more labor and less capital to produce a given quantity of output. So lower wages result in a substitution in the direction of labor. The less expensive or the output that is declining in relative price. So, for two reasons, the costs of doing business are lower and labor is now relatively less expensive, lead to an increased quantity of labor demanded as the hourly wage falls.

Let’s look now at the other side of the market, that is, the supply curve for labor. And the supply curve for labor is a controversial graph. How do we know how workers are going to respond when the wage goes up. There are two things that workers might do. As the wage increases, we might expect first of all what we call in Economics a substitution effect. That is, with higher wages now, workers are inclined to work more hours, because the cost of leisure has increased. That is, if I have to choose between spending an hour sitting in the park reading a book and an hour working at the factory, well, I’m going to be considering the quantity of goods and services that I could purchase with the wage that I earn from that hour’s worth of work.

If the wage goes up now, then I’m giving up more goods and services to enjoy the hour sitting in the park. The cost of leisure to me, the opportunity cost of leisure, is rising. Therefore, I’m inclined to consume less leisure and enjoy more of the goods and services that I can purchase from that hour of work in the factory. Therefore, the substitution effect leads workers to work more hours when the hourly wage increases. On the other side of the argument, there is what we call the income effect. The income effect is the effect on the worker’s behavior when the worker becomes richer.

After all, if the wage rises then you can now afford to pay your bills by working fewer hours. If I have to pay $500 worth of bills every week, then if I can earn $500 from 20 hours worth of work instead of 40, then I'll work my 20 hours and spend the rest of the time having fun. That’s because at a higher wage, I’m now wealthier, I’m now richer for a given amount of work and I’m inclined to purchase more leisure. That is, as my wealth increases, as my income rises, as I get richer, I’m going to want more of all goods, including leisure, and therefore I will cut back my hours at work. Ask yourself this question. What would you do if your boss at work doubled your wage? Would you work more hours or fewer hours?

Well, it depends on whether the substitution effect is stronger for you or whether the income effect is stronger for you. If the income effect is stronger, then you would cut back your hours and enjoy more leisure or study time. If the substitution effect is stronger, you would work longer hours so that you can afford to buy more toys and goods and services and other things that you enjoy that money can buy. So, the effect of a higher wage on household behavior...
Understanding Markets

Understanding the Problem of Minimum Wages in Labor Markets

is ambiguous. It depends on whether the substitution effect is stronger or whether the income effect is stronger. But for the sake of our argument, we're going to draw the supply curve for labor as if it is upward sloping. That is, I'm going to draw this picture as if the substitution effect were stronger. That is, as if higher wages lead workers to work longer hours.

So, here I go, I'll make sure this line is real clear. There's my substitution effect dominating. There's my supply curve for labor. So, the demand curve represents the behavior of firms. They demand a larger quantity of labor when the wage falls and the supply curve represents the behavior of households. They supply more labor hours as the wage rises. Suppose we have an unregulated market. In a free market, equilibrium would be established where the quantity supplied and the quantity demanded are equal. That is, at the intersection of the blue and the red curves. If we drop our line down here, we'll see this is the quantity of labor that would be supplied and demanded in the labor market at equilibrium and the wage at which the market clears would be W star.

So this might be something like $5.00 an hour and this might be 100 hours a week, in a particular labor market. Well, suppose we decide, as a society, that $5.00 an hour is too little for a worker to be paid. That in order to make a living wage, workers have to earn at least $6.00 an hour and we calculate that through studies and logical reasoning and looking at the prices of food and housing and medical care. And we want to establish then a minimum wage that's higher than the equilibrium wage. Now, the minimum wage is a floor below which the wage is not legally permitted to fall. That is, if a minimum wage is established, then all businesses have to pay a wage that is at least as great as the minimum wage.

So, we'll put the minimum wage up here. We'll call this W_m for minimum wage. And it's important that the minimum wage is above the equilibrium. After all, if it's below the equilibrium, then it's irrelevant, because market forces are going to bid the wage up to at least $5.00 an hour. It's only if your minimum wage is greater than $5.00 an hour that it makes any sense at all. So, suppose now that we have legislated a minimum wage. That is, all workers can expect to receive this wage if they have a job and all employers are required by law to pay it. What would the consequence of a minimum wage be? Well, if we follow the minimum wage over to check on the behavior of the players in this competitive market, we see first of all that the prospect of earning $6.00 an hour draws a lot more households into the labor market.

The quantity of labor supplied increases with the minimum wage. That is, the amount of labor that households want to supply at this higher wage is greater than it was before. So, we'll call this QS_m. That's the quantity of labor supplied under a minimum wage. On the other hand, firms are less interested in hiring workers when they have to pay this higher wage. Therefore, the quantity of labor demanded is reduced. And it's reduced for two reasons in this story. The same two reasons that the demand curve for labor is downward sloping. The first is that the minimum wage raises the overall cost to the firm of doing business and therefore, inclines the firm to cut back its output. Therefore, it needs fewer workers. The second problem for the workers is that the higher wage leads firms to substitute away from labor in the direction of capital or other factors of production.

With the higher wage and less labor being demanded, or a smaller quantity of labor demanded, then we have Q equals the quantity of labor demanded under the minimum wage being less than the quantity supplied. Ah, here's the unintended consequence. With the minimum wage above the equilibrium, the quantity of labor supplied exceeds the quantity of labor demanded. We have therefore, an excess supply of labor and that has a special name. We call that unemployment. There are a lot of people who want jobs at the minimum wage who are not going to have the opportunity to get them because firms want hire the workers. That is, the quantity of labor supplied exceeds the quantity of labor demanded and the distance between these two points constitutes unemployment. Workers who are willing and able to work at the going wage, but cannot find jobs because the firms are not hiring. This is the unintended consequence.

Well, what's going to happen then? Who's going to actually get these jobs? If there are only 50 jobs being offered but 150 applicants, how are we going to sort out which workers actually get them? Well, all kinds of things could happen. First of all, notice that we've got 50 workers who are willing to work for a much lower wage down here. If we look at the actual supply curve, we can get 50 workers for a wage down here of maybe $3.00 an hour. Now, legally we have to pay those workers $6.00 an hour, but we could ask those workers to do other things. We could make sure that they wash their uniforms themselves. We can make sure that they get transportation to and from work, perhaps riding with
Understanding Markets

Interfering With Markets

**Understanding the Problem of Minimum Wages in Labor Markets**

their parents or other workers or whatever. So, ultimately firms might be able to pass costs on to workers that lower their effective return from the job down to something like $3.00.

These workers wouldn’t apply in those circumstances, because they couldn’t make enough money to cover their opportunity cost. They’d rather work at home or be out of the labor force, perhaps taking care of children or some kind of self-employment. But these workers down here would work for the equivalent of $3.00 an hour and therefore even though they’re being paid $6.00 an hour, the firm is able to impose extra duties on them that reduce their effective pay down to $3.00 an hour. The economist has been concerned because look, all of these jobs that should be created where the benefit to firms is greater than the cost to the workers, all of these jobs are shut down by the minimum wage. The price regulation creates a deadweight loss and a kind of non-price competition. Those workers who are willing to stand in line the longest, wash their uniforms the cleanest, make sure they get transportation to work, or any other arbitrary conditions that the employer might like to impose, they’re ones who get the jobs. And their effective return is lower than the minimum wage.

So, to summarize. The minimum wage was adopted in order to help low income workers or low wage workers, but in fact there are two unintended consequences that work against the very people that the policy is designed to help. The first is that with a minimum wage, employers may reduce their demand for labor and when the quantity demanded of labor falls, unemployment is created as there is an excess supply. The workers want to work at the higher wage, but the jobs aren’t available. The second consequence is that there may be some kind of non-price competition. Employers imposing extra burden on workers so as to lower their effective wage down to somewhere even below the equilibrium wage that we started with, due to extra duties. Like cleaning uniforms or providing transportation or other things.

So, in summary, the minimum wage is an example that economists like to use to show how policies that are well meaning may actually have unintended consequences that actually hurt the very people they are designed to help. So, why then, are labor unions often in favor of minimum wages? We’ll see this in the next unit when we consider the relationship between wages and productivity.
Understanding Markets

Interfering With Markets

Understanding How an Excise Tax Affects Equilibrium

We’re still talking about taxes. Now the question is how does an excise tax affect the market equilibrium? We’re going to use the supply and a demand diagram to show how the existence of an excise tax influences the price and quantity traded and how it effects the amount of economic value that’s created in the market. We start with our usual demand curve and supply curve. Now if we had a free market, a market that had no interference, no taxes in it, the equilibrium would be where the curves cross. The equilibrium price would be here and the equilibrium quantity would be right here.

However, if we introduce an excise tax that’s going to mess things up a bit. You can show the excise tax in this diagram in either of three ways. Let me go through those three ways in turn. The first way that you can represent an excise tax is you can think of it as an additional cost that the sellers have to bear. If the seller has to pay the excise tax, say the baker has to give the government $2.00 for every loaf of bread that he sells, then that’s like an extra cost. It increases his opportunity cost of providing the bread for sale. So, for every loaf of bread we have to go through and add on $2.00 to the opportunity cost of providing it. What that’s going to do is it’s going to shift up the supply curve. It’s going to shift up the cost curve, the opportunity cost curve by the amount of the excise tax. So, I could do that like this.

It’s a parallel shift. The whole curve moves upwards by $2.00 at every point to represent the excise tax as an extra cost. In that case, the new equilibrium is going to be where this curve, the tax distorted curve, where the supply curve plus tax crosses the demand curve. The equilibrium price will be here, the equilibrium quantity will be down below. And the difference between these two curve is the difference between the price the buyers pay and the price that the seller gets to keep, after the tax. So, we can show the excise tax as a shift in the supply curve. That is, you can think about it as an extra cost for sellers.

The second way that we could represent the excise tax is we can show it as a cost that’s imposed on buyers. Suppose buyers have to pay the excise tax. That is, suppose when you go in to buy a loaf of bread you have to report it to the government and pay $2.00 to the government for every loaf of bread you buy from the baker. In that case, what happens is that the buyer lowers the price that he’s willing to pay the baker. If the loaf of bread is worth $4.00 to you and you’ve got to give $2.00 to the government every time you buy a loaf of bread, then you’re only willing to give $2.00 to the baker himself. The government got $2.00 of your $4.00 reservation price.

So we can think of this excise tax as a cost that’s imposed on the buyers of bread and we could show that by moving the demand curve downward by $2.00 at every point. By moving it downward we’re saying that the buyer is only willing to give the baker an amount that’s $2.00 less than what he’s willing to pay, because the government is going to take $2.00 of his price. So we could show it this way. The demand curve shifts downwards by $2.00 to give us what we might call a tax distorted demand curve, or the demand curve minus the tax. This is what the buyer is willing to pay to the seller.

In this case, our equilibrium price would be here at the intersection, and the equilibrium quantity traded would be here. Notice that this tells us the price the seller actually gets to keep when the buyer comes in to buy the bread. The buyer, however, is paying $2.00 more for the bread. It just so happens that $2.00 is going to the government.

Now before I show you the third way, let me point out something interesting about these first two ways of representing an excise tax. And that is, they’re completely equivalent. It doesn’t matter whether you shift the demand curve downwards by $2.00 at every point or whether you shift the supply curve upwards by $2.00 at every point. From the point of view of geometry they’re exactly the same thing. It doesn’t matter whether you’re moving this curve downward or this curve upward, you’re still going to have the same point of intersection. You’re going to have the same quantity traded and you’re going to have the same price paid by the buyer and received by the seller.

We call this result the irrelevance of the legal incidence of a tax. What that fancy phrase means is this. It doesn’t matter whether legally the tax is imposed on the sellers so that it shifts the supply curve upwards, or whether the tax is legally imposed on the buyers so that it shifts the demand curve downwards. It doesn’t matter who legally is responsible for paying the tax. In the end the equilibrium price, quantity and division of economic values are going to be the same, no matter who legally pays the tax.

So, since it doesn’t matter whether we shift the supply curve upwards or the demand curve downwards there’s a third way that we can show the tax that works in any case. That is, that points out and reminds us that the legal incidence
Understanding Markets

Interfering With Markets

Understanding How an Excise Tax Affects Equilibrium

is irrelevant. And that is, we can think about this tax as a wedge. It's a wedge that has to fit between these two curves. The only trades that can take place in this market, the only trades that can take place in this particular market for bread are those trades where the buyers are willing to pay enough so that after the tax is subtracted the sellers can cover their opportunity cost. What I do then, is I look for the place where the tax wedge just fits between the demand curve and the supply curve and that is going to be my equilibrium. That is, the end of trades that can take place when the excise tax is imposed.

So I slide my wedge right here until it just fits between the two curves. And I put a dot here, and a dot here to represent the equilibrium. The equilibrium, with this tax wedge, what we have is buyers paying this high price here, we'll call this the demand price, sellers receiving this lower price, $P_s$ we'll call this the seller's price and the gap or the difference between the two, is going to be the tax wedge or the amount of money that's raised in taxes. The equilibrium quantity traded with the tax, we'll call the $Q_t$ is the quantity that just leaves room for the tax wedge to fit between the buyer's price and the seller's price.

Now, that's the same thing, the equilibrium when I calculate it here, is exactly the same thing as if I had shifted the demand curve downwards by the $2.00 tax, my intersection would be at this point. The sellers would get this price, the buyers will be paying this price plus the $2.00 tax and the quantity traded would be at this point on the axis. This equilibrium is also the same thing as if I had shifted the supply curve upwards by $2.00. In that case, we'd get this as our equilibrium price, what the buyers have to pay. But what the sellers get to keep is $2.00 less because they are responsible for the tax.

So, to summarize, there are three ways that you can represent an excise tax in a supply and demand diagram. You can shift the supply curve upwards and think of it as an extra cost for sellers. You can shift the demand curve downwards and think of it as a cost that's imposed on the buyers, if the buyers have to pay the tax. Or you can simply think of it as a wedge that's imposed between the buyers and the sellers that separates them, so that the price that the buyers pay is greater than the price that the sellers get to keep and the difference between those two prices is the amount of the excise tax. Notice however, no matter which of the three methods you use you're going to get the same outcome. The same quantity traded the same price that buyers pay and the same price that sellers receive.
Understanding Markets

Agriculture Economics

Examining Problems in Agricultural Economics

Consider the peculiar characteristics of a market where the demand is price inelastic. We’ve already seen that when the demand is price inelastic, there is a direct relationship between the price of the good and the total revenue that is earned in that market. Consider, however, in addition to this inelastic demand that supply is highly variable; that is, the supply curve shifts around a lot, say in response to changes in the weather. When you’ve got inelastic demand, plus shifts in the supply curve, you wind up with a situation where total revenue itself becomes highly variable, and this can become a problem, particularly if you’re a farmer.

Most agricultural products have price inelastic demand; that is, people are going to buy their food products whether they’re expensive, or whether they’re cheap – you don’t really have much choice. So if we take an agricultural market, this is the situation that characterizes it. What’s more, not only is the demand for food price inelastic, it’s also income inelastic as well. When you get richer, you don’t tend to spend a lot more money on food; you’ll spend your money on vacations, or health care, or automobiles, or something else instead. So things that shift the supply curve – say changes in productivity, technological advance, other stuff like that, that tend to make people wealthier – don’t have a big effect on the demand curve for food. So shifts in the supply curve, then, cause the price of agricultural products to vary quite a bit. And with this variability in price, comes variability in total revenue, and that means variability in the income of farmers.

From the point of view of a farmer, this is a problem, and it’s a cause for government intervention. We’ve had agricultural markets forever, and back well before there were well-functioning insurance markets, farmers faced this problem of price and income variability, so they encouraged governments to step in and implement policies that would stabilize income, or failing that, at least stabilize the price of their products. We have now in the United States a vast array of agricultural policies designed to deal with this problem, and that’s why economists like to talk about agricultural markets when it comes to a discussion of government interference in the market – its causes and its consequences, both intended and unintended.

So let’s look now at how an agricultural market works, and how particular government policies may end up actually making the situation worse. Consider, first of all, the demand for corn. When the price is high, the total revenue is going to be price times quantity – the area of this green box. If the price is low, on the other hand, we get a smaller total revenue – the area of the purple box. Because demand is inelastic, price and total revenue will move in the same direction.

Now let’s consider what’s happening in the market that causes the price to change. That’s going to be caused by shifts in the supply curve. Farmers are going to make their plans based on some average supply curve, where the market would presumably through the bidding mechanism give us some average price and some average quantity. But the average thing almost never happens. You can have a bad year. A bad year means a year where weather is unfavorable – floods and draughts – which shifts in the supply curve. At any given price, you get less corn than the farmers expected to get. The odd and paradoxical thing is that a year that’s bad for weather is actually a year that’s good for total revenue. When farmers have a bad year, they actually end up making more money. It’s almost like a monopolist that restricts the quantity so as to jack up the price. Well, competitive farmers can’t collude like that, but if the weather forces a scarcity on them, it becomes an occasion for the bidding mechanism to provide a higher price and a bigger total revenue.

On the other hand, if the farmers had what we might call a good year, the supply curve shifts outwards. That is, a bumper crop, favorable weather, lots of corn – the market is flooded with food. However, this is bad for farmers, because the lower price winds up giving us a smaller total revenue in this market where the demand is inelastic. So the first paradox about agricultural markets is that a good year is bad for farmers, and a bad year is good for farmers if farmers are really concerned about their total revenue, or their earnings.

So what can we do about a situation like this? What can we do about a situation where shifts in the supply curve create price instability, and price instability creates instability in total revenue? Well, one thing that we could try to do is to stabilize the price. Suppose we decided that the price is going to be stabilized at some notion of average price of corn. That is, let’s draw in here a supply curve that represents the average of the good years and the bad years, and let’s look at what the price would be in a case that we had this outcome. So supply and demand crossing at this point gives us this, and we’ll call this $P^*$, which represents our average price. And I won’t clutter the diagram by putting in the average quantity, but it’s certainly going to be down here on the horizontal axis.
Examining Problems in Agricultural Economics

Well, of course, no year is truly average. So what’s going to happen is that in a bad year, a year with bad weather, when the supply curve would shift inwards – so we have here a supply in the bad year, so I’ll call this “bad” – we have a higher equilibrium price. But if the government is going to support the price – that is, keep the price at $P^*$ – what’s going to happen in these years is we’ve got an excess demand. That is, farmers are only supplying this small quantity down here, but people who want corn flakes and corn meal and the agricultural product are going to want this larger quantity at the supported price.

So the government has to stand ready to provide enough corn to make up for the excess demand. Where is the government going to get the corn that it sells to keep the price constant? It’s going to get it by buying corn in years of bumper crops. So in the years of the bumper crop – here is our supply curve in a good year – if the government keeps the price at $P^*$, then the farmers are going to find that they’re not able to sell corn at $P^*$ to all the people that they have corn for. In fact, the quantity supplied is going to be much larger than the quantity demanded – we have an excess supply of corn at a price of $P^*$. So the government takes that excess supply, fills up its barns and silos, and stands ready to pay farmers this price – $P^*$ – which farmers can now count on.

Now we’ve got a bunch of corn sitting around and we don’t know what to do with it. We could give it away in the United States, but that would push down the price of corn and complicate our farmers’ problem. What we do then is we export it; we send it to some other market, like Africa, and we can even send it there as food relief. That sounds great and humanitarian until you consider the fact that all of this free food flooding into Africa distorts incentives in those agricultural markets. Farmers counting on free food from the U.S. and Europe aren’t inclined to plant food themselves, because they know they can’t make a profit for it, because their price is going to be low because of all of this free food aid. And then, anytime the free food aid is disrupted, there isn’t food there to feed the local population, and famines can result. So our attempts to get rid of these chronic surpluses by dumping them in other markets creates problems for other countries – countries where the needs may be more desperate.

So let’s suppose then we don’t go with a flat-out price support – a guaranteed price for farmers – but instead we try to restrict the quantity supplied so as to keep the price higher. How can we restrict supply when we’ve got thousands and thousands of competitive farmers out there each trying to make a profit by selling more food? It’s kind of like the problem OPEC has of controlling oil producers. Well, one possibility would be set-aside programs, where we tell farmers you can only plant corn on so many acres, and the farmers agree that this is a condition of their support from the government – that they have to go along. Well, if they do, they’re just going to take their least productive land out of agriculture, and hypercultivate the most productive land. That is, they’re going to use a lot more labor, a lot more fertilizer, to try to grow more corn on limited parcels of ground.

Well, this just raises the cost of producing any given quantity of corn, because we’re forcing farmers to substitute away from land, and instead to use more expensive labor and fertilizer that they wouldn’t use if it weren’t for this government policy. So set-aside programs introduce another distortion, which are inefficient techniques of production. Another possibility besides set-aside programs are marketing arrangements, where farmers agree only to bring so much corn to market each year; that is, they agree to collude and shift the supply curve in so as to keep prices up.

Well, these agreements are subject to all kinds of cheating; people are going to look for other ways of selling their products on the side, so they’re kind of hard to enforce. But even if they were perfectly enforced, that means there’s going to be a lot of waste, because if the amount that farmers are allowed to sell is limited to this amount right here – $QH$ – which gives us the high price, then in a year of a bumper crop, we’ve got a lot of oranges and corn and other stuff that we just simply have to destroy, because otherwise it would be floating around in the market, depressing the price. So these marketing agreements can be very wasteful, unless farmers want to get into the business of storing their products, which may or may not be feasible.
Examining Problems in Agricultural Economics

A final way of dealing with the problem that farmers face is to directly support their incomes, and this is what the economists would say to begin with. The problem is, farmers' incomes are too variable. If we want to help the farmers, then let’s stabilize their incomes. Well, the economists’ first solution would be to have these farmers buy insurance; the insurance market can solve this problem. But short of that solution, if the government is going to provide some kind of insurance substitute, then it should give farmers money to supplement their incomes in years when prices are low, and tax the farmers to cover the cost of this program in years where prices are unexpectedly high. And this is the basic idea behind the Federal Agricultural Improvement and Reform Act of 1996. This act is designed to get the government out of the business of price supports and running storage programs, and instead to provide income for farmers to help them stabilize the difference between years when output is low and prices are high – and farmer incomes are typically good – and years when bumper crops push prices down and provide farmers with less income.

So although we can be critical of the act – I mean, there are some provisions that economists would say are not promoting efficiency – it is a step in the right direction by getting the government out of the business of tinkering with prices, and putting it in the business of redistribution of wealth, which is something that the government can do in a society where we want to protect the incomes of people who are otherwise at the mercy of the weather. So agricultural markets are a great example of the government interfering to try to make the world work better. But, as is usually the case, there are unintended consequences of each of these government policies, and the unintended consequences, in many cases, wind up creating more problems than the problems they were originally designed to solve. The moral of the story is: If you've got a distortion in the market, try to find the policy that's crafted so as to directly address that distortion, without introducing new ones.
Utility Theory

Understanding Utility Theory

Utility theory is the tool economists use to describe the way households allocate their limited budget across the array of goods and services that provide satisfaction. Why is it that most consumption bundles are balanced? They contain some apples, some oranges, some milk, some eggs. Why don’t consumers spend all of their income on one particular good or service that’s especially satisfying, and buy very little of anything else?

Utility theory can help explain the balance in people’s consumption bundles. Utility theory has at its root a unit of measure called the “util,” and the util is an increment of satisfaction. For instance, if I have a potato chip – and I do like them – I might measure my satisfaction by creating this fictional measure of happiness called the util, and the first potato chip that I eat adds ten utils to my satisfaction.

Think of this as a kind of psychic payoff from the pleasure of eating a chip. The marginal utility of consuming a potato chip is the additional utility – the extra number of utils that are added to my total satisfaction when I consume that first potato chip. So its marginal, or psychic satisfaction is equal to ten. That means if I eat one potato chip, my total utility is ten utils – it’s so good. Now, suppose I decide to have a second potato chip. I’m sorry – it’s so good, but not quite as good as the first one. I mean, after all, marginal utility does diminish, and the second potato chip is never quite as good as the first one, even though it’s still quite delightful. The total utility that I get from eating two potato chips is 10 for the first plus 8 for the second, for a total utility of 18. Of course, no one can eat just two, right? The third chip adds only 6 utils to my total satisfaction, bringing the complete level up to 24. I mean, it’s better than nothing, but not quite as good as the first three. My total utility, however, is still increasing, for a while. I mean, it was there, right? So I ate it, but I can’t say it was especially enjoyable.

Okay, can I have a glass of water? I mean, is the point clear? More is better, up to a point. But each additional unit provides a little less satisfaction than the unit before, after some point. Consumption is subject to diminishing marginal utility. Each additional unit that you consume in a set period of time – other things held constant – is going to provide less satisfaction than the unit before. This means that, while more of everything may be better, eating more potato chips at any given point of time, after some point, becomes a bit of a chore. Marginal utility can fall to zero, and even become negative.

So now we’ll show how to apply this tool as we calculate the way consumers choose to spread their income out over its competing uses. But first, I’m going to take a break.
We’re ready to tackle the problem of consumer choice. That is, how will a household allocate its scarce income across the goods and services that provide it with satisfaction? For instance, how much of our snack budget will we spend on apples, and how much will we spend on candy bars? The answer to this problem hinges on utility theory. And it turns out that the rule the household uses to allocate its income across competing uses is this one: The household will continue to shift its money between spending on apples and spending on candy bars, until the marginal utility for extra dollars spent is equal across the two goods.

Let me give a simple example to show how this principle works. Suppose we’re going shopping and we have a limited budget – say $6.00 – to spend on apples and candy. How much will we spend on each? How many candy bars will we buy, and how many apples? Suppose we start putting candy bars into our shopping basket. It cost $1.00 to buy a candy bar, and the first candy bar that we put into our shopping basket gives us a psychic satisfaction of 10 utils. So right now we’re getting 10 utils per dollar from our purchase of chocolate. Suppose we also put an apple into our shopping basket, and the first apple that goes in provides us with psychic satisfaction of 8 utils. Now since apples only cost $.50, these 8 utils average out to 16 utils per dollar. That is, at the moment, we’re getting more marginal utility per dollar spent on apples than we are on chocolate.

So let’s go ahead and put another apple into our shopping basket, since they’re such a good deal at providing us with satisfaction. The second apple also costs $.50 and it provides us with a little bit less psychic satisfaction than before – only 6 utils. Well, 6 utils for $.50 comes out to 12 utils per dollar, which is still more than we’re getting from chocolate, so we’ll go ahead and buy another apple. This third apple provides us with an extra utility of only 3 utils. So, 3 utils for $.50 amounts to 6 utils per dollar. Now chocolate looks like a better deal, so let’s go ahead and buy another candy bar. For $1.00 spent, this next candy bar provides us with psychic satisfaction of 8 utils; 8 utils per dollar is better than 6 utils per dollar, so chocolate still looks like the best buy. Keep spending on chocolate.

Here we spend another dollar, and the third chocolate bar provides us with a marginal utility of 5. So now we’re getting 5 utils per dollar from chocolate, 6 utils per dollar from apples; that means apples are looking good again. Let’s put another apple into our basket. And this apple is going to give us a psychic satisfaction of 1 util; 1 util for $.50 spent is 2 utils per dollar, so now chocolate is now the preferred way of spending our snack money once again. Spend another dollar on a candy bar, and that gives us a psychic satisfaction of 2 utils; 2 utils per dollar on chocolate is equal to 1 over $.50, or 2 utils per dollar on apples.

So now at the margin an extra dollar spent on either snack is providing us with the same amount of satisfaction. Over here we’d get 2 utils from spending $1.00 on apples, and over here we’d get 2 utils from spending $1.00 on chocolate. Good thing that we equilibrated the two, because now we’re out of money. That is, the consumer is going to keep shifting money between the two until marginal utility divided by price is equal for the two goods. There’s no reason for us to do without an apple so we can have more chocolate, because then apples would be providing more satisfaction than chocolate at the margin; we’d want to go right back to what we were doing before. And there’s no reason for us to do without a candy bar to have an extra apple, because the candy bar we’d be giving up is providing us with more satisfaction per dollar spent than an extra apple would.

If we add up the money spent – $.50, $.50, $.50, and $.50, plus $1.00, $1.00, $1.00, $1.00 – we’ve now spent our entire snack budget of $6.00, and we’ve allocated it across apples and across chocolate in such a way that the marginal utility for an extra apple divided by the price is equal to the marginal utility of an extra chocolate bar divided by the price. We spent our limited income, and we spent it across two goods in such a way that an extra dollar spent provides us with the same extra satisfaction, whether we’re buying apples or chocolate. This is the rule that households use to maximize satisfaction from a limited amount of income.
Imagine that you are in the coolest store in the world, with all of your favorite stuff. I mean, we've got the Ping-Pong ball pistol, you know, we've got, umm cupcakes. What are these called? Butter cream. We've got moon pies. We even have a crocodile whistle. What are you going to do? You say, “I want it all.” Well, of course you do. You're an insatiable consumer. But, there's a problem. The problem is you've only got so much money to spend and this stuff ain't free. You either have to pay three bucks to get the punching monkey. You're going to have to a buck for a package of circus peanuts.

How are you going to spend your limited budget, when you have to choose between Krispee Cremes and squeaky nuns? Well, it's tough. So we're going to help you by introducing some economic tools that we use to describe consumer choice. The first set of tools describe what's possible, or feasible. These are the tools that describe the amount of money you've got to spend and the combinations of toys and snacks that you can afford on your limited budget. That is, we're going to start with what we call “The Budget Constraint.” Then we're going to move to a representation of your preferences. That is, how you would make tradeoffs between bunny ball and snacks. That is, we will come up with a set of indifference curves that represent your preferences for toys and snacks.

Then we'll put the curves together with the constraints, we'll put what you want together with what you can get and come up with a prediction of how you'll shop. That is, we're going to derive a demand curve and show how you respond when you're given a set of constraints and a set of opportunities and we'll use that to connect it with a tool, the demand, curve that you're already familiar with. So, let's get started.

The first thing we're going to do is clear away these infinite possibilities and focus on what's feasible for you, given the constraint of your income and the price of the products that you want to buy. So, let me do that. Get a little space here. Do a little regular math work. And the first thing we're going to is derive a budget constraint. So, let me put some axes here in my picture. Whenever you're doing consumer choice, you're going to be drawing graphs that we describe as goods space. Goods space means you're measuring quantities of goods on each of the two axes. So on the vertical axis we're going to measure snacks and on the horizontal axis we're going to measure choice. And this is how you make your tradeoff between toys and snacks, is you choose a point in this space.

For example, this point here might represent two toys and two snacks and this point here might represent 12 toys and four snacks. Sorry. I'm wrong aren't I? Twelve snacks and four toys. So we've got to be careful. The horizontal coordinate tells you how many toys you've got and the vertical coordinate tells you how many snacks you've got. Now that we've got this picture of goods space, we're going to use it to represent your budget constraint. That is, what are all of the combinations of snacks and toys that you can afford to purchase given your limited income?

Well, if you want to put together a budget constraint, you're going to need some information. The first piece of information you need is how much money do you have to spend in the store. Let's suppose you have $12.00 to spend, $12.00 to spend on whatever combination of snacks and toys you choose to purchase. The second piece of information that we're going to need is the price of a toy. And let's say the toys are selling for $3.00 apiece. So the price of a toy is $3.00. Let's say that the price of a snack is $1.00. That's the third piece of information you need, how much do the snacks cost. Well, now that you've got these three pieces of information, your total income or budget, your price of a snack and your price of a toy, you can describe all of those combinations of toys and snacks that you can afford to purchase.

I like to start by finding the endpoints on my budget constraint. Suppose you spend all of your income on snacks. How many snacks will you be able to buy? Well, you have $12.00 worth of income and snacks cost a dollar apiece and that's going to give us a total snack purchasing power of 12 snacks. If you converted all of your income into snacks, you could buy 12 snacks. Now let's suppose, instead you bought only toys. How many toys could you afford? Well, with a budget of $12.00 and a price per toy of $3.00, then you could afford to buy 12 divided by three equals four toys. So the endpoints are 12 snacks and four toys.

Now, what about combinations that have some of each in them. Let's suppose you decided you wanted to buy two toys. How many snacks could you afford to buy? Well, two toys would cost you two times $3.00 equals $6.00, subtract six from 12, that leaves you $6.00 to spend on snacks and since snacks are a dollar apiece you can afford to buy six of them. So one point on your budget constraint would be two toys and six snacks, for a point like this. Let's calculate another point on the budget constraint. Suppose now we only buy one toy. How many snacks can you afford to buy? You calculate it.
Constructing a Consumer’s Budget Constraint

One toy costs $3.00. Subtract $3.00 from your budget of $12.00 that leaves $9.00 to spend on snacks. And since snacks cost $1.00 each, you can afford to buy nine snacks, another point on the budget constraint. One more point. Suppose you buy three toys. How many snacks can you afford to purchase? You calculate it. With three toys, you can afford to buy three snacks. Three toys cost you, three times $3.00 is $9.00, subtract it from $12.00, leaves you $3.00 for snacks and since snacks are a dollar apiece, you can afford then three snacks.

If we kept doing this we would be generating more and more combinations of toys and snacks that you can afford given their prices and given your $12.00 worth of income. Eventually, when we connect these dots we will get the budget constraint. The budget constraint would be the points, the budget constraint is the collection of points that shows all the combinations of toys and snacks that the consumer can afford to purchase given their fixed income and the prices of the goods. That is, all of these points along this line represent feasible combinations of toys and snacks. Combinations of toys and snacks that you can afford.

The way to represent this line is to manipulate the budget constraint. Let’s write the budget constraint down and then let’s change it algebraically. The budget constraint says that the income that you have to spend, N, will be equal to the price of toys multiplied by the number of toys that you buy plus the price of snacks multiplied by the number of snacks that you buy. Well, if snacks are on the vertical axis, we want to isolate our snack variable in terms of all the other variables. The first thing that we’ll do is we’ll move toys to the other side of the equation. N minus the spending that you do on toys is equal to the amount of money that’s left to spend on snacks.

Now, divide both sides of this equation by the price of snacks and I’m also going to move snacks to the other side of the equal sign and I get this. The total number of snacks that I can afford to buy is equal to my income, divided by the price of snacks minus the price of toys, divided by the price of snacks, multiplied by the number of toys that I buy. This equation that I’ve drawn is the equation that I’ve graphed here in goods space. This is the equation for this line, which is the budget constraint. The way to understand this equation is this. The total number of snacks that you’re able to afford is equal to what you would be able to afford if you spent your entire income on snacks, that is, 12 divided by the price of snacks or the dollar, for a total of 12 snacks, minus the number of toys that you buy, multiplied by the relative price of toys and snacks.

That is, since toys are $3.00 apiece and snacks are only $1.00 apiece, each toy that you purchase is going to cost you three snacks. You have to give up three snacks to get each toy. The negative sign lets you know that the line is downward sloping. But this relative price trend, the price of toys divided by the price of snacks is important, because it’s the opportunity cost of getting another toy. The opportunity cost of buying another toy is this number right here or the three snacks that you have to give up to get that toy.
Consumer Choice and Household Behavior
Budget Constraints and Indifference Curves

Understanding a Change in the Budget Constraint

We’re considering how to represent your company in the world’s coolest store. You have a limited amount of money
to spend on toys and snacks and you want to spend it so as to get the most satisfaction. Last time, we looked at this
equation. This is the equation for the budget constraint. It tells you how many snacks you can afford to buy given
your income, the price of snacks, the price of toys, and the number of toys that you choose. When we graphed this
we got the vertical intercept, which is the purchasing power of your income measured in snacks, that is, the total
number of snacks you can afford to buy given your income and prices. And we got the horizontal intercept, which is
the total number of toys you can afford to buy. The slope of the budget constraint, which is this number right here,
tells you the opportunity cost of an extra toy measured in terms of the number of snacks you have to give up.

What I want to do now is show you how this constraint changes when we change some of the things in the economic
environment. In particular, what happens to your opportunities when your income changes or when the price of the
products changes? Let’s start with your income. What happens to your feasible set? What happens to your budget
constraint or the stuff you can buy when your income is reduced? Let’s suppose instead of having $12.00 worth of
income, you only have $6.00 worth of income. In that case, how many snacks could you afford to buy? That’s right,
six. And how many toys could you afford to buy? Six divided by $3.00 per toy is only two toys.

Now the relative prices of toys and snacks are still the same. We haven’t changed them, only your income. The way
we show a change in income in this picture is the budget constraint will shift. In this case it shifts inward because your
purchasing power has shrunk. You now have a smaller set of toys and snacks that you can afford because of a
reduction in your purchasing power. If your income should increase, the feasible set shifts outward. That is, you can
afford to buy more toys and more snacks. These intercept points, again, represent your purchasing power, what you
can afford to buy, how well off you are measured in terms of toys or snacks. The slope of the line, however, again,
still refers to the opportunity cost of toys measured in terms of snacks.

Let’s look at a second possibility. Suppose now instead of changing your income we change the price of toys. And
let’s cut the price of toys in half. Now, instead of costing $3.00 to buy the punching monkey, you can get the same
punching monkey for only $1.50. Now let’s suppose the same is true for the vibrating sea lion and the squeaking nun.
If all of our toys are now $1.50, how many toys can you buy with a budget of $12.00? The answer is, you can afford
to buy eight toys. Eight times $1.50 equals your $12.00 income.

Now your purchasing power in terms of snacks has not changed. Snacks still cost $1.00 each and you only have
$12.00 to spend, so we show a change in the price of toys by rotating the budget constraint like this until we reach its
new intercept. This is the budget constraint when toys cost $1.50 apiece, snacks cost $1.00 apiece, and you still have
$12.00 to spend. A movement in the budget constraint, therefore, represents a change in your opportunity, a change
in what you can buy given your income and the price of the goods. If your income changes, then the budget
constraint shifts—outward for an increase in income, and inward for a decrease in income. But the slope remains the
same. As long as prices don’t change, the slope remains the same.

If we change the price, on the other hand, of one of the products, say toys, then we have to pivot the budget
constraint. The budget constraint moves outwards in the direction of the product that’s become less expensive. Now,
instead of only being able to afford four toys on your $12.00 income, you can afford eight toys, and that’s why the
horizontal intercept has moved out. Notice because toys are less expensive now, you can afford to buy more stuff
than before. The budget constraint has pivoted outwards.

In the future then, when you change the income, shift the budget constraint parallel. Keep the slop the same. But if
you change the price of one of the goods, make sure that whichever good is becoming less expensive, that you’re
shifting the budget constraint outwards in that direction in the same way that you would be shifting it inwards if one of
the products became more expensive. This is how you manipulate the budget constraint to describe a change in the
consumer’s opportunities, a change in the feasible set of things they can afford.
Consumer Choice and Household Behavior

Budget Constraints and Indifference Curves

Understanding Indifference Curves

Think fast. Which would you rather have, four toys and three snacks or six toys and two snacks? Would you be willing to give up your animal crackers if it means that you could enjoy the dinosaur pistol and the bunny bowl? Which would you rather have, four and three or six and two? Maybe this is hard to figure out real quickly, maybe it isn’t. Maybe you’d just as soon have one as the other.

Well, now I’m going to explore a tool that economists use to represent consumer preferences. This tool is called the indifference curve and it becomes part of a map that represents all the combinations of goods that you find more, less or equally desirable. Once we’ve drawn a map of indifference curves, we can put the indifference curves together with the budget constraint and derive the consumer’s demand curve. That is, once we know what your constraints are and what your preferences are, we can use them together to predict your behavior.

So, let’s get right to it. How do we draw an indifference curve? An indifference curve is defined as a set of points representing combinations of two goods, like toys and snacks that a consumer finds equally desirable or equally satisfying. So, we have to begin with a list that the consumer provides us that tells us those combinations that he or she finds equally desirable. Suppose we have a consumer named Chris, and we quizzed Chris about toys and snacks and he provides us with this list. This list tells us that Chris would just as soon have two toys and six snacks as 12 toys and one snack, or four toys and three snacks, or six toys and two. All of the combinations represented on this table provide Chris with an equal amount of satisfaction. That is, Chris is indifferent among each of these alternatives. He’d just as soon have this one as this one or this one.

Let’s represent now, Chris’ preferences in a graph. Here we are back in goods space, and let’s remember to label our axes, that’s always important. The vertical axis is going to represent the number of snacks that Chris has and the horizontal axis, just like before, is going to represent the number of toys that Chris has. Now let’s take the information from the table and represent it in the picture. So, we’ll start with the combination of two toys and six snacks, and that we’ll represent with this red dot, right here. Next, we find the combination of four toys and three snacks equally satisfying, so we can put that point right here. Then six toys and two snacks gives us this point and finally, 12 toys and one snack gives us this point.

All of these points represent combinations of snacks and toys, and all of these points, according to Chris’ report provide him with an equal amount of satisfaction. Now, you can imagine that there are a lot of other combinations in between these dots that represent combinations of toys and snacks that Chris finds just as satisfying. We can represent these infinite possibilities by connecting the dots and drawing a curved line that we call an indifference curve. So here I go, connecting the dots and there’s my indifference curve. A collection of points representing combinations of snacks and toys, all of which Chris finds equally satisfying.

Looking at this indifference curve, you’ll notice I drew it pretty carefully. I’m trying to make sure that the way I draw the indifference curve is consistent with what economists believe about people’s preferences. Let me be specific. The first thing you’ll notice about this indifference curve is that it’s downwards sloping. The downwards sloping indifference curve means that both of these goods are desirable. That is, if you take away some of Chris’ snacks, the only way that we can leave him just as happy as he was before, is to give him some extra toys. Anytime we take away one of the goods, we have to compensate by increasing the quantity of the other, and that gives us a downward sloping indifference curve. Every time we lower snacks, we have to increase toys.

It’s another way of saying that more is better. If we had a combination of toys and snacks that represented an increase in both, well Chris would prefer that, so it would have to be on another indifference curve, with another set of combinations. For instance, suppose we consider a point like two, four, six snacks and two, four, six toys. Since that represents more toys and more snacks together than in any of these points in here, that would be a point that would be way out here on another indifference curve somewhere. So a point that has more of both goods than a point on this curve would be more satisfying and therefore it would lie on another curve somewhere, a curve that represents increased satisfaction.

Another thing to notice is I’ve drawn this curve sagging downwards. We would call this a convex curve because it’s bowed downwards towards the origin. A convex curve is a curve that has a decreasing slope as you move down it. Notice up here, the slope is very steep, then the slope gets a little flatter and finally down here towards the bottom, the slope of the curve is very flat. Now, you’ll recall, from our early math review, that the way you find the slope of a curved line is to take the slope of a line that’s tangent to the curve. So for instance, if we want to know the slope of
this indifference curve, at this particular point, with four toys and three snacks, we would look for the straight line that is tangent to or touches the curve at that point. And, that tangent line’s slope is the slope of the indifference curve at that particular point.

Notice that the slope of the tangent line, as we move down the indifference curve, is getting flatter. The slope is decreasing. The line is getting flatter as we move down the curve. Well, that’s just a mathematical statement about what convexity means. But actually, this convexity has an economic interpretation. Notice that when we’re way, way up here, whenever Chris has a lot of snacks and very few toys, the curve is steep. That means, if you take away a lot of Chris’ snacks you only have to give him a few toys to compensate. Or, to put it another way, if we take away one of his precious toys, way up here, we have to give him a whole lot of snacks to leave him as happy as he was before. Just one toy less means a whole lot more snacks to leave him at the same level of satisfaction.

Well, once you’re down at a point like this, and you take away one of his toys, you hardly have to give him any snacks at all to compensate. You can take away a toy with very little compensation because he has lots of toys. The slope of an indifference curve is called the marginal rate of substitution and the marginal rate of substitution will be different as we move along the indifference curve. Up here, where the slope is steep, the marginal rate of substitution between snacks and toys is very large. If I take away one of your toys, I have to give you a lot of snacks to compensate. The marginal rate of substitution tells you how many snacks you would require in exchange for one toy.

As we move down the curve the marginal rate of substitution gets smaller and smaller and smaller. When you get way down here, where the curve is flat the marginal rate of substitution is very small. Which means, if I take away one of your toys, then I don’t have to give you very much snack at all to compensate you. That’s because you’ve got a lot of toys. You’re not as valuable to you in terms of snacks. So a quick summary, the indifference curve slopes downward to represent the fact that both goods are desirable and that you have to have more of one to compensate you for less of the other. Also, if you have more of both goods, you’ll move up to a higher curve somewhere else, somewhere up here to the northeast.

Finally, the slope of the indifference curve is called the marginal rate of substitution and it’s like a kind of internal price, a psychological price that you put on toys. If the marginal rate of substitution is big, then you want a lot snacks to compensate you for the loss of a toy. Your price of toys, in terms of snacks, is high. But, if you move down the curve and the marginal rate of substitution is small, then the price of toys for you, in terms of snacks is small. That is, you are willing to give up a toy with very little compensation, with very little extra snack in return.

This indifference curve then, represents a set of combinations that Chris finds equally desirable. Well, let’s look at another set, another set of combinations that Chris finds equally desirable, but a set combinations that provides him with more satisfaction than the combinations on this curve. Here’s another table that represents a set of indifferent points for Chris. Chris would just as soon have two toys and 12 snacks as four toys and six snacks, eight and three, or 12 and two. Now, all of these combinations are more desirable than the combinations on our original curve.

How do you know that? You know that because, right here, we have a point that represents 12 toys and one snack. On this new list, we have the point 12 toys and two snacks. Since this involves more snacks and no fewer toys, that is, it’s got more than the original bundle it must be more satisfying, because of our assumption, more is better. Any time you’re moving to the northeast, directly to the north, directly east or to the northeast, in this diagram, you’re moving to points that are strictly preferable, points and combinations that are strictly better.

So let’s go in and draw the picture that represents the combinations in this graph. Let’s start with two toys and 12 snacks, that is, two toys and 12 snacks will take us up to this point. Four toys and six snacks will take us to this point. And eight toys, two, four, six, eight toys and three snacks will take us to this point and finally, 12 toys and two snacks will take us to this point. Again, these are only a few of the combinations that provide Chris with this new higher level of satisfaction. If we connect these dots, we’ll get a new indifference curve.

All of these points are equally desirable and all of these points are strictly more desirable than the points on the old indifference curve, since it’s to the northeast, and more is better. Well, now that we’ve got a couple of indifference curves, we’re going to need some labels to keep them straight. So let’s call this original curve $U_0$. $U_0$ represents a number, and that number means the amount of satisfaction you get. We can call that utility. Utility is an economic term that means the amount of satisfaction that a consumer gets from consuming a bundle of goods. So $U_0$ is the
Consumer Choice and Household Behavior

Budget Constraints and Indifference Curves

Understanding Indifference Curves

original level of utility or satisfaction, and we can label this new higher Curve U₁, which represents a higher level of satisfaction. Now notice, I didn’t put in numbers like 10 or 20 or 100 because, I mean, what would that mean, numbers of utility, measuring warm, fuzzy feelings in numbers? How do you do that? All I’m telling you is that however we measure satisfaction, the points on this curve represent a higher level of satisfaction than the point on this curve.

One more thing to say about these indifference curves. Notice there are still points out there to connect, to connect with other bundles that are equally satisfying. I could spend all day drawing indifference curves, and these indifference curves would eventually fill up this graph. Indifference curves are everywhere against. Every point on this graph has some other points that are equally satisfying, and I could spend all day drawing the lines connecting them. I would eventually have an indifference map, a complete representation of Chris’ preferences.

What can we say about this indifference map? We can say a couple of things. The first thing we can say is that the curves are everywhere against. Everywhere you put your pencil in this graph, you are touching an indifference curve. The next thing we can say is that indifference curves can have all kinds of shapes. They don’t have to be concentric circles. These indifference curves can be kind of real blobby things, as long as they’re downward sloping, and as long as the slope decreases as we move down the curve.

One final thing. Indifference curves can never, ever cross. Can you figure out why? I’ll give you a hint. Remember, more is better. Let me show you. Suppose we had a situation up here where an indifference curve like this crossed over another indifference curve at a point like this. Now, how do we know that something like this can’t happen? Well, we know from the following arguments.

Let’s say we’ve got three points on this particular indifference curve and we’ll call them A, B, and C. Now let’s take a couple of points that are on the other curve, and we’ll call those points D and E. Now the thing that I want you to notice here is that point B is on both curves. That’s what it means for the curves to cross. Well, according to this logic, D is northeast of A. That is, D has more stuff than A, so that must mean what? That’s right. D is better than A. On the other hand, if you go down here, you’ll see that since C is to the northeast of E that must mean that this bundle, which has more stuff, is better. C is better than E. Well, if D is better than A, and C is better than E, we have a contradiction, because A is just as good as B, which is just as good as C. And if C is better than E, then A must, by transitivity, be better than E.

However, we’ve already said that E is just as well liked as B, which is just as well liked as D. That’s what it means for them to be on the indifference curve. Hold on. If A is better than E, and D is as good as E, but better than A, then we’ve got a mess. We’ve got some kind of contradiction. How can A be better than E and worse than B, which is the same as E? That doesn’t make any sense. It is illogical for these indifference curves to cross. You can’t have more is better and have cross in indifference curves. So we’ll put a big circle slash through there. This is not a possibility. Indifference curves never cross.

Well, we’ve done a lot in this lecture. We’ve taken a set of preferences that Chris expressed, and we’ve represented them in a curve called an indifference curve, a collection of points that represent equally well-liked combinations of two goods. Then we saw that that indifference curve slopes downward, has a slope that represents a kind of psychological price called the marginal rate of substitution, and we saw that this indifference curve is part of a big indifference map that represents Chris’ preferences for toys and snacks.

Now, we’re going to use this indifference curve along with the budget constraint to make some predictions about Chris’ behavior, and eventually, you’ll be able to derive Chris’ demand curve.
Consumer Choice and Household Behavior

Consumer Optimization

Locating the Consumer’s Optimal Combination of Goods

What would you do if you had Chris’ preferences and $12.00 worth of income? How would you allocate your limited income over all the possibilities available to you? Would you buy more toys, more snacks, a lot of one or the other, or a balanced combination? How do economists figure out what consumers are going to do, given their preferences and the constraints of income and price? We’re going to put the indifference curves that represent preferences together with the budget constraint, which represents opportunities to show you how economists think about consumer optimization.

Consumer optimization is the behavior of households trying to maximize satisfaction subject to the constraints of income and prices. So let’s take a look now at the graph that we’ve developed. First, we have this dark line, which you’ll recall represents the budget constraint. That is, it represents all the combinations of toys and snacks that the consumer can buy given the limited income and the price of the two goods. My axes aren’t labeled here, so it doesn’t look like I have meaningful graph. Someone is going to have to give me a label. Let’s see. Ah. There’s a snack, so we’ll label the vertical axis with snacks, and here’s a toy, so we’ll label the horizontal axis with toys. Make sure that your axes are always labeled; otherwise your graphs are meaningless.

Now, here we have the representation of the consumer’s feasibility set. That is, the opportunities that the consumer has given income and prices. So, if you spend all of your income on snacks at a price of $1.00 per snack and an income of $12.00, you can buy 12 snacks. If you spend all of your income on toys at a price of $3.00 per toy, you can buy four toys. And the slope of this budget constraint represents the relative price of toys and snacks, which is the price of toys at $3.00 apiece divided by the price of snacks at $1.00 apiece for a number of three. Three is the relative price of toys and snacks and it’s the slope of this green line. No. I guess it’s a blue line. It’s the slope of this line.

Notice one more thing. This line is downward sloping, which means that its number is negative. The trend slope of this line is the negative of the relative prices, but let’s ignore the negative sign for just a minute and focus on the intuition. How do you think about relative prices? How do you think about the slope of the budget constraint? This number three says this, “If you will give up one toy, if you give up one toy, you’re going to have enough extra income to buy three snacks.” Three is the relative price of toys measured in terms of snacks.

Now, if this blue line represents your possibilities, what are you going to do? Well, economists imagine that consumers do one and only one thing. That is, they seek the combination of goods that maximizes satisfaction. Let’s find that combination. Look at all of these red indifference curves that I’ve drawn in here for Chris. All of these red indifference curves represent combinations of goods and services that are equally desirable. Here’s a low level of satisfaction, this level of satisfaction on this indifference curve to the northeast represents an improvement and you can do even better by going way, way up here to the far northeast. The further to the northeast you go, the better off you are, because more is better in this model.

Now, I could go in and label these indifference curves. It looks like I’ve labeled this one U₀, which represents our baseline level of satisfaction and an improvement we might call it U₁, which means this curve represents more satisfaction. And I might label this curve, underneath here, I might call it U₂ because it’s worse. See the numbers don’t really matter, as long as the numbers increase as satisfaction increases. Which point is going to be the most satisfying to Chris? The answer is simple. Chris is going to choose that point on his budget constraint that lies on the highest obtainable indifference curve. Chris is going to find that point that gets him to the highest level of satisfaction and that is the point that he will choose.

In this picture, the highest indifference curve that Chris can reach is the one that I’ve labeled U₀, and the combination of toys and snacks that maximizes Chris’ satisfaction is going to be two, four, six snacks and two toys. With six snacks and two toys, how much money is Chris spending? Well, let’s see. Six times $1.00 apiece is $6.00 on snacks and two toys times $3.00 apiece is $6.00 on toys. Six and six makes a total of twelve and that, of course, is Chris’ income, which is why this point is on the budget constraint. The total spending is equal to $12.00, the amount of money Chris has to spend.

So, to find the point that gives Chris the most satisfaction, find the budget constraint and then look for the indifference curve, the highest indifference curve that that budget constraint touches. Now, there’s something to notice about how to identify that highest attainable indifference curve. Notice that the budget constraint is tangent to that indifference curve. That is, the budget constraint touches it in one and only one place. Any other budget constraint, any other indifference curve—Well, let me make sure I say this clearly. Any other indifference curve that touches this budget
Consumer Choice and Household Behavior

Consumer Optimization

**Locating the Consumer's Optimal Combination of Goods**

A constraint will touch it in two places. Only at this highest attainable indifference curve do we have a single point of tangency.

Let's look at some other possibilities. Suppose we look at this lower indifference curve. Notice this lower indifference curve touches Chris' budget constraint up here and it touches it again, down here. The thing to notice about these two points of interception is that at both of these points the indifference curve has a slope that's different from the slope of the budget constraint. That is it's not a tangency, it's an interception. It's a point where the budget constraint and the indifference curves have different slopes. The indifference curve cuts into and goes through the budget constraint.

Here's a little bit of intuition to let you know why this can't be the best point for Chris. Think about this. At every point on the budget constraint, the slope is three. That means Chris always has the opportunity to trade one toy for three snacks. Down here, at this particular point, the slope of the indifference curve is flatter. That is, the marginal rate of substitution is less than three. That is, Chris would be willing to give up a toy for less than three snacks, maybe just for two snacks. Ah, think about what that means. Chris is willing to give up a toy for only two snacks, but the market will give him three snacks. What's Chris going to do? Obviously, he's going to trade with the market. Obviously, he's going to give up his one toy and take his three snacks.

Two snacks would make him indifferent. Three makes him better off. That's an improvement on his situation. So, Chris starts moving up his budget constraint, trading toys for snacks until he finally gets to the point at which his desire to trade is equal to his ability to trade. That is, at this point, where the slope of the indifference curve, the marginal rate of substitution is also three, Chris is willing to trade one toy for three snacks and that's exactly what the market will allow. At that point, Chris can no longer improve his situation by trading away his toys. He stops. He has reached an optimum.

Look at this point up here. At this point, the marginal rate of substitution is very high, higher than three. That is, that Chris would be willing to give up more than three snacks to get an extra toy. Aha. But the market will only charge him three snacks. He's willing to give up four. So what's he going to do? He'll give up the three snacks and take an extra toy and he's better off. He's better off because he's willing to give up four snacks and he only had to give up three. That is, he pockets an extra snack and he moves down the budget constraint. Chris makes himself better off. He continues to trade until finally the marginal rate of substitution is equal to the relative prices in the market.

So, there you have it. The conditions of consumer optimum, the conditions of maximizing satisfaction is that the marginal rate of substitution is equal to the relative prices of the goods in the market. At that point, an indifference curve is tangent to the budget constraint and that's the highest level of satisfaction that Chris can afford.

Let's do a quick summary here. The budget constraint represents your opportunities and its slope is the relative price in the market. The indifference curves represent your preference and their slope is the rate at which you will trade off the goods, one for the other. The consumer gets the most satisfaction by finding the highest attainable indifference curve. That's going to occur at the point where the budget constraint is tangent to or touches the indifference curve at one place. The interesting thing about that point is that's the place where the relative prices are equal to the marginal rate of substitution, the slope of the indifference curve. That's the point at which the rate at which the consumer is willing to trade is equal to the rate at which the market will allow him to trade. Now, let's mess with that budget constraint and see how the optimum point moves around as prices change. This will get us a demand curve.
We’re back with our indifference curves of budget constraint, and we’re getting ready to derive the consumers demand curve. Remember what we did last time. We showed that if you have a set of opportunities created by your income and prices, and a set of preferences, you’re going to look for the most satisfaction you can get given your constraint. That point of maximum satisfaction will be where your budget constraint touches the highest attainable indifference curve.

Now, what’s going to happen if your opportunities change? How will your behavior change if the constraint changes? In particular, how will Chris behave if we reduce the price of toys? Let’s cut the price of toys in half from $3.00 apiece to $1.50 each. What will Chris do in this circumstance?

First, let’s show in our graph the new budget constraint. With a price of $1.50 each for toys the vertical intercept remains unchanged. We haven’t done anything to the price of snacks or to Chris’ income. So if he spends all of his income on snacks, he can still afford 12 snacks. However, if he spends all of his income on toys now at a price of $1.50 per toy, he can afford to buy eight toys, instead of four, like before. So here is Chris’ new budget constraint. A blue line here represents his opportunities with an income of $12.00 and a price of toys of $1.50 each.

What’s Chris going to do? As usual, he’s going to look for the combination of toys and snacks that gives him the most satisfaction subject to this budget constraint, and that means looking for the highest attainable indifference curve. Well, it’s clear that the highest attainable indifference curve in this picture is the one that’s labeled U₁. And the combination right here, two, four, six snacks and two, four toys—so it’s four toys and six snacks, this particular combination represents Chris’ optimal choice. Let’s make sure that point’s attainable. Of course, it’s on the budget constraint so we know that it is. Four toys times $1.50 each gives us a total of $8.00, and four snacks at $1.00 apiece gives us a total of $4.00. Four plus eight is 12. Once again, Chris satisfies his budget constraints.

How do I know that this particular combination is the best Chris can do? Because it represents a point of tangency between an indifference curve and the budget constraint. This is the point at which Chris’ marginal rate of substitution, that is, his willingness to trade toys for snacks is equal to the relative price of toys and snacks, that is, what the market will permit. You always find the optimal point where the indifference curve is tangent to the budget constraint. That’s always the best that he can do.

So even though there are other indifference curves that are way up here to the northeast, Chris can’t reach them because he doesn’t have enough money. Even though there are other indifference curves that are below, they’re irrelevant, because they intercept there because, they’re lower. Chris can get to a high level of satisfaction, the highest he can attain, by finding the point where the indifference curve touches the budget constraint at only one point. U₁ is the highest level of satisfaction he can get.

Notice what I do here. I changed the price of toys and I looked at how Chris’ optimal choice changed. He buys more toys than before. Before he was only buying two. Now that the price of toys has fallen, Chris buys four toys. What I’m getting to is a demand curve. When the price changes, the quantity demanded responds, and that’s what a demand curve is. In the next lecture, we will put the indifference curve budget constraint diagram that we have here before us, we’ll put it on top of another axis and use this apparatus to derive a demand curve. What we’re going to be seeing is the foundation for the demand curve, that is, where it comes from, using this more basic theory.
Consumer Choice and Household Behavior

Consumer Optimization

Deriving the Demand Curve

We’ve put indifference curves and budget constraints together to find the consumer’s optimal choice under the constraint of the budget and prices. Last time we showed how when the prices changed, the consumer’s optimal choice changes. Now we’re ready to use the tool we’ve developed to show where the demand curve comes from.

Where does the demand curve come from? It comes from consumers making optimal choices subject to their constraints.

Look at the graph that we had before. Here I’ve reproduced it in a smaller form. The blue line represents the budget constraint, that is, all the possible combinations of toys and snacks that a consumer can buy with limited income given the prices. The red line represents the highest attainable indifference curve, the most satisfaction that our consumer can get. And here’s that optimal point. What I’ve done is taken the graphs we derived and put another graph beneath it. And this graph is our familiar price/quantity graph that we used for representing a demand curve. Notice on the horizontal axis I’ve represented the quantity of toys, just like on the axis above it. Quantity of toys is measured on the same scale on both of these axes. So a quantity that comes down to this axis can be dropped on down into the graph below it. These two graphs are measuring the same thing horizontally.

Vertically, this graph down below measures the price of toys. So originally, our original price of toys, which was $3.00 per toy, I call that $P_0$, the original price. At a price of $3.00 per toy, then our consumer wanted to buy two toys—two toys a week say. So here’s a point on the demand curve. At a price of $3.00 per toy, the consumer chooses to buy two toys a week. Now, let’s change the consumer’s constraints by lowering the price of toys from $3.00 per toy down to $1.50 per toy. When we do that we have to redraw the budget constraint in the diagram above. So here I go—gonna redraw that budget constraint and I’m going to draw it very thick and beautiful so you can see it on the red. There you go. There’s the new budget constraint.

Now with this new budget constraint we’re going to be looking for a new indifference curve, and that’s the indifference curve that’s going to be the highest one we can get to given the constraint of the blue line. And as we saw before, in the case of Chris’ preferences for toys and snacks, that combination is going to be at a point like this. Here’s the new indifference curve—call it $U_1$, the highest attainable one, and here, right here at this point of tangency is the combination that Chris would choose to purchase in that situation. So when the price of toys drops from $3.00 a toy down to this new lower price of $1.50 per toy. When the price of toys falls, the quantity of toys that Chris consumes increases from two toys a week to, in our case, four toys.

So let me take those four toys a week, and with a dotted line, I’m going to go down below the axis, down into my downstairs diagram, and here’s my new quantity, $Q_1$, or four toys a week. So let’s put a new price/quantity dot right there at $1.50 per toy and four toys a week, and there you have it. Hey, once you’ve got two points, you can connect them, right? And that gives you a demand curve. So you can imagine if we kept doing this exercise, we’d get an endless combination of price and quantity points, and that becomes the demand curve for toys for Chris.

So I’ve labeled this curve D, that’s demand. And this curve does all those things that you’ve seen demand curves do before. What’s the purpose of this lesson? The purpose is to show you where a demand curve comes from. And where does it come from? It comes from consumers who have preferences, and who have constraints imposed by prices and income, doing the best they can with what they’ve got. As their circumstances change, that is, as prices change, the optimal choice will change. And the demand curve is a record of those optimal choices. As the price rises and falls, the optimal point will move around upstairs in this graph, and as it does, we record it downstairs in this graph, the demand curve.

The demand curve shows you what the consumer will purchase, the quantity of toys Chris will buy as the price of toys changes, holding constant everything else, holding constant income, holding constant the price of snacks, holding constant Chris’ tastes and preferences. That’s where the demand curve comes from, and that’s why we spent that time developing the indifference curve tool and putting it with the budget constraint and learning how to do consumer optimization.
Production and Costs

The Basics of Production

Understanding Outputs, Inputs, and the Short Run

Suppose you have a business and it produces something, maybe hamburgers, maybe television sets, and you’d like to know, “How much should I produce so that to make the most profit I can from my operation?” We’re now entering a set of lectures that’s going to equip you with the tools economists use for describing the behavior of profit maximizing firms. In order to get to the question of what a profit-maximizing firm should do, we have to back up and start with some basics.

The first thing we’ll do is describe the technology of the firm. That is, what can the firm do with a given set of inputs to produce a certain amount of output? We’re going to describe the technical or technological possibilities that a firm has for producing output from input, whether it’s for assembling television sets or making hamburgers. Once we have a description of the firm’s technology, we will then look at the cost to the firm of producing output. What we do is we take the technology and combine that with knowledge of the prices that the firm has to pay for its inputs, like labor, raw material, capital and come up with the cost per unit of making hamburgers or televisions. Finally, we’ll combine the information we have about the cost of production with information about the price at which the firm can sell its product.

Once we’ve combined those pieces of information, we’ll be able to predict the best course of action for a profit-maximizing firm. We’ll be able to answer the question, “How much output should a firm produce if it wants to maximize its profits, subject to its technology, the price it pays for its inputs, and the environment, the competitive environment, in which it finds itself?” So, let’s back off then of this big picture and focus on the first set of specific questions we’d like to ask. That is, what are the technological possibilities of the firm? That is, what can the firm do? We’re looking now at a firm’s technology and a technology simply means a catalog of things that a firm knows how to do.

What we are most interested in is how much output a firm can produce with a given amount of input. Now, this sounds a lot like engineering, doesn’t it? It’s because it is. Economics, when you describe production, begins with a purely engineering relationship. That is, a relationship that has nothing to do with Economics or money and stuff like that, but has more to do with technological possibilities. What can the firm do, according to the laws of physics, according to its know-how to turn labor and raw materials into its finished product?

Now, we’re going to start with a careful description of the firm’s technological possibilities and then we’re going to look at how to represent those technological possibilities in a graph. Finally, we’ll introduce some concepts that economists focus on when they describe the technology that a firm uses to produce its output. Let’s start then with a chart that shows the technological possibilities that the firm faces. It’s kind of like a production possibilities chart for the firm and we will call this production possibilities a “total output relationship.” Total output is defined as the total amount of output that a firm can produce using a given amount of input subject to its technological know how and any other factors that constrain its production.

What a firm can do, given its know-how, with a certain amount of input. Total output, therefore, is the amount of output a firm can produce with a given amount of input. Let’s look at the technological possibilities of a particular firm; let’s say this firm produces television sets. So their output in this case is TVs and their input, let’s say, is labor. Now, when we describe this relationship, total output, we’re talking about what labor can produce, what a given amount of labor can produce in terms of television sets. But clearly, we need other inputs to make a television. We need raw materials, we need tools, we need a place to work. Whenever we describe the total output relationship in Economics, we’re usually talking about the short run.

The short run is a period of time that is so short that you can only vary one or a few of your inputs. A plant that’s producing televisions might decide over time that it wants to expand its operations, add another conveyor belt, get some more tools, build another warehouse. Well, you can do all that stuff in time, but in the short run, if you’ve got a big order for television sets, you could only meet that big order by hiring more workers to come in and work with the factory as it exists now. The short run is a period of time during which some inputs are fixed and other inputs are variable.

So, when you describe the total output relationship, the first thing you want to do is to distinguish between fixed inputs and variable inputs. Let’s suppose now that we’re in the short run in our television production process. We’re building TVs and suddenly we get an order to increase the number of televisions that we’re making. Maybe we get a big order from a department store. In the short run, let’s suppose our only option for increasing output is to add more labor.
Production and Costs

The Basics of Production

**Understanding Outputs, Inputs, and the Short Run**

Everything else is fixed in the short run. It's too expensive to alter quickly the number of conveyor belts, the size of the factory, the number of tools, and so on.

So, for all practical purposes, our only variable when we're altering our output is labor, and we will call labor the variable input. Now, let me go back and define total output one more time. The total output of the variable input labor is the total amount of output that a given amount of labor can produce, holding constant technology, that is, the firm's know-how, and holding constant the quantity of all of the other inputs that the firm uses, capital, the size of the factory, tools, anything else. We're going to consider those, the fixed inputs. And, whenever we write a schedule, like the one in front of me here, we're assuming that those fixed inputs remained fixed in the short run.

So, we've got a certain factory size, certain number of conveyor belts, a certain pile of raw materials to work with, but in the short run we can only produce more televisions by adding more labor. Let's look then at what the possibilities are in the short run. In the short run, if you have one worker you can make two television sets per week. If you add a second worker with the same tools and the same technological know-how, you can make ten television sets in a week. Until a third worker, and the three of them can produce a total of 30 television sets in a week. Four workers have a total product of 40 television sets per week. Five workers will give you a total output of 45 and with six workers, you can make 48 television sets per week. As you see, as the numbers get larger, up to a point, the total output of labor increases.

Oddly enough, if you go past a certain point, with eight workers, in this story, the total number of television sets actually drops by one, from 49 to 48. This table of numbers then, is a total output schedule. It shows you the total amount of output you can produce with given amounts of the variable input labor, holding constant technological know how and holding constant the amount of all other inputs used.

The next thing I’d like to do is represent this information graphically. So, in the next lecture we’ll take this chart of numbers and put it into a picture. That picture will represent the production possibilities of the firm in the short run and we will call that the “total product curve.”
Production and Costs

The Basics of Production

Explaining the Total Product Curve

We’re here in the factory because we’re talking about production possibilities. Remember, we’re talking about what a single firm can do in the short run, when it’s constrained by having some of its inputs fixed. The factory space is fixed. The number of conveyor belts is fixed. The tools are fixed. Maybe even our inventory of raw materials is fixed. But in the short run we can vary our output by varying the amount of labor that we hire. Let’s look now, at the firm’s production possibilities in the short run.

Last time, we looked at this chart, which showed what a given number of workers could produce in a week, holding constant technology and the fixed input. Now we’re going to take the information that’s in this chart and transform it into a graphical snapshot of the firm’s production possibility, and we will call the picture that we draw the total product curve. Let’s start now, by labeling the axes. Remember, you always label your axes first, or you’re not drawing an economic graph, you may be making art, but nobody knows what it is.

Let’s start by labeling the vertical axes. The vertical axis will measure the total amount of output that the firm produces. The horizontal axis, in this picture, will measure the total amount of labor that we are using to produce television sets. So, here we have one, two, three, four, five, six, seven, eight workers and on the vertical axis, we'll have 10, 20, 30, 40, 50 television sets. Let’s now take the information from the total output chart and translate it into a curve. So, the first point on our curve would be one worker and two television sets and I’ll put a red dot right here, at this point, to represent this possibility. If the firm employs one worker, in a factory of a given size, with a given quantity of tools and so forth, they will get two televisions per week.

If you add a second worker, your total output per week would then go up to ten televisions, according to the chart. The third worker increases total output to 30 televisions per week, the fourth worker adds a total of ten more televisions to give us a total output of 40 televisions per week. Notice now, we’re getting combinations of quantity of workers and quantity of televisions produced per week. The fifth worker makes it possible for us to produce a total of 45 televisions per week. With a sixth worker, we can produce a total of 48 televisions per week. Seven workers means we get 49 televisions per week and eight workers puts us back down to 48 televisions per week.

These, of course, are not the only possible quantities of labor that we could hire. We could hire fractional workers. We could hire fractional workers by hiring people for only part of the day, or to only work part of the shift. If we connect these dots, we’ll get a smooth relationship between the amount of labor that the firm employs and the number of televisions that are produced per week. So, let’s connect the dots then and get a total product curve. Here it goes, connecting the dots and I’m going to label this curve TP because it is the total product of labor, in my television factory. Every time I add a worker, I calculate the total number of televisions that my entire crew can assemble per week, and that gives me a point on the curve.

So, one, two, three, four, five workers can produce a total of 45 televisions per week. Now, when you look at this snapshot, you’ll notice a few things about the production technology, production possibilities. What do you see, in looking at this snapshot, what do you see that catches your attention? Well, let’s notice a few things. The curve has an S shape and this kind of S shape is made up of three characteristics. The first thing you notice is that whenever you have small crews, adding an extra worker, that is, going from zero to one and then one to two workers, output increases at an increasing rate. That is, each additional worker adds more to total output than did the worker before. The curve is increasing, that is, the slope of the curve is increasing, the quantity of televisions produced are increasing at an increasing rate.

So, in this region down here, the curve is convex. Convex, the curve is convex. The slope is increasing. The next thing you notice is that over a range, up to here, the total number of televisions produced is increasing at a decreasing rate, as we add extra workers. That is, each additional worker adds less to total output than did the worker before. We’re adding more workers, but the extra televisions that we’re getting are fewer and fewer and fewer. Output is increasing at a decreasing rate. Finally, we notice that this curve has a maximum point, that there is some kind of capacity to our factory. If we hire too many workers, that is, beyond a certain point, the total quantity of televisions produced, actually begins to shrink. That is, if we go from seven workers to eight workers, the volume of televisions that we ship out the door actually begins to contract.

Now this requires some explanation. Economists usually draw the total product curve for a particular firm as having this S shape, because it represents assumptions about technologies that we believe match pretty well the reality of a lot of production processes. Certainly, production processes that involve manufacturing or making things. Let’s then
Production and Costs

The Basics of Production

Explaining the Total Product Curve

define a concept that will be useful to us in describing the shape of this curve and that is the concept of the marginal product of labor. The marginal product of the variable input, in this case, the marginal product of labor, is the change in total output that results from increasing the amount of variable input by one unit. It is the change in total product that results from a change in the amount of the variable input, the amount of labor that’s used.

Let’s calculate now the marginal product along this curve. The first worker takes our total television production from zero to two. The marginal product of the first worker then, is two television sets. That’s the total amount of output that he adds by being hired. The second worker, now that we have a shift that has two workers instead of one, brings the total up to ten televisions produced. That is, the second worker is adding ten minus two or eight television to total production. The marginal product of the second worker hired is eight televisions. The third worker now brings the total up to 30, 30 minus 10 is a total additional production or an additional production of 20 units. The third worker has a marginal product of 20 televisions. If it weren’t for the third worker, we’d be back with two workers and 10 televisions. So, the third worker adds 20.

The fourth worker then adds 40 minus 30, is ten. The marginal product of the fifth worker is 45 minus 40 televisions or five. So, the change in output that results from the change in labor is what we call the marginal product of labor. Remember, we’re calculating this marginal product holding constant the firm’s know-how and holding constant all other inputs, like the size of the factory, the number of tools, and so forth. Here’s a little chart that we’ve made up that shows the marginal product of labor. I could have labeled it marginal product, but I decided instead to label it with the definition of marginal product, the change in total output or televisions that results from a change in the labor input.

If we do this, I can go over and lay this chart next to my other charts and we see that the first worker has a marginal product of two televisions, he adds two to total output. The second worker has a marginal product of eight, as we calculated before, and you can go right on down the chart. The third worker has a marginal product of 20, the fourth has a marginal product of ten, the fifth has a marginal product of five and you can look at the rest of the chart, which we’ll move over next door, so that you can study it. Here’s what you want to notice about this firm’s technology.

We can describe a firm’s technology by changes in its marginal product. Back there, whenever we were looking at the shape of the curve, we were really talking about changes in the firm’s marginal product of labor. Let’s look at that curve again. Down here, in this region of the curve, where the curve is convex, the marginal product of labor is increasing as we add additional workers. That is, down here in this region each additional worker adds more additional output than the worker before. The first worker has a marginal product of two and then eight, and then third worker adds 20, and so forth.

Now, why does the economist believe that at small scales, down here at whenever the firm was first adding variable input that the marginal product would be increasing? Why is that? Do you have an idea? This is why. Economists believe that increasing marginal product comes from specialization and tools. Whenever you add additional workers to a production process, the workers are able to divide the tasks or divide the production process into lots of separate tasks, which they can do separately.

For example, the famous example that Adam Smith gave was the production of straight pins, pins used for pinning clothes. He said the production process involves a lot of distinct tasks, stretching the wires, putting a point on the pin with a grinder, putting a head on the pin with a different tool. And he said, “If you notice the process of producing pins can be divided up into a lot of different tasks.” And when each worker can concentrate on a particular task, perhaps the task that he or she is particularly good at even, has comparative advantage in, in that case, he said, “the productivity increases remarkably.” Three workers can make 50 times as many pins as a single worker working alone. That’s because they can divide the job into different tasks and save on the time they would otherwise spend doing one task, getting out a new set of tools, moving to another task, getting out a new set of tools, moving to another task, the so-called setup costs.

Another thing that you want to think about when you’re thinking about moving from one worker to two workers to three workers, is what we call teamwork. Teamwork is when larger groups can use a different technique of production. Think about moving furniture. You can spend all day trying to move one room of furniture, but if you can get a buddy to help you, you could do the job in less than a quarter of the amount of time. That’s because two people can move furniture using different techniques than one person. One person moving furniture pretty much has to squat and push
Explaining the Total Product Curve

and grunt and try to move that sofa across the room, but two people can each pick up an end and move the sofa much, much more productively. So teamwork and specialization are always tending to increase the productivity of labor. Add more workers and you have more scope for teamwork and specialization. After a certain point, however, you notice that scope for teamwork and specialization is no longer driving the story. We eventually reach a point where additional input results in output decreasing, but at a decreasing rate. This is what we call the problem of congestion.

Imagine, you’re having friends over and you’re making dinner in the kitchen in your apartment. What do you do? You give everybody a task. You cut up the salad, you make the sandwiches, you go over and stir the soup, and so forth. Well, everybody’s got something to do, but before too long you’ve got so many people crowded in your kitchen that you’re beginning to get in each other’s way, and that reduces your productivity.

As you crowd more and more television workers into a factory of a given size, eventually they’re having to share tools, they’re getting in each other’s way, there’s too much labor for the fixed amount of capital. When you’ve got too much of a variable input congested into the fixed inputs, then your productivity begins to fall. The workers begin to have accidents, step on each other, get in each other’s way. Output begins to increase at a decreasing rate. Finally, after this point right here, after our point of maximum output, the losses from congestion are so great that additional workers actually reduce total output. There’s so much congestion that it’s actually reducing total output. No firm in its right mind would ever operate over in this region of its total product curve. If you can produce as much output with fewer workers, that’s going to be less expensive and more profitable.

So, as a quick summary, when you draw this S shaped total product curve, you are summarizing the assumptions that economists usually make about technology. That for a firm that has fixed inputs and alters its output by changing only one input, say labor, you usually have a region of teamwork and specialization, whereas marginal product is increasing and the curve is convex, followed by a region where marginal product is decreasing, but still positive. Output is increasing at a decreasing rate because of congestion. This is our region of diminishing marginal product. And finally, after you get past this top of the curve, marginal product becomes negative. Congestion is so strong that productivity has actually become negative. This S shaped curve represents the technology possibilities of the firm in the short run.

Next we’re going to represent marginal product in its own curve and we’ll show another curve that summarizes another measure of productivity.
Production and Costs

The Basics of Production

Drawing Marginal Product Curves

Remember, we’re talking about how economists describe the productivity of a firm. Eventually, we’re going to use this information when we prescribe what a firm should do to maximize its profits. But for now, we’re concerned with the task of representing a firm’s technology. The last time we looked at the total products curve, which tells us what a firm can do in the short run, when it can vary only one of its inputs. We also described the marginal product of a firm, that is, the additional output that the firm can produce by hiring an extra worker, an extra unit, of the variable input.

What I’d like to do now, now that we’ve discovered that marginal product is simply the slope of the total product curve, is I would like to draw a picture that puts the marginal product curve in a graph by itself. Let’s look now at this diagram, and we’ve got another one of these now famous two-story diagrams. Remember, the trick with a two-story diagram is the graphs are related and they have the same variable measured on the horizontal axis. So in this case, we are measuring labor input on the horizontal axis in both pictures. On the upstairs picture we’re measuring total product on the vertical axis and we’re using those same numbers that we’ve been using all along to represent the total products, and those numbers are next door.

In the graph below we’re going to now represent marginal product. And I’ve shown you those numbers. I’ve calculated them before. It’s the change in output that results from a change in the variable input. And these are my marginal product numbers, they’re also next door as well. What I’m going to see is that when the curve is convex, that is, when the curve is increasing at an increasing rate, the marginal product will be increasing, that is, the slope is getting steeper, the marginal product is going to be increasing. When the slope is decreasing, that is, when the curve is concave, when the total product is increasing at a decreasing rate, then the marginal product will be decreasing. This marginal product curve would be downward sloping. Finally, when the total product curve finally maxes out and heads south, that is, when extra workers actually decrease, the total product of the operation, then the marginal product will itself be negative.

So, I’ve put the points from this chart—I’ve put the points down here in the marginal product curve and now I’m ready to connect them and draw the marginal product. So for one worker the marginal product is two, for the second worker it’s 10, for the third worker it’s 20, and so forth. So here is two, here’s 10. I made a mistake there. I don’t know why I had that point for 10 because I’m using this for 10 and this for 20—so there’s 20. Yeah, you’ve got to be careful about this. And that means when I go back to a marginal product of… Oh, I know why, because that isn’t a marginal product of 10, it’s a marginal product of eight. I got ahead of myself, so I can mark that point off.

Here’s what the curve looks like—2, 8, 20, 10, 5, 3, 1, and -1. And hey, I’ll save my mistake here by making it part of the picture. I’ll use that as the letter P for my marginal product curve. Hey, there’s a little lesson. Sometimes you can turn the mistakes into part of the arc, and that’s what I’ve done here. Now the marginal product curve shows me the slope of the total product curve at every point, and by connecting the dots you can see where the curve is convex, the marginal product is increasing. Where the curve is concave, the marginal product is decreasing. And where the curve goes negative, the marginal product is below zero. So this curve represents the slope of this curve at every point.

I want to throw in two little points of information here that you might find useful. Sometimes we write total product with an L and parenthesis afterwards to indicate that total product, that is, the total output that you could produce in the short run, given your technology and given your amount of fixed inputs, as a function of labor. So we write this as a function of labor. I could also write marginal product also as a function of labor. I could write MP of L, because the marginal product changes as I add additional workers to my factory.

The next thing that I want to point out is that you’re talking here about the number of workers who are working in the factory at a particular time. It’s not like you’re hiring one worker or then a second worker later in the week, or a third worker even later. We’re talking about two workers working together for a week, or three workers working together for a week. See, the change in the marginal product is part of the technology. There are things that two workers can do to produce television that one worker can’t do by herself. There are things that three workers could do that two workers can’t do. Teamwork and specialization become possible as you add extra workers working together at the same time. So when we talk about the marginal product of technology, we’re talking about how productivity changes as you add additional workers that are working at the same time, at once, together, over a given period of time, like a week.
Production and Costs

The Basics of Production

**Drawing Marginal Product Curves**

Finally, the last thing that I want to point out is I’ve been using the term convex and concave. I sometimes forget which is which, so I’ll show you a little trick, a little memory trick. A memory trick, by the way, is called a mnemonic device. Some of my students were telling me the other day—I used the expression mnemonic device, and they started telling me that they liked these demonic devices, they thought they were helpful. So here’s a little “demonic” device to help you.

Let’s suppose you’re standing on the axis looking up at a curve, and if you’re looking up and you see a cave, then we say that the curve is concave. But if you look up and you’re looking for a cave and what you see doesn’t look like a cave, it’s bowed downwards, and you’re looking for a cave, well then you would be frustrated or vexed, so this is convex. So here is where you’re vexed, and here is where you actually see a cave. If the curve is concave, you look up and see a cave, if it isn’t, then you’re vexed, the curve is convex.

So now we’ve got a picture of the marginal product curve, which is the slope of the total product curve. And what we’ll add next is another measure of productivity that you may often hear about in the newspaper, that is, output per worker, and that’s called the average product of labor. So we’ll do it next.
Production and Costs
The Basics of Production
Understanding Average Product

We’ve been describing the technology of the firm, and we’ve been focusing on the marginal product, the change in output that results from the change in the variable input when you hold everything else constant. Now we’re going to look at another measure of productivity that becomes important later in the story. I'll give a little hint about where we’re going. Remember, our main concern is being able to recommend to a firm how much output to produce if it wants to maximize its profits. Well, the marginal product turns out to answer that particular question. The average product, on the other hand, the concept that we’re about to introduce, tells the firm whether it’s able to make a positive profit or not.

That is, when a firm is doing the best it can, is it making a profit or a loss? So let’s go ahead and introduce now the concept of average product of labor. That is, the total output divided by the total labor input. When you read the newspaper or hear discussions of labor productivity on the TV news, what people are talking about is average product of labor, how much output is a typical worker at a firm producing, the average worker. We define the average product of labor, again, as total output divided by the total number of workers employed. So, if we look back at the tables we've been looking at, the particular television firm we’ve been considering, if we have one worker producing two televisions then that means on average, workers are producing two TVs apiece. Two workers producing ten televisions gives us an average product of labor of five TVs per worker. Three workers producing 30 has an average product of labor of 10, 4, 40, also an average product of 10, 5, 45 gives us 9 TVs per worker and so forth, on down the list.

These numbers represent the average product of labor for different levels of employment inside the firm. That is, for different quantities of workers hired and different numbers of televisions produced. I’ll move all these numbers now, next door and we will draw a picture of the average product of labor. We will draw a graph below the total product curve that represents the average product of labor. Now, I use the red for product. I kind of like red for output for some reason and I used the red a minute ago to draw the marginal product curve. So in order to avoid confusion, I’m going to switch to green here, to represent the average product of labor.

So, let’s do that. The average product again, is total output divided by the total number of workers. So, up here we see that a total output of two televisions when we have one worker gives us an average product of labor of two TVs per worker. So, I’ll put this in at two TVs per worker and here we’ve got two workers producing ten TVs for and average product of labor of five TVs per worker. Then three and 30 gives me an average of ten per worker, which is exactly what I get with four workers producing 40 TVs and five workers producing 45 takes my average down to nine TVs per worker and then eight and then seven and then six. So, if I connect the dots here, I’m going to get an average product of labor curve that looks something like this. First it’s increasing, then it levels off, and then it’s decreasing. And, I’ll label this APL for the average product of labor. It’s the output per worker or the average—the amount of output that the average worker produces, output per worker or labor productivity.

Now, before, remember we said that the marginal product of labor was the slope of this curve at any given point? There’s also a geometric way to calculate the average product of labor, but it doesn’t make as much intuitive sense. It’s actually the slope of a line from the origin up to the curve. The rise would be output and the run would be labor. So, output divided by labor would be the average. If you’re interested, you can find the average product of labor by taking the slope of a line that goes from the origin up to any point on the curve and as the points on the curve change, the slope of the line from the origin also changes and that changes your average. That seems a little complex and not a lot of, you know, intuitive assistance do we get from that. So, I’m going to leave that alone and go back to the curve that I drew from the numbers that I got from my table, that’s just as easy a way of doing it.

I want to notice a few things about this average product of labor curve. I want to notice a few things about it. First of all, again, it’s kind of U-shaped. That is, first the curve slopes upwards and then it slopes downward. You reach a point of maximum labor productivity at some point. At some point, average product of labor reaches a maximum. That is, at some point we are getting the maximum number of televisions per worker that we can get out of this factory, given its size, given the number of tools, given our technology and so forth. That’s thing one I want to notice. The average product of labor simply gives us, like an upside down U. At first, it increases and then it decreases. It’s like a hill.

The next thing I want to notice is, I’m going to put the marginal product curve back in this picture. I’m going to draw it again and it looks like this, you’ll recall. So here’s my marginal product of labor, and I’ve got it in red. Notice again that the marginal product curve first slopes upwards then the curve is convex and downwards when the curve is
Production and Costs
The Basics of Production

Understanding Average Product

concave and then it goes below zero when the curve, the total product curve tips downwards. Well, notice something about the average product of labor. This is something kind of interesting, and that is, when the marginal product of labor curve is above the average product the average product curve is rising. And, when the marginal product of labor curve is below the average product the average product is falling.

Now here’s the intuition. If you’re adding workers that are more productive than average you’re pulling up the average. Kind of like, when you have a really good semester at school and you make a grade point, an overall grade point average that’s higher than your cumulative GPA, you pull up your cumulative average, by adding a semester that’s above average. Over here, when we’ve got workers that are less productive, than average, we’re pulling the average down. Kind of like when you make a bad exam grade, an exam grade that’s below your class average, you pull your class average down. You pull your own average for the course down by a below average score.

When the marginal is above the average, it pulls the average up. When the marginal is below the average, it pulls the average down. That means that the marginal product curve is going to intercept the average product curve when the average product curve is at it’s maximum. That is, the marginal pulls it up and then pulls it down and it intersects it, therefore, at the maximum of the average product. The intersection of the marginal product with the average product occurs at the point where the average product curve is at a maximum. Now, that piece of information is going to become important, later, knowing where the marginal product curve cuts into the average product curve. It cuts into it at the point where workers are producing, on average, the most televisions per worker that can be produced, from this particular factory.

This is the point when labor productivity is at its highest. There was a joke that we told at my school growing up about a rival school. Did you here about the guy from our school who left to go to a rival school? He raised the IQ at both schools. Now, this is a joke about marginals and averages. That anybody who would leave our school, must be someone whose IQ is below average, but he was still smarter than the average at the other school. So, he raised the IQ at both places. Now when you here that joke and you laugh, what you’re really laughing at is mathematics. You’re laughing about this funny relationship between averages and marginals, that makes it possible for you to take someone who’s below average, move him over into another pool that has still a lower average and raise the average in both pools.

That’s what this relationship between marginals and averages is all about. Well, that then completes our steps of product curves. We’ve now done as much as we can to explain the technology of a firm in the short run. What we’ve done here is we’ve described the possibilities of a television factory that produces television sets using labor and some other fixed inputs. The firm can increase its output, in the short run, but it can only do so by adding more labor. And we’ve described the productivity of labor as two concepts, the marginal productivity of labor, which depends on teamwork and specialization and on congestion and the average productivity of labor, which is what we usually call labor productivity or output per worker.

We also described the relationship between marginals and averages. When the marginal is above average, it pulls it up. When it’s below average, it pulls it down. What’s next? Next we’re going to take these notions of productivity and couple them with some information about input prices and we’re going to come up with a set of cost curves, a representation of the firms cost of producing any given amount of output.
Production and Costs
The Basics of Production
Relating Costs to Productivity

We’re now entering a series of lectures that a lot of students consider very difficult. So we’re going to take our time and try to do this extra clearly. We’re going to be talking about the costs of production. Now, you’ll recall, we’ve just done a series of lectures on productivity, the firm’s technology. Now we’re going to talk about how you take the firm’s technology, what they know how to do, and combine it with the cost of hiring inputs, like labor and capital, and come up with the cost of producing output, in our case, the cost of producing television sets. This information is going to be very helpful to us when we move to the next phase—finding the profit maximizing output choice for a firm.

So here we are—technology, costs and profit. We’re in the middle of this stream of questions that’s going to lead us to the answer, what is the best choice for a firm to make if it wants to maximize its profits. Students find this material difficult because it involves a lot of different tools used at once. We’re going to be looking at charts and some simple mathematics, then we’re going to be looking at curves that represent the information in those charts. So I’m going to try to do this very carefully. I’ll show you some charts, then show you the curves that go along with it, and at each stage I’ll try to remind you of the intuition behind all of it.

Let me start with what I think is a very simple explanation of the relationship between costs and productivity. The first relationship we want to begin with as we start our discussion of the costs of production is the relationship between cost and productivity. In particular, cost and productivity are inversely related. That is, when productivity is increasing, the cost of production is falling. And when the productivity of your workers is falling, the cost of production is increasing. Let me see if I can make that very, very clear.

Let’s suppose we had a television factory, and in this television factory we have a certain number of fixed inputs, like the size of the factory, the number of conveyor belts, the number of tools. And we also have variable input, that is, the number of workers that we hire to assemble television sets. Let’s suppose we start with an assumption about productivity. Let’s suppose we assume right now that one worker given her know-how and the tools available to her, can assemble one-quarter of a television in a day. If one worker can assemble one-quarter of a television, then how many workers will it take to make a whole television? The answer is, it will take four workers. Four workers, each assembling one-quarter of a television, will produce a whole television.

Now notice the point that I’m making. If one worker has an average product, or an average output of one-quarter of a television, one over four, then the reciprocal of that, four workers, is what it takes to produce a single television set. Notice productivity and costs are reciprocal. If one worker can produce one-fourth of a television, then it takes four workers to produce one television, and you’ve got to pay those workers a wage, and that’s what determines the cost of getting that television put together. One-quarter is the reciprocal of four. If one-fourth is the productivity of your worker, then four is the number of workers it takes, which becomes your cost of producing the television. This idea, this simple explanation, is the basis of the relationship between cost and productivity. The more productive your workers are, the fewer workers you’ll need to produce your output.

Now, what I’m going to do, is I’m going to step-by-step, do a description of the different notions of cost of production. We will look at the costs associated with hiring labor. We’ll look at the costs associated with having a factory and tools ready for those workers to use. And finally, we’ll look at the cost of producing another television set. What is the cost on average of producing a single TV, and what is the cost to produce one extra television set? The notion of average cost and marginal cost.

Along the way, after I give you an example of each of these concepts, I’ll be taking a break and then drawing the curve that goes with each concept, so that when you look at the picture you won’t just see a line on the page. Perhaps you’ll also see a story that relates productivity to costs. That way, when we finally come to the point of using these curves to describe the firm’s profit-maximizing choice, maybe you’ll have some intuition that makes this more than just a technology exercise. I hope so.

So we’ll move now to the first concept of cost, and that is the cost of hiring the labor to produce your output, the variable cost of production.
Production and Costs

Variable Costs

Defining Variable Costs

We're now beginning a series of lectures where we introduce the concepts that we use to describe the cost of production for a firm. The first concept that we're going to introduce is the concept of variable cost. The variable cost of production is the cost of hiring the variable input needed to produce a given amount of output. That is, in our case, the amount of money the firm has to spend to hire the labor that it needs to make a given quantity of television sets.

In order to find the variable costs for a particular firm, you're going to need two pieces of information. The first piece of information we've already discussed. This is the amount of labor that the firm needs to produce a given quantity of output, and this is an idea about productivity. Remember, the total product of the firm. It told us how much labor we needed to produce a given quantity of televisions. So the first piece of information you need is a piece of technological information. How much labor do we need to hire to produce a given quantity of television sets?

The second piece of information that you need is a piece of economic information. That is, how much do you have to pay to get a unit of labor? In other words, what is the wage rate that you must pay your workers in order to get them to work for you assembling television sets? Once you know how many workers you need, and what you have to pay each one of them, then you've got the variable cost, or the labor cost of producing a given quantity of television sets.

Let's suppose, for the sake of our example, that the wage for a television set assembly worker is $1,000 per week. That means every time a firm hires another television assembly worker, they've got to pay that laborer $1,000 per week. Let's now use that information to find the variable cost of producing certain quantities of television sets. Go back to the chart that we began with. This chart describes the people of the firm. That is, how many television sets a certain number of workers can assemble in a week. What we'll do now is take the number of workers in each case, and multiply by the wage that must be paid those workers to get the total amount of money the firm has to spend on labor to produce output.

For instance, if you hire one worker, you've got to pay that worker $1,000 a week for a total of $1,000. That's the variable cost of producing two television sets. Because, after all, one worker can produce two television sets, that means $1,000 is the labor bill for turning out those two television sets. If you hire two workers, you can produce 10 television sets, and since each of those workers will cost you $1,000 per week, the variable cost of producing 10 television sets is $2,000. Keep going. To produce 30 television sets, you need to hire three workers, that's $3,000 a week. To produce 40 television sets, you have to hire four workers, that's $1,00 per week, and so forth.

All of the numbers in this column are calculated by multiplying the number of workers you need by the wage of $1,000 to get the variable cost of producing a given quantity of television sets. Now, you might ask yourself, "Well, what's it going to cost us to produce 35 television sets a week?" Well, we can't say for sure, because we don't have enough detailed information here to answer that question. But we know that it takes between three and four full-time workers per week to produce 35 television sets. So the number lies somewhere between those two, which means the variable cost will be somewhere between $3,000 and $4,000.

Again, you can imagine a fractional number of workers to give you a particular number of television sets, and then multiply that for $1,000 per week to get the variable costs. Once again, the variable cost is simply total spending on the variable input needed to produce a given volume of output. What we'll do next is draw a picture of this concept.
Production and Costs

Graphing Variable Costs

We’re ready now to draw the first of our cost curves. We’re going to be drawing a lot of these cost curves so I’m going to take special care with this one to kind of show you how they work. Next door you can see the numbers that we used last time whenever we derived a relationship between cost and output. This is our variable cost series. It tells us how much money we have to spend on labor in order to produce a given number of television sets.

What I’m going to do now is draw a picture of that relationship, and this relationship will be called the variable cost curve. The variable cost curve shows you how much money you have to spend on labor in order to produce a given number of television sets. Let’s look at the way this graph will be set up. On the horizontal axis I’m going to measure the number of television sets that we’re producing in our factory. On the vertical axis I’m going to be measuring dollars, the amount of money that we’re spending to hire the labor to produce those television sets. So use the numbers from over there, and we’ll put them in a graph and come up with a variable cost curve. Ask the question, how much does it cost us to produce two television sets, how much do we have to spend on labor? And the answer, as you can see next door, is $1,000. So we’ll put a dot here to represent two television sets and $1,000 worth of spending on labor.

Next, how much does it cost us to produce 10 television sets? Well, you can see it costs us $2,000, so put a dot there at 10 televisions and $2,000. To produce 30 televisions we’re up to $3,000 worth of spending on labor. To produce 40 televisions we’re up to $4,000 worth of spending on labor, and we can produce 48 televisions—sorry, 45 televisions by spending $5,000 on labor; 48 televisions by producing—with six workers, that costs us $6,000 a week; 49 televisions with seven workers for a total labor cost of $7,000, and finally, we’ve got this odd point up here, this odd point that says if we hire eight workers, which costs us $8,000 a week, we can produce only 48 televisions.

Remember, before I told you that this point is inefficient, because you can produce the same number of televisions with less money spent on labor. That’s why I’m putting an X on this point. We’re not going to consider this point on the variable cost curve, because no firm in its right mind would ever operate at that point. That is a point of inefficiency. So we’re going to leave that point off the curve. Now, are these the only number of television sets we might be interested in producing? No. We might be interested in producing any number that lies along this axis. That means that we could have labor expenses that vary probably continuously from zero all the way up to and beyond along the curve. So all the points in between are possible, and we represent that continuity by connecting the points and forming a curve.

So here’s what we get. We go from zero up to $1,000, up to $2,000, then on up to here, and so forth. And eventually, my variable cost curve simply becomes vertical. And the variable cost curve, when it becomes vertical, represents the fact that you can spend all the money that you want to on labor, but you can’t produce any more television sets. Once the total product of our factory has reached the maximum, the firm can spend more money if it wants to on labor, but it can’t produce any more televisions. And as we looked at the numbers before, we saw that the maximum total product was 49 television sets. That’s the most we can ever produce. So at that point, our variable cost curve will go vertical. You can hire more labor, but you can’t get any more output.

One final thing. Let me label this curve “variable cost,” VC. This curve represents the amount of money that you spend on labor to produce a given quantity of output. Now that we’ve got the variable cost curve, I want to show you a little trick using another set of graphs that represents the relationship between productivity and cost. Remember, I told you that relationship is inverse; productivity and cost vary inversely. I want to show you how that looks comparing our product graph with our cost graph.
Graphing Variable Costs Using a Geometric Trick

We’re taking a time-out here to do a little trick to show you the relationship between productivity and cost. I want to recall a diagram that we spent a lot of time developing a little while ago. That was the total product curve. Remember, the total product curve shows you the relationship between the amount of labor your firm hires and the number of television sets you produce in a week. For instance, if you hire two workers you’ll be able to produce—according to the number we’ve been using—you’ll be able to produce 10 television sets a week. If you hire three workers, then you’ll be able to produce 30 television sets a week. That’s your productivity.

Now, think about this diagram viewed from the other direction. Suppose you ask yourself, “How much will it cost my firm to produce 10 television sets in a week?” So instead of starting on the horizontal axis and going up to the total product curve, we’re starting on the vertical axis and going over to the curve. We’re saying, “If we want to produce 10 television sets a week, how much are we going to have to hire—how much labor will we have to hire to make that happen?” And the answer is, we’re going to have to hire two workers. If we want to, however, increase our output up to 30 television sets a week, going over to the curve, we’re going to have to hire three workers.

Now, think about this for a minute. Once you know how many workers you have to hire, you know how much it’s going to have to cost you in labor expense. Remember, those two workers cost you $1,000 each for a total of $2,000. If you want to make 30 television sets a week, you’ve got to hire three workers, that’s going to cost you three times 1,000 or $3,000 a week. The point I’m making here is that you can use the information in the total product curve to derive your costs only you’re just reading the curve in a different way. Rather than starting with the number of workers and finding their product, you’re starting with the product you want to produce, and figuring out the number of workers that you’ll need.

This is exactly the information that we’re looking for when we draw the variable cost curve. Only the axes are reversed. Instead of putting product on the vertical axis and labor on the horizontal axis the way we do in our products curve, we flip things around. We put the total amount of output on the horizontal axis and not just labor, but its cost, on the vertical axis. Let me show you how to turn a product curve into a cost curve. Start with our product curve, and the first thing you want to do is flip the axes around. Well, I’m going to pick up my product curve like this. I’m going to pick it up and flip it over and lay it right here on my cost graph. Notice what I have now is a kind of inverse, or mirror image, of my original curve. Now, if I want to produce 10 television sets in a week, I know that I have to hire two workers, and I multiply those two workers by $1,000 and I’ll get by variable costs.

So in order to turn a product curve into a cost curve, you need two steps of geometry. The first step is take your product curve and reverse the axes, taking the curve with it. So we’ll just flip the curve over and put the axes, the horizontal axis on the vertical, and the vertical axis on the horizontal. The second step is, you take this curve, and you stretch it upwards by multiplying it by the wage. Each of these points represents a number of workers that you’ll need. Now, multiply that by $1,000 a week, and it becomes your labor bill. So what I do is I take this curve and stretch it upwards, and I get the variable cost curve that I drew a moment ago, and it looks like this.

Notice this variable cost curve has the same shape as my inverted total product curve. It’s just stretched upwards because I multiplied all of those numbers by the wage of $1,000. Now, that’s kind of a cool little trick, and if you want to try this trick at home, you’re welcome to. I always like to draw my total product curve and then hold it up to the light, you know, look at it through my paper and see its inverse. It’s a neat way, a neat geometrical way of making it clear to you that cost and productivity are inversely or reciprocally related. It’s because we’re changing the axes and stretching the curve. There’s the relationship.
Production and Costs

Marginal Costs

Defining Marginal Costs

We spent the last lecture looking at the relationship between output and the cost of production, the concept of variable cost. How much do you spend on your variable input, labor, in order to produce a given amount of output, say a certain number of television sets. Now we’re going to introduce two other concepts of cost that will prove useful to us down the road when we’re looking at a firm’s profit-maximizing decisions. That is, how much output should we produce if we want to make the most profit we can.

The first concept that we’re going to introduce is how costs change when you change your output. If you think back to our discussion of productivity, we introduced a concept called marginal product. How much additional output do we get when we hire an extra worker? Now we’re going to ask a similar question about costs. How much extra money do we have to spend in order to produce an extra unit of output? This concept is called marginal cost. And you’ll see over there that we’ve defined marginal cost very carefully. The marginal cost of production is the additional cost incurred when the firm produces an extra unit of output. Or to define it mathematically, it’s the change in variable cost that results from a change in output, or a change in production.

Let’s look at how to calculate the marginal cost of production. If we go from one worker to two workers, the number of televisions that we produce goes from two to 10. That is, the marginal product of this second worker is eight additional television sets. Now, how much is each one of those televisions sets costing us? Well, in order to get those extra television sets, we’re having to increase our spending on labor from $1,000 to $2,000. The change in variable cost then is 2,000 minus 1,000, or $1,000. That’s our change in variable cost. The change in output is 10 minus 2 or 8. So if we divide 1,000 by eight extra television sets, we get the marginal cost of those extra television sets over that range, and that would be $1,000 divided by 8 equals $125 per set. That’s the marginal cost, the extra money that must be spent to get an additional television set over that range of production.

We can continue this exercise. If you add another worker, you can go from 10 to 30 television sets. That’s 20 additional sets for an extra labor expense of an extra $1,000. You hired an extra worker. $1,000 divided by those 20 extra sets, is a marginal cost of $50.00. So $50.00 a set will get you an extra television set as you go from two to three workers, as you go from 10 to 30 televisions. Keep calculating. In each case, look at the change in variable cost and divide that by the change in the number of television sets. What you will get in each case is the marginal cost, the cost of adding another television set to your output.

What I’ve done here is come up with a chart that shows the marginal cost of production. How much does it cost to produce an extra television set in each of these circumstances? If you go from one to two workers—well, let’s start from the beginning. If you go from zero to one worker, the marginal cost of those first two sets if $500.00 apiece. If you go from one to two workers and go from two to 10 television sets, those next eight sets are costing you $125.00 apiece. The next 20 television sets, when you add the third worker, are costing you $50.00 apiece. In each case the marginal cost tells you how much you’re paying to get an additional television set in a particular case.

Now, a good question is, why is the marginal cost of production changing? Why is the marginal cost of the first two sets only $500 each, the marginal cost of the next eight set is only $125.00 each, the marginal cost of the next 20 sets drops down to $50.00, then the marginal cost of the next 10 sets is $100.00 each? Why is the marginal cost changing as output and employment change? What do you think? If you had to come up with a reason why marginal cost is changing as output and employment change, what would you think is the answer? Here’s a hint. Go back to our original lectures on the total product.

The answer is this—marginal cost is changing because productivity is changing as we hire more workers. At first, the marginal product of labor is increasing. You add extra workers, they add more to output than the worker before. Remember, that second worker adds eight TVs, the third worker adds 20 TVs. Adding extra workers is making the workforce more productive. When the workforce gets more productive, the cost of producing extra televisions drops, the inverse relationship between cost and productivity.

Marginal cost varies as the productivity of labor varies. And as we’ve seen before, the productivity of labor changes as you add more and more workers. At first, you have teamwork and specialization. Over that range the marginal cost of production is falling, because workers are getting more productive. After a while you get congestion of your fixed input and the marginal product of labor begins to fall. When that happens the marginal cost of producing extra televisions begins to rise. When productivity is increasing, marginal cost is decreasing. When productivity is shrinking, marginal cost is increasing. Once again, we have the inverse relationship.
Production and Costs

Marginal Costs

**Defining Marginal Costs**

A quick summary. Marginal cost is defined as the change in variable cost that results from a change in output. When productivity is increasing, marginal cost is falling. When labor productivity is decreasing, marginal cost is increasing.
Production and Costs

Marginal Costs

Deriving the Marginal Cost Curve

In this lecture we’re going to draw a marginal cost curve and I’m going to show you first the easy way to draw the marginal cost curve and then I’ll show you a relationship between cost and productivity that you can see by looking at the curve.

First, let’s go back to the numbers that we derived before—the marginal cost of producing another unit of the good. And I’ll put these numbers over there so that you can work with them. And what we’ll do is plot the points in a space where we have output on the horizontal axis and cost on the vertical axis. So the first point will be two television sets and the marginal cost of producing an extra television set when you’re making two is $500. So we’ll put a dot right here at two televisions and $500. One, two, three, four, $500, two televisions. Ten television sets takes the marginal cost down to $125. So I’ll put a dot there. If we’re up to 30 television sets, the marginal cost is down to $50 a set, and if we are at 40 television sets we have $100 for our marginal cost. Forty-five gives us a marginal cost of $200, 48 gives us a marginal cost of $333, so I’ll put a dot here. And finally, at 49 television sets we have a marginal cost of $1,000.

Now, of course, these are not the only quantities of televisions we could be producing. These are not the only cost numbers we could have. So to represent the points in between, we will connect the dots with a line and that will become the marginal cost curve. So let me do that now. I connect my dots with a green line and the green line becomes the marginal cost curve. There we go—marginal cost, the cost of producing an extra unit of output given the number of units that you’re already producing. When you’re only producing two units, the marginal cost of an extra unit is high. But if you’re producing 30 televisions, the marginal cost of an extra unit is low.

Now, you might ask yourself, “Why is that?” Why does the marginal cost of production vary as you change the quantity that you are producing? What do you think? The answer is this. Cost and productivity are reciprocally related. If it takes more workers to produce a television set, you’re going to have to pay more for that television set. And if the additional worker, if the next worker that you hire produces only a small fraction of a television set, then you have to hire a lot of workers to produce a whole television set. Again, it’s the reciprocal relationship between costs and productivity. If the next worker produces only one-quarter of a television set, it takes four workers to produce a whole one. But if the next worker produces half of a television set, it only takes two workers to produce a whole one. The inverse relationship between cost and productivity.

You can see that inverse relationship in the graph. Does this graph look like something that we’ve seen before? It certainly does. Let’s take the graph now and flip it over and look at it from the other direction. What does this look like? Well, let me first of all fill in the little chunk that I lost due to the tape. And you can see you’ve got this upside down, U-shaped curve, which is exactly the shape of the marginal product curve. I need to re-label my axes. If I put output on this axis like we did before, and if I put labor down here on this axis, then this curve becomes the marginal product curve. I can just put an MP right here. See, the marginal cost curve and the marginal product curve are just mirror images of each other. That’s because cost and productivity are reciprocally related. When productivity is increasing, when the marginal product is increasing due to teamwork and specialization, then the marginal cost of an extra unit of output is falling. When the marginal product is falling, the marginal cost of producing an extra unit is rising, because you’re having to hire more and more workers at the margin to produce an extra television set.

So the point that I’d like to make here is that marginal cost and marginal product are mirror images of each other. They are reciprocals of each other. Now, I can show this same point using some simple mathematics, and I’m going to do that in just a minute. But first, I want to show you another way of deriving the marginal cost curve based on the definition of marginal cost. Remember, the definition of marginal cost is the change in variable cost that results from producing an extra unit of output. So what we can do is look at the variable cost curve and show how to derive the marginal cost curve from it, because the two concepts are related by definition. Let’s look now at the variable cost curve. The variable cost curve shows us how much we have to pay to hire the variable input labor that’s necessary to produce any given quantity of television sets.

Now, look at the variable cost curve and notice that the variable cost curve has a very steep slope close to the origin. And the slope of the variable cost curve gets flatter and flatter until we reach this point of inflection. At that point the slope stops getting flatter and begins to get steeper again. The curve gets steeper and steeper and steeper, and finally becomes completely vertical when we reach the capacity of our plant. Remember, in the short run you reach a point where you simply cannot produce any more television sets, no matter how much labor you hire. At that point,
Production and Costs

Marginal Costs

Deriving the Marginal Cost Curve

You can hire all the extra labor you want, but you don't get any extra TVs. At that point the variable cost curve becomes vertical, it goes straight up.

Well, think now about what this slope tells you. The slope of the variable cost curve is the rise over the run. The rise in this diagram is a change in variable cost. That is, an increase in the vertical coordinates mean that the variable cost is increasing, you're spending more money on labor. The run in this case is an increase in the production of television sets that is, producing more television sets as you hire more labor. So the slope of the curve, the rise over the run, is the change in variable cost that results from a change in output.

Well, that's the definition of the marginal cost. So if you want to graph the marginal cost curve and you want to do it in a kind of intuitive way that relates it to the definition of marginal cost, you can work from the variable cost curve itself. The marginal cost curve is a graph of the slope of the variable cost curve at every point.

Well, if I want to draw a marginal cost curve down here in the diagram below the variable cost curve, I’ve got to do this kind of carefully. First thing I have to do is make sure that I’ve got the same variable on the horizontal axes. That is, if you’re going to stack one graph on top of another, you’ve got to make sure that the axis that they share are measuring the same things. So I’m measuring output on both of these horizontal axes. Upstairs I’m measuring variable cost on the vertical, and downstairs I’m measuring marginal cost on the vertical.

So next thing I have to do is I’ve got to draw carefully a line down here below that shows the changing slope of the variable cost curve as we move along it. The most important point in this graph is going to be this point right here. That’s the point of inflection, the point at which the variable cost curve stops being concave and starts being convex. This is the point at which the slope stops diminishing and starts increasing. So if I’m real careful about this, I can indicate the quantity of television sets at which that inflection occurs and drop that on downstairs into the marginal cost diagram and make that the minimum point on my marginal cost curve. Marginal cost is decreasing, that is, the slope is decreasing up to that point, and then beyond that point it’s going to be increasing again. So I’ve identified the minimum point on the marginal cost curve.

Now if I wanted to be totally mathematically precise, I’d have to figure out exactly what the slope is at that point and put my dot down here at the number that represents that slope. But I’m not going to be that careful. I’m just trying to represent the main idea here. We know that this is the inflection point. This is the point of minimum marginal cost. Well, that’s also going to let us be able to draw the curve now, because over here, on this side, as output is increasing, marginal cost is diminishing, so I can draw marginal cost that is falling down to this point. So the marginal cost will be decreasing as this slope is decreasing, until I reach the inflection point.

Now, beyond the inflection point, marginal cost starts to increase again. As you see, the slope gets steeper and steeper. So beyond that point I draw an upward sloping marginal cost curve that looks like this. And this curve heads on up to infinity as the variable cost curve becomes vertical, because you know, the slope of a vertical line is infinity, so the marginal cost curve is shooting on up so that it has a reading of infinity by the time the variable cost curve goes vertical. Well, I can label this curve now “marginal cost.”

It’s the same curve that I draw before when I was carefully plotting these points from over there. But now I’m doing it intuitively. I’m using the definition of the marginal cost curve, the slope of the variable cost curve, and I’m using that now to derive the marginal cost curve graphically. I like this derivation because it makes clear the connection between marginal cost and variable cost. This curve represents the slope of this one.

What I’m going to do next is show you a mathematical relationship between cost and productivity that I think will give you more insight into what marginal cost is all about.
Production and Costs

Understanding the Mathematical Relationship Between Marginal Cost and Marginal Product

Here we are in the middle of a series of lectures on productivity and costs, and this stuff sometimes is pretty heavy sledding. So let's take a moment and review what we've done and what we're going to do next. We're talking about marginal cost and marginal product. Remember, marginal product is the extra output that can be produced when you hire another worker. That is, anytime you increase your variable input, the input that you can change in the short run, how much extra output do you get? This is a matter of your technology, and we call that measure the marginal product of labor.

Also, marginal cost is the cost of producing an extra unit of output. How much do your costs change when you hire the labor that you need to produce one extra television? Now, I'm stopping right here in the middle of these lectures to make a point about the relationship between costs and productivity. I've talked about this before. The relationship is reciprocal. When productivity goes up, costs fall. And when productivity goes down, costs increase.

I'm going to write a few steps of math out right now that I think will make this very, very clear. For those of you who get kind of nervous when I start doing mathematics, just take a moment and take some deep breaths and we're just going to go right ahead with this. I think it will actually help if you'll just be patient with me.

So let's start with marginal cost. Marginal cost is defined as—so I'm just putting an equal sign here because by definition this is true. The change in variable cost that results from a change in output. So we'll call that output total product. Remember, that's the total number of television sets that you produce. So if you change your output, how does variable cost change, that's what marginal cost is. Well, how does variable cost actually change? If you hire another worker, how does your variable cost change? Well, you're going to have to pay that worker the wage, and we're going to imagine, for the sake of the stories we're telling here, that the wage is constant. You can hire as many workers you want to at $1,000 per week. So I can write in here, wage, which is $1,000 a week, times the change in labor, and that's my variable cost. We'll have to divide that now by total product.

Does this term right here look familiar? What is this term? What is this little fraction that I have downstairs? It's the marginal product of labor, the extra TVs that you get when you hire extra workers. So I can finally rewrite my expression this way—that the marginal cost is equal to the wage divided by the marginal product of labor.

Ah, what have I just proved? I've just proved with a few steps of logic and math that the marginal cost is equal to the wage divided by the marginal product. Now, you can go back and check the numbers that we've used before and see, in fact, that this is true. I'll do that in a little bit. But I want you to notice that marginal product and marginal cost are reciprocals. Marginal product is over here in the reciprocal or the denominator of this fraction that is defining marginal cost—they're reciprocals. Now, that's kind of "mathy," that's not real intuitive, but at least it's precise.

Now, what I'd like to do is to try to make this very, very intuitive. Let's suppose that I've got workers producing television sets. And let me put my television set over here. And let's suppose right now that the marginal product of labor is such that one worker can produce one-fourth of a television set. So the marginal product of labor, the extra worker hired, can produce one-fourth of a television set. If we want to produce a whole television set right now, we're going to have to hire four workers. Why? Because the marginal product of labor is one-fourth of a television set. So to get a whole television set, you've got to hire four workers. Each of them can produce one-fourth, and then you've got a whole television set. And you've got to pay each one of those four workers the going wage.

Now look over there at the numbers that we just derived. Look over there at the formula that we just set up. If one-fourth is the marginal product of labor, that is, one worker can produce one-fourth of a television set and the wage is $1,000 a week, then 1,000 divided by one-fourth means it's going to cost you $4,000, four workers that you're going to...
Production and Costs

Marginal Costs

Understanding the Mathematical Relationship Between Marginal Cost and Marginal Product

have to hire to produce one whole television set. The extra cost for your factory of producing a television set right now is four workers times $1,000 equals $4,000. That’s the marginal cost of a television set.

I hope I’ve made this clear that cost and productivity are reciprocals. The marginal cost is the reciprocal of the marginal product. In fact, it’s quite precise. The marginal cost is equal to the wage divided by the marginal product of the next worker. Well, now that we’ve got it established that marginal cost and marginal product are inversely related, that is, they are for all practical purposes reciprocals, it should be very clear why the curves have the similarity that they do. Look at the curves for marginal product and marginal cost. We’ve drawn them before, but I’m putting them here together in a double-decked graph.

Now here’s a warning. In this case the double-decked graph is not measuring exactly the same thing on the two horizontal axes. I’m not saying that these graphs can be mapped perfectly into each other. I’m just setting them beside each other for comparison. This is an important difference. Because I have labor on the horizontal axis up here and I have televisions on the horizontal axis down below. But you know that as you hire more labor, you’re going to be producing more television sets, so in some ways, these two graphs are not incompatible. I can set them on top of each other and be measuring the same idea, even if I’m not measuring the same variable.

Well look, up here, when the marginal product is increasing, when each additional worker is producing more output than the worker before, what’s happening down below? The marginal cost of production is falling. Here, it takes, let’s say one worker can produce one-fourth of a television set, so it takes four workers to produce one television set and that’s expensive. Up here, one worker can produce half of a television set, so it only takes two workers to produce a whole television set and the marginal cost is lower. Up here maybe one worker can produce a whole television set by himself. In that case, down here you only have to hire one worker to produce a television set and the marginal cost is at its minimum. As long as productivity is increasing at the margin, cost is falling at the margin. As long as workers can produce more television than the workers before, the cost of adding extra televisions to your output keeps falling. But when productivity at the margin begins to fall, then costs at the margin begin to rise.

One final thing to note. When marginal product is at its maximum, marginal cost will be at its minimum. Remember, marginal product is at its maximum when? When is it going to be that an extra worker is going to add the most television sets? That will be the case when teamwork and specialization are maximized. When the scope for teamwork and specialization has been fully enjoyed. Beyond this point congestion starts to take over. Workers are crowded, using the given number of tools and the fixed production space. Over here you get congestion and marginal productivity begins to fall as the workers get in each others’ way and end up sharing the tools, and that gets hard, and we don’t have enough fixed inputs here to accommodate all these workers. Well, once you’re into declining marginal productivity, then the marginal cost of production starts to rise.

So what has this lecture been about? This lecture has been about the inverse relationship between marginal cost and marginal productivity. We saw first, mathematically, that marginal cost and marginal product are reciprocally related. Then we took an example to explain the intuition. And finally, we used that intuition to explain why the graphs look a lot alike, why they look like mirror images. Anytime marginal product is increasing, marginal cost of production is falling. Anytime your workers are getting less productive, the cost of adding extra televisions is going up.
Production and Costs

Average Costs

Defining Average Variable Costs

So we’re back for another lecture in the continuing series on cost curves and productivity. We’ve been talking a lot about marginal cost, the cost of producing an extra unit of output. It’s going to turn out that the concept of marginal cost is very important when it comes to a firm deciding how much output to produce. You don’t want to produce another unit of output if the cost is greater than the money that you can earn by selling it. If cost exceeds price, you don’t want to produce the unit of output. Marginal cost becomes a guide to a firm that’s trying to decide how much output to produce.

We’re now going to talk about the concept of average cost. Average cost isn’t a guide to the firm in how much output to produce, but it does tell the firm that is choosing how much to produce whether they can make a profit or not. Marginal tells you how much to produce, average cost tells you whether you’re going to be profitable doing it or not. Let me see if I can make this clear by first defining average and then talking a little bit about what it means for a firm that’s trying to maximize profits.

Let’s begin with a definition. The average variable cost is the cost of labor per unit of output produced. That is, it is your variable cost, your total wage bill, divided by the total number of television sets you produce. Let’s look now at some numbers that will make this clearer. Suppose we have one worker producing two television sets, and we’re paying that worker a total of $1,000 a week. That’s his wage. Well, if you divide that $1,000 variable cost by two television sets, you get an average variable cost of $500. There’s $500 worth of labor going in to each of the two televisions that you’re producing. So the average variable cost or labor cost per unit is equal to $500. If you hire two workers, you can produce a total of 10 television sets. Those ten television sets are going to cost you $2,000 worth of labor expense because two workers times $1,000 each is $2,000. Divide $2,000 by ten televisions, and the average variable cost is now $200 worth of labor per television set produced. Keep going. Three workers make 30 television sets, $3,000 worth of labor expense for an average variable cost of $100. And that continues. Four producing 40 televisions, $4,000, again, $100 per television set. Five producing 45 television sets, $5,000 worth of labor expense divided by 45 televisions is about $111 of labor per television set, and so forth. You can go ahead and complete the chart.

I’ll now move these numbers over there so that we can use them in a minute whenever we draw the curves. The point that I want to make is this. Average variable cost is the total amount of money you were spending on labor divided by the number of television sets you make. If, for example, you can only sell a television set for $150, then it doesn’t make any sense to be producing 10 television sets with two workers, because you’re spending $200 on the labor alone, not to mention the cost of your factory and tools and everything else. If this average variable cost is greater than the price of the television set, you know that you’ll be making a loss producing the televisions. See, you can use this number to help you figure out whether you can make a profit or not in this business.

Down here, however, if you can sell your television sets for $150 apiece, then it might make sense to make 40 television sets a week with four workers. That’s because the labor cost per television set is only $100. Now, that leaves some room for other costs of production associated with the fixed inputs, that is, paying the overhead on your factory, the tools, the utilities, all that other stuff. If that comes to less than $50 per set, then your cost of production will be less than the price, and in that case you can make a profit.

We’ll see later how these things all fit together, but I want to get you ready thinking about average variable costs as a guide to whether your operation is going to be profitable or not. So the definition of average variable cost is the labor cost per unit produced, and notice that it varies as we change the amount of output that we’re producing. Do you have any idea why? Think for a moment. Why would average variable cost change as you change the amount of output that you produce? Of course, the answer has something to do with productivity, right? The more productive your workers are, the fewer workers you’ll need to produce television sets and therefore, the lower the cost on average. The more productive your workers are, the lower the cost. And as productivity changes, the cost of production will be changing.

Next, we’re going to draw a picture to show what the average variable cost looks like in a graph, and then we’ll talk about the inverse relationship between average variable cost and it’s related productivity concept, the average product of labor.
Understanding the Relationship Between Average Variable Cost and Average Product of Labor

Let’s focus on the relationship between cost and productivity one more time. This time we’re going to look at how average variable cost is related to the average product of labor.

Now, stop for just a minute and make sure you’re clear on what these concepts are. The average product of labor is the amount of output that your factory produces divided by the total number of workers you’re hiring. That is, average product of labor is the output per worker. Average variable cost is your total spending on labor divided by the output that you produce. That is, it is your labor cost per unit of output. Now, how are these two concepts related? I’m going to do a little bit of math like I did before when I discussed the marginal product and the marginal cost. I’m going to show you mathematically precisely what I mean by the relationship between average variable cost and average product. Then I’ll give an intuitive explanation and then I’ll show you some graphs that summarize this relationship.

First, let’s look at some math. The average variable cost is equal to what? It’s equal to the variable cost, that is, the total amount of money that your firm spends on labor divided by the total amount of output that you produce, the total product of labor. So here is spending on labor, here are the number of televisions you produce.

Well, the variable cost, as we have said before, is equal to the wage multiplied by the number of workers you hire, and we can divide that by the total product of labor. Well, if that’s the case, then I can divide the numerator and the denominator of this fraction—I can divide both parts of the fraction by the number of workers that we hire and we get this: \( W \), that is the wage, divided by the total product of labor over the number of workers. Now, what gives this term—what is this term right here in the denominator of my fraction? It's the average product of labor, output per worker. So the average variable cost is equal to the wage divided by the average product, that is, the output per worker at your firm. Let’s see if I can make this a little bit more intuitively clear with an example. And again, I’ll use the example we had before.

Let’s suppose that one worker can, on average, produce one-fourth of a television set. If one worker can, on average, produce one-fourth of a television set, then it’s going to take four workers, on average, to produce one whole television set. If the average product of labor is one-fourth of a television per worker, then it takes four workers to produce a television set. That means you’ve got to hire four workers at $1,000 each to produce a television set and that’s going to be the labor cost per television set, \( \$4,000 \). If, however, one worker can produce half of a television set, if one worker can produce half of a television set, then it’s only going to take two workers to produce a whole television set, and that means your average labor cost is only going to be \( \$2,000 \) per television set. The more workers it takes to produce a television set, the higher your cost and the lower the labor productivity, that is, the less that a given worker can produce, the more workers you have to hire to produce a whole television set.

This is the relationship between productivity and cost. When average product is higher, then it takes fewer workers to produce a television set and the average labor cost is therefore lower. Let’s look now then at the relationship between the average variable cost curve and the average product curve. And here I have it drawn. When the average product of labor is increasing. That is, when it takes fewer workers, on average, to produce a television set. When one worker can produce more of a television himself or herself, then the average variable cost is falling. As workers become more productive, it takes fewer workers to produce a television and the labor cost per television is falling.

Finally, the average product reaches its maximum, and after that point, if you hire extra workers to produce more television sets, the average product of labor is falling, that is, the typical worker produces less and less and less. That means you’ve got to hire more and more and more and more workers in order to produce a television set, and the cost per television is going up. The labor cost per television increases as the average product of labor falls.

Now, you can look at these two diagrams again and see that they are very, very, very closely related. When average product is increasing, average variable cost is falling. When average product is at its maximum, the labor cost per unit is at its minimum. When average product begins to fall again, average variable cost begins to rise. The curves are mirror images of each other. This looks a lot like the same kind of picture we saw a few minutes ago when we compared the relationship between the marginal product curve and the marginal cost curve. Now, what we’re going to do is put all of this together and look at the relationship between marginal and average product, on the one hand, and marginal and average variable cost on the other hand.
Production and Costs

**Understanding the Relationship Between Marginal Cost and Average Variable Cost**

Now we’re going to put together the things that we’ve done in the last few lectures. In particular, we’re going to draw one graph that has in it the marginal cost curve and the average variable cost curve. And when we draw the two curves together in one graph, we’ll see a relationship between marginal cost and average variable cost. The relationship that we discover will be the same relationship that we discovered between the marginal product curve and the average product curve.

Marginals and averages always have the same relationship. That is, when the marginal is above average, when you’re adding things at the margin that are above average, you pull the average up. When you’re adding things at the margin that are below average, you pull the average down. We’ll see that same relationship now as we draw the marginal cost and average variable cost together in one graph.

Let’s start with a picture of the marginal cost curve. This is a review. Remember, the marginal cost is just the slope of the variable cost curve at any point. So when you’re going to draw the marginal cost curve, what’s the first point that you should look for on the variable cost curve? The first point that you look for is the point of inflection. That’s the point where the slope stops decreasing and starts increasing. And that would be the point where the marginal cost curve reaches a minimum. So I’ll go ahead and put that curve here.

Now, if I want to draw a marginal cost curve I’ll look over here and see the region where the slope of the variable cost curve is decreasing, and that will give me a decreasing marginal cost. So over in this region marginal cost is decreasing. Over in this region marginal cost is increasing and it’s headed towards infinity as the variable cost curve goes vertical. So if I’m careful here and finally that variable cost curve is going vertical, I can draw a marginal cost curve that represents the slope of the variable cost curve. And now I can label this curve marginal cost just like before. The marginal cost curve shows me the slope of the variable cost curve at every point.

Now, how do I draw in the average variable cost curve? How do I draw it in? Well, here’s what we know about average variable cost. We know that average variable cost is going to hit a minimum point where marginal cost curve cuts through it. Why? Because if the marginal cost is above average variable cost, it’s going to be pulling it up. And if marginal cost is below average variable cost, it’s going to be pulling it down. So in that case, what we get is a curve that looks something like this.

Now how did I draw this curve the way I did? I drew it because I know that any time marginal cost is above average variable cost, it’s pulling the average up, the AVC will be increasing. Anytime marginal cost is below average variable cost, it’s pulling the average down, the average will be decreasing. That means that if average variable cost is it’s usually U-shape, the bottom of the U, the minimum point has to occur where it touches the marginal cost curve, where it cuts through it.

Now the only thing that I haven’t explained in my geometry here is how did I choose this particular point for this intersection. How did I choose that point? And the answer is not easy. The answer is going to involve knowing something else about geometry. In fact, the way I got this was I used a ray from the origin upstairs to touch the variable cost curve. A ray from the origin has a slope that’s equal to variable cost divided by output. Notice, I didn’t label my axes here. Now, why did you let me get by with that? If you don’t label your axes you’re not measuring anything. Back to my ray from the origin. My ray from the origin here goes up and touches the variable cost curve at this point.

Now what does the ray from the origin tell you? Well, the slope of the ray from the origin is rise over run, and the rise is the variable cost and the run is the output. Well, what is variable cost divided by output? That’s right. It’s the average variable cost. So at this point the ray from the origin is at a minimum. That is, if we take this ray from the origin to other points on the variable cost curve, it’s going to have a higher slope. Here it’s going to be at a minimum. If I go beyond that, the slope goes back up. That’s a little geometric tidbit for you that lets you know why I picked that particular point. Your teacher is probably not going to require that you understand that, but I always like to know, “Well, why did he pick that point instead of some other point?” The point that you pick for the minimum average variable cost is the point where the ray from the origin would just touch the variable cost curve. That’s the point where the average is lowest.

Well, the point that you do need to understand that your teacher will want you to understand and that really is important intuitively, is that anytime the marginal is below the average, it’s pulling it down. And anytime the marginal...
Understanding the Relationship Between Marginal Cost and Average Variable Cost

is above the average, it’s pulling it up. One more thing I’d like to say and that is, if I flip this diagram over and look at this picture—does this look like anything you’ve seen before? Well, you may have trouble seeing it because the green is kind of pale, but I’m going to go over it again in red and hope that it looks familiar to you. This is the marginal cost curve turned upside down, its mirror image. What’s that going to be? What’s the mirror image of the marginal cost curve? The marginal cost is reciprocally related to the marginal product. So I can label this MP.

And here’s the average variable cost, I’ll go over it in red. The average variable cost is, of course, the mirror image of the average product of labor. So if I flip those graphs upside down, I’ve got a picture that should look very familiar to you. This is the picture we drew a few lectures ago where we put labor on the horizontal axis and output on the vertical axis, and we showed how marginal product and average product were related to each other.

Marginal product and average product are the reciprocals of marginal cost and average variable cost. And again, the intuition is, the more workers it takes to make a television, the more it’s going to cost. And the lower a worker’s productivity is, that is, the less output a single worker can produce, the more workers it takes to produce an extra unit of output. That’s the marginal idea. Or on average, if one worker, on average, produces one-fourth of a television, then on average it takes four workers to produce a whole television. That’s the average concept.

Productivity and cost are inversely related. You’ve seen it mathematically, you’ve heard a story about it, and you’ve seen how these graphs resemble each other. They’re mirror images of each other. Cost and productivity are reciprocally related. That’s the most important relationship in this series of lectures. However, there’s one other relationship that’s important also, and that is the relationship between marginals and averages. When the marginal is above average, it pulls it up. When the marginal is below average, it pulls it down.
Production and Costs

Defining and Graphing Average Fixed Cost and Average Total Cost

Remember, we're talking about a firm's decision in the short run. The short run is a period of time when some of the firm's input is fixed and some are variable. We're imagining that we're in a television factory, a factory that turns out television sets by using labor and other inputs. And those other inputs include a factory, the size of the factory, space, a conveyor belt, tools, anything else that they might use. But all of these other inputs are fixed in the short run.

Now if you want to calculate the cost of production, you have to not only consider the variable costs, which we've been studying, the cost of labor, but you also have to include the fixed costs. Now the interesting thing about fixed costs is they're very easy to calculate because they don't vary with anything. That's the definition of fixed costs. Fixed costs are costs of fixed inputs. They are fixed. They do not change. No matter how many television sets you produce, no matter how many workers you hire, fixed costs are fixed.

So you can look at a table like this one right here, and no matter how many television sets we produce, the fixed costs are going to be a constant $10,000. That's going to be our assumption. $10,000 is the cost of all of the inputs that we have to pay for that we cannot vary in the short run, including our tools, our conveyor belt, our factory space, etc. So whether we produce two television sets, 10, 30, 45, 48, 49, whatever, our fixed costs are all going to be $10,000. So fixed costs are constant, they are fixed.

However, the more television sets we produce, the more we can spread our fixed costs out over the individual television sets. If you only make two television sets, then the fixed cost per television is going to be $5,000. $10,000 divided by two is $5,000. You can only spread your overhead, your fixed costs, over two sets, and that's very expensive. But the more sets you produce, the more thinly you can spread your fixed costs out. If you produce $10,000 television sets, your fixed cost per set will go down to only $1.00. So the more your produce, the more your fixed cost per unit shrinks, shrinks, shrinks.

This fixed cost per unit has a special name. Can you guess what it is? It's the average fixed cost, of course. The fixed cost per unit of output produced. In order to calculate the average fixed cost, what you want to do is take the fixed cost, divide by the number of television set produced, and that gives you the average fixed cost. $10,000 divided by 40 televisions is $250 per television set. $10,000 divided by 49 television sets is $204 per television set. Notice that as long as you're producing more television sets, your average fixed cost is getting smaller and smaller and smaller.

If we ever went into an inefficient production where we added more workers and the number of television sets produced actually decreased, well, in that case, our average fixed cost would be rising again. But as long as you produce more sets, average fixed cost is shrinking.

Let's draw a picture of average fixed cost, and I'll move all these numbers over there so that you can use them for reference. If you want to draw a picture of average fixed cost, what you'll find is that average fixed cost is what we call in geometry a rectangular hyperbole. Now, that's a mouthful, and it refers to a curve that approaches both axes asymptotically. If you're producing very, very few television sets, then your fixed costs per unit are going to be very high. But the more television sets you produce, the more the average fixed cost is going to get closer and closer to zero. So I can label this curve "average fixed cost." It's very high when you're not producing many television sets. It then shrinks, and the more television sets you produce, the smaller it gets. This is spreading your overhead out over more and more television sets.

Now, let's take this average fixed cost and let's put it together with another diagram, another graph that we're used to drawing, another curve, and that is the average variable cost curve. So let's draw the average variable cost curve in here. And remember, it looks like this, AVC. So now we've got two curves together, the average fixed cost, which is the cost of your tools and the other fixed inputs, and the average variable cost, which is the cost per unit of the labor that you need to hire. Well together, your variable cost and your fixed cost make your total cost of production. So if you have average variable cost and you add it to your average fixed cost, you're going to get your average total cost. The total amount that you have to spend on labor and everything else per television that you produce.

Now, how do you calculate this average total cost curve? Where do you get it? Well, you can get it a couple of different ways, but since we're in a graph, let me show you how it looks graphically. If you take the average fixed cost of producing a given number of television sets right here—I'm sorry, the average variable cost of producing a given number of television sets—I've got to be careful here because I've got two curves, so I've got to make sure which one
Production and Costs

Total Costs

**Defining and Graphing Average Fixed Cost and Average Total Cost**

I’m looking at—and you add that to the average fixed cost—if you add these two together, then if I take this amount and put it on top of this and add them together, I get the average total cost.

The average total cost is just average fixed cost added to average variable cost. And if I add the two together, then I’m going to get a curve. So by vertical addition, I’m going to get a curve that looks something like—let me find a couple of other points on it—here’s one, add those two together—and eventually, I’m going to get a curve that looks like this. And there’s my average total cost curve.

The average total cost curve is the sum of the average variable cost and the average fixed cost. This is how much it costs our factory to produce television sets per unit, including the cost of labor and the cost of other inputs.
Production and Costs

**Calculating Average Total Cost**

The cost of production for a firm in the short run have two components: first there’s the variable cost. That’s the cost of hiring the variable input. In our story, the variable input is labor. And labor can be varied in the short run when you want to increase or decrease your production. The second component of costs in the short run is the fixed cost. That’s the cost of having a factory, conveyer belts, tools, all of the inputs that cannot be changed in the short run. The total cost of production in the short run, is the addition of variable costs and fixed costs. Let’s look at some numbers and calculate the total cost of production for our television factory.

If we hire one worker and produce two television sets, the fixed cost would be $10,000 and the variable cost would be $1,000 for a total cost in the short run of $11,000. And the same if we increased our input of labor. The number of televisions that we can produce increases, and with it variable costs, and with it total cost. Let me move my chart over a little bit so I can show you fixed cost and variable cost added together to give me total cost of production. In each case, the number in this column is the sum of these two numbers. The total cost is the cost of your labor plus the cost of fixed inputs. So the total cost of production will continue to increase as we produce more and more television sets.

Now, how do we draw a curve to represent the total cost of production? Well, it wouldn’t surprise you to learn that the total cost curve looks a lot like the variable cost curve. The only difference is the total cost has this $10,000 added on to it to represent not only the cost of labor, but also the cost of all the other inputs. So in order to get the total cost curve, we take the variable cost curve and shift it upwards by the amount of the fixed costs. I’ll move these numbers over there and do the rest of my work here in a graph.

Now, to my variable costs I want to add fixed costs, and fixed costs will be a straight line. That is, fixed costs—I’ll move this curve off for a moment—fixed costs are going to be a straight line that’s horizontal at the amount of the fixed costs. So in our example the fixed cost curve will be a horizontal line, meaning that no matter how many televisions you produce, you still have the same fixed cost, and that would be equal to $10,000. Now, how do you find the total cost of production? We find the total cost of production—oh-oh, I’ve got an unlabeled curve here. See, I’m not sure I can continue to work with it. So I’m going to make a temporary label, and my temporary label will say VC, variable costs. So there’s my variable cost curve labeled temporarily.

Here’s a graph with costs on the vertical axis measured in dollars, and television sets produced, or output, measured on the horizontal axis. And I’ll plop in this curve, which by now has the familiar shape to you of the variable cost curve. Notice it has a decreasing slope, a point of inflection, and then an increasing slope. Now ordinarily I would quickly label this curve VC for variable costs. Because you don’t want to have a curve sitting here with no label on it or you don’t know what it is. So, for the moment—well, I guess what I’m going to do is I’m going to make a temporary label here, because it’s driving me nuts not to have a label on this curve. I’m not sure I can continue to work with it. So I’m going to make a temporary label, and my temporary label will say VC, variable costs. So there’s my variable cost curve labeled temporarily.

Now, how do I find total costs? I find total cost by adding fixed cost and variable cost together. And I’m going to show you a little trick about how to do that graphically. It’s called vertical summation of graphs, right? For each given amount of output I want to add the variable cost of production on to the fixed cost of production. So I’ve got my variable cost curve here and I’ve got my fixed cost curve right here. And if I add the two together, I’ll get my total cost curve. That means I need to start with a given amount of output, let’s say, zero televisions. If you’re making zero televisions, you don’t have any variable costs because you’re not hiring any workers, but you still have your fixed costs because you’ve already got a factory rented and you can’t change that in the short run. So add on to this fixed cost zero cost of labor, and you get a point on your total cost curve. It will be right here were my finger is.

If you want to find the variable cost of producing, let’s say, 30 televisions, well, the variable cost is going to be $3,000 because you’ve got to hire three workers, and the fixed cost is $10,000, so add them together and you’ll get a point that is like this—add your variable cost on top of your fixed cost and you get a point on the total cost curve, which will be $13,000. This is the method of vertical summation. Add this point to this point to get a new point that’s further up the page. Well, what this looks like is this. We take the variable cost curve then and we shift it up by the amount of the fixed cost. So now the variable cost curve is sitting on top of the fixed cost curve. And now, it’s no longer the variable cost curve at all. It’s now the total cost curve, so I’m going to label it such. This is now my total cost curve, and the total cost is simply the sum of the variable cost and the fixed cost. I took my variable cost curve and I shifted it upwards by the amount of the fixed cost, which is the same thing as adding the fixed cost onto the fixed cost at every point on this axis. I just picked the fixed, added the variable, and I got my total cost curve.
Production and Costs

Total Costs

**Calculating Average Total Cost**

Now let’s notice a couple of things about the total cost curve. The first thing is the slope of the total cost curve at every point is going to be the same as the slope of the variable cost curve. That’s exactly right. This curve is parallel to the variable cost curve. I just shifted it upwards so the slope at every point is the same as the slope of the variable cost curve. And what is the slope of that curve going to be? The slope of that curve at every point is the rise over the run, the change in cost that results from a change in output. The slope is just a marginal cost curve. So the slope of the total cost curve is equal to the slope of the variable cost curve, which is equal to the marginal cost of production. That’s the first thing to notice, and that’s really the most important thing to notice.

So there’s your total cost curve. It tells you the variable cost plus the fixed cost added together. This is how much it’s going to cost you, in total, variable and fixed, to produce a given number of televisions. This is the summary of the firm’s cost constraints in the short run. If, in the short run, you want to produce a given number of television sets, you’ve got to go up to this curve to find the cost, the total cost, of labor and everything else that you need to produce that number of television sets. If this number is 40 and this total cost $14,000, that tells you how much you have to pay in the short run to produce those televisions. There’s the total cost curve.
Putting the Cost Curves Together

We’re going to put everything together now as we wrap up our lectures on the cost curves. Again, a lot of students have to kind of fight the idea of cost curves because it just looks like a bunch of lines floating around on a page. Things are going to go better for your in your Economics class if instead of just seeing lines on a page, you have some sense of the intuition of the story behind them. And that’s what we tried to do in these last few lessons. I’m going to wrap everything up now and draw the picture that we’re going to take with us into the lectures on profit-maximization. And as I wrap up, I hope I’ll be reinforcing some of the intuition that we’ve covered in the previous lectures.

So let’s look now at two curves that we were drawing in the last segment—the average variable cost curve and the average total cost curve. Now, what should you think about when you see these two curves? Well, let’s look at average variable cost first. Remember, the average variable cost curve is the mirror image of the average product curve. As workers become more productive, it takes fewer workers to produce a television set, and the cost of producing a television set, the labor cost, begins to fall. When the workers’ average productivity begins to fall, the labor cost per television goes back up.

So when you look at the average variable cost curve, you’re getting a summary of the productivity, the average productivity, of the workers of the firm. We are experiencing here increasing average product, down to this point of minimum average variable cost, then the average product of a worker starts to fall and the labor cost per unit goes back up. That’s what you’re thinking about when you’re looking at that curve.

The average total cost curve is just the sum of average variable cost and average fixed cost. I want to take a little break here to remind you about where these numbers come from. Remember, we have the average fixed cost, which is the fixed cost per unit of output, and we also have the average variable cost, which is the variable cost per unit of output. This is the labor cost. And if you add the labor cost on to the cost of the other inputs, then we’re going to get the average total cost, the total cost per unit of output. Labor cost added to the cost of all of the fixed inputs gives you the average total cost.

And notice that average total cost is falling at first, but eventually starts back up again. It’s falling at first for two reasons. It falls at first—I’m going to put these numbers next door so we can talk about them. It falls at first because you’re able to spread your fixed costs out thinner and thinner and thinner over your increasing output. It also falls at first because increasing productivity means it takes fewer workers, on average, to produce a television set. After a certain point, however, even though the average fixed cost continues to fall, even though you’re spreading your overhead out thinner and thinner, we’re now working against decreasing labor productivity. After a while, your workers start to get less productive on average, and that then tends to raise the costs up.

So we’ve got these two forces in balance. Your overhead is spreading thinner and the average fixed cost is falling, but your workers are getting less productive, so the average variable cost is rising. Eventually, the effect of the average variable cost takes over and begins to pull the average total cost up. But the minimum point of average total cost will be further out. That is, it will be further out than the average variable cost minimum, because even after you’ve reached the maximum of labor productivity, you still have some more of that overhead spreading to keep pushing your costs down for just a little bit longer. Then the falling labor productivity takes over and starts to pull up the average total cost.

The thing to notice about these two curves is that the difference between them, the difference between the average total cost and the average variable cost, at every point, is the average fixed cost. Here, the average fixed cost is a big gap. Here, the average fixed cost is smaller, and that gap shrinks and shrinks and shrinks and shrinks. As these curves move out, as we increase the production of television sets, the average total cost curve and the average variable cost curve get closer and closer together. Why is that? Why do these curves get closer and closer together as we move further out, as we increase our output? These curves get closer together because the vertical distance between them at any point is simply the average fixed cost. And since the average fixed cost is always shrinking as output increases, these curves get closer and closer together. They never quite touch, but they will approach each other asymptotically.

All right, so there’s the average total cost curve, which is the average variable cost curve added on to that average fixed cost curve. And there’s one more curve we need to draw, and that is the marginal cost curve. Now, how do we draw the marginal cost curve in this picture with no other points of reference? We draw it by remembering the
Production and Costs

Putting the Cost Curves Together

relationship between averages and marginals. When the marginal is below the average, it’s pulling it down. When the marginal is above the average, it’s pulling it up. That means that every marginal cost curve will cut through average cost curves at their minimum point. Because when the curve is below it, the curves continue to fall. When the marginal cost curve rises above it, it pulls the average up. So we know that the marginal cost curve will look like this.

Remember, the marginal cost curve tells you about the marginal product of labor. It tells you about the workers that you have to hire—the additional workers that you have to hire to produce an extra television set. Here, the marginal product of labor is increasing due to teamwork and specialization. The marginal product of labor reaches its maximum when the marginal cost is at a minimum. Remember, marginal product and marginal cost are reciprocally related. Then the marginal cost curve cuts through the average variable cost curve at the minimum of average variable cost. Over here it’s pulling the average down, and once it crosses it, it begins to pull the average up. The same with the average total cost curve. As long as the marginal is above the average total, it’s pulling the average total cost down. But when the marginal cost curve cuts through it, it starts to pull it up.

With these three curves we are ready to go into a discussion of how the firm can maximize its profits. These three curves become very important in describing the profit-maximizing choice of a firm. Anytime you look at these curves what you’re really looking at is information about the firm’s productivity and the price the firm has to pay for its inputs. That’s the information that’s summarized in these curves. Now we’ll take that information and use it to predict profits.

But before I end this lecture, let me ask you one more question. If you were a firm that wanted to maximize its profits, which level of output would you choose for your firm in the short run? Which level of output in the short run would you choose? Would you choose that level of output that maximized marginal product, that is, where teamwork and specialization were at a peak? Would you choose that level of output that minimizes marginal cost? That’s your first choice. Or perhaps you’d choose that level of output that’s at the bottom of the average variable cost curve, where the labor cost per unit is at a minimum? Would you choose that point? That’s the point where the average product of labor has reached its peak, where you’re getting the most output per worker possible. Is that the point you’d choose? Or finally, you might choose to produce at the point where average total cost is at a minimum. This is where the total cost of producing a television, on average, is the lowest it’s going to get, where the combined labor and fixed input costs per television are the lowest they’ll ever be. Which point would you choose? Minimum average total cost, minimum average variable cost, or minimum marginal cost. Which one is best?

The answer is, none of the three. You don’t have enough information in this picture to answer the question, what is the profit maximizing choice for a firm. You know why? Because the only information that’s in this picture is information about costs, and costs is only one side of the story. What’s the other side of the story? If you want to maximize profit, what else do you need to be concerned about besides your costs? The answer is, you also have to be concerned about revenue. The revenue that you earn from selling television sets. See, there’s no information in this picture at all about the price of a television set. And until you know what the price of a television set is, you don’t know how many televisions you want to produce. If the price is high enough, you’d be willing to incur very high costs. You cannot tell from looking at this diagram alone what the profit maximizing choice is. This diagram tells you only about costs. That is, it summarizes the firm’s technology and the price of the firm’s inputs to give you information about the cost of producing televisions.

Now I’ll reel out information about revenue in the next lectures to begin to discern the profit-maximizing choice of a firm.
Production and Costs

Long-Run Production and Costs

Defining the Long Run

We've been talking for the last few lessons about the behavior of the firm in the short run. Now, at long last, we're going to talk about the long run. The long run is the period of time in which all inputs become variable. The firm has no fixed inputs, no fixed costs; everything is to be decided in the long run. You can decide which inputs you want to use, how you want to combine them, how big you want your firm to be. The long run differs from the short run in two important ways. By that I mean, there are two decisions that the firm has to make in the long run that it doesn't have to worry about in the short run, and that's what I'm going to be talking about in this lesson.

First, let's do a quick review. A firm always maximizes its profits by producing where price equals marginal cost. Whether the firm is in the short run or in the long run, the firm will always keep expanding its output until the marginal cost of adding more begins to exceed the price. That's always going to be true. The difference between the long run and the short run is how the costs are determined. In the short run, the firm has a fixed quantity of tools to work with, a factory of a fixed size, a certain number of conveyor belts, and a certain number of tools; they're fixed. The firm pays for these tools, that's the firm's fixed costs, and the workers work with them which determines the worker's productivity, but in the short run, the quantity of tools is fixed. The way the firm can increase its output in a short run is by adding more workers. It can have one, two workers, it can hire a third, it can hire a fourth, but the only way the firm can expand its output in a short run is by adding more workers. Eventually those workers begin to congest the fixed inputs, there aren't enough tools to go around, the workers are getting in each other's way, productivity begins to fall, and of course, marginal cost goes up. In the long run, the firm gets to make two decisions separately. Two decisions separately that are linked together in the short run. Let me point this out, if the firm has a fixed quantity of tools, then the only way the firm can expand its operations is by increasing the ratio of workers to tools. That is, the firm can hire more workers, but as it hires more workers then each worker will have fewer tools to deal with. The more the firm expands its output, the more it increases the size of its operation, the more the workers will be crowded onto the tools, or the fewer tools there will be per worker. In the long run, the firm can make these two decisions separately. It can choose the size or scale of its operation independently of the ratio in which it uses workers and tools. Let me make clear the two decisions that the firm faces in the long run.

The first decision that the firm faces in the long run is the choice of technique. In economics when we refer to the technique of production, we're referring to the way in which the firm combines its inputs. That is, a technique where you use a lot of workers with very little capital, that's one possibility. Say, this is the way rice is grown in rural China, a lot of labor is used with very few tools, that's one technique. An alternative would be to use one worker and a lot of tools, and this is a way rice is grown, say, in Louisiana where more capital is used, more combined more irrigation, more tools are used, and fewer workers are used. So the technique is capital intensive, that is, it uses a lot of capital and very little labor, or whether the technique is labor intensive, that is, it uses a lot of labor and very little capital.

That's something the firm can decide in the long run, how it wants to combine these inputs so as to minimize the cost of production.

The second decision that the firm faces in the long run is once it's decided how it wants to combine labor and capital; it has to decide how big it wants its operation to be. Does it want a small operation that uses just a few workers and a few tools to produce output, or does it want a big operation with a lot of workers and a lot of tools? The bigger your operation gets, of course, the higher the cost could get or the lower the costs could get. The scale of operation, that is, the size or your operation influences the costs of production, and therefore, influences profit. So, the firm decides whether it wants to be big or whether it can make more money by being little. Let me quickly review, as we move ahead into an analysis of the firm's behavior in the long run, we're going to be considering two decisions that the firm can now look at separately. Two decisions that are somehow meshed together in the short run that can be made separately in the long run.

The first is the choice of the cost minimizing technique, how to combine labor and capital to produce output at the lowest cost. The second is the choice of scale. Is a small operation or a larger operation going to give us more profits? We'll look first at the question of cost minimizing techniques, how to combine labor and capital to produce the given quantity of output. Then we'll consider the question of scale, is the firm going to do better by being large or by being small.
Production and Costs

Long-Run Production and Costs

Determining a Firm's Return to Scale

In the last lecture, we discussed the first problem that the firm faces in the long run. That's the problem of choosing the cost minimizing technique, the way of combining capital and labor so as to minimize the cost of producing a given quantity of output. In this lecture, we're going to consider the second problem that the firm faces in the long run. And that is the choice of the optimal scale, the right size of operation for maximizing your profits.

In the short run, the firm can't help but change its scale and change its technique at the same time. If the firm wants to produce more output in the short run, they've got a fixed amount of capital and they can only increase output by adding more labor. That means, that in order to increase the size of their operation, or their output of television sets, they've got to add more workers, which increases the intensity of labor. That is, they're using more labor per unit of capital. So, they're also choosing scale, the size of their operation at the same time as they're choosing the technique of production. Three workers and one tool may be the only option that's available to the firm for increasing their output in the short run. The firm is simultaneously choosing the technique of production and the size of their operation.

In the long run, the firm gets to choose these two things separately, which is one of the reasons why long run costs of production are always lower than the short run costs. The firm has more flexibility. If the firm wants to increase the size of its operation, it doesn't necessarily have to use more labor per unit of capital. It can also increase the amount of capital that each worker has, thereby keeping the labor capital ratio constant even as its firm grows bigger. That is, it keeps the cost minimizing technique that it prefers while allowing the scale of operation to grow larger.

The definition of the scale of operation is as follows. Scale is said to increase if the firm has a proportional increase in all of its inputs. So, if we start with one worker and one unit of capital, the scale of the firm is said to increase if we add a second worker with a tool of his own, and a third worker with a tool of her own. To have an increase in the scale, it is required that all inputs increase by a particular proportion. That they all increase by the same proportion. If we just increase labor and hold capital constant, that is not a change in scale. If we just increase capital and hold labor constant, that is not a change in scale. To change the scale of operation, we must have a proportional increase in all of the factors of production.

Now, here's the question that we want to answer. How does the firm's productivity change as we change the scale of operation? In order to answer that question, we're going to need to define a concept very specifically and that's the concept of economies of scale. What happens to the firm's productivity when the scale of its operation changes? Suppose we go from one worker with one unit of capital to two workers and two units of capital. What happens to output? Well, you can imagine that one of three things would happen. The first is probably the most reasonable assumption. And that is that if both of our inputs double, our output would double as well.

This is the case when the firm is said to have constant returns to scale. Constant returns to scale is defined as a technology where increasing all inputs by a given proportion increases output by the same proportion. The reason this seems like a reasonable assumption is the possibility of replication. If we've got a guy over here making television sets by using a particular tool and he can make ten television sets a day, if we have another worker who has the same toolkit, we would imagine that that worker could also make ten television sets a day. Replicating this operation, would allow us to double the amount of output we get in a single day from our factory.

If we add a third worker, also with a tool, we should be able to replicate the operation again. Tripling inputs, we would expect to be able to triple output. That is the idea of constant returns to scale. And it seems like a reasonable assumption because of the possibility of replication, having someone else do what another worker with his tools is already doing successfully. Now, there are two other possibilities. The first is the possibility that things are actually better than constant returns to scale. That adding a worker with a tool, adding a second worker with a second tool allows us to more than double our output.

That is, if we have one worker with one tool, producing ten television sets, two workers with two tools might actually be able to produce 30 television sets in a day. That would be the case if somehow there's some kind of teamwork and specialization between these workers, if working together, they can operate in a way that's more productive then if both of them worked separately. Having the workers work together and cooperate, using their labor and tools together, may actually allow them to be more productive than two workers simply replicating what each other are doing. This is the case with increasing returns to scale.
Production and Costs

Long-Run Production and Costs

**Determining a Firm’s Return to Scale**

The final possibility is the possibility of decreasing returns to scale. It may be that by doubling all of our inputs, we actually increase output by a smaller proportion, maybe only ten percent or 20 percent. Now, you might ask yourself, “Why would we ever have decreasing returns to scale?” What could ever lead to a situation where two workers with two tools would do less productive work than they could do if they were working separately? And it has to do with the problem of a growing organization. As an organization gets larger and larger, you have more and more activity under a single manager and that manager is responsible for making decisions about how this process in organized, how the workers are paid, how they are scheduled, how their activities are coordinated.

When you’ve got a manager who’s over this whole operation, it may be that the individual workers say to themselves, “Hey, rather than trying to make my money by working harder, producing television sets, I’ll go and lobby the manager for a raise. I’ll go and ask for favors. I’ll ask for a better schedule. I’ll try to use my influence to improve my position in this organization. This is a form of rent seeking and it goes under the name of influence activities. Influence activities are efforts on the part of the workers to influence the organization to their advantage and the cost of these influence activities is this. When workers believe that they can get more benefits by lobbying the boss, they’ll spend less time actually producing televisions for the firm.

Because of that, as the organization grows larger and the manager has more and more power, it is a greater and greater temptation for the individual workers to engage in influence activities, rather than to produce television sets. And the cost of production then goes up as the firm has to employee monitors to make sure that people stay on the job and don’t spend their time taking too many breaks, going and lobbying the boss and such things like that. This is the idea behind decreasing returns to scale. As the firm gets very large, the problem becomes managing workers and preventing them from engaging in influence activities.

So, this is the idea of scale economies. The firm wants to consider, what is the potential for teamwork and specialization? What is our scope for enjoying increasing returns to scale? Because as long as we’re enjoying increasing returns to scale, it will probably pay for us to expand our operation. Also, we want to know when are we likely to have constant returns to scale. Because any time a firm has constant returns to scale, replication is a possibility. And finally, there’s the problem of decreasing returns to scale. If the firm grows too large, its workers are tempted to engage in influence activities, thus raising the cost of production. In the next lecture, we’ll see how economies of scale translate into the costs of the firm in the long run.
Production and Costs

Long-Run Production and Costs

Understanding Short-Run and Long-Run Average Cost Curves

We're back with cost curves and we're going to take it slow here to try to clear up some confusion that I know I suffered in my economics class that maybe we can spare you from. And this is the confusing question of the relationship between the short run and the long run cost curves. All right. Now, let's think for a moment, before we draw any curves, about what we're talking about. The long run, is the period over which all inputs are variable. When I draw a long run cost curve, I'm assuming that the firm can change anything that it wants to, labor, capital, the combination of labor and capital, the scale of output, any of that stuff can be change when we draw the long run average cost curve.

The short run cost curves, on the other hand, and you'll recall them from our earlier lessons, these short run cost curves assume that some inputs are fixed. Now, how are these cost curves with fixed inputs related to this cost curve that has no fixed inputs, this cost curve where anything goes? The relationship works like this. Let's suppose we have a firm that has a long run average cost curve. And let's suppose the long run average cost curve for this firm has a U shape. What this U shape means is this. It means that a first the firm experiences decreasing returns to scale, or increasing returns to scale that cause the average costs to decrease. Over this region right here, the firm has increasing returns to scale and the average cost of production is falling.

At the bottom of this curve, it's temporarily flat, maybe just for an instance, and that's the point of constant returns to scale. Finally, if you go beyond that point, you have increasing average costs, which is a signal of decreasing returns to scale. So here the firm has increasing, then constant, then decreasing returns to scale. So we have decreasing average cost, a moment of constant average cost and then increasing average cost. As we move along this curve, remember the firm can alter anything it wants to. It can change the way in which it combines the inputs. It can change the scale of operation. Anything goes, there are no fixed inputs.

Suppose, however, we pick a point on this curve and the point on this curve that we pick will involve a certain combination of labor and capital to produce a given amount of output and that combination of labor and capital will have a particular cost and produce a certain amount of output. Suppose, now we freeze capital. We freeze capital at the amount that the firm is using at this point on its long run curve, a particular square footage of factory, a particular number of tools and conveyor belts. Now if we freeze capital now and allow this firm to alter its output by changing only labor, we're back in the short run. We're back in the short run, where the firm can only change its behavior by changing its variable input labor.

All of its costs are going to be higher in that case. Do you know why? Why are all of the costs in the long run going to be lower than the costs in the short run? Why in the short run does it always cost a firm more to change its output than it costs the firm to change its output in the long run? The answer is, in the long run the firm has more options. The firm can combine labor and capital so as to minimize costs. In the short run, the firm can't change its capital. It's stuck with a given amount of capital. It has to use it, therefore its options are reduced and its costs are going to be higher necessarily. If there were better options with lower costs, the firm might not be able to get to it in the short run because of its fixed inputs.

In the long run, of course, the firm can get there because none of the inputs are fixed. Everything is variable. So, if we want to look at the short run costs they're going to lie above the long run costs. The short run costs of production are going to lie above the long run costs at every point except the one point that touches the curve. At this point, where the curve touches the amount of capital that's being used is cost minimizing for that level of output. There's one point, that the short run average cost curve has in common with the long run average cost curve. That's the point at which the firm is operating with the cost minimizing combination of labor and capital for that particular amount of output.

If you want to change output in the short run, you can hire more labor, but you can't get down to the blue curve in the short run, because you have fixed inputs. Therefore, in the short run you're stuck with higher costs. Well, that would be the situation if we froze capital at the amount associated with this point. Suppose we freeze capital at the amount associated with the point at the bottom of the curve. In this case, you're going to get short run cost curves that look like this. If you're at the bottom of the curve, using the cost minimizing combination of labor and capital, you will be at the point where the short run curves intersects the long run curve.

However, if you freeze capital at that particular quantity, and want to change your output in the short run, all you can vary is your labor. And since you have less flexibility, your costs are necessarily going to be higher. That's why the
Production and Costs

Long-Run Production and Costs

Understanding Short-Run and Long-Run Average Cost Curves

short run and average cost curve lies everywhere above the long run average cost curve. I’ve labeled this the average variable cost. It’s not average variable costs. In fact, it’s average total cost. It’s the cost of all labor and capital combined. So make sure you labeled your curves correctly. I made a mistake and called it average variable, which would only refer to the labor costs. In fact here, we’re referring to the combined costs of labor and capital. That’s why I need to call that the average total cost.

I hope I’m making my point clear. The point is, that in the short run, your average costs will always be higher than it will be in the long run, to produce any given quantity of output. That’s because in the long run you have more flexibility. You can change the combination of labor and capital to get to the cost minimizing technique. Now, one diagram that you’ll see in your Economics textbook, is one that has the long run average cost curve with a lot of short run cost curves all along it.

So, what you get is, this, and that, and finally, if we go over here into a region of decreasing returns to scale, and increasing average cost, you get that. This is a picture that shows the relationship between the long run average cost curve, the blue curve, and all of the short run cost curves that are associated with it. The blue curve becomes the envelope. It becomes the envelope that contains all of the green curves. All of the green curves fit on top of it, because the average costs, and here again I’ve got to make sure I’m saying average total cost. The average total cost will always be greater in the short run than it is in the long run and that’s quite simply because in the long run, you have more flexibility.

By changing labor and capital together, by finding the cost minimizing techniques, you can always lower your costs in the long run. Unless you happen to be right at this point of tangency, right at the point where the firm has just the right amount of capital to be cost minimizing even in the short run. If you start at a point on this curve and then change only labor, costs will be higher. But if you happen to be at the point that’s tangent to the blue curve that means that you’ve got just the right combination of labor and capital in the short run to be cost minimizing. In the long run, you wouldn’t want to change anything to produce that particular quantity of output.

One more thing to notice that’s kind of a technical thing that students sometimes have trouble with, this material. And that is, when you’re drawing the short run cost curves, in relation to the long run cost curve, notice that it’s only at the very bottom that the minimum average total cost in the short run is the same as the point of tangency. It’s only at the very bottom that that happens. Otherwise, if you looked up here, at a point where the long run average cost curve is downward sloping, the bottom of the green curve is not going to be where it touches the blue curve. That’s because the tangency is at a point where the blue curve is downward sloping.

The point where the two curves coincide is going to be at a point where the green curve and the blue curve are touching and they share a downward slope. Over here, they touch and they’re sharing an upward slope. So, when you draw these curves, you want to be careful. So, what’s the intuition of that? What’s the intuition? Why is the point where they touch not always the minimum point? And I guess the answer has something to do with the way fixed costs work. You can always decrease your costs a little further by increasing labor productivity in the short run, or over here, if you’re at a point where labor productivity is now, you’ve got diminishing returns to scale. You’ve want to back off a little bit to get to the point of minimum.

But, I’m not sure that that’s—it’s a bit of a subtle point and it’s not something that you probably need to fret too much over right now. The point that I’m trying to make is, the key to understanding this diagram is the relationship between costs in the long run and costs in the short run. Costs in the short run are always higher, because in the short run, you have fixed inputs and you can’t necessarily get to the cost minimizing technique. Costs in the long run are always lower, because you have more options and can therefore more likely get to the cost minimizing technique.

We’ve looked at three extreme cases of the long run average cost curve. In one case, we had increasing returns to scale for every level of output. In another case, we had constant returns to scale for every level of output. And finally, in the third case, we had decreasing returns to scale. Now, it will seem apparent that not every technology is going to have constant, increasing or decreasing returns throughout. In fact, it’s probably unreasonable to expect that each technology would not have some range of each of these. That is, for most technologies, whenever a firm is small and growing, it will be experiencing some increasing returns to scale, as workers and tools are used with teamwork and specialization.
Production and Costs

Long-Run Production and Costs

Understanding Short-Run and Long-Run Average Cost Curves

At some point, constant returns to scale will occur and as the operation grows still larger and problems with coordinating and management begin to arise, then decreasing returns to scale will then kick in. If this is the case, and this is what economists usually believe is true with firms, that there is increasing returns to scale at the beginning then constant returns to scale at some point and finally decreasing returns to scale as the firm grows larger. Then the long run average cost curve will have the U shape that I’m showing right here.

The U shaped long run average cost curve represents a technology that has increasing returns to scale then constant returns to scale and then decreasing returns to scale. As a quick review, let’s see how the scale economies relate to the shape of the average cost curve. Whenever there are increasing returns to scale, due to teamwork and specialization, the average cost of production is diminishing. If the firm is producing only a few units, there’s not a lot of scope for teamwork and specialization. But as output expands, workers and tools can be used in conjunction with each other in a kind of complimentary way, with teamwork and specialization, and as productivity increases, the cost of production is falling.

Finally, at some point, constant returns to scale will kick in. That is the scope for teamwork and specialization has played out, it’s been exhausted. And now, here, at the bottom of the average cost curve, where it’s flat for just an instant, the firm has constant returns to scale. That is, expansion leaves the average cost constant. However, if the firm continues to grow, that is, if output continues to expand, the capital and labor needed to expand output is now making the organization so large that the firm begins to have some of those management problems we talked about earlier, influence activities and rent seeking within the firm that take away from productivity.

Beyond this point, we find then that the average cost to production begins to rise again because proportional expansion in all inputs results in a less than proportional expansion in output. Which means the average cost to production is rising. So, the region of increasing returns to scale, where the average cost is falling, constant returns to scale at the bottom, where the curve is flat, if only for an instant, and then decreasing returns to scale as the average cost begins to rise again.

Let me introduce one more concept that will become important, later when we talk about the long run and how firms enter and leave the market. And that is this point right here, at the bottom of the long run average cost curve, this point where constant returns to scale exists for jus a moment, where the curve is flat. If we put a straight line at the bottom of this curve it will be tangent at the very bottom, that is, at the point at the bottom of the long run average cost curve. We call this point, where long run average cost is at its minimum, we call this the point of efficient scale. This is the size of the firm’s operation. This is the quantity of output or the number of television sets that the firm produces when it is in some sense the right size.

The right size means, the point at which the cost of producing a television set is at its minimum, the average cost is at its minimum. And in the long run, as we’ll see a little later, the firm is going to wind up at this point. So, the point where the long run average cost is at its minimum, this point where the long run average cost curve is flat, right at the bottom, that moment of constant returns to scale, that has a special name in Economics. We call it the point of efficient scale and we’ll be returning to that point in just a little bit. Next, we’re going to look at how the long run and the short run cost curves are related.
Another question that students ask is "When do you move along the cost curve and when do the cost curves shift?" Let's see if I can answer that one. Over here, we have the long run and the short run cost curves and I've reproduced them over on the board so that we can talk about them. First, anything that changes the price of the firms product is going to cause a movement along the cost curves. Remember, you can never shift the curve by changing what's on one of its axis. You can only shift the curve by changing something that isn't in the picture. You can only shift a curve by changing something that you're holding constant when you draw the picture.

So, if the price of the firm's product changes you will move along the curve. In the short run the firm will move along the short run marginal cost curve and in the long run, the firm will move along the long run marginal cost curve as the firm expands its output. However, if you change one of the things that's held constant when you draw this picture, then you'll shift the whole set of curves. For instance, let's suppose that the price of the firm's inputs increase. Let's suppose that the price of labor and capital increases. In that case, what will happen to the cost curve is this. The whole set of cost curves will shift upwards, representing an increase in the cost of producing any given quantity of output.

If the price of the inputs falls, then the cost curves will shift downwards, representing a reduction in the cost of producing any quantity of output. The next thing that could shift the cost curves would be a change in technology. If the firm's technology improves, that is, if it can make more output with a given amount of input than before, then the cost curves will shift downwards because now the firm can produce its target level of output with less input and therefore at lower cost. So the thing that will shift the cost curves will be a change in the price of the firm's inputs or a change in the firm's technology.

One more thing to point out. I told you a moment ago that when the firm moves into the long run and changes its quantity of capital there is a shift in the short run supply curve. That shift occurs not because the short run marginal cost curve actually shifts, but because the firm moves from one short run marginal cost curve to another short run marginal cost curve. Remember, any time you change the amount of the fixed input you have to change the short run marginal cost. The short run marginal cost is based on labor productivity, which depends on the amount of tools or capital that the workers have to work with.

Any time you pick a point on our blue long run average cost curve, any time you pick a point, you could hold capital constant and draw a new green curve that's tangent at that point. If firms acquire more capital in the long run, they're going to be moving from this set of short run cost curves to another set of short run cost curves at a different point on the long run average cost curve. So, it's not that the green curve actually shifts for a single firm. It's that you move from one set of short run cost curves to another set of short run cost curves.
Production and Costs

Isocost/Isoquant Analysis

Constructing Iso-Cost Lines

The first thing that the firm can do in the long run that they can’t do in the short run, is to find the best way of combining capital and labor to produce output. We’ve been talking about the technological possibilities of the firm. That is, the ways in which a firm is able to combine capital and labor to produce a given quantity of television sets. Now, we’re going to consider which combination of capital and labor is in fact the best combination for the firm in the long run and best means cost minimizing. In order to maximize its profits, the firm needs to minimize the cost of producing whatever level of output it chooses. In this section, we’re going to be looking at how a firm minimizes its costs in production.

Let’s look first at a way of representing the costs of production in a firm. A firm is going to be combining capital and labor to produce output and it’s going to be looking for a way of combining these inputs to produce a given quantity of output at minimum cost. Let’s begin by representing the cost of production for a firm. I’ll start with an equation, all of this will be reproduced over on the board for you to use throughout this lecture, but I want to do it carefully at first so that the math makes sense. We’ll use the expression TC to stand for total cost and the total cost of production for the firm will be equal to the wage it pays its workers times the number of workers that it hires, plus the price it pays to rent a unit of capital time the number of units of capital that the firm actually rents.

So, the labor cost plus the capital cost is equal to the total cost. Here we’re assuming that the firm has only two inputs, labor and capital. If it has other inputs, you would just simply add those on in this equation. Now, you’ve seen the graph that we’re going to be using to graph the firm’s costs. In order to draw a line that represents the firm’s costs in that graph, and since capital is on the vertical axis, we need to isolate capital on one side of the equation. That is, we’re finding the intercept slope formula for writing a straight line and that straight line is going to be the total cost. So what I’ll do first is I’ll move everything to the other side of the equal sign except for the capital term.

So, let’s do that. I’m going to have now total cost minus labor cost is equal to the cost of capital. And now let’s divide both sides through by the price you pay for a unit of capital, and that gives us total cost divided by the price of capital, minus the wage divided by the price of capital times labor is equal to K. Now we’re used to writing the vertical intercept, the variable, over on the left hand side of the equation, so let me just switch these two expressions and we get, K is equal to the total cost divided by the price of capital minus the wage divided by the price of capital times the amount of labor that the firm hires. This again, is just another way of writing the total cost of production for the firm, only now we’ve isolated that in terms of capital. We’ve got here the firm’s budget expressed as the amount of capital and labor that the firm chooses to hire.

Notice that any values of capital and labor that we’ve plugged into this formula have to satisfy this equation and that equation means that every combination of capital and labor that satisfies this equation, will have the same total cost. Now, I could represent this equation with a line in the diagram. So, if we look at our diagram, and we’ll keep the equation over on the board for reference, we can draw a straight line and we call this line an isocost line. The definition of the isocost line is this. The isocost line is a collection of points representing all combinations of capital and labor that have the same total cost. The intercept of the isocost line here is the total cost expressed in terms of capital. So, if the total cost along this line were $100, and the price per unit of capital were $10, this would be ten units of capital. Ten times $10 apiece equals our total cost of $100, which is another way of saying, that the firm’s budget expressed in terms of capital is the equivalent of ten units of capital. Down here, at this intercept, we have the firm’s total cost expressed in terms of labor. So, if the total cost, again, is $100 and the price of labor is $20 per worker, then $100 divided by $20 is five workers. So, the intercept here is just the firm’s budget spent all on labor or expressed in terms of labor. Finally, and this is most interesting and most important to our analysis, the slope of this line is equal to the wage divided by the price of capital. That is, it’s the wage relative to the price of capital or the relative price of the two inputs.

Now, here’s how to think about the slope of this line. The slope of this line is the opportunity cost of hiring an extra worker. If you want to hire an extra worker, you’re going to have to pay the wage, W. But in order to pay that wage, you’re going to have to reduce your employment of capital in order to free the money that you would use to pay the worker. Well, how much capital do you have to lay off in order to be able to afford hiring an extra worker? And the answer is one over the price of capital. If the price of capital is equal to $10 and the price of labor is equal to $20, then you’re going to have to get rid of two machines in order to be able to hire one extra worker. So, the slope of this line tells you the number of units of capital that you have to give up to be able to hire an extra worker, the rise over the run, the slope of the isocost line.
Production and Costs

Isocost/Isoquant Analysis

**Constructing Iso-Cost Lines**

So, a quick summary. The isocost line has a position that’s determined by the total budget. If you have more money to spend, your isocost line will be out further from the origin. If you have less money to spend the isocost line will be shifting inwards and if you have very little money to spend, the isocost line will be close to the origin. If however, you change the price of the factors, then you’re dealing with the slope of the isocost line. A change in the price of labor will change the slope of the isocost line. If labor becomes less expensive, the isocost line will pivot outwards. That is, if you spend your whole budget on labor, after the wage falls, you can afford to hire more labor than you were hiring before. When the wage is higher, then the isocost line becomes steeper. That is, if you spend your whole budget on labor, you can afford less labor than before the wage went up. If the price of capital changes, you get the same kind of analysis. A lower price of capital causes the curve to become steeper. A higher price for capital cause the curve to become flatter.

What we’ll be looking at, in a moment, is how to find the cost minimizing technique of production. Given the price of labor and the price of capital, that is, given the slope of the isocost line, we’ll be looking for a production plan that let’s us shift this line as close to the origin as we can get it. Getting the isocost line close to the origin, shifting it down close to the origin, means spending less money on labor and capital, minimizing the cost of production. So, what we’ll do now is we’ll take this tool, the isocost line and put it over in the diagram with our isoquants and identify the cost minimizing technique for producing a given quantity of output.
Isocost/Isoquant Analysis

Understanding Isoquants

We’re back with the cost curves in the short run and the long run. We’re going to do one more short lecture to show you the relationship between short run and long run cost curves. I hope you’ll enjoy this lecture, I know I would have found it useful. What we’re going to do in this lecture is show you the relationship between the short run cost curves and the long run average cost. The next thing we’ll do is use those cost curves to show you how an industry adjusts in the long run when something changes.

First, let’s go back and look at the short run cost curves that we derived in an earlier lecture. I don’t know if you can see these curves, so let’s take advantage of perspective and increase the size of them and make the graph a little bit more legible. Here you’ve got a diagram with the short run marginal cost curve and the average total cost curve that we’ve looked at in a previous lecture. Remember, the average total cost curve is u-shaped and the short run marginal cost curve intersects the average total cost curve at its minimum point. What I want to do now is put this curve in the long run average cost curve diagram. So, here we have it again with the long run average cost curve here in blue and the short run cost curves in green. The average total cost curve touches the long run average cost curve at the point where the amount of fixed input being used on these green curves is cost minimizing for a particular amount of output. So, if we use this amount of output right here, we can call this Y*, and we look at the blue curve, we’re looking at the minimum cost, the minimum average cost of producing that amount of output in the long run. The firm has chosen the cost minimizing combination of capital and labor. Now, if we freeze capital at the amount that’s cost minimizing for this level of output, and then change only labor, we trace out the short run cost curves. Remember, short run is defined as the period of time over which some inputs are fixed and others are variable. In our story, the only variable input is labor, so if the firm wants to change its output in the short run, the only way it can do that is by changing labor, and therefore, you get this set of cost curves. Now, you’ve seen this before, the point that I want to make is this. We can put in the short run marginal cost curve, that marginal cost curve that we drew in earlier lectures, and notice that it intersects the average total cost curve at the minimum of that average total cost curve. Notice also, that that minimum point is not the same point as the point of which the average total cost curve is tangent to the long run average cost curve. So, whenever the firm is operating in a region of increasing returns to scale, we get something that looks like this. Take a moment, how do you know that the firm is operating in a region of increasing returns to scale in this part of the diagram? How do you know that increasing returns to scale are going on here? The answer is increasing returns to scale is associated with declining long run average costs. So anytime the blue curve is downward sloping, the firm is experiencing increasing returns to scale associated with teamwork and specialization and better use of capital and labor and different kinds of combinations. So what that means is a proportional increase in all inputs results in a more than proportional increase in the output, and therefore, the average cost is declining.

Now, I can go ahead and put in the short run cost curves associated with other points on the long run average cost curves. If we look at the bottom of the long run average cost curve, we get a set of green curves that looks like this. Notice, the interesting thing about this set of curves at the bottom of the diagram is the short run marginal cost curve cuts through the average total cost curve at its minimum point, which is also the point at which the average total cost curve is tangent to the long run average cost curve. Things like this occur only at the point of constant returns to scale, only at the bottom of the curve. So when you find these three curves having one point in common, it can only happen at the bottom, the point of minimum long run average cost. Take a moment and recall, we had another name for the point of minimum long run average cost. What is that other name? The answer is this is called the point of efficient scale. The point of efficient scale is the output level at which the long run average cost is at a minimum.

Finally, we can put in one more set of cost curves, and here our average cost curves and short run marginal cost curve, we are putting them over on the other side of the long run average cost curve. What’s going on here? Remember, long run average cost always refers to the blue curve. What’s happening here is the point of tangency between the blue curve and the green curve occurs past the point of minimum average total cost, which is always the point where the short run marginal cost cuts through it. If we freeze capital at the amount that’s cost minimizing to produce this particular level of output, this particular quantity of output, this particular number of television sets, and then we change only labor in the short run to alter our output, we’ll trace out the green curves as our set of cost curves. Again, this is kind of technical stuff, but you’re seeing pictures like this probably in your textbook and it doesn’t hurt to keep reminding yourself what the story is behind the curves. All of these short run cost curves refer to a situation in which the amount of the fixed input is held fixed. The amount of capital is held fixed. The blue curve always refers to a situation where capital and labor are variable and can be combined to achieve the cost minimizing technique. Notice, with the green curves that the marginal cost in the short run, the SMC, always cuts into the short run average total cost curves at their minimum point, and that minimum point will not necessarily be the point where the blue curve is tangent to the average total cost curve in the short run. That only occurs whenever you are at the
Production and Costs

Isocost/Isoquant Analysis

Understanding Isoquants

point of efficient scale, when you are at the point where long run average cost is at a minimum. Also, the point of constant returns to scale. Over here, the firm has increasing returns to scale and diminishing average cost. Right here at the bottom, for an instance, the firm has constant returns to scale and constant average cost and finally, if you go over into this region, the firm has decreasing returns to scale and increasing average costs. Well, now that we’ve looked at these curves pretty carefully, and seeing the relationships among them, we’ll take them now and use them to tell a story about the way firms adjust in the long run and the short run when there’s a change in their industry.
Production and Costs

Isocost/Isoquant Analysis

Finding the Cost-Minimizing Combination of Capital and Labor

We've talked about isoquants, representing the firm's technology. We talked about isocosts, representing the firm's budget. Now we're ready to put the two together and find the cost-minimizing technique of production in the long run. Remember the long run is where the firm can combine labor and capital in whatever form minimizes the cost of production.

Let's suppose now, for the sake of argument, that the price of capital is $30.00 per unit; that is, every unit of capital that you hire, you have to pay $30.00 to employ. Let's suppose that the price of labor is $10.00 a unit, so each worker costs you $10.00. The important thing here is that capital costs three times as much, a unit of capital costs three times as much as a unit of labor. And that's all we need to say what the slope of the isocost line is. Remember the slope of the isocost line is $wage \over price \ of \ capital$. In this case, that would be $10.00 \over 30.00$, or $\frac{1}{3}$.

So here we've got an isocost line that has a slope of $\frac{1}{3}$. How do we find, in this case, the production strategy that's going to minimize our firm's cost of production in the long run? Well, the first thing you need to know is how much output does the firm intend to produce? Let's suppose our firm is going to produce 12 television sets a week. What combination of labor and capital gives this firm the lowest cost of producing this output? Well, here's my isocost line and, if I shift it out, I'm spending more money and, if I shift it in, I'm spending less money, but the slope of this line is definitely going to be $\frac{1}{3}$. We know that, because we know what the wage is and we know what the price of capital is.

So let's keep shifting this line in, until we can just barely afford a combination of labor and capital that will allow us to produce 12 television sets. Well, is that combination going to be two workers and six units of capital? And the answer is no, because we could shift the isocost line in further and still be able to afford some combinations on the line. So keep shifting. Will that combination be 1, 2, 3, 4 workers and 1, 2, 3 units of capital? The answer is no, because you can still see the green line down here below. The green line is still there, which means there are combinations of labor and capital that are still less expensive. So finally, if we shift this curve carefully down and we keep the slope at $\frac{1}{3}$, we finally wind up with a point where we are just barely touching this isoquant. We have shifted the isocost line so close to the origin that we are just touching the one remaining point on the isoquant, and that is the combination 1, 2, 3, 4, 5, 6 workers and two units of capital. That's the last combination that we're going to touch. If we move the isocost line anymore, notice now the isoquant lies completely above it. There are no combinations that are feasible if we're spending less money. If we're spending less, we're not going to be able to touch any point on the isoquant. But if we spend just this amount right here and employ six workers and two units of capital, this is the minimum cost of producing 12 television sets.

Now notice there are a lot of other combinations that are possible for producing 12 television sets, but those combinations all cost more money. How can you tell, by looking at this graph, that these combinations all cost more than this particular combination that touches the isocost line? How can you tell by looking? You can tell because all of these dots lie strictly above the isocost line. In order to get to them, we'd have to shift the isocost line outwards, and that means spending more money on labor and capital. However, if we shift down to this point right here, we find the cost-minimizing technique is the last dot that we touch before we lose contact with that isoquant. So look for the point where the isocost line is tangent to the isoquant. Look for the point where the isocost line just touches the isoquant that represents your target output. That's the cost-minimizing technique.

In this case, the cost-minimizing technique involves six workers and two units of capital. What is the total cost? Why don't you take a break and calculate that?

All right, total cost is going to be the wage, $10.00, times six workers is $60.00, plus the cost of capital, which is $30.00 per tool times two units of capital. That's another $60.00. $60 + $60 = $120. That's the total cost of production.
Production and Costs

Isocost/Isoquant Analysis

Finding the Cost- Minimizing Combination of Capital and Labor

Now, if you wanted to prove to yourself that this was the minimum cost of production, you could try calculating the total cost with some of these other combinations. For instance, if you use 12 workers and one unit of capital, what’s the total cost going to be? Take a break and calculate that.

12 workers times $10.00 per worker is 120, plus $30.00 for a unit of capital is 150. So, as our graph points out, this dot is more expensive than this one. This combination of labor and capital is more expensive or more costly than this combination. You can, if you want to convince yourself, do the same analysis for this dot and this dot. All of them involve higher expenditure to produce the same 12 television sets.

Now, it’s important for you to know that the cost-minimizing technique depends completely on the relative price of labor and capital. Suppose the relative price of labor and capital, instead of being 30 to 10, or in this case, a slope of $\frac{1}{3} \frac{10.00}{30.00}$, suppose we have the price of labor being, let’s say, $30.00 per worker and the price of capital being $40.00? Well, in that case the slope of the line is going to be $\frac{30.00}{40.00}$, or $\frac{3}{4}$. The line is going to be steeper in this case.

Well, in this case, if we pull the line backwards, we’re going to lose touch with our isoquant, we’re going to drop below it, but not until we’ve touched this point right here. With the slope of $\frac{3}{4}$, the cost-minimizing technique is going to involve 1, 2, 3, 4 workers and 1, 2, 3 units of capital. Notice that as the wage rises and the price of capital falls relative to the wage, as the isocost line gets steeper, because labor is getting more expensive, notice what happens is this: the firm finds that it’s cost-minimizing to use less labor, because labor is getting more expensive, and more capital, because capital is, in relative terms, getting less expensive. As the isocost line gets steeper, the firm moves a long the isocuant to find a new cost-minimizing technique. If labor got even more expensive, we might see the firm using an even more capital-intensive technique.

The rule is this: when the price of labor increase, the firm minimizes its costs by using less labor and more capital. When the price of capital increases, the firm minimizes its costs by using more labor and less capital. The firm always substitutes, as long as substitution is possible, the firm always substitutes in the direction of the factor that’s becoming relatively less expensive.

Now, let’s go back to our original situation. Labor costs $10.00 per worker and capital costs $30.00 per machine. Let’s suppose now that our firm wants to increase its output. It wants to increase its output from 12 television sets a week to 24 television sets a week. Now notice something: if we were in the short run, when capital is fixed, where in this picture would the firm have to go to increase output to 24 televisions? Find the place in this diagram where the firm would be producing, that is, find the combination of labor and capital that the firm would have to use in the short run. The answer is if capital is fixed, the firm is only going to have two units of capital. If it started at this point with two units of capital and six units of labor, now it has to go over to this point here with two units of capital and 12 units of labor. Now, that’s going to be expensive. We’ve got to shift this cost line all the way out here until it passes through that particular point, 12 workers, two units of capital. How much does that cost? Take a moment and calculate that.

12 workers is $120.00, two units of capital is $60.00, for a grand total of $180.00 to produce those 24 television sets. Is that correct? 12 workers times $10.00 a piece, 120, plus 60 is 180. That’s right. However, if this firm is careful, they’ll be able to substitute in the long run; that is, well, I don’t mean careful, I mean if they have more options, that’s what I mean. If, in the long run, the firm isn’t limited to having two units of capital, but can choose any combination of labor and capital that they want, what will they do? The answer is they’ll use more capital and less labor. How can you tell? You can tell, because if you look at this particular isoquant line, you can see that there’s part of the isoquant for 24 televisions that lies below it. This firm can economize by shifting the curve in and finding a combination of labor and capital that’s tangent to the curve; that is, by shifting the isocost line in and finding another point that touches the isocost line, but is still on the isoquant, the firm can lower its costs. In the long run, once the firm is able to buy more capital, it can then use more capital and less labor to produce the desired number of television sets. In the long run therefore, substitution becomes possible and costs will be lower.
Production and Costs

Isocost/Isoquant Analysis

**Finding the Cost-Minimizing Combination of Capital and Labor**

Well, we've covered a lot in this lecture. Let me go back and quickly recap the main points.

The first point is that the firm minimizes its cost to production by finding the point where the target isoquant is tangent to the isocost line. It's at that point that you've shifted the isocost line in as close to the origin as you can get it. It's at that point that you've got costs at the minimum you can possibly attain.

The second point is if relative price of labor and capital changes, the firm will move along the isoquant to keep its costs at a minimum. If the price of labor increases and the price of capital falls, the firm will move up the isoquant, using a more capital-intensive technique. The firm will always move the technique of production in the direction of using more of the factor whose relative price has fallen. So when labor gets very cheap, the firm uses the labor-intensive technique. When capital gets relatively cheap, the firm uses a capital-intensive technique. The firm always operates at a point where the isoquant is tangent to the isocost line, and that point will depend on the relative price of labor and capital.

Finally, the third point from this lecture is if the firm wants to increase its output, then, in the short run it's going to have a higher cost, because it can't adjust its fixed input. If capital is the fixed input, the firm has to go all the way out to this point that use two units of capital, that is, if two units is the fixed amount of capital. In the long run, however, the firm can move the isocost line inward and substitute and pick a different point on the isoquant. In the long run, the firm has the flexibility to combine labor and capital in whatever combination minimizes costs.

So, this is the first important thing to know about the long run. In the long run, the firm can combine labor and capital, in order to minimize costs. And therefore the cost of production in the long run of any particular quantity of television sets will be lower in the long run than it is in the short run when some of the inputs are fixed.

Now, we'll look at the second thing you need to know about the long run, and that is in the long run, the firm has to choose its scale carefully.
Perfect Competition

The Basic Assumptions of Competitive Markets

Understanding the Role of Price

We’re going to talk about the profit-maximizing strategy of a firm. But before we get into the details, we need to make one thing very clear. The definition of profit is total revenue minus total cost, and total revenue, that is, the revenue that a firm earns from selling its product, is equal to the price of the product times the quantity that the firm sells.

Now, here’s what we need to make clear. Does the firm have the power to set its own price, or is the price given to it by the market? That is, is the firm a price taker or a price maker? Are they a firm that accepts the price as given, or does this firm have the power, by altering the amount of output that it produces, to manipulate its price?

Later in this series, we’re going to look at the decision that faces a firm that has the power to set price. We call that power market power. A firm that can influence the price of its product by changing the amount of output that it sells is a firm that is said to have market power. And an extreme example of market power is the monopolist.

However, right now, to make the problem very simple, we’re going to look at the other extreme, a firm that has no market power, a firm that takes prices as given. When a firm is unable to influence the price of its product, when the firm takes as given the price of its product that’s set by supply and demand in the market, we call that firm competitive. The term “competitive,” in economic theory, refers to the behavior of a firm that cannot influence prices, the behavior of a firm or any agent that must take prices as given.

So in all of the analysis that we’re about to do, all of our discussion of profit-maximization, we’re going to assume that we’re dealing with a competitive firm. As far as our television factory is concerned, it cannot set the price of its televisions. The price at which it can sell its televisions is given to it by the market, supply and demand out there in the big market determines that you can sell a television set for, say, $500. And if our firm wants to sell televisions at $500 apiece, we will imagine that they can sell as many as they can produce. If they try to charge $501, they will not be able to sell any televisions, because competition will drive them out of business. On the other hand, why should they sell their televisions at $499 when they can sell all they want to at $500?

If the firm is competitive, that is, if the price is set in the market and our firm is small enough so that its actions don’t influence the market price, then the firm will always charge the market price for its product. It won’t charge more because it will lose all of its customers. It won’t charge less, because then it would be giving up revenue that it doesn’t have to give up. It will always charge the competitive price.

So keep in mind now, as we march into our discussions of profit-maximization, that for the time being, we’re dealing with a competitive firm, a firm that takes the price of its product, and the price of all of its inputs as given. No matter how much output our television factory produces, it is unable to influence the market price.
Perfect Competition

The Basic Assumptions of Competitive Markets

Understanding Market Structures

We’ve done an awful lot of work to represent the firm’s cost structure, that is, what it costs a firm to produce a certain number of television sets. Now comes the payoff. We can put information about costs together with information about revenue, and make predictions about the firm’s profit-maximizing output. In this lecture we’ll describe the strategy that a firm uses in order to maximize the profit that it earns from its operations.

Let’s start with the definition of profit. Profit is defined as the difference between total revenue and total cost. That is, total revenue, the total amount of money that the firm earns from selling its product, minus total cost, the total amount of money that the firm spends to produce its product and bring it to sale.

Let’s begin by examining the concept of total revenue. The total revenue that a firm earns is simply equal to the price at which it sells its product times the quantity of its product that it sells. Suppose now that television sets sell for $500 apiece. That is the price of a television set. If so, we can write down a table of numbers that represent the firm’s total revenue from different quantities of televisions sold. Let’s go back and use numbers that we’ve seen before. If the firm sells two television sets, two times $500 is equal to a total revenue of $1,000. If the firm sells 10 television sets, 10 times 500 is equal to a total revenue of $5,000, and so on. Each time we multiply the number of televisions sold by the price of the televisions, $500, to get the total revenue.

Now, we can represent total revenue in a graph. Let’s do that graphically now. Suppose we have a graph with money measured on the vertical axis, that is, cost and revenue. And on the horizontal axis we have output, number of televisions produced and sold. Total revenue will be drawn as a ray from the origin, a straight line, that has a slope equal to the price of the product. That is, each time the firm sells another television set, that is, the run, it adds $500, that is, the rise, to its revenue. So each time the firm sells another television, total revenue increases by $500, and that is the rise over the run, or the slope of this total revenue line. So let me draw in that ray from the origin that’s equal to the firm’s total revenue function, that is, total revenue as a function of output. I’ll try to draw this line in very carefully. Notice it’s a ray from the origin. That is, it’s a straight line that begins at the origin. And notice also that it has a slope equal to the price of the product. I will label this curve TR. TR stands for total revenue, and this is the firm’s total revenue. If they sell one television, $500, two televisions, $1,000, 10 televisions, $5,000, 100 televisions, $50,000, and so forth—the total revenue line.

Now total revenue is half of the profit relationship. The other half is total cost. That is, the firm is going to earn this revenue, but they only get to keep what’s left over after they’ve covered their costs. That would be their profit. So total cost then comes back into play. You’ll remember these visuals. We’ve looked at them before. Total cost is the cost of labor as well as other inputs that the firm uses to produce television sets. And in the short run—and we’re still in the short run—in the short run the firm has some fixed inputs—its factory, conveyor belts, tools, and so forth, and some variable inputs, that is, labor. So you’ll recall that if you add variable cost, the cost of labor, to fixed cost, that is the cost of other inputs, you get the total cost of producing a given quantity of television sets. And as you review, you can look at these numbers. The fixed cost in each case was $10,000. No matter how many televisions you produce, fixed cost is going to be $10,000. It doesn’t change. And the variable cost depends on the number of workers you have to hire to produce that quantity of televisions. If you want to you can go back and review that information from the lectures on cost.

But here’s our total cost relationship. Two televisions cost $11,000, 10 televisions cost $12,000, 30 televisions cost $13,000, and so forth. If we plot these numbers in our graph, we get the total cost curve. And we did that at the end of our series of lectures on costs. So let me go back now and draw the total cost curve in our diagram. Remember, the total cost curve has an intercept that is above zero. The line begins on the vertical axis at some distance from the origin. Do you know why that is? It’s because you have to pay the fixed cost even if you don’t produce any television sets at all, because the fixed cost is a positive number. Even with a zero production, the firm still has some costs. They have to pay their rent, they’ve already acquired tools. All of that stuff is fixed.

So now I’ll draw in the variable cost curve that we did from last time. And notice that at first the slope of it is decreasing, and then the slope is increasing, and here we have the total cost relationship. Now remember, the slope of the total cost curve is the same as the slope of the variable cost curve. It’s just the labor costs shifted upwards by the amount of the fixed costs. And the slope of the total cost curve at each point is the marginal cost of production. And remember, the marginal cost depends on labor productivity. All of this we covered in great detail in the last series of lectures.
Perfect Competition

The Basic Assumptions of Competitive Markets

Understanding Market Structures

So I can label this curve the total cost curve. Total cost at first is increasing at a decreasing rate because of teamwork and specialization, and then it’s increasing at an increasing rate because of congestion of the fixed inputs, that is, too much labor working with too few tools.

Now where do you see profit in this picture? Profit is the difference between total revenue and total cost. We can go back to our numbers now and find the point at which profit is at a maximum. Take total revenue, which we have calculated as price times quantity of televisions, subtract total cost, which we’ve calculated as equal to fixed costs plus variable costs, and if you take the difference between total revenue and total costs, you’ll get profit. And you’ll notice here that I’ve labeled my table of numbers with the Greek letter Pi, which is a symbol that we sometimes use in economics for profit. So the profit is going to be total revenue minus total costs. Now notice at first, that if you subtract the total costs from the total revenue, you get a negative number. The firm is not making enough money selling televisions to cover its costs. Why is that? It’s because your fixed costs are so large, that if you’re only producing a few television sets, you’re not able to spread the fixed costs out enough to be profitable. If the firm has a big fixed cost, it doesn’t have a lot of revenue, and therefore, its profits are negative—negative $10,000. If the firm is producing 10 televisions, they’re still going to be in the red, that is, making a negative profit. $5,000 in total revenue minus $12,000 in total costs leaves you with negative $7,000 profit.

Now, when you’re producing 30 televisions, for the first time you break above the surface of profitability. $15,000 is the total revenue, minus $13,000, the total costs, gives you a profit of $2,000. So now the profit is beginning to be positive. Notice the profit increases as you produce 40 television sets, and increases further when you produce 45 television sets. Finally, when you’re producing 48 television sets, the profit is at a maximum. That is, $24,000 total revenue minus $16,000 total cost, leaves you with $8,000 in profit, and that’s as high as the profit’s going to get. So you might put a little star by this particular outcome, because it’s going to be especially interesting to us. This is the point at which the firm is maximizing its profits, when it produces and sells 48 television sets to earn revenue of $24,000 with costs of $16,000. Profit of $8,000 is as good as this firm will be able to do. Or should I say as well as this firm is going to be able to do.

Forty-nine television sets pushes us back down to $7,500 worth of profits. Notice that profits are beginning to fall. And finally, if we go on to hire eight workers and production begins to decline, well, of course, profit is going to fall as well, because our costs are increasing, our productivity is now negative, so we’re pushing ourselves down into some bad territory here.

We’re thinking now about how a firm maximizes its profits. If you understand the definition of profit, and if you know what the firm’s total revenue and total costs will be for several different levels of output. That is, if you’ve got a table like the table of numbers I have here, finding the point of profit maximum is pretty easy. However, it may be that we don’t have all of these numbers, and we might need to try to find some intuition behind these numbers that could guide us in searching out the profit-maximizing outcome. That is, what does profit maximization look like if you don’t have a whole table of numbers like this, but rather have to deal with intuition about productivity, costs, and revenue?

I’m going to be doing that in the next lecture, but let me go back to the graph for a moment and show you something that will get us started thinking in that direction. If we go back to the graph, we’ve got the total revenue curve and the total cost curve. Now, what is the point of profit maximization going to look like? Well, let’s look at a few points in this diagram. What about this point right here? I can mark this point with a big dot. What’s going to be true at this point? What’s special about this particular point? How much profit is the firm making? The answer is, the firm is making zero profit, because revenue is equal to cost. Profit is zero. The same is going to be true of this point up here. We can call both of these points break even points, because these are points at which the firm makes zero profit.

Now, the interesting points are two other points that we can identify. We can find the points at which the green line and the blue line are farthest apart, and those points are going to be especially interesting. One of those points is going to be this point right here. That is, an output level—we’ll call this Y⊥ and we’ll call this point here, where the two lines, again, are furthest apart—looks like it’s going to be about right here, we’ll call this point Y*. Now what’s special about these two points that I’m right now identifying in our graph? Y⊥ and Y*?

What’s special about those points? Those are points where total revenue and total cost are furthest apart. Now, how do I know they’re furthest apart? Well, I just eyeball the diagram, right? I said, “well, here’s where this gap is bigger,
Perfect Competition

The Basic Assumptions of Competitive Markets

Understanding Market Structures

and here’s where this gap is the biggest.” Right here, total cost is as far as it gets above total revenue. That is, this is the point where the gap is the biggest with total cost on top and total revenue on the bottom. What does that mean? It means, of course, that this is the point where the firm is maximizing its loss. That is, the point where the cost is the most above revenue that it gets. Here’s the firm making a huge negative profit.

What about this point where total revenue is above total cost by the largest amount? What’s true in this point? This point is, of course, the point of maximum profit, the point where revenue exceeds cost by the greatest amount. These points are interesting. And I want to end by making one geometric observation about this graph. Look, suppose we start at this break even point right here. As we increase output, as this firm produces more and more television sets, notice that total revenue is increasing by $500 for every set that it produces. That’s the slope of the blue line.

Notice, however, that the slope of the green line is at first less than the slope of the blue line. But at first this green line is flatter than the blue line is. But then, the green line slope begins to increase and increase and increase, until finally it’s equal to the slope of the blue line. At this particular point, the line tangent to the green curve has the same slope as the blue line. This is the point at which the slope of the total cost curve is equal to the slope of the total revenue curve. Think about this as an alligator’s mouth opening around this point. And as long as the green curve is flatter than the blue curve, the distance between the two lines is growing greater and greater. As long as the blue curve is rising faster than the green curve, the gap between the two of them is getting bigger. At this point right here, where the slopes are the same, at that point we know that the gap between them is no longer getting bigger. In fact, the gap is at its maximum. I can put a couple of dashes here and here to represent that at this particular point, at $Y^*$, the blue line and the green line have the same slope, and that’s the point where the gap has reached its maximum size.

Now, what is the slope of the green line called? Do you remember? The slope of the total cost curve, the slope of the total cost curve is just the marginal cost. And the slope of your blue curve, the total revenue line, the slope of that line is equal to the price of the product. So the point at which profit is at a maximum is the point at which the marginal cost and the price are the same. The point at which the green line and the blue line have the same slope.

Now notice, that’s also going to be true down here. Down here when this gap is the largest, the blue line and the green line also have the same slope, marginal cost is equal to price at $Y_\perp$ as well as $Y^*$. The difference, however, is this. At $Y^*$ marginal cost is increasing, the green line is getting steeper. Down here, at $Y_\perp$, marginal cost is increasing; the green line is getting flatter.

So you know you’re at a point of maximum profit if you’re at a point where marginal cost and price are equal. That’s the point where the gap is the biggest. And to make sure you’re at a maximum profit instead of a maximum loss, make sure that at that point the green line slope is increasing, not getting flatter. In the next lecture, we’ll see how to apply this geometry and come up with an intuitive way of describing the profit-maximizing choice of a firm.
Finding Economic and Accounting Profit

You want to know how your business is doing. Are you going to be around a year from now? -- two years from now? -- five? What numbers should you look at to gauge your progress? How would an economist advise a business owner to figure out whether the business is really in some long-run sense earning a profit?

Well, here's the idea: Economic profit is the difference between the revenue that a firm earns, and the sum of the opportunity costs of all the resources that are employed at that firm. As long as revenue is greater than the sum of the opportunity costs, the business owner can afford to pay all of the resources -- a salary that is better than their next best alternative. That means they'll be willing to stay and work with this business as long as revenue exceeds opportunity costs. But if revenue falls below the combined opportunity costs of the employed resources, then the business is in trouble. One by one, the resources will seek alternative employment where they can earn more money. This is the notion of economic profit, and why it is that it's important that a business owner pay attention to it.

Now, let's consider the difference between the way economists reckon profit and the way accountants reckon profit. It turns out that if you use accounting profit as a guide, you can sometimes get a faulty picture of how your business is doing, and that could lead you to make mistakes. Let's consider, then, the way accounting profit is calculated.

Accounting profit is the difference between the revenue that a business receives and all of the expenses that the business pays out. So suppose now we are running a retail television business. You're selling television sets to people who come in off the street, you've advertised in the newspaper, you've made a name for yourself, and now customers come in, interact with your salespeople, and decide whether or not they want to buy a TV set.

Let's suppose that in one particular year, the revenue earned by this television business is $100.00. How would an accountant calculate the profits of this business for that year? Well, the first thing we'd do is make a list of all of the resources that are being written paychecks by the manager, and we would add up the expenses. So one of the first expenses is going to be the cost of goods sold -- that is, all of the television sets that we've bought from some wholesale operation that we're now selling retail. And let's suppose that the cost of goods sold for this operation is $30.00. So out of our $100.00 worth of revenue, we've got to take $30.00 of it to write checks to our wholesale provider of television sets.

Let's suppose that the other expense that we have is all of the help that we have to pay wages to. And let's suppose that for the sake of this example, our workers have combined wages of $30.00. So that's another $30.00 worth of expenses that this business has. And let's suppose that that's all the checks that we are writing out of our checking account for this TV operation. That means that $100.00 worth of revenue minus $30.00 worth of wages minus $30.00 for cost of goods sold leaves an accounting profit of $40.00 -- $100.00 minus $60.00 in total expenses -- so the accounting profit in this case is going to be what's leftover, and that is equal to $40.00.

Now what happens to that accounting profit? The accounting profit then is divided among other resources that are employed in this business that don't figure explicitly into the costs. Who are they? Well, for one, there's the manager. If the manager is not paying himself a salary explicitly, then he gets to take a share of the profits. Who else gets a share of the profits? Well, any investors that put their money into this TV store, maybe so that the TV store could afford to buy inventory, so they could afford to accumulate other kinds of capital that are important to their operation. Inventory is a really good tangible thing to think about why you may need to borrow money in order to run a TV retail store. So the money that's leftover -- the accounting profit -- is available to be divided between the manager and the stockholders.

Now, can we tell by looking at this picture whether or not our firm is going to be around a year or two? And the answer is no, because we have no idea from this reckoning how much the manager could earn somewhere else, and what the shareholders could earn somewhere else. And for that matter, what the workers could earn somewhere else. It's only when we see that all of the resources employed are earning more in this business than they could earn in their next best opportunity that we can conclude this business is going to be a going concern. That's what economic profit is about.

So what we do next is we add up the opportunity costs of all the resources that are employed, and we figure out whether this firm is earning enough revenue to cover those. Let's start with the cost of goods sold. These television sets that we are selling to our customers, what's their opportunity cost? Well, if we imagine that the wholesale television business is a very competitive business, you could turn around, and rather than sell these television sets to the public, you could sell them right back to the wholesaler, and probably get your original $30.00 back. So the
Perfect Competition

The Basic Assumptions of Competitive Markets

Finding Economic and Accounting Profit

opportunity cost of the televisions that we're selling to our customers is probably exactly what we paid for them – $30.00 – because we could always sell them back to the people who sold them to us.

So the opportunity cost of our cost of goods sold is probably pretty much the same number that was on our accounting books. The workers, however, probably are another story. We're paying them a total of $30.00 in wages, but their next best opportunity is surely lower than that. If the market for labor is very tight and very competitive, then the $30.00 might be their opportunity costs, but chances are, they're doing better in this operation than they would be doing at their next best alternative; that's why they've chosen to work in TV retail.

Let's suppose that the next best alternative for the workers would pay them not $30.00 but only $20.00. In that case, these workers are delighted to have the retail TV job, because their opportunity cost is lower than what they're actually being paid. Think about it, these workers are making, in some sense, a profit; they're earning $30.00 here while their next best alternative is $20.00. We call the extra $10.00 that they're earning at the TV store "economic rent." Economic rent is defined as the amount that a resource is paid in excess of its next best alternative – in excess of its opportunity cost.

So all we know here is that we've got to pay our workers at least $20.00; we happen to be paying them $30.00. But $20.00 would be enough to keep them in this employment. The big difference between economic profit and accounting profit is that economic profit includes the opportunity costs of all the resources employed, not just the ones that explicitly get paychecks. For instance, the next step in calculating economic profit is going to be to include the opportunity cost of capital. What could our investors earn if they put the money that they've presently invested in our television retail business into some alternative?

Let's suppose that a comparably risky investment would earn a 20% rate of return. That means that if our capital owners have put $50.00 into this business, say $50.00 that's lent to us so that we can acquire an inventory, if they'd put that money instead into the stock market, or into a bank account, they could earn 20% – 20% times $50.00 equals $10.00. So $10.00 in this case is the opportunity cost of capital – what our capital owners could have earned if they had invested this same amount of capital in a comparably risky investment. So let's go ahead and put $10.00 here as the opportunity cost of capital. Of course, that means that if we don't pay our capital owners at least $10.00 they're going to start looking elsewhere for investment alternatives.

Another resource that's not explicitly accounted for in accounting profit is going to be the manager, particularly if the manager is not paying himself a salary, he may not show up anywhere over here on the books. But certainly he has an opportunity cost; he could take his talent and go manage some other business. And suppose his next best alternative would allow him to earn $20.00; so $20.00, then, is his opportunity cost. If he doesn't get paid at least $20.00 for a year's worth of management expertise, then he can't be expected to stay with this firm. So what we've done here is we've added up the opportunity costs of all the resources that are employed at this business. Whether they're explicitly receiving a paycheck, or whether they're getting bonuses or dividends or whatever and not actually on the accounting books, they are considered here in this reckoning. Add up the opportunity costs, subtract the opportunity costs from the revenue that's earned, and we get what's called the economic profit. Economic profit is the amount at which the revenue exceeds the combined opportunity costs of all the resources that are employed. And this economic profit, then, will be divided among the employed resources in the form of rents.

We've already talked about how the workers are getting $10.00 worth of economic rent; they're being paid more than their opportunity costs. Maybe the manager is being paid $25.00, which would be $5.00 worth of rent for him, and maybe the shareholders are getting $15.00, which would be $5.00 worth of economic rent for them. So the economic profit, then, is going to be divided among the stakeholders in this business, whether they are capital suppliers, the managers, the workers, or anyone else. The sum of the economic rents would be equal to the firm's economic profit.

So we can conclude by saying that if the firm were not making an economic profit it wouldn't expect to be around very long, because the various resources – the manager, the workers, the capital – all these resources would go off in search of other alternatives, other opportunities, that had higher returns. But as long as the firm is making a positive economic profit, it can afford to pay all of its resources enough money to keep them interested in working for this business.
Perfect Competition

The Basic Assumptions of Competitive Markets

Finding Economic and Accounting Profit

If the opportunity cost of capital is greater than the accounting profit, well, there you have it. The firm looks like they're doing okay, but in reality, they're winding down, because the economic profit will be negative. Negative economic profit means, in the long run, you're out of business. Even though the books may look good, because you haven't taken account of other factors of production that aren't being explicitly paid paychecks, you can make a mistake in management. Economic profit is a notion that helps you consider whether or not your firm is actually doing well enough to stick around.

Now there's one more point I want to make about the difference between accounting reckonings and economic reasoning, and that's the issue of sunk costs. Sunk costs are amounts of money that you pay that you will never get back – the acquisition of inputs that immediately have zero opportunity cost. Here's a parable about sunk costs that makes clear why economists are so eager to resolve the confusion about them.

Suppose you join an expensive health club, and you have to pay all of your membership dues up front. And the first day you enthusiastically go to the health club and play a vigorous game of tennis. Now, being out of shape, you're probably going to really hurt your tendons, and the next morning you wake up with tennis elbow, and you think to yourself, "You know, my arm hurts. But since I already paid my dues, I really ought to go work out again." And you do, and the arm gets worse. And the next day, and the next day, until finally you show up at the doctor, hardly able to bend your arm, and the doctor says, "Gosh, you're being really foolish. Every day when you go to that health club, you're hurting your arm worse. Why do you keep showing up?" And you say, "Because I already paid my dues."

Well, this is the problem of the fallacy of the sunk costs – you're never going to get back the money that you've paid for your membership. So every day when you wake up, you should make the decision fresh. Because the membership is prepaid, it costs you zero to go down to the health club. Are you getting enough benefit to cover that cost? Not if you're hurting yourself – in that case, the benefit is actually negative, and you are shrinking your profit by showing up. The way you should make a rational decision about costs that are sunk is, "What would it cost me to bring this project to completion?" And "What is the benefit that I can expect once the project is completed?"

If the benefit is greater than the required additional costs, go for it. But if not, it doesn't matter that you've already put money into this project. That money is sunk, you're not going to get it back anyway. It shouldn't influence the decision about whether to go forward. Even though costs show up on your accounting books, some of them that are sunk should be ignored for the sake of decisions that are forward-looking. Always consider how much it will cost to bring a project to completion, and how much benefit will be derived. If the benefit is greater than the cost, it makes rational, good, profitable sense to carry the project out. If not, don't let sunk costs become a kind of neurotic guilt that influences your actions in the present in a way that makes you do unprofitable things. What's already paid is sunk, and the economist advises you to ignore sunk costs for the sake of rational, profit-making decisions.
Perfect Competition
Calculating Profit and Loss

Finding the Firm's Profit-Maximizing Output Level

Imagine you're running a television factory and you want to make maximum profits. You're trying to decide how many television sets to produce. What rule are you going to use to decide whether to produce one more television set? The answer is, you'll produce an extra television set any time that extra set adds more to your company's revenue than it adds to its cost. You're never going to produce a television set that costs you more, at the margin, than the revenue that it adds. That would be subtracting from your profits. But the rule of producing up to the point at which marginal cost is just equal to the price. That's the rule that a competitive firm uses to maximize its profits.

Let's look at how that rule looks in a graph. We're dealing with a firm, in the short-run, that's trying to maximize its profits and we've made a particular assumption about that firm. We've assumed that it is a competitive firm. Now, I haven't used this term before, but now is a good time to introduce it. A competitive firm is a firm that can take any action it wants to. It can produce as many television sets as it chooses to without influencing the price of its product. A competitive firm is a firm that can take any action without having an effect on the price of its product.

Therefore, the price of your product is a given, a constant, something determined in the market that you have no control over. We've imagined that our competitive firm, in this example, can produce television sets at a price of $500 per television sold. That is, the price is constant at $500, no matter how many televisions you produce and sell. If that's true, will you produce another television set if you want to maximize your profits? The answer is, it depends on the cost of producing that television set. The cost of producing an extra television set is called the marginal cost of producing that set.

The marginal cost of production, if it's less than the price, signals you to produce the television set and thereby increase your profits. But if the marginal cost is greater than the price, producing that television set would actually reduce your profits, and therefore it's not something you should do. Let's look now at this story in the picture. Here's a picture that you're familiar with from our last discussion. The total revenue curve is a straight line, away from the origin, which represents the situation of a competitive firm. You can produce all the television sets you want to and sell them at a constant price of $500 a piece. The total cost curve now, has a shape that reflects the productivity of the firm. First, you have diminishing marginal cost, as marginal product increases due to teamwork and specialization, and then you have increasing marginal cost as congestion in the fixed inputs occurs.

We now want to represent this picture, this same information in this graph below. I'm using another one of my double-decker graphs, which means I have to be clear that on the horizontal axis I'm measuring the same thing. In this case, I'm measuring output on both of these axes and I'm using exactly the same scale. Now the first line that I'm going to draw in the diagram below, is the firm's marginal revenue. Marginal revenue is defined as the change in total revenue that results from a change in output. Now, how does marginal revenue change? How does the firm's total revenue earn change if it sells one more television set? The answer is, it changes by $500, the price of the television set. It doesn't matter whether it's your first television sold, your tenth or your 100th. Each television sold would always add $500 to your total revenue. $500 is the marginal revenue, which in this case, the case of the competitive firm, is equal to the price of the television set.

A quick note. For a competitive firm that cannot influence the price of its good, marginal revenue is always equal to the price of the product. So, let me draw in a line that represents marginal revenue. In this case, marginal revenue is the price of a television set and the price of the television set will always be equal to $500. So, here I have a line with a slope of zero. It represents a constant, and this line is representing the price of a television set, in my story that's $500. Now, let's represent next the marginal cost of producing a television set.

Where do I get information, from this picture, on marginal cost? Where in this picture do I see the marginal cost of production? The answer is, it is the slope of the total cost curve. Remember, marginal cost is defined as the change in total cost that results from a change in output. It's the rise over the run, the slope of the total cost curve. So I want to graph a picture of the total cost curve and, let me see where the inflection point is. It looks like the inflection point might be right about here. This is the point where the total cost curve is no longer concave and becomes convex. That is, the slope stops decreasing and starts increasing instead. So if this is my point of inflection, this gives you the point at which marginal cost is at a minimum. The slope has decreased and decreased and decreased and it's just not going to get any lower. So here's the point of minimal marginal cost.

Now, two other things that I want to notice. First of all, this point right here, this output level, in our last lecture we called this output level $Y^*$. This is an output level at which the slope of the total cost curve is the same as the slope of
Perfect Competition
Calculating Profit and Loss

**Finding the Firm's Profit-Maximizing Output Level**

the total revenue curve. This is an output level at which price is equal to marginal cost. So marginal cost at this point would be equal to price and I can call this point $Y^*$, just like I did before. That’s $Y^*$. Over here, is another point where the slope of the total cost curve is equal to the slope of the total revenue curve, and this point before I called it $Y_\perp$. $Y_\perp$ is another point where marginal cost is equal to price. $Y_\perp$. That’s the point at which losses are maximized, by the way.

All right now, if I keep the marginal costs diminishing until I reach the inflection point, and then have it rise after that I would be drawing the slope of the total cost curve in this downstairs graph. So, let me do that. Here’s diminishing marginal cost. Bam, then marginal cost increases, bam, and there’s your marginal cost curve. Again, what have I drawn here? This marginal cost curve is simply the slope of the total cost curve at every point, and I had some points that I knew had to be on the curve, the point of inflection, where marginal cost is at a minimum, and these two points where marginal cost is equal to price. $Y_\perp$ and $Y^*$.

Now, now, I can use this diagram down below to make an intuitive point about profit maximization. I argued before that this point, $Y^*$, was the point of maximum profit for the firm. How did I know that? Well, if I look in the upstairs diagram, I know that because that’s the point where the gap between total revenue and total cost is the biggest. That’s the point where profit is maximized. But downstairs, I can tell the story a different way, using a different intuition. Let’s suppose now that you run that television factory and you’re producing, at this point, $Y^*$. Is that a good point for you to be at? The answer is, yes.

If you increase your output beyond $Y^*$, what’s happening? You’re adding $500 for that television. If you produce another television at that point, you’re adding $500 to your revenue, but the amount that you’re adding to your cost is greater. What does that mean? It means that that next television is actually shrinking your profits. Because it costs more to produce than it earns you in revenue. So you don’t want to go past the point $Y^*$. You also, however, don’t want to decrease your output. If you decrease your output, you would lose $500 by not selling a television set. If you back off and produce one less television set, you’re going to cost your firm $500 in lost revenue.

How much are you spending on the labor that you’re not employing? Well, notice your savings is less than $500. The savings is less than $500. So you’re sacrificing $500 in lost revenue, but you’re only saving yourself maybe $450 in costs that you don’t incur. It’s not worth it to save $450 if it means you’re throwing away a $500 sale. Therefore, you don’t want to move backwards from that point either. If you’re at this point, $Y^*$, where price is equal to marginal cost, you do not want to produce an extra television set. You also do not want to produce one fewer television set. This is the point that is un-improvable. This is the point at which profit is maximized.

Suppose we were back here at another point. Let’s say this point of inflection. Just pick it, for example. Is this a good place for your firm to operate? The answer is no. How do I know? How could you improve your situation if you were at this point, if you were producing this smaller quantity of television sets? How could you improve this situation? You could produce more television sets. If you produce more television sets, notice, each television set that you produce adds how much to your revenue? $500, the price of the television sets. Each one of them costs how much to produce? Well, some amount less than 500. Notice the green line is below the blue line at that point.

If marginal cost is less than marginal revenue, if marginal cost is below the price of the good, for a competitive firm, you can always increase your profit by increasing your output. As long as the cost is less than the price, by all means, make more television sets, drive up your profits. If you’re over in this region and producing a large quantity of television sets and the marginal cost is greater than the price, by all means, back off; produce fewer television sets. Because the money that you save by not producing those television sets, the money you’d save by not employing the extra workers is greater than the money you lose from the lost sale.

You economize more by not hiring the workers than you would lose by not producing the television sets. So back off, save more than you lose in revenue. Save more by economizing on costs. It’s only at the point where marginal costs equals price, it’s only at this point that the firm cannot improve its situation. Now, one final note. Look over here at $Y_\perp$. $Y_\perp$ is another point where price is equal to marginal cost. $Y_\perp$, on the other hand, is a bad point, because anything you do, any movement away from $Y_\perp$ makes you better off. If you’re at $Y_\perp$ and you increase your output, what happens? Well, look, you’re adding television sets that earn $500 apiece and the cost is less than 500. Get away from $Y_\perp$, increase your output, because price is greater than cost. As you increase your sales, that’s going to add to your profits. The televisions are bringing more in in revenue than they cost to produce.
Perfect Competition

Calculating Profit and Loss

Finding the Firm's Profit-Maximizing Output Level

Look. Even if you reduce your output, you're improving your situation. Because if you back off here, each of those televisions is bringing in $500, but they're costing more to produce at the margin. So, if you reduce your output away from \( Y_\perp \), then you're saving more in labor costs than you're losing in lost sales. \( Y_\perp \) is an awful point. In fact, if you'll look at the picture, you can see it's the worst you could possibly do. It's the point of minimum profit, the point where cost is the highest above revenue that it ever gets.

So, let me summarize. The point at which profit is maximized is the point where price is equal to the marginal cost of production. This is the rule, the profit maximization for a competitive firm. Any competitive firm to maximize its profits must be producing at the point where price equals marginal cost. There's one additional condition and that is, it's not enough to be at the point where price equals marginal cost. You also have to be at a point where the marginal cost curve is sloping upwards. That is, you have to be at a point where marginal cost is increasing. That's what distinguishes \( Y^* \) from \( Y_\perp \). \( Y_\perp \), where marginal cost is decreasing, is a terrible point. It's the point of minimum profits. \( Y^* \), where marginal cost is increasing, is actually the point of profit maximization.

So, our two conditions, our first, price equals marginal cost and next, that the marginal cost curve itself must be upwards sloping at that point. Marginal costs must be increasing. If those two conditions are satisfied, then the firm is maximizing its profits in the short run.
Perfect Competition

Calculating Profit and Loss

Proving the Profit-Maximizing Rule

We just developed the firm’s rule for profit maximization. The rule is produce where price is equal to marginal cost and make sure that marginal cost is increasing. What I want to do in this his short segment is apply that rule to the numbers that we were looking at a moment ago and make sure that it jives with our calculations of profits in a real example. Then I want to prepare you for the next step in analyzing a firm’s profit maximization decision.

So let’s look at our real numbers. In our real numbers from a moment ago, we saw that profits were maximized when the firm was producing 48 television sets. Not surprisingly, notice that if you calculate the marginal cost of production, that when you’re producing 48 television sets, the marginal cost of production is $333. That’s just under $500, which is the price of a television set. If you go beyond to produce the 49th television set, the marginal cost of the television set has risen to $1,000. Marginal cost is now greater than price, and the firm can improve its position by cutting back on its output. Anytime marginal cost is greater than price, you can reduce your marginal cost by reducing output and getting back in the direction of profit maximization.

Notice here, that the firm is maximizing profit, profit continues to increase, as long as marginal cost is less than price. But after marginal cost exceeds price, profit begins to decline again. One more thing to notice about these numbers is this. Marginal cost is equal to price when we’re producing two television sets. Marginal cost is $500 per television set. But notice, however, at that point, marginal cost is declining. It’s going from something to $500 to $125 down to $50. Marginal cost is decreasing at that point. And as we talked about a moment ago, if you produce where price equals marginal cost and marginal cost is decreasing, then you’re going to be minimizing profit, and that’s exactly what we wind up with. Price equals marginal cost gives us the smallest profit number or the biggest loss that we earn in this whole table. So our two rules then to repeat are—price equals marginal cost and make sure marginal cost is increasing to make sure you’re at a profit maximum instead of a profit minimum.

Now all this talk about price and marginal cost and marginal cost increasing—this is the rule for a firm to do the best it can with what it has. This is the way a firm maximizes its profit. However, just because a firm is maximizing its profit, doesn’t mean that the profit is positive. Maximizing a profit and minimizing a loss amount to the same thing. How do we make sure that when a firm is maximizing its profit, when it’s doing the best that it can, that it’s actually making a positive profit instead of a loss, a negative profit? Well, this is where those average cost curves come back in. The average cost curve doesn’t tell us about what’s happening at the margin. It looks at the whole picture. It summarizes everything going on in the firm to tell us the cost, on average, of producing a television set. The average cost tells us the per unit cost of producing a television set. As long as the price is above the per unit cost, the firm is maximizing its profit and making a positive profit. If the price is below the per unit cost, the firm is making a loss.

So when you find that point of price equals marginal cost, and the marginal cost curve is increasing, once you’ve found the best the firm can do, then compare the price at that point with the average total cost. And that lets you know whether the firm is making a positive profit or if price is less than cost, making a loss. We’ll do that next when we look at the curves.
Perfect Competition

Calculating Profit and Loss

Calculating Profit

Just because a firm is maximizing its profits, doesn't mean it's actually making a profit. It could be minimizing its losses instead. How do we show the total profit that a firm earns when it's doing the best it can? I'll use now, the graphs that we've developed, the cost curves, and use them in conjunction with the price of the product to show you what profit maximization looks like when the firm is earning a profit and what it looks like when the firm is making a loss.

Let's start with a fortunate firm that's making a profit. Now, what is profit? Profit is the difference between total revenue and total cost, and the profit per unit that a firm earns is the difference between the revenue per unit, which is equal to its price and the cost per unit, which is equal to average total cost. Let's look now at the diagram and bring those concepts into clearer focus. We'll start with the cost curves that we developed in the last series of lectures. Here’s the marginal cost curve, which tells us the cost of producing an additional television set. Here’s the average total cost curve that tells us the cost of producing an additional television set and notice that varies as labor productivity varies.

Here’s the average variable cost curve that tells us the labor cost per televisions produced. Notice again, varying as labor productivity varies and the difference between the average total cost curve and the average variable cost curve, what is that? It’s the average fixed cost. Average variable costs plus average fixed costs equals average total cost. And why is it that the average total cost curve and the average variable cost curves get closer and closer as output increases? It’s because average fixed cost is headed towards zero as we spread the fixed costs out over more and more units of output.

Now, let’s see what’s going on in this firm. Let’s suppose we have the price of the product. In our example, we’ve been looking at television sets that sell for $500 a piece. We have a competitive firm here, that can produce as many televisions as it wants to and sell them at $500 a piece. It’s competitive, which means that it cannot influence the price of its product, whether it produces a lot or a little it can still sell its televisions for $500 a piece. Now, if the price is $500, how many television sets will this firm produce if it wants to maximize its profit? Let me give you a chance to answer that question.

Where in this graph will the firm be operating if the price is here and the firm wants to maximize its profit? The answer is, take the price and go over to the marginal cost curve. Remember, the rule for maximizing profit, for a competitive firm, is the firm produces up to the point at which marginal cost equals price. Don’t go past that point or your costs will be greater than revenue for each additional television set. If it costs more to produce the television than you can make by selling it, then you’re cutting into your profit. But don’t stop shy of that point. If you stop short, then you’re giving up profits, because the revenue on those televisions is greater than the cost of producing them. Keep producing up to the point at which price equals marginal cost. That will maximize your profits.

That means our firm wants to produce an output, and we’ll call this output $Y^*$. That’s what we’ve been calling it all along, the output that maximizes profit. Now, we’ve got a firm here that’s doing the best it can. A firm that’s always going to produce where price equals marginal cost and the marginal cost is increasing. That’s the best the firm can do. That’s the point of profit maximization. Now, however, we’re going to ask ourselves, “When this firm is doing the best it can, well, how well is it actually doing?” Is it making a profit? Is it making a loss? What? How is this firm doing?

Well, in order to answer this question, let’s again think about profit. Profit is total revenue minus total cost. If you look at the picture that we’ve drawn here, you can actually see the firm’s total revenue. Where is it in this picture? Where in this graph can you see the firm’s total revenue? Remember the definition of total revenue, price times quantity. Total revenue is seen in this picture as the area of this rectangle. The height of the rectangle is the price the firm charges for its product and the width of the rectangle or the base, is the quantity that the firm produces. Price times quantity, gives you the area of this rectangle that I’ve drawn here as a black box with dotted lines here and the axes has the lines on this side. This rectangle, right here, is the firm’s total revenue from selling this quantity of television at this price per unit.

Think of this as a big chunk of money that the firm takes in. Now, here’s the question. What does the firm do with the money? Who gets this? Who is entitled to the revenue that the firm takes in from its sales. Well, let’s look at the diagram here. The first thing we notice, if we look at this quantity $Y^*$, let’s imagine that this is 48 television sets. If the firm is producing 48 television sets, we can go up from 48 television sets, we can go up from $Y^*$ and find this point on
Calculating Profit and Loss

Calculating Profit

the average variable cost curve. What does this point tell us? If we follow this point over to the axis, it gives us a dollar value. The question I’m asking you to answer is this. What does that dollar value tell us? 48 television sets, here’s a point on the average variable cost curve, what is this amount of money on this axis going to tell you?

It tells you the labor cost per television produced. It tells you how much you have to pay your workers for each television they produce. Now, if you take the average variable cost, the labor cost per unit produced and you multiply it by the number of units that are produced, what do you get? You get the variable cost, the total money that you’re spending in your firm to pay your workers who are making your 48 television sets. So, if you multiply the number of units that are produced, the base of this rectangle, times the average variable cost, that is the labor unit per television, multiply that by 48 televisions and this little area down here at the bottom is the firm’s variable cost. Now, we can think about this as kind of a parfait, one of those deserts that has lots of different layers. The bottom layer of our parfait is the variable cost, the amount of money that the firm has to pay its workers who are producing the 48 television sets.

So, if I want to and I do want to, I can label this particular chunk of the parfait with an AVC. This chunk down here at the bottom, and I could shade in the whole thing, but I don’t want to make my graph too cluttered, this whole bottom chunk of the parfait is what’s going to the workers. This is the chunk of total revenue that goes to the workers to pay them for their labor. Now, if we follow our 48 television sets, further up we hit another curve. The next curve that we run into is the average total cost curve. So, if we follow this box over, to an amount of money on the axis, what’s it going to tell us? We’re producing 48 television sets and this is the point that we touch on the average total cost curve. What does this amount of money tell us?

It tells us the average total cost per television sets produced. That includes all the workers, as well as all the capital we’re using, tools, factory space, resources, everything else. All those costs that are fixed in the short run. So, this little chunk of the parfait, now, that’s sitting on top of variable costs, what is that? That’s average total cost after you take out the variable costs. Well, what’s left? What’s the difference between total cost and variable cost? The answer is, fixed cost. The difference between the average total cost curve and the average variable cost curve, at every point is the average fixed cost. If you multiply this average fixed cost times those 48 television sets, times the number of units that you’re producing, what do you get? Well, average fixed costs times the number of televisions you’re producing, you’d get fixed cost, the total fixed cost.

Let’s put fixed cost, now on this picture? And I could shade in this whole area, but again I don’t want to confuse the diagram. So, this whole area right here, the second layer of the parfait, is the fixed cost. All right, quick review. We’ve got this big black rectangle, which represents the firm’s total revenue, price times quantity. We went up to the average variable cost curve and identified average variable cost when we’re producing 48 televisions. We multiplied that by our 48 televisions and we got variable cost, the total wage bill that this factory has to pay. That comes out of the revenue. Next, we looked at the difference between average total cost and average variable cost and what’s left over is the average fixed cost. We multiplied that by our 48 television sets and that gave us the fixed cost, the fixed cost of production.

Now, if we add fixed cost and variable cost together, what do we get? We get the total cost of production. And, if we subtract the total cost of production, these two bottom layers of the parfait, if we subtract them from the big black rectangle, what’s left? What’s total revenue minus total cost? The answer is, it’s the profit of a firm. That is, the layer that is left on top after we subtracted fixed and variable cost, the layer that’s left on top is the firm’s profit and I’ll use the symbol Π to indicate that. So, by way of review, here’s the way we examine the firm’s profit maximizing decisions and their profit position.

First, we took the price of their good. They’re a competitive firm, they take the price as given. We then let the firm maximize profits by continuing to produce until marginal cost was equal to price, that gave us Y*. Y* is the profit maximizing output of this competitive firm. Notice it’s where price equals marginal cost and it’s where the marginal cost curve is sloping upwards. This is the best that our firm can do. Then, we broke down total revenue into some chunks. First, we found average variable cost, using the AVC curve and multiplied it by output to get the bottom chunk of our parfait, which is the chunk that goes to the workers. Then we found the difference between average total cost and average variable cost, that’s average fixed costs. We multiplied that by our 48 television sets and got the fixed costs. What’s left over is the difference between total revenue and total cost and that is, by definition, the profit of the firm.
Perfect Competition
Calculating Profit and Loss

Calculating Profit
Now, I’m going to do the same set of steps, the same analysis, on a different firm. This firm is not going to be as fortunate, because it’s going to turn out that they are in fact not making an economic profit in the short run. They’ll be making a loss. Let’s examine now a different situation.
Perfect Competition
Calculating Profit and Loss

Calculating Loss

Let's look at a different situation now. Here's a firm, maybe even the same firm. I've drawn the curves exactly the same. But now the price of the product is lower, maybe due to a change of supply and demand out in the market. The price of televisions is no longer $500 apiece. The price has now slipped down to, say, $300 per television. That means our firm can sell all the televisions it wants to at a price of $300 per television.

Now, what are you looking at here? The answer is, you're not looking at anything because the axes are not labeled. So make sure you always label your axes. Output on the horizontal axis here and cost and revenue measured in dollars on the vertical axis here. Now, see now I can relax, I can feel better, because all my curves and axes are labeled. Make sure you always label your curves and axes.

Now, here's the price at which our firm can sell its television sets. Let's say this is now a lower price of $300 per television set. How much output will our firm produce if it wants to maximize its profit? The answer is it will follow the price over to the marginal cost curve. Price equals marginal cost where the marginal cost curve is sloping upwards, that is the point of profit maximization. And we'll call this Y* because here's where our firm is maximizing its profit. Maybe this is 30 television sets. Now if our firm is producing 30 television sets, 30 television sets times $300 per television set gives us total revenue of $9,000, and that's shown right here in this graph. The area of this rectangle is price times quantity. Once again, this is the firm's total revenue.

We're going to make a parfait, just like we did before, and we're going to find out who gets this money. Well, the first thing we do is we follow 30 television sets up to the average variable cost curve. And this point tells us the total amount of money that the firm spends per television set on labor. This is the labor cost per unit, per television produced. Well, if this is the average variable cost or the labor cost per unit, then if we multiply that by the number of units produced, we're going to be calculating what? You tell me. What is the area of this rectangle right here? It's the variable cost. If we multiply the average variable cost or the labor cost per unit times the number of televisions we produce. If you multiply the average by the number of televisions, we get the variable cost. So we'll write in down here variable cost. And let's see, I used blue before for that, so let me try to be consistent with my colors. And this bottom chunk of the parfait then, represents the total amount of money that's paid for the workers.

Now, notice the way I've drawn this, the fixed cost seems to include two layers of the parfait, these two layers. But these really aren't two layers. This is just one layer in the fixed cost. What happens now is the layer that used to be on top, the profit layer, has collapsed down into the fixed cost layer. That is, our profit layer has collapsed down into it. The price is now less than the average total cost.

Now if you take the difference between the average total cost and the average variable cost, what is that going to be? Once again, it's the fixed cost per television produced. Now, if we multiply the average fixed cost, or the fixed cost per unit times the number of televisions we produce, we're going to get a rectangle, and here is that rectangle. The rectangle begins down here at the average variable cost curve, goes up to the average total cost curve and gets multiplied by the number of units of output produced. So here's our fixed cost. Fixed cost, I used red last time, so let me continue to use red. It's this rectangle.

Now, notice the way I've drawn this, the fixed cost seems to include two layers of the parfait, these two layers. But these really aren't two layers. This is just one layer in the fixed cost. What happens now is the layer that used to be on top, the profit layer, has collapsed down into the fixed cost layer. That is, our profit layer has collapsed down into it. The price is now less than the average total cost.

What does that tell you? If the firm is paying more per unit in costs than it's earning, that it sells a television at a price of $300, what's going on here? The firm is losing money on each television. They're average total cost in my example was $350 per television. The price is $300 per television. The firm is making a loss of $50 on every television they sell. Multiply that loss by 30 televisions and you get this top layer here now—is total cost minus total revenue. It's a negative profit. That is, it's a profit that's less than zero. Another name for a profit that's less than zero is a loss. And if I wanted to, again, I could shade this in, maybe with dots this time to try to make it clear that this very top layer of the parfait is the difference between total revenue and total cost. In this case it's negative, because total cost is this big box, variable cost right here, plus fixed cost added together. Total revenue is this small box—price times quantity. Because cost is greater than price, because total cost is greater than total revenue, this firm is earning a loss.
Perfect Competition
Calculating Profit and Loss
Calculating Loss

Now, here’s a question for you to think about. If you were the manager of this firm—in the short run—you can’t change your fixed costs in the short run—in the short run with a certain factory size, a certain number of conveyor belts, a certain number of tools, and you can hire more or less labor, labor is still variable... If you were the manager of this factory, would you keep your factory open in the short run or would you shut it down? Think about that for a moment and make a decision.

If you were the manager of this factory you’re going to do better to keep your factory open in the short run. Let’s see if I can make that clear. Suppose you shut your factory down in the short run. What would you be losing? Well, you would save on all of your variable costs, right? You’d be able to economize on them. They’d go away because you wouldn’t be hiring any workers. You wouldn’t have to pay them. However, you’d lose price times quantity. You’d lose your revenue rectangle. If you lose the revenue rectangle and you save on the costs, you still have to incur the fixed costs. If you shut down in the short run, you’re still going to lose the top two layers of this parfait. All those fixed costs, the rent on your factory, the tools you bought, all of that’s gone, so you would still be losing, in our example, $10,000 if you shut the factory down in the short run.

If you continue to operate, as long as you can afford to pay your workers, at least you’re earning some of that fixed cost back. At least you’re earning some of it back. As long as the price of the product is above the average variable cost, you can afford to pay all of your workers’ wages and still earn something towards covering your fixed costs. Earning something towards covering your fixed costs is better than earning nothing towards covering your fixed costs. If you shut down in the short run here, you’ll be making negative $10,000, a loss of all of your fixed costs. If you continue to operate in the short run, in my example here, you would only be losing the difference between total revenue and total cost, which in this case is less than your fixed costs. That’s because the price is more than enough to cover your wages. It still leaves you a little bit of money to pay towards those fixed costs.

And that brings us to the rule that a manager uses to decide whether to shut down his firm when it’s making a loss. A manager will shut down the firm anytime the price drops below minimum average variable cost. If the price drops so low that the firm can’t afford to pay its workers, then the firm is better off to shut down. I’ll be back in a moment with a new graph to make this point as clearly as I can.
Perfect Competition

Calculating Profit and Loss

Finding the Firm's Shut-Down Point

We're going to take one more look at a profit-maximizing firm doing the best it can. In this particular case, it's going to turn out that the best the firm can do is to shut down. Let me start with a question. Under what circumstances would you, a manager, find it optimal to shut down your factory in the short run? The answer is this. You're only going to shut down your factory in the short run if you're making a smaller loss by shutting down than you'd be making by staying open. The firm's making a loss, of course, but sometimes you can minimize your loss simply by sending all of the workers home and shutting the factory down in the short run. Let's see how that works.

Suppose the price of television sets now is very low. Let's suppose it falls down to maybe $200 per television set. In that case, what's the best the firm can do? If the firm continues to produce, what output will it choose to produce? Use the same rule that we've been using for the past four segments and find the point that the firm will choose to maximize its profits. That's going to be going over to the marginal cost curve and finding the place where price equals marginal cost. That's the profit-maximizing point of output. And if we go down here, we identify $Y^*$, the output level, at which the firm is maximizing its profit. Oh-oh. Look. I forgot to label my axes again. If you don't do this on a test, you might lose points. So let me try to encourage you to develop a good habit. Now that my axes are labeled, once again this picture means something, and we may proceed.

Now, here's the firm's total revenue—price times quantity. However, notice in this case, the total revenue happens to be exactly equal to variable cost. If you take $Y^*$ and go up to the point where you touch the variable cost curve, notice you're right at the point of minimum average variable cost. Here's a good review question. How do you know that this point that I've touched here is the point of minimum average variable cost? How do you know that? Here's the answer. You know it because the marginal cost curve cuts through the average cost curve at its minimum. And since we're at the point where price equals marginal cost, and we're also touching average variable cost, we must be at the point of minimum average variable cost.

Here we are at the bottom of the average variable cost curve. So if we multiply average variable cost times output we get, once again, the familiar term, variable cost. So look, in this case our total revenue is just equal to our variable cost. What does that mean? It means that we're spending all the money that we are earning selling television sets to pay the wages we owe our workers. Nothing is left over. All of our revenue is consumed covering our variable costs.

And unfortunately, those aren't the only costs we have in the short run. In addition to the variable cost, if we go up to the average total cost curve, we're now including the fixed costs, the difference between the average total cost and the average variable cost curve is average fixed cost. And if we multiply average fixed cost times the total quantity of television sets we produce, we get another layer here of parfait—here's our fixed cost. Unfortunately, for this firm now, however, all this fixed cost is lost. The firm has no revenue to cover the fixed cost, it's spent all of its revenue down here covering its labor cost. So fixed cost, all of the fixed costs are a loss. In this case, the firm's loss is exactly equal to its fixed costs. It has no money to cover its fixed costs.

So should the firm shut down? Anytime a firm shuts down it loses its fixed costs. It makes no money to recover them. However, if this firm were to continue to operate, doing the best it can, it's going to lose all of its fixed costs anyway. It's going to spend all of its money just covering its labor costs, and there's nothing left to go towards the fixed costs. So when the price falls to the point of minimum average variable cost, in the short run, the firm is indifferent between continuing to operate and shutting down.

Suppose now things get even worse. They can get worse, you know. The price could fall even lower. And if the price falls even lower now, lower than the firm, doing the best it can, cannot even afford to cover its variable costs. In that case, if the firm operated, not only would it be losing the amount of its fixed costs, it would also be paying workers and incurring a variable cost that's greater than its total revenue. If the price falls below minimum average variable cost, this firm would be making an even bigger loss. It would be losing all of its fixed costs, plus it would be paying workers more than it was taking in selling its television sets.

That would be completely ridiculous. No manager would ever operate a firm that could not cover its variable costs. You'd be paying workers to come in and drive your firm into the ground. You'd be paying waiters who weren't serving enough food to cover the cost of their wages, much less the rent on the building, the overhead costs, all of this. You would shut your restaurant down rather than charge a price that doesn't cover the cost of the waiters' services.
Perfect Competition

Calculating Profit and Loss

Finding the Firm’s Shut-Down Point

Same thing is true of our television sets. If the price of a television set is not enough to cover the labor costs, much less the fixed costs—if it’s not enough to cover the labor costs, you’d do better in the short run by shutting down. So here’s the rule a manager should use. If the price of the product drops below minimum average variable cost, in the short run, that manager should shut the firm down. You minimize losses by shutting down the factory. If, however, the price is greater than minimum average variable cost, if you can operate and cover variable costs, then continue to operate in the short run. At least you’re earning something towards your fixed costs rather than losing them all with no revenue to show for it. That’s why we call this point the minimum of average variable cost, the shutdown point.

Now notice, if the price is below the shutdown point, the firm is going to be shut down. If the price is above minimum average variable cost, the firm is going to be producing the price equals marginal cost. That is, for all prices above the shutdown point, the marginal cost curve tells us the quantity of output that it will be maximizing for the firm to produce. For all prices above the shutdown point, this marginal cost curve tells us how much output the firm will be producing when it maximizes its profit. That’s why we call this region of the marginal cost curve, the marginal cost curve above the shut-down point, that’s why we call this the short run supply curve for a competitive firm. And I could label this SS. The short run supply curve for a competitive firm.

Let me repeat. The short run supply curve for a competitive firm is just the firm’s marginal cost curve above the shutdown point. In the next lecture we’re going to be looking now at how this analysis of an individual firm translates into a picture of the whole market. Since all firms, all competitive firms, are making these same decisions, coming up with short run supply curves, when we add all those individual short run supply curves together, we get the market short run supply curve. We’ll be looking then at how the goings on in the market influence the situations of individual firms. So in the next lecture we go from the individual firm to the market in the short run.
Perfect Competition

Deriving the Short-Run Market Supply Curve

In the last few lectures we've been looking at the profit-maximizing behavior of a competitive firm. That is, what a firm does when it wants to maximize its profit and it cannot influence the price of its product. Well, this raises an interesting question, and that is, where does the price of a firm's product come from anyway if the firm is taking it as given, well, who's determining it, who is setting it?

We've seen before in our earlier lectures, that the price of a firm's product is determined by the interaction of supply and demand. What we're going to do in this particular lesson is show how we can move from the competitive profit-maximizing firm to the supply curve in the short-run for the market.

Now, let's remind ourselves of a couple of things before we launch in to deriving the short-run market supply curve. The first thing is, we are dealing with competitive firms, and competitive means that the firm takes the price of its product as given. It cannot influence the price of its product no matter how much or how little output it chooses to sell. The second thing to keep in mind is that the profit-maximizing firm will always produce where the price of its product is equal to the marginal cost of production, that's the conditions of profit-maximization. And remember, the firm will also be producing on an upward sloping portion of it's marginal cost curve so that it's maximizing profit instead of minimizing profit.

Finally, one more thing to keep in mind, and that is a firm will shut down in the short-run if it cannot cover it's variable cost of production. That is, anytime the price of its product drops below the average variable cost, that is, the cost of labor per unit of output, the firm will find it better to shut down in the short-run.

Keeping those three things in mind, we are now able to move from the individual firm to the market, and that's what I'm going to be doing in this lesson. Remind ourselves that we're dealing with the short-run. In the short-run, some of the firm's inputs are fixed, like the size of its factory, the number of conveyor belts and tools and all of those things, and the way that it can vary its output is by varying its variable input. That is, in our story, labor. By hiring more or less labor, the firm can modify its output.

All right, given that we're in the short-run, we have competitive firms, and they're maximizing profits, what does the short-run supply curve for the market look like? Have a look here at these graphs. In these graphs, what I've done is take the situation of three firms—we'll call them Firm 1, Firm 2, and Firm 3, and written down the marginal cost curve and the average variable cost curve of each firm. Well, you'll notice I've already done something a little careless. That is, even though I've indicated that the vertical axes are measuring things in dollars, I haven't put any kind of indicator on the horizontal axis. And if you don't label the horizontal axis, you don't have a graph, because you're not telling people what you're measuring. So let me indicate here that I'm measuring the output of Firm 1, television sets, the output of Firm 2 in television sets, and here's the output of Firm 3, also measured in television sets. So what I'm measuring on the horizontal axis in each of these graphs is the output of a television factory.

On the vertical axis, notice I'm measuring the same thing in all three graphs. I'll be looking at prices and costs all measured in dollars, and the scale on my vertical axis is the same in all three graphs. That's why I'm able to put them side by side and take something from each graph and add them together to form a big graph over here on the far right. If you're putting graphs side by side and are planning to use them together, it is important that they are measuring the same thing with the same scale on the vertical axis.

So here I have Factory 1, Factory 2, and Factory 3. Over here, in this graph that is vacant for the moment, I put price and quantity on the axes, and that lets you know that I'm going to be deriving what? What kind of graph goes in with price and quantity on the axes? Of course, that's going to be the supply curve or a demand curve. In this case, supply, because we're looking at the behavior of suppliers or producers.

All right. So given that we have information on these three individual firms, how can we use that information to derive the market supply curve for the short-run? Well, let's remind ourselves what a supply curve is. A supply curve tells us how much output producers will offer for sale at a given price. We assume when we draw a supply curve that producers are responding passively to the price of the product. That is, we're assuming competitive behavior. Well, if we're assuming competitive behavior, we can go over to our individual graph and say, at a particular price, at a given price, how much output will this firm produce? And we know what the answer is. At any given price, how much output will a firm produce? The firm will maximize its profits by producing where price equals marginal cost. We can follow
Perfect Competition

**Deriving the Short-Run Market Supply Curve**

any particular price over to the marginal cost curve and down to the quantity axis to find the output of an individual firm at any given price.

So let's take an example. Suppose the price of our product—and I'll use my clear ruler here to move around and represent different heights in my graph—suppose the price that we're dealing with is a low price, way down here in the bottom part of my graph. Let's say the price starts out at this level right here, and I'll use my little blue arrow to make clear where we are. How much output will each firm produce if the price of the product is down here where the blue arrow is? Well, notice that none of the firms are going to produce any output in the short-run. Why is that? How can you be sure that if the price is down here where the blue arrow is, if we have a low price, that none of my firms will be producing any output at all? The answer is this—because the price is below the minimum average variable cost for each of the firms. The price is below the minimum average variable cost, and we call that point the “shut down point.” If a firm can't cover its variable costs of production, there's no point in producing. It's already losing all of its fixed costs. It would just be adding a loss on top of that. So the firm would be shut down. Firm 1 would be shut down, Firm 2 would be shut down, Firm 3 would certainly be shut down if the price were way down here below the shut down point for the firm.

So as long as the price is below the shut down point of all three firms, the total market output is going to be zero. So for all prices down here in this region, until the price rises to the point where some firm can cover its variable costs, we get no output at all. So for all prices over here on this axis—we can start to shade in our market supply curve—we're going to get no output at all until the price rises to a point that's high enough to cover the variable costs of Firm 1.

Now, I'm going to do a little bit of observation at this point. Notice that Firm 1 has got the lowest shut down point, Firm 2 has got the next highest shut down point, and Firm 3 has got the very highest shut down point. Of all three firms, Firm 3 is going to not be able to operate it won't be able to cover its costs until the price rises way up here. So Firm 3 has got the highest shut down point, or the highest price required before it will produce any output.

Let me take a moment here before I go back to deriving my market supply curve and ask why are the shut down points for these three firms different? Why do my three firms in my example have different shut down points? What makes these three firms different? Well, remember, the cost curves summarize two pieces of information. They summarize, first of all, information about the technology of the firm, how the firm is able to turn output into inputs. It also summarizes information about the price the firm has to pay to hire its labor or to obtain its other inputs. So each of these sets of cost curves tells you something about the technology and the input prices that each of the three firms is dealing with.

Now it could be that Firm 1 has a superior technology and is able to produce more output with less labor. It could be. And that Firms 2 and 3 just have less productivity for their workers and therefore need more workers and have higher costs to produce output. That's a possibility. It could also be that Firm 1 happens to have a bigger factory, more of the fixed input. And with more of the fixed input, they may need less labor to produce a given number of television sets. They're using tools to substitute for labor. That's a possibility also. So in the short run it may be true that each of my factories has a different amount of the fixed input. The final possibility would be that the different factories are paying different prices for labor, that is, they're able to pay different wages to their workers. Now, that's unlikely in a competitive market. In fact, by definition, in a competitive market, everybody can hire labor at the going wage. If they were paying different wages, of course they would have different sets of cost curves, but again, it's kind of hard to believe that these three firms, if they are in competition with one another, would actually be able to pay different wages. They would all end up paying approximately the same price for their labor.

All right, so enough of that little digression. Let's go back now to the question of how to get a market supply curve from these individual cost curves. If the price rises—if the price of the good rises to the point where Firm 1 can cover its variable costs, then for the first time we have a factory producing television sets. That factory is Firm 1. Notice Firms 2 and 3 are still below their shut down points, they're out of the picture, as it were. And Firm 1 is able to produce television sets. So if we take the output that Firm 1 produces when the price is equal to its minimum average variable cost—I'm going to see here, it looks like one, two, three, four, five, six, seven, chunks on my grid takes me up to here. So let's put this point on the graph. We'll call this P1. And P1 is the price where Firm 1 first begins to operate. So I could draw a dotted line over and find the output of Firm 1 at a price of P1 and we can call this Y*1.
Perfect Competition

**Market Supply**

**Deriving the Short-Run Market Supply Curve**

This is the profit-maximizing output for Firm 1 when the price is P1. And then I can take P1 on over, moving horizontally across my graph until I get to the market supply curve.

And what I do is I take this quantity that Firm 1 is producing, and I put it over here in the market supply diagram. That is, when the price is P1, the total market output is going to be Y*1. That’s all the output that we get produced at that particular price.

Now, what happens as the price of the products continue to rise above T1? Well, anytime the price rises, Firm 1 is going to maximize its profit by going over to the marginal cost curve and choosing the output level where price is equal to marginal cost. So as price rises, Firm 1 increases its output by moving along its marginal cost curve. It hires more workers, their productivity is diminishing due to congestion, marginal cost is rising, and that higher price entices Firm 1 to produce more output. It allows them to cover the rising marginal cost associated with diminishing labor productivity.

So what I do now over here in this diagram, in the market diagram, is as the price rises, I’m tracing out Firm 1’s marginal cost curve. I’ll be tracing it out here. So let’s just take this marginal cost curve and draw it here. As output increases, Firm 1 increases its—as the price increases in the market, Firm 1 will increase its output of television sets. So this segment right here of the supply curve will be simply the marginal cost curve of Firm 1.

Remember, the supply curve of an individual firm is what? How do you find the supply curve of an individual firm? Remember, it’s just the marginal cost curve of that firm above the shut down point. That’s the supply curve for an individual firm. Now notice what happens eventually, is the price rises to a point that’s high enough that Firm 2 can now cover its cost of production. Firm 2 can now cover its variable cost and it enters the market. When the price rises up to this point, P2—we’ll call this P2 because that’s the price at which Firm 2 can now afford to enter the market, that is, it’s no longer shut down. At the price of P2, and I won’t clutter up my diagram here too much. I’ll just draw my dashed line over to this diagram and then down. Then we’ll call this Y*2. This is the price and quantity at which Firm 2 finally enters the market, P2 and Y*2.

Now notice, this is a jump. This is a discreet jump, a jump in total output, because suddenly we’ve added another factory that’s producing a certain amount of output. So what happens then is I add, to the amount that Firm 1 was producing at that price, I add the amount that Firm 2 enters with. So I have to scoot my market supply curve over and create a new point. I’ve got a jump here, and that jump is equal to the amount that Firm 2 produces when they first start production. So I take this horizontal distance and add it on to my market supply curve.

Now what happens as the price continues to increase? As the price continues to increase, Firm 1 continues to move along its marginal cost curve and Firm 2 continues to move along its marginal cost curve. As long as the price is below this high price here, Firm 3 is still out of the market. But Firm 1 and Firm 2 continue to move along their marginal cost curve. So let’s go over to our diagram then and add together Firm 1 and Firm 2’s marginal cost curves. This line segment shows how much output is produced by the combination of Firm 1 and Firm 2 at any of these prices, at a price that’s somewhere between P2 and this higher price, we take the amount produced by Firm 1, add it to the amount that’s produced by Firm 2, and we get a point on this segment.

Now, as the price continues to rise, eventually what happens is Firm 3 finds that it’s profitable for it to produce output in the short-run. That is, the price will get so high that Firm 3 can now cover its variable cost. We’ve passed Firm 3’s shut down points. And that will occur at this high price. We’ll call it P3. At P3 Firm 3 can finally enter, and what it does is it produces certain amount of outputs right here, and if we go down to its axis, we can call that amount of output Y*3. So Firm 3 now enters at this high price with its own output. So what we do is take the amount of output produced by Firm 3 when it enters at P3 and add it on to the amount that’s already being produced by Firm 1 and Firm 2. And from there on it’s just a matter of moving on up the marginal cost curve of all three firms. As the price continues to rise above P3, we move along the marginal cost curve of Firm 1, Firm 2 and Firm 3, and that gives us our market supply curve for the higher prices.

So that’s what things look like. Notice a couple of things about the market supply curve by way of summary. First of all, it has some jumps in it. The jumps occur anytime the price rises high enough to pass the shut down point of a particular firm. So we have a jump here, a T1, when Firm 1 is first able to profitably produce in the short-run. We have a jump at P2 when Firm 2 is first able to profitably produce. And finally, we have one more jump up at P3, that is
Perfect Competition

Market Supply

**Deriving the Short-Run Market Supply Curve**

when Firm 3 is able to cover its variable cost when we finally reach the shut down point of Firm 3. So we have these discreet jumps, and the discreet jumps occur whenever a new firm is able to cover its variable costs and resume operations in the short-run.

Between those jumps we have a smooth curve that represents the sum of the marginal cost curve of all producing firms. In this region, only Firm 1 is producing. In this region, we're adding together the marginal cost curve of Firm 1 and Firm 2. And finally, in this region up here, we're adding together the marginal costs of Firm 1, Firm 2, and Firm 3. This then is the short-run supply curve for the market. It starts here on the axis, because nothing is being produced. It then jumps to the marginal cost of Firm 1, the sum of 1 and 2, and the sum of 1, 2, and 3.

I've done this example with only three firms because, you know, if you use any more firms, the thing is going to start to get really, really messy and complicated and too small to draw on the web. So what you can do then, is you can imagine that instead of only three firms, we have a very, very large number of firms. With a very, very large number of firms, the shut down points are encountered more frequently. Anytime the price is rising a little bit, you're likely to be drawing in another firm, rather than having these big gaps, the shut down points will occur closer together.

Another thing that happens if the firms are small relative to the market, is the amount that they produce when they first start operating is also small. You don't get big jumps, you get smaller jumps, and they occur more often. So if we have not three firms, but say, 100 firms or 1000 firms, what we begin to get is a curve that doesn't have all this jumpiness in it. The curve begins to be a smooth curve, and you can imagine, it might look something like this. And we call this then the short-run supply curve for the market. This is how we derive it.

The short-run supply curve for the market is composed of the sum of all of the marginal cost curves of the individual competitive firms above their shut down points. To get the short-run supply curve of the market, add together the short-run supply curve of all of the firms that are available for production. And those short-run supply curves will simply be the marginal cost curve above the shut down point for each firm. The more firms you have, the smoother this curve will be.

Now what we'll do is we'll take this information about the short-run supply curve, put it together with the demand curve, and show you how the market adjusts to changes in the short-run, and how those adjustments influence what individual firms in the market choose to do.
Perfect Competition

Market Supply

Relating the Individual Firm to the Market

We’ve been talking about how an individual firm responds to a market price to maximize its output. Well, that still leaves the question unanswered, how is the market price determined? What we’re going to do in this lesson is put together the supply and demand stuff that we did early in the course, with all the stuff about cost curves that we’ve been doing in the recent lessons. I remember when I took this course I found this stuff eventually became very confusing. We have all these cost curves flowing around, and all of them have different names and how does that relate to supply and demand? Here’s a lecture where we stop for a moment and start to put a lot of things together. Now, I hope you’re not confused, at this point, about the different cost curves, what they mean and how they’re labeled. If you are, you can go back and review some of the earlier lessons. I’m going to do a quick review before we actually begin the supply and demand analysis here.

The question we’re going to be answering here is how does the market influence the individual firm, and how does the individual firm, in turn, along with a bunch of other individual firms influence and actually make up the market? So, let’s look then at these two diagrams side by side, I put them side by side because they are measuring the same thing, price and costs on the vertical axis. Over on the right, I have the situation for an individual firm. Over on the left, I have the situation for the market. Now, the individual firm is defined by its marginal cost curve, its average variable cost curve, and its average total cost curve. This is what the individual firm looks like in the short run. Now, all right, a few quick review points. Remember, first of all, all the short run refers to the period where the firm cannot change its fixed inputs, it has a factory, conveyor belts, tools, all that stuff that’s fixed. All these curves refer to what the firm can do if it modifies only its variable inputs, if it changes only the amount of labor that it hires. The marginal cost curve is eventually upward sloping because labor productivity begins to fall because of congestion of the fixed inputs. The average variable cost tells you the labor cost per unit. If the price isn’t higher than the minimum average variable cost, then the firm cannot profitably produce in the short run. The firm will minimize its losses by shutting down. The average total cost curve tells you the total cost per unit produced. If the price is below average total cost, the firm is making a loss in the short run. If the price is above the average total cost, the firm is making an economic profit in the short run. So these curves summarize the technology and the input prices of the firm to tell you how much it costs to produce the given amount of output. Once we put a price into this picture, we can tell you how much the firm will be producing and how much profit the firm may be making.

Over here we have the situation for the market. The market supply curve is simply the sum of the short run supply curves of the individual firm. Take a moment and ask yourself, what is the short run supply curve for an individual firm? Where do we get it? The answer is the short run supply curve for an individual firm is just the firm’s marginal cost curve above the shut down point, above the point of minimum average variable cost. So, if we add together all those short run supply curves the way we did in the last lesson, we get this blue curve, the short run supply curve for the market. Now, here’s the market demand curve that’s determined by the behavior of consumers. If you put the short run supply curve together with the demand curve for the market, you can find the point of equilibrium. The point of equilibrium is the price at which the quantity supplied in the short run is equal to the quantity demanded. If the quantity supplied and the quantity demanded are not equal, the price will adjust. Remember the bidding mechanism? The price will rise if there is a shortage and the price will fall if there’s a surplus until finally, we reach the equilibrium price, the price at which quantity supplied and quantity demanded are equal.

Now, let’s look at this for a moment. We know that the supply curve represents the sum of a bunch of small firm’s supply curves, the supply curve of a bunch of competitive firms in a short run. That’s where we got this blue curve. We’re given that this is the summary of the behavior of all of these firms, given that this summarizes the behavior of all of the firms in the market. We know that the equilibrium price is going to be right here where the curves cross, that is, the equilibrium price will be, we’ll call this \( P^* \), and the quantity that will be produced in the market will be \( Q^* \). This is exactly what we did early in the course when we played with supply and demand curves. Now, how does an individual firm respond to this particular price? That is, what is our little firm doing here when its playing its own part in the market? Well, what it’s doing is this: take the price that’s set, follow it over to the individual firm’s marginal cost curve and put a dot there. This is the output that our firm will be producing in the short run, the output is, we’ll call this \( Y^* \), the profit maximizing response of this firm to a market price of \( P^* \).

So, the first thing is how much output is our firm producing? And the answer is, where price equals marginal cost and the marginal cost curve, or course, being upward sloping at that point.

The second question is how much profit is our firm making at that point? How much profit? Well, remember the trick for calculating that. Take the price, go over to the curve and look at how much of its revenue the firm is spending on
Perfect Competition

Market Supply

Relating the Individual Firm to the Market

its various inputs. If we take $Y$ and go up to average variable cost curve and over to the axis, we can see that the firm is spending this chunk of its revenue on labor, and if we go up to the average total cost, we can see that the firm is spending this chunk of its revenue, the top portion of this box, on the fixed inputs. Add them together and it looks like, at this particular price, price being equal to average total cost, our firm is just breaking even, it is making zero economic profits. Now, remind yourself, zero economic profits doesn’t mean that you don’t want to be in business. Zero economic profits mean that the firm is just able to cover the opportunity cost of all of its inputs, labor, capital, etc. Everything including the value of the manager’s time. Everything that’s relevant here, all of those costs are included in these cost curves, and so, if the firm is covering its total costs, it is in fact covering all of the opportunity costs of all the inputs that it’s using.

Well, in the case, the firm would be completely satisfied and happy about producing. Let’s suppose now that there’s a change in the market, we’re starting with a situation that is kind of nice and stable, the firm is breaking even, its covering its cost, its making zero extra profit, zero economic profit, and the market is in equilibrium. Let’s suppose now that this firm is producing television sets like before, and let’s suppose that the demand for television sets increases because it’s summer and the Olympics are going to be broadcast and a lot of people run out to buy a new television set. Well, how would we show that in this picture? How would we show an increase in the demand for television sets because of the Olympics? The answer is the demand curve would shift outward representing an increase in the quantity of televisions demanded at every price. So, our demand curve would shift outward, showing a change in the market. Let me go ahead and draw that new demand curve, it would look something like this, and I’ll put my label up here just to try to keep my graph uncluttered, $D'$, represents the new demand because of the change in the economy, people want televisions to watch sports. Well, given that the demand curve has shifted outwards, we have now got a situation of dis-equilibrium. At he price of $P$, say $300 per television set, the quantity demanded exceeds the quantity supplied, that is, here’s the quantity supplied. All those firms trying to make a profit, producing where price equals marginal cost, they are not able to keep up with the quantity demanded, we have an excess demand at the old price. Therefore, the bidding mechanism pushes up the price of television sets, and as it pushes up the price of television sets, we get to a new equilibrium point. Now, we've done this before, this is just review. What happens as the price of television sets rises, is this: some buyers decide that they will put off the purchase of the television set, they will go watch the Olympics with a friend, and sellers produce more television sets. We move up along the market supply curve until we reach the new equilibrium point. The equilibrium point in this case has a higher price per television sets, $P'$, and it has a larger quantity of television sets actually produced and traded, and I can label this $Q'$, an increase in the quantity. What we're interested in this particular lesson is going behind this picture, looking in more detail at how the supply actually ends up changing. How does the quantity supplied actually increase in response to the higher price? In order to see how the quantity supply actually changes, in order to put this movement under a microscope, we’re going to go over and look at the behavior of an individual firm.

As the price rises to $P'$, if we take that price on over into our individual firm’s diagram we see that that individual firm now facing a higher price, we’ll respond by producing a larger quantity of output. Why is that? Why does the firm choose to produce a larger quantity of output when the price goes up? The answer is that at that higher price, the firm can now afford to cover a higher marginal cost. Before, this firm’s marginal cost might have been equal to $300, in fact, price equals marginal cost, but if the price of televisions goes up to $400 a television, this firm can now afford to hire some extra workers and turn out more televisions. Maybe these extra workers are brought in by being hired at the going wage, certainly, but the point I want to make is, why weren’t those workers hired before? Why weren’t those extra workers hired before? Why weren’t those extra televisions produced before? The answer is at a price of $300 per television set, the firm can only afford to hire so many workers, those extra television sets here are going to cost the firm up to $400 to produce because labor productivity is diminishing, if you crowd more workers into your factory and make them work with a given set of tools, and a given factory space, eventually that productivity starts to decline as you push more workers in, but if the price of televisions goes up, then it’s profitable for you to bring the workers in even though the productivity is falling. That’s the thing you need to know, the firm responds to the higher price by hiring more workers even though the worker’s productivity is diminishing at the margin, even though the cost of producing televisions is rising with a higher price for televisions, this firm now finds it profitable to hire those workers anyway. It wouldn’t have hired them when the price was $300, but it will hire them when the price is $400 per television set.

All right, so given that the firm now finds it optimal to hire more workers, given that it's now profit maximizing to produce more television sets, what happens to the firm’s profits? The answer is this: before the firm was making zero economic profit, now when it's producing more televisions at a higher price, the firm is actually going to be making
Perfect Competition

Market Supply

Relating the Individual Firm to the Market

some economic profits. See here, that the average variable cost when they're producing this larger quantity of television sets is right here. This is what it’s going to cost per television set to pay your workers. If you look at the difference between average variable cost and average total cost, this is the average fixed cost. This is what your having to pay to cover your overhead per television produced. So, if we wanted to draw our little parfait, again, we would find this chunk of the revenue goes to workers and this chunk of the revenue goes to the fixed costs. Well, that leaves us with a nice little chunk here on top. If we go over to the average total cost curve, we find that it only takes this bottom rectangle to cover the cost of production. All of this revenue on top is economic profit because of the increased demand for television sets, the firms that are producing television sets will find that in the short runs, they are able to earn an economic profit equal to the amount of this area. The higher price for television sets leads them to produce more television sets, and because the total cost of production is less than the total revenue they earn, the firm is making an economic profit in the short run. So, I could label this \( \pi \), remember \( \pi \) stands for profit, \( \pi \) is greater than zero, the firm is making a positive economic profit.

Will this situation last? Well, it will last in the short run, but in the long run what will happen and we’ll be talking about the long run very shortly. In the long run what will happen is new firms will enter this market. Hey, look, this profit is sitting there like a bunch of raw meat on the surface of shark infested waters, and the sharks are going to start to gather until there’s a great crowd of them all fighting over the meat and the meat is gone. So what you have here in an economy is: you’ve got this chunk of raw meat, this chunk of tasty raw meat of economic profit that draws in the sharks, and not to say that sharks aren’t doing what they do when they go after the raw meat. It’s what companies do when they see an opportunity for extra profit, they enter, and what happens when new companies start up in this market to enter? The short run supply curve begins to shift outwards, and when it shifts outwards, the price of the product falls and eventually the firms will be pushed back into the situation where there are no extra profits to be made. There is no economic profit; everybody just covers their opportunity cost. This is what happens in a competitive market in the long run.

I want to summarize with this question, how do we adjust from this original price quantity equilibrium to this new one? What happened behind the scene that gave us the increase in quantity supplied after the demand shifted outwards? We can see in this individual firm’s diagram that an individual firm responded by increasing its output. An individual firm moved along its marginal cost curve as the price rose allowing it to cover the increasing marginal costs associated with expanding its output. Put simply, when the price rose, this firm could afford to hire more workers even though those extra workers were less productive and marginal cost was rising. But that’s only one part of the picture; that’s all we can see here in this little diagram that I’ve drawn. Something that you can see here, we looked at more carefully in the last lesson. What’s another reason why the quantity supply is increasing in the short run? It’s not just that firms are moving along the marginal cost curve, but what else is happening in the short run that’s increasing the quantity supplied as we move along this blue curve? What else is happening in the short run? The answer is in the short run. Firms that were previously shut down are starting up again. Some firms were not able to cover their variable costs, some firms had average variable cost curves that had minimum points up here in this region and they were shut down before when the price was $300 per television. When the price goes up to $400, some firms that were previously shut down, now resume operation and they add their output to the market total. So when you move along the short run supply curve, two things are happening; individual firms are expanding their output and firms that were previously shut down are resuming production. That’s the short run adjustment process, you can see some of it in this picture, but some of it you can’t because we are only looking at one firm.

Anytime the demand curve shifts, there will be this kind of adjustment. An outward shift in the demand curve will increase economic profits for firms in the short run; an inward shift in the demand curve will lead to perhaps economic losses or a reduction in economic profit in the short run. However, in the long run, the supply curve itself will shift as firms enter or leave the market in response to changing conditions. We are going to be going to the long run next.
Perfect Competition

Examining Shifts in the Short-Run Market Supply Curve

There was a lot to cover in this diagram and I find that when I give this explanation, students always have questions. I’m going to try anticipate some of your questions and answer them in this segment. The first question that students often ask is what causes the short run supply curve to shift? What causes this green curve to change its position? What leads to a larger quantity of output at any given price?

Now, over on the board I have redrawn the short run supply curve by itself and I am going to talk about the things that would cause the short run supply to change its position. In particular, let’s talk about what would cause it to shift outwards. To answer this question, start with recalling what the short run supply curve is. The short run supply curve is the sum of all of the firm’s individual short run marginal cost curves above the shut down point. If you take all of the firms that are in the market at a given time and you add up their short run supply curves, that is, the amount of output they will produce at a particular price, you get the short run supply curve for the market. It’s the sum of the individual firm’s short run supply curves above the shut down point. Now, what would cause those curves to shift outwards? Well, there are two things that would cause this shift to happen. First, would be a change within the firms that are producing currently. The second would be adding more firms. Either of those would cause the curve to shift. Let’s start with a change in the existing firms.

Suppose the existing firms suddenly had a shift from a few trucks in the short run to a lot of trucks in the long run. If the existing firms acquire more capital, then the marginal product of labor is going to change. If you have five workers working with a single truck, the marginal product of labor is going to be very low, but if you have five workers working with three trucks, then the marginal product of labor is going to be high. Remember, the marginal cost of production is the reciprocal of the marginal product of labor. So, when the firm gets more capital, when the firm moves from the short run to the long run, and is able to acquire more trucks, then the marginal product of labor increases and as the marginal product of labor increases, the marginal cost of production decreases. So, we can show it this way: we’re moving from one set of short run marginal cost curves with only a few trucks to another set of short run marginal cost curves that have more trucks, and in that case, the marginal cost of production is going to be lower. It’s like we’re moving from this particular set of marginal cost curves to this other set of marginal cost curves that are further up along the long run average cost curve that have a lower marginal cost because the marginal product of labor is higher. Now, that’s a round about way of saying that one thing that can shift out the short run supply curve is if the existing firms get more capital. As we move from the short run to the long run, the short run supply curve will move outwards because the existing firm may acquire more capital and when they acquire more capital, labor becomes more productive and the marginal cost falls for the individual firms. All right, that’s reason number one. Short run supply curve shifts outward if the existing firms get more capital.

The second reason that the short run supply curve will shift outward is adding more firms. In the long run, when there is free entry, the existence of profit in this market will attract new firms into the industry. When new firms are attracted into the industry, those new firms add their short run supply curves to the market, and that cause the market short run supply curve to shift outwards. It shifts outwards because we are adding new firms. It’s that simple. So, what causes the green curve to shift around? There are two things. The first is a change in the fixed inputs of our given firms, our existing firms. The second is adding new firms. You can consider the process in reverse. If firms leave the market, the short run supply curves shifts inward and if firms cut back on their capital, that is, in the long run when they’re able to lay off some of their trucks or sell them or return them to the people from whom they’ve rented them, then the capital stock shrinks, and when the capital stock shrinks, the marginal productivity of labor falls and the marginal cost rises again. So, the same two ideas are at work whether the curve is shifting outward or inward. A change in the capital stock changes marginal product and that changes marginal cost, that’s reason number one, and a change in the number of firms causes the short run supply curve to shift, that’s reason number two.

Another question that students ask is when do you move along the cost curves and when do the cost curves shift? Let’s see if I can answer that one. Over here, we have the long run and the short run cost curves, and I’ve reproduced them over on the board so that we can talk about them. First, anything that changes the price of the firm’s product is going to cause a movement along the cost curves. Remember, you can never shift the curve by changing what’s on one of its axis. You can only shift the curve by changing something that isn’t in the picture. You can only shift a curve by changing something that you’re holding constant when you draw the picture. So, if the price of the firm’s product changes, you will move along the curve. In the short run, the firm will move along the short run marginal cost curve, and in the long run, the firm will move along the long run marginal cost curve as the firm expands its output. However, if you change one of the things that’s held constant when you draw this picture, then you’ll shift the whole set of curves. For instance, let’s suppose that the price of the firm’s input increase; let’s suppose that the
Market Supply

Examining Shifts in the Short-Run Market Supply Curve

price of labor and capital increases. In that case, what will happen to the cost curve is this: the whole set of cost curves will shift upwards representing an increase in the cost of producing any given quantity of output. If the price of the input falls, then the cost curves will all shift downwards representing a reduction in the cost of producing any quantity of output. The next thing that could shift the cost curves would be a change in technology. If the firm’s technology improves, that is, if it can make more output with a given amount of input than before, then the cost curves will shift downwards because now the firm can produce its target level of output with less input, and therefore, at lower cost. So, the thing that will shift the cost curves will be a change in the price of the firm’s input or a change in the firm’s technology. One more thing to point out, I told you a moment ago that when the firm moves into the long run and changes its quantity of capital, there is a shift in the short run supply curve. That shift occurs not because the short run marginal cost curve actually shifts, but because the firm moves from one short run marginal cost curve to another short run marginal cost curve. Remember, anytime you change the amount of the fixed input, you have to change the short run marginal cost. The short run marginal cost is based on labor productivity, which depends on the amount of tools or capital that the workers have to work with. Anytime you pick a point on our blue long run average cost curve, any time you pick a point, you can hold the capital constant and draw a new green curve that’s tangent at that point. If firms acquire more capital in the long run, they’re going to be moving from this set of short run cost curves to another set of short run cost curves at a different point on the long run average cost curve. So, it’s not that the green curve actually shifts, for a single firm, it’s that you move from one set of short run cost curves to another set of short run cost curves.
Perfect Competition

Market Supply

Deriving the Long-Run Market Supply Curve

This is the last lecture in our series in productivity, costs and profits. In this lecture, we'll be looking at the effects of the expansion of an industry and we'll be deriving a tool that we call the long run supply curve. Let's start with revisiting the relationship between the market and the individual firm that we derived last time. So, here we have the market equilibrium, the demand curve, which is the behavior of households and consumers, and the short run supply curve, which is the sum of the marginal cost curves of the individual firms. In the short run, the supply curve and the demand curve intersect and give us an equilibrium point, we'll call this P₀ and a market quantity, we'll call this Q₀.

Now, if we take this price over into the diagram for the individual firm, we see that in long run equilibrium, the firm is making zero economic profit and the firm is producing at the point of efficient scale. This is a long run equilibrium. The only way we can have a long run equilibrium is if no firm wants to change its behavior. In this case, the firm is maximizing profit by producing where price is equal to marginal cost and the firm is breaking even by producing where price is equal to average cost. There is zero economic profits and therefore no tendency for entrance or exit. Everything is stable in this market. Let's upset the equilibrium now, by changing the demand for the product. If we change the demand for the product, and the demand curve shifts up to this new point D prime. Then we had an excess demand for the product at the old price and the bidding mechanism then, will push the price up, so that firms supply a larger quantity and the quantity demanded decreases. So that the price goes up to this point here. Now, that would be the short run equilibrium and the firm would move along its short run marginal cost curve over here in this diagram.

I'm not going to draw the short run marginal cost because my focus here is on a different story. I want to know what happens as new firms enter the market. What happens to the cost of production? As new firms enter the market and the short run supply curve shifts outwards, representing an increase in the number of firms offering trucking services, it may be that as these new firms enter the market that they drive up the cost of production for all firms. Suppose that truck drivers are in short supply. If truck drivers are in short supply, then as trucking companies try to expand their operation, they're going to be bidding up the salaries of truck drivers. And, as the salaries of truck drivers rise, then the cost of production for every firm in the trucking industry are going to increase.

When this happens, the cost curves shift upward and the costs that firms can afford to operate at will be rising. That is, the point of minimum long run average costs will be shifting upwards. So, what we have then, as the short run supply curve shifts outwards and firms enter the industry and compete for a strictly limited resource, like talented truck drivers, the cost of production will shift upwards and we will get something like this. A new long run average cost curve at a higher level. If expansion in an industry causes the cost curves to shift upward, like this, we say that this is an increasing cost industry. If the expansion of the industry causes the cost curves to shift upwards, we call it an increasing cost industry.

What we're saying is, as the industry expands, competition for a strictly limited resource, will in fact increase the costs of all firms in the industry. This was the case in the computer industry in the late 1970's, early 1980's, when there was a limited supply of talented computer programmers. It took awhile for students to go to school and learn computer programming and of course, those students were attracted by the promise of higher salaries. But in the short run, as computer companies tried to expand their operation, competing for a pool of talented programmers, they pushed up programmer's salaries through competition and that raised the cost of production for all of the companies.

Now what does this mean? It means that there will be a limited amount of entry before the price falls to the point at which firms can just afford to cover their costs. The break-even point for firms has shifted upward as the cost of production has increased. That means the price for the product cannot fall back to its original level, but can only fall as low as the new point of minimum average cost. That new point of minimum average cost might be not eight cents a mile for trucking service, but 12 cents a mile for trucking service. The increase in costs limits the amount by which the short run supply curve can shift outwards. This new price P₁, has to be the new equilibrium price, it's the new price that's given by the intersection of short run supply and demand, it's the new price that we have in long run equilibrium. Firms are breaking even and the quantity supplied equals the quantity demanded.

If I connect these two dots, I get what we call in Economics the long run supply curve. The long run supply curve shows what happens to minimum average cost as an industry expands. In our case, we were talking about an increasing cost industry. When an increasing cost industry expands, the minimum average cost of production rises as firms compete for a strictly scarce resource and push up the cost of production. The long run supply curve in this...
Perfect Competition

Market Supply

**Deriving the Long-Run Market Supply Curve**

case is upward sloping. It is possible that you have what’s called a constant cost industry. A constant cost industry would be one, where as the industry expands, the costs of production do not change for the individual firms.

In that case, the short run supply curve would keep shifting outward, until our price fell back to eight cents per mile, the original minimum long run average cost. In that case, the long run supply curve would be a horizontal line. The short run supply curve would keep shifting out, until it intersected the demand curve at the original price. The original minimum long run average cost. One more possibility. What if the expansion of an industry actually decreases the cost of production for individual firms? What if we have a decreasing cost industry? What if a large number of firms in an industry actually lowers the cost for each firm in that industry?

That could be the case, say in software design, where you have a bunch of software engineers who work in a market and with more people trying to solve the same kinds of problems, they’re able to solve them faster. They share ideas. Everyone’s working on the same problem and therefore it takes less time to solve it. In an industry like that, a large number of talented software designers lower the cost for each other, by working on common problems and sharing insights with each other. That leads to what’s called a decreasing cost industry. As the industry expands, the costs of production actually fall. And in that case, the long run supply curve would be downward sloping. So, a quick summary.

The long run supply curve tells us about the relationship between the size of an industry and the cost of production for an individual firm in that industry. If expansion of the industry pushes up costs for individual firms, we call it an increasing cost industry and the long run supply curve slopes upwards. That will be the case if expansion in the industry heats up competition for a strictly scarce resource, like talented programmers or certified truck drivers. It’s also possible that expansion of the industry leaves costs unchanged. That’s what we call a constant cost industry and that would give us a long run supply curve that’s horizontal. Finally, it could be that expansion of an industry actually reduces costs for the individual firms. This will be the case if growth in the industry allowed communications among the participants or other ways of economizing on costs. External economies of scale we call it sometimes in Economics, where the growth of an industry lowers cost for everyone. In that case, the long run supply curve would be downward sloping.
Perfect Competition

Competitive Firms’ Responses to Price Changes

Examining the Firm’s Long-Run and Short-Run Adjustments to a Price Increase

Suppose you’re running a trucking firm and you suddenly discover that you can charge more for your trucking services than you were able to before. How would you change your operation and how would your response in the short run differ from your response in the long run? In this lecture we’re going to answer that question and we’re going to do it first by drawing out all of the cost curves that we’ve studied up to this point. Then we’re going to look at how the firm maximizes its profits in the short run when it can make a small adjustment and in the long run when it can make a large adjustment.

Let’s cut off first the cost curves that will be associated with the firm’s choice. And, I’ll start with the long run cost curve. I’ll first draw the long run average cost curve and the long run average cost curve; I’ll draw with its usually U shape. This long run average cost curve tells us about the firm’s scale economies, increasing returns to scale, when long run average cost is downward sloping, constant returns to scale at the bottom of the curve and decreasing returns to scale when long run average cost is increasing. And I’ll label the curve LRAC, long run average cost. Now, of course associated with the long run average cost curve is the long run marginal cost curve. And the long run marginal cost curve tells us about the cost of increasing the firm’s output in the long run when the firm can achieve the cost minimizing technique.

Remember, the long run curve, the blue curves, always represent the firm’s possibilities when it can combine capital and labor in whatever proportions it wants. That is, the blue curves represent cost when the firm can get to the cost minimizing technique, the cost minimizing combinations of capital and labor. So, let me label this curve long run marginal cost, or LMC. Let’s suppose now that the firm is facing a price that’s equal to $P_0$ for its trucking services. Maybe this is eight cents a mile and that’s what the firm is able to charge right now for providing its trucking services to its customers. What would the firm do, in the long run, if it could sell its services at eight cents a mile? How many miles worth of trucking services would the firm choose to offer?

The answer, as always, is find the place where price is equal to marginal cost. So, follow the price line horizontally across the diagram until we get to price equals marginal cost, until we get to the long run marginal cost curve. If the price had been higher, say at a price way up here, then the firm would have gone over to a point like this on its long run marginal cost curve. If the price had been lower, say down here, then the firm would have looked for a point like this. The firm always, always seeks to make price equal to marginal cost as it tries to maximize its profits and in the long run, that means being where the price touches the long run marginal cost curve.

Now notice, I’ve set up this example in a special way. I wanted to make sure that the firm was operating here at the bottom of its long run average cost curve. That is, it turns out that this firm happens to be at $Y_e$, the point of efficient scale. In our previous example, this was 50,000 miles a day of trucking services provided, with 20 trucks and 30 drivers. This is the situation at which the firm maximizes its profits. How much profit is the firm making in this case. I’ll give you some options to choose from and you pick the correct one. Looking at this diagram, in the long run, how much profit is the firm making?

The answer is, the firm is making zero economic profit, because the price of its service is exactly equal to the average cost of providing the service. The firm is breaking even. If the price were higher, the firm would be making a positive economic profit. If the price were lower, the firm would be making an economic loss. But as it is, when the firm is operating at the point of efficient scale, the firm is making zero economic profits. The firm is breaking even. It’s making just enough money to cover the opportunity cost of all of the resources that it’s using to produce its trucking services.

Now, let’s consider the short run. Suppose this firm wants to alter its pattern of production in the short run. Remember, the short run is the period of time over which the number of trucks is fixed. We’ve got a firm with 20 trucks and it wants to change the number of trucking services that it provides or the number of miles that it travels, but it can only do so by changing its labor input. It cannot change the number of trucks it uses in the short run. So, I’ll use my green curve here, to represent short run costs and in this case, I’ll get another U shaped curve and this U shaped curve will be the average total cost curve for the short run.

Now, where did I get this average total cost curve? I held constant the number of trucks the firm used, equal to 20 and I modified labor in order to produce extra quantities of trucking services. If the firm wants to run more routes and provide more service, it’s got to hire more drivers and use those trucks more intensively. Or, if it wants to cut back its output, it’s going to have to lay off workers, but it’s still stuck with the trucks. So, here’s the average total cost curve.
Perfect Competition

Competitive Firms’ Responses to Price Changes

Examining the Firm’s Long-Run and Short-Run Adjustments to a Price Increase

associated with changing output in the short run when you’ve got 20 trucks. And 20 trucks is the number of trucks that minimizes cost for providing 50,000 miles of service a day. So, let’s put a little note on here, we’ll call this average total cost. And I’ll just make it understood that this is the average total cost that’s associated with having the number of trucks that’s cost minimizing for the point of efficient scale. So, in our case, that’s 20 trucks, in my story.

Also, there’s a marginal cost curve that’s going to cut through this average total cost curve and it’s going to cut through it at its bottom. We know, from our previous lecture that the bottom of this curve will also be the bottom of the long run average cost curve. So, there’s going to be here an intersection that involves four curves touching at once. And here’s the short run marginal cost curve. Let me put SMC here, where the short run marginal cost curve is the one that’s associated with the point of efficient scale. Well, I’ve got two more questions to answer before I feel like this diagram really makes sense.

The first one is, why can you be sure that the green average total cost curve is everywhere above the blue long run average cost curve? Why is the average total cost in the short run, always higher than the long run average cost? You give the answers. The answer is, in the long run the firm can always change capital and labor both in order to get to the cost minimizing technique. In the short run, the firm does not have that flexibility; it can only modify its labor input. It can only change one input, therefore the costs are always higher in the short run, because you can’t get to that cost minimizing technique. In the long run, of course, you can. So the costs of production are always lower, in the long run.

Next question. How can I be sure that the short run marginal cost curve lies above the long run marginal cost curve? Why does this short run marginal cost curve go above the blue long run marginal cost curve when my firm expands its output? The answer is, in the short run, if you want to expand you can only do so by hiring extra labor. In the long run, if you want to expand, you can hire more labor and more capital and do so in the combination that minimizes the cost of expansion. See, if you can’t get more trucks, you may have to use a whole lot more workers then you’d otherwise like to, just to use those trucks more productively or really load workers in, so they’re always running at maximum peak productivity.

However, in the long run, when you can modify both workers and truck, you have the flexibility to combine them to find the lower cost way of expanding your output. In the long run, you have more options and more options equals lower cost of expansion. Now, here’s one more point and this point’s kind of subtle. So, if I don’t make it clearly enough for you, the first time, don’t worry about it and just go on. But give it a shot and see if this works. If we reduce our output, why is it in the short run, that costs will be reduced less than they will be in the long run? If we cut back our output from 50,000 miles of trucking service down to 40,000 miles of trucking service, how can you be sure that in the short run our costs will fall less than they would fall in the long run?

The answer is, in the short run, when we try to cut back our output, we can only lay off workers, we can only reduce the labor input. And we may have to only lay off two or three workers before we’ve got trucks that can’t be used productively. We’ve got to be careful in the short run about how many workers we lay off or our output could start to fall very rapidly. In the long run, however, if we want to cut back our services, we can lay off both labor and capital and lay them off in those combinations that allow us to maximize our cost savings. In the short run, we have less flexibility and less flexibility always means less cost savings when you’re trying to reduce your output and cut back.

In the long run, whenever you can both lay off workers and reduce capital input, then you get bigger cost savings. Your marginal cost is higher, that means you can lower your cost faster or faster is the wrong word. You can lower your cost by more if you have more time to adjust both inputs than if you can only make one adjustment. That is, adjusting labor. All right. So, what I’ve done here is I’ve explained this important point of intersection. I’ve explained why the green curves are above the blue curves and we’ve discussed this particular point of efficient scale. Now, we’re ready to do an experiment.

Suppose instead of eight cents per mile, our firm is suddenly able to charge a higher price, maybe due to increased demand for trucking services. If that’s the case, then what happens is the firm now faces a higher price for its product. And with a higher price for its product, the firm is going to want to expand its output. Let’s figure out now what the firm is going to do in the short run. In the short run, with this higher price, the firm would like to get a piece of this action, expand its services. However, it’s limited in the short run. What’s its limitation? It can’t get more trucks. It can only
Perfect Competition

Competitive Firms’ Responses to Price Changes

Examining the Firm’s Long-Run and Short-Run Adjustments to a Price Increase

expand its output by hiring more workers and crowding more workers into the trucks to try to use those trucks more productively, keeping them running all the time, scheduling workers between trucks, so that the trucks are never idle.

In the short run we can only get increased output by expanding the variable input, labor. That means we’re going to be on the green curve in the short run. So, where’s the firm going to go? How much extra output will the firm produce in the short run with this higher price? Locate the point in the diagram that represents maximum profit in the short run. The answer is, follow the price over to the short run marginal cost curve and that gives you the profit maximizing output in the short run. What’s going to be happening? The firm is going to be trying to provide more trucking services by using more workers, hiring more workers. Instead of running one-person crews, we may have two person crews. Two drivers in a single truck, so the truck is used more productively and we get a few extra miles out of it every day.

We may have more workers working between trucks, eight hours in one, eight hours in another. The firm may incur the extra cost of flying workers around so that they meet up with trucks that need a replacement driver. All that stuff will happen. And what’s going to happen when we start crowding more workers into our given number of trucks? What’s going to start happen when we start packing more workers into those 20 trucks, is that labor productivity continues to fall. And as it does, the marginal cost of production rises. See, this is how it works. The higher price gives the trucking firm the latitude or the flexibility to hire the extra workers that push up marginal cost. Those extra workers are expensive and getting extra miles out of your given fleet of trucks, means using labor packed into trucks at a higher labor truck ratio and that’s going to lower labor productivity. As labor productivity falls, the cost of production rise and if it weren’t for this higher price the trucking company couldn’t afford to pack workers into trucks the way they’re doing. But the higher price encourages the trucking company to hirer the extra workers in the short run. So, output increases and output increases maybe to, let’s say, to 50,000 or 55,000 miles of trucking service provided a day. So, we get a new higher output and we’ll call this new higher output Y1S. So, in the short run, we get an increase of output.

Now, in the long run what’s going to happen? Oh, by the way. Notice also, is the firm making a profit in the short run? Sure they are. Go from the price down to the short run average cost, the average total cost in the short run and this gap right here multiplied by the number of miles of service provided is the firm’s profit in the short run. There’s a short run profit because price is greater than average total cost. As we see by looking at the difference between price and the average total cost curve above 55,000 miles.

What’s going to happen in the long run? In the long run, of course, the firm is going to say, look, if we can charge 10 cents a mile for our service, we’re going to get a bigger fleet of trucks. And as the firm employs a bigger fleet of trucks they’re going to find it optimal to change their labor to match the bigger fleet and they’re going to find the cost minimizing technique for providing a larger quantity of service, for serving more mile a day. So, as we move from the green curve to the blue curve in the long run, the firm is changing both labor input and capital input, more trucks and more drivers. Changing the two in whatever combination gives us the minimum cost of providing that quantity of service.

So, notice in the long run, the firm will end up producing more output than they do in the short run. Because in the long run, the firm has more flexibility and they’re going to be able to make more profit. Notice, profit is bigger in the long run, because costs in the long run are always lower. This gap is bigger in the long run than it is in the short run. So, in the long run, when the firm is able to modify both its capital input and its labor input, the firm does two things. It produces a larger quantity of output and it makes a bigger profit. It’s always good to have the flexibility that comes with the long run. Because in the long run, you can modify both capital and labor in order to achieve the cost minimizing technique. And any time your costs are lower, you’re probably going to want to expand your output.

So, quick review. I started with the short and the long run cost curves together in one diagram and I discussed the relationship among these curves, what determines their positions relative to each other. Then, I did an experiment. I raised the price that our trucking company can charge for its services from eight cents a mile to 10 cents a mile and I looked at the response of the trucking company in the short run and the long run. In the short run, the trucking company can only expand its output by hiring more workers. So, it does so, but the costs rise very rapidly because labor productivity declines quickly if we’ve got to keep packing workers into the same old trucks.
Perfect Competition

Competitive Firms' Responses to Price Changes

Examining the Firm's Long-Run and Short-Run Adjustments to a Price Increase

In the long run, however, costs rise more slowly, because the firm can modify both its capital and its labor. It can get more trucks and more drivers. So, because it takes longer for the costs to catch up, the firm is able to increase its output more in the long run, because it takes a bigger range of output for the marginal cost to catch up with the price. In the long run, the firm produces more output and the firm makes a bigger profit. In the next lecture, we'll look at how a single firm responds to a change in the market.
Other Market Models

Monopolies

Defining Monopoly Power

Suppose you’re a firm that has a highly desirable product and you’re the only one that offers it. How would you price your product? How much would you sell? The more you sell, the lower the price would be according to the law of demand. But, the higher you raise your price, the less of the product people will buy. How do you as a firm, who has market power make the decision about how to price your product? We’re going to be studying this problem in the coming lessons, and we will begin this lecture by introducing the concept of market power and talking about an extreme case, the case of the single seller, a monopoly. Let me begin by defining market power. A firm is said to have market power, if it is able to influence the market price of its good or service, that is, if it has the power to set a price for its product, rather than take the price as given in the market. Market power is a contrast to a competitive firm; a firm that is competitive is a price taker. A firm with market power has the ability then to set its price. An extreme case of the firm with market power is a monopolist. A monopoly is a single seller of a good or service. Now, you may be hard pressed to come up with an idea quickly of a firm that really exists in the real world that’s a true monopoly. In fact, monopolies are quite rare and monopolies usually arise due to one of three factors.

The first factor that can create a monopoly is a firm that has sole ownership of a particular resource that is important to the production of a good or service. An example of this would be the case of the market for diamonds. There was a time when the DeBeers Diamond Cartel owned almost eighty percent of the world’s diamonds. Many of the competitive diamonds were in the Soviet Union and were not available on the market at the time. Because one single company owns so many of the diamond mines, they were able to have considerable influence over the price that was charged for diamonds. So, one thing that can create a monopoly is if the company has sole ownership of a particular resource that is important in the production of the good. Another example would be if we were at an airport, for example, and one company had sole ownership over the space that was available for selling concessions. If they owned that space, they would then be the only restaurant in the airport and would, therefore, be a monopoly.

This brings us to the second situation that can create a monopoly and this is monopoly that is created by government action. Government action meaning patents, copyrights, and in some cases special licenses. In these cases, the government creates a situation whereby entry into an industry is prohibited. For example, one of the important driving forces in the pharmaceutical market is company’s desire to obtain a patent for a new effective drug to treat a certain kind of illness or condition. The company that developed the new drug first, will be granted a patent, and with this patent, they will have the sole right to produce and sell this drug in the United States for a period usually as long as fourteen years. In this case, a drug patent grants a monopoly to a company allowing it to have sole right to sell this product, and because there are not competitors, at least while the patent is in effect unless the company wants to license production to other companies. While this company has the patent, they have the right to be a monopolist, and being a monopolist allows them to earn extra profit and thereby recover the costs that they sank into developing this new drug. Other things that can give companies patent granted by the government, or other things that can give companies monopolies granted by the government, are called sole licenses. It has been the case throughout history that governments have granted single licenses to particular companies to provide particular types of service. For example, the Hudson Bay Company, which was licensed by the King of England to develop trade in the New World, had a monopoly right to, on behalf of England, go and provide imports and exports to the New World.

A third case in which we get a monopoly is what we call a natural monopoly. A natural monopoly arises because of the interaction between the size of the market and the efficient scale of operation of a single firm. Let’s look at a picture that will make the idea of natural monopoly a little clearer. You’ll recall from our earlier discussions that the bottom point of the long run average cost curve determined the efficient scale of operation or the right size for a firm in a long run in this market. Firms that want to be competitive in the long run will be operating at the bottom their long run average cost; that is, they will be minimizing the average cost of production and thereby maximizing their profits. If it’s a competitive market, firms will continue to enter until the price of the good drops down to the bottom of the average cost curve. Now, if the long run average cost curve is relatively close to the axis, that is, if the long run average cost curve points to a point of efficient scale that is small, relative to the amount of the product that the market wants to buy, that is, if the demand curve is way out here and the point of efficient scale is way back here, then there’s room for a lot of other firms to enter this market. We’ll get several of these blue curves fitting in adding up to the total amount of the product that the market is actually buying. That is, if the efficient scale of operation is small, relative to the size of the market, it takes a lot of firms to meet demand at this minimum average cost. However, if scale economies are very great, if the economies of scale are large, relative to the size of the market, then we can get a situation like this, with a long run average cost curve for each of its minimum point at a point that might actually be on its intersection with the demand curve. In this case it only takes one firm to meet the entire market’s demand for this product at minimum average cost. If we look at the point here at minimum average cost, it’s beyond the demand.
Other Market Models

Monopolies

**Defining Monopoly Power**

curve; this firm can produce everything that the market wants to buy at a price equal to minimum average cost. This is a case when an industry has very, very great economies of scale. Where the opportunities for teamwork and specialization, increasing returns to scale, and diminishing average cost go on for a long time, that is the economy or the industry grows very, very large before it runs into problems with management and communication, all of those other things that lead to decreasing returns to scale. When the economy has a lot of scale economies, when the firm has a lot of scale economies relative to the amount of the good that people want to buy, then we have what we call a natural monopoly. When the point of efficient scale is enough to provide all of the goods that the market wants to buy, from a single firm producing it, then we have a situation of natural monopoly. Some examples of this that have been sited in the past were electrical utilities. Electrical utilities provided power to your house by stringing up electrical wires that led to a large generating plant that used coal or water power to produce electricity that has been wired to your house, and there was really no need for competition in this market because one utility taking care of the wires and the generation could provide all the power that a particular market might want to buy. So, we have a situation of natural monopoly. The alternative would be for another power company to come in, build a large tower plant and string their wire to your house so you could then choose whose electricity you wanted to buy. As long as we think of power generation as coming from utilities that then send the power to your house through electrical wires, it really seems that it is a natural monopoly. We don't want a lot of firms out there stringing their own wires and building their own plants. Now, we are going to talk more about electrical utilities a little bit later when we talk about the possibility that monopolies might be deregulated, but for now, let me summarize.

The three things that create monopolies are, first of all, sole ownership of some strategic resources, as in the case of the Diamond’s Cartel. The second is government action, usually in the granting of patents or special licenses, and finally, natural monopolies which occur when there are a lot of economies of scale in an industry so that the point of efficient scale has a lot of production relative to the demand in the market when all the demand in the market can be met at very low cost or a diminishing cost by a single firm. Given what we have in the monopoly or supposing that we have a monopoly. How then will this monopoly behave? Will it charge high prices and sell little output, or will it charge lower prices and sell more output? In the next lesson, we take up the question then of the monopolist’s problem. How to make the trade off between higher prices and larger sales.
Imagine for a moment that you are the owner and manager of the only restaurant in the airport. You have a problem. What price should you charge for meals that you serve to your customers? If you charge a higher price, then customers will just fly hungry and avoid your product. If you charge a lower price, you'll bring more customers into your restaurant, but you won't make as high a margin on the meals that you serve. Do you want to charge a higher price and bring in fewer customers or a lower price and bring in more customers? Let's analyze this problem now and look at what we call in economics, your marginal revenue. Marginal revenue is defined as the change in total revenue that results from a change in quantity sold; and we'll begin to develop this concept of marginal revenue, which we'll see is very useful in helping you make a decision about the profit maximizing quantity of meals to serve, and of course, the profit maximizing quantity of meals implies the price you should charge for your product. So, let's begin our analysis with an example. Suppose you get information about your restaurant's market from a consultant who's kept careful records about what's happened in the airport.

The price you can charge for your meal ranges from $11.00 down to $3.00, and we are going to look at data in $1.00 intervals. So, suppose you charge $11.00 for your meals, or $10.00, $9.00, $8.00, and so forth, all the way down to $3.00. Remember, you are the only restaurant in the airport. So, if people want food, they've got to come to your establishment, and when the price is higher, more people will do without. When the price is lower, you will bring in more customers. Suppose the following numbers are observed. You observe that when you charge a price of $11.00 for the meals that you serve in your restaurant, no body comes in. If the price goes down to $10.00, then you will get typically one customer per hour. If the price is $9.00 per meal, you get typically two customers per hour, and so forth. The numbers of customers increasing each time you reduce the price, and finally, if you charge $3.00 per meal, you'll bring in eight customers per hour.

Now, the question is how do these changes in price and quantity influence the revenue that you earn? Well, if you multiply price times quantity, you get the total revenue that your restaurant is earning. So, let's calculate now the total revenue associated with different prices that you might charge. Here's the total revenue table, which is reproduced, on the board. Eleven times zero is a total revenue of zero. You got a high price, but you're not selling any meals, so your revenue is nothing. If you lower the price to $10.00, the quantity of meals goes to one for a total revenue of $10.00. With $9.00 for your price, you serve two meals an hour for a total revenue of $18.00 per hour, and so forth. You can see that the total revenue is increasing to $24.00, $28.00, and finally, you reach a maximum revenue of $30.00 when you have a price of $6.00 and five meals served. You get the same total revenue here with a price of $5.00 and six meals served per hour. Now, that's the highest your total revenue is ever going to get. If you lower the price further, the number of meals served is not enough, you don't serve enough extra meals to compensate for the lower price, and your total revenue begins to fall.

Now, we'll put in one more column, and this next column is going to be useful because it is going to summarize the effect on your total revenue of a change in the quantity of restaurant meals sold, and this is what we call the marginal revenue. The marginal revenue of the firm's operation is the change in total revenue that results from selling an extra unit of its product. So, as we go from zero to one unit, our total revenue goes from zero to ten and the marginal revenue of that first restaurant meal is $10.00. If you sell a second restaurant meal, in order to do that, you are going to have to lower the price by a dollar. So lower the price by a dollar, sell two meals for total revenue of $18.00, and that gives you a marginal revenue of $18.00 minus $10.00 or $8.00. If you lower the price to $8.00 per meal, you're going to serve three meals an hour, so the total number of sales that you make is increasing as the price falls and the total revenue goes up to $24.00. That's a marginal revenue of $6.00 or $24.00 minus $18.00. You can look through and see that down here, where the total revenue stops changing, our marginal revenue would be recorded as zero. As we move from $6.00 to $5.00 per meal, and the quantity you can sell goes from $5.00 to $6.00, your total revenue remains unchanged. So the marginal revenue is measured at zero. If you try to lower your price further and sell more, total revenue is going to begin to fall. So, marginal revenue is a negative number, you have a reduction in revenue as you increase your sales. Hold on, how can you have less revenue when you're selling more? Because in order to sell those extra units, you have to lower the price, and the lower price hurts you by subtracting money that you could earn on meals that you are already going to sell anyway. Some of these customers are going to come into your restaurant anyway. So to lower the price just to bring in another customer may eventually begin to hurt you, because the customers who were going to come in anyway would have paid the higher price, and that's revenue that you give up in order to attract an additional customer. So, let's put these numbers now over on the board and graph them in a diagram with price and revenue on the vertical axis and quantity on the horizontal axis. When the price of the product is $11.00, you sell no restaurant meals, so you get a point on the graph that looks like this: a price of $11.00 and a quantity of zero. If the price drops to $10.00 per meal, you sell one meal per hour, and a price of $9.00
Other Market Models

**Defining Marginal Revenue for a Firm with Market Power**

per meal, you sell two meals per hour, and a price of $8.00, three meals, and you can continue to fill in the chart like this. What we are doing is writing down the demand curve for your restaurant. Whenever the price is $5.00 per meal, you sell six meals per hour, $4.00 per meal, seven, and so forth. So, eventually we wind up with a demand curve that shows the quantity of meals that people in your market are willing and able to buy at different prices. We can now connect these dots and produce a demand curve, and I’ll do that in just a moment.

But first, let’s write in the marginal revenue curve. The marginal revenue curve shows you the amount of money you earn by selling an extra unit of the product. So for instance, the marginal revenue from the first unit is $10.00, so we can put a dot here, indicating that when we sell that first restaurant meal, we add $10.00 to our total revenue. When we sell the second restaurant meal, our marginal revenue is going to be only $8.00, so notice here that marginal revenue is less than the price at which we are selling the second meal. The third meal has a marginal revenue of only $6.00, the fourth meal has a marginal revenue of only $4.00, the fifth meal has a marginal revenue of $2.00, and the sixth meal has a marginal revenue of zero dollars. That is, we have the same total revenue whether we’re selling five meals at $6.00 a piece or six meals at $5.00 apiece, so the marginal revenue from the sixth meal is zero. Now, things get interesting because if we sell an additional meal, our marginal revenue is going to go below the horizontal axis, that is, we now have negative marginal revenue on the seventh and eighth meals. So those dots will be down here below the horizontal axis. So, if I connect a couple dots, I’ll have my marginal revenue curve. Marginal revenue telling me to change in total revenue that results in selling that marginal meal, whether it’s the first, second, third, fourth, and so on.

Now, let me go ahead and draw in the demand curve by connecting the red dots and the demand curve tells me the price at which I can sell any given quantity of restaurant meals. The demand curve is also what we call the average revenue, the average revenue is total revenue divided by the number of units of the product that you sell, and the average revenue will always be equal to the price at which you’re selling the product. So, the definition of average revenue is total revenue divided by the quantity of restaurant meals that you’re serving, and that will always be equal to the price of the product. So, for example, if we’re selling five meals at $6.00 apiece for total revenue of $30.00, $30.00 divided by six meals if $5.00, the price at which we’re selling each meal. So the demand curve represents the average revenue.

Now, the marginal revenue curve lies below the demand curve, and if I connect the purple dots, I’ll get the marginal revenue curve. The marginal revenue curve always lies below the demand curve, that is, for any given quantity that we sell, the marginal revenue from selling that last unit will always be less than the price at which we sold that unit. Now, this requires a little bit of explanation. Why is it that marginal revenue is always less than price? And the answer is this: marginal revenue is always less than price, because in order to sell that last unit, in order to sell that extra unit, that marginal unit, the firm has to lower its price; thereby losing revenue that it could have earned on the unit it was going to sell anyway. That’s the key. Marginal revenue for a firm with market power, marginal revenue is always less than price, because in order to sell the marginal unit, the firm has to lower its price to attract customers, and lowering its price, it’s giving away money or losing money that it could’ve earned on the customers that it was going to attract anyway.

So, we can summarize then by saying the demand curve is a downward sloping line for a firm with market power. The firm can choose between a higher price and fewer customers, or a lower price and more units sold. That’s the choice or the problem that the monopolist faces. That’s the problem that you face whenever you’re the owner of the only restaurant in the airport, and marginal revenue will always be less than price, because in order to sell an extra unit, you have to lower the price you’re charging, and when you lower the price, you’re giving away revenue that you could’ve earned by keeping your price high and serving fewer customers. Next, we’ll look at how the monopolist takes into account marginal revenue and price whenever it sets the price at which it maximizes profits.
Determining the Monopolist’s Profit-Maximizing Output and Price

We’ve been talking about the special problem that faces a firm that has market power. The problem is this: if you charge a higher price, you’ll sell fewer units, but if you want to sell more units, you have to lower the price. It’s a kind of dilemma. So how do you balance these competing forces? Now we’re ready to answer the question, “How does a profit-maximizing monopolist choose the price at which it sells its product and the quantity that it sells?” We’re now ready to put together what we’ve discussed about the marginal revenue of the monopolist together with the cost curves that we analyzed in earlier lectures to identify the point at which the monopoly maximizes profits.

Let’s start by a quick review of marginal review. If we draw the demand curve that a monopolist faces, the downward sloping demand curve, we’re showing that as the price of the good rises, the monopolist will be able to attract fewer customers. If the monopolist charges a lower price, he or she will be able to sell more units. So imagine again we’re in the airport and you are running your restaurant. If you charge a higher price for your meals, you’ll attract fewer customers. But if you want to have your restaurant full, you’ll have to charge a lower price for all of the meals that you serve. This dilemma is summarized in the marginal revenue curve. The marginal revenue curve tells you how much revenue you add to your total by selling an extra unit of output. Remember that marginal revenue is always less than the price you’re charging for the good; that is, the marginal revenue curve always lies below the demand curve. The reason is this: in order to sell an extra unit of output, you have to lower the price that you would have charged on all of the units that you could have sold at a higher price. If you’re selling 10 units at a price of $1.00 a piece and you want to sell and eleventh unit, you’ve got to lower your price, perhaps down to 0.80 per unit. That means 0.20 that you’re giving up on the 10 units you could have sold at the price of $1.00. This is why marginal revenue is always less than the price of the product, and why the marginal revenue curve always lies below the demand curve. So let me go ahead and label the curve that I’ve drawn here in purple MR for marginal revenue.

Now, let’s suppose that we are a monopolist and we produce a product that has no cost. Suppose we have spring water that bubbles freely out of the ground, and that there is no cost at all for us to sell this to customers who bring their own buckets and scoop it up. What price would be appropriate to charge for a good that is freely produced? Think for a moment. If we were in a perfectly competitive market and water bubbled freely out of the ground, what would the price be for a bucket of water? The answer is the price would be zero. If the water is truly free, then no one could charge a price for the water that’s above zero, or they would be undercut by competition. Someone else would go uncork a spring somewhere else and offer the water at a lower price, until finally, in equilibrium, the price of the water would fall to be equal to its marginal cost of production, in this case, by assumption equal to zero.

So in a perfectly competitive market, a good that is freely available will sell at a price of zero. However, if we were in a monopolized market – suppose you’re the only person who has a spring, you’re the only person who can provide water to people and people cannot look elsewhere for substitutes. Then you have market power and, in the case of market power, you can maximize your profits not by charging a price equal to marginal cost, but rather by charging a price equal to what the market will bear. Now there’s a phrase for you – the price that the market will bear. What are people willing to pay for a bucket of water?

Well, consider that we have this diagram working and if you are looking to maximize your profits, you’re looking for an area that is going to give you the maximum total revenue, price × quantity. Pick a point on the demand curve, any point that you like, and if you pick a point on the demand curve, let’s say this one right here, with a price equal to P₀ and a quantity equal to Q₀ - and here, for a moment, I’m considering only the red curve. I’m imagining that the purple curve doesn’t even exist. Price × quantity, the area of this black rectangle, is the total revenue that you earn by selling your water at a price of P₀.

Now, in this case, the total revenue is not maximized, reason being that if you lowered the price and attracted more customers, you would actually increase the size of this rectangle. You would maximize your profits not by charging price of P₀, but by charging whatever price makes this rectangle the largest. And it so happens that that rectangle is the largest at the point where the marginal revenue curve cuts through the horizontal axis. At this point right here, where the purple curve cuts through the horizontal axis at a price of P_max and down here, we can call this point of intersection Q_max at this particular point with a price of P_max and a quantity of Q_max; this rectangle is the largest that it can possibly get. Any firm maximizes its revenue by pricing at the point where marginal revenue equals zero. As long as you can continue to sell buckets of water and add to your total revenue, keep selling them. Keep lowering the price and adding to your sales until finally you have made your revenue as large as it can possible get. The rule for maximizing your revenue is to price the product where marginal revenue is equal to zero. At that point, you’ve made your particular chunk of revenue as big as it can get.
Determining the Monopolist’s Profit-Maximizing Output and Price

Now, you don’t necessarily want to maximize your revenue as a firm. You may, in fact, want to maximize your profits. But in our particular case of the water that bubbles freely from the ground, revenue and profit are the same. That’s because there are no costs.

Let’s look now at the problem from a different angle. Suppose that instead of simply wanting to maximize your revenue, you want to maximize profit when, in fact, production is costly. Well, that’s going to require a different diagram. If your product has a cost to production, you need to put the cost curves in the diagram, along with the demand curve and marginal revenue curve. Let’s go ahead and put in this picture the marginal cost curve. The marginal cost curve, as usual, is an upward sloping curve that tells us how much it costs our firm to produce an extra unit of output.

Now, what will the rule be for maximizing profit for a monopolist? Well remember the definition of profit is this: profit, and I’ll use here the Greek letter \( \pi \) for profit, is equal to total revenue minus total cost. So we want to keep increasing output until the change in profit is zero. We want to keep increasing our output until our marginal profit, the amount of profit we make from selling one additional unit, has dropped to zero. And, at that point, we can’t push our profits any further. So if we keep producing output until the change in profit that results from a change in output is equal to zero, that’s going to require that we keep producing output until the change in total revenue that results from increasing output minus the change in total cost that results from adding output is equal to zero. And over on the board I will summarize this equation by plugging in the definitions that we’ve introduced earlier. The change in total revenue that results from a change in output has a name. What is it? It’s called the marginal revenue, what’s represented here by the purple curve. The change in total cost that results from the change in output also has a name. What is it? It’s the marginal cost. So here we have marginal revenue minus marginal cost equaled to zero; that is, when we’re at the point of maximizing profits, marginal revenue minus marginal cost equals zero. Or, to put it another way, marginal revenue equals marginal cost. The extra revenue that we add from producing one more restaurant meal is exactly equal to the cost that it adds to our production. At the margin, we’ve reached the point to where the extra revenue is equal to the extra cost. And, at that point, we should stop if we want to maximize profit.

Now, you can see that point here in the diagram. The monopolist is always doing well to increase output as long as producing an extra restaurant meal adds more to his or her revenue than it adds to his or her costs. Think about your restaurant in the airport now. If you want to sell more output, you’ve got to lower the price. And lowering the price means that you’re cutting in the revenue you could have earned on other units. So marginal revenue is less than the price you’re charging for your meals. As you lower the price of your product, you sell more output, but marginal revenue keeps falling as you keep lowering the price to bring in additional customers. Marginal revenue is still positive, as long as the purple curve is above the zero line. But eventually marginal revenue falls below marginal cost. After you’re selling so much output, lowering the price to sell an additional unit makes it unprofitable, because selling that extra unit adds more to your costs than it adds to your revenue. Back here for these units, extra sales add more to your revenue than they add to your cost. Say this is the tenth restaurant meal that you serve, when the prices are down to $5.00 a meal. Well, the marginal revenue on that unit might be, say, $3.00, and as long as the cost of that unit is less than $3.00, maybe $1.00, then that unit is adding profits to your restaurant. But eventually you’ll get to the point to where the marginal revenue and the marginal cost equal. At that point, if you go beyond and try to sell more units by lowering the price, you’re actually cutting into your profits, because the marginal cost will be greater than the revenue that its adds to your coffer.

Therefore, the rule for maximizing profit for a monopolist is to find the intersection of marginal revenue and marginal cost. Keep producing output until the revenue added by an additional unit is exactly equal to the cost of that additional unit. You’re still making a lot of profit maybe. You’re still producing all of these restaurant meals that are adding more revenue than cost, but that’s within the margin. At the margin, that last restaurant meal that you serve is adding zero to your profits, and that’s when you know that it’s time to quit.

So let’s look then at how many restaurant meals you’ll actually be serving. Find the point where the two curves cross and then drop down here. And we can call this point down here \( Q^*M \). This is the monopolist’s profit-maximizing quantity of restaurant meals, maybe 30. Now, what price are you going to be charging for those meals? Well, to find the price that you can charge, you’ll want to go up to the demand curve. The demand curve tells you the price that the market will pay, or the price that the market will bear when you’re willing to sell those 30 restaurant meals. So if we go up here, we find the price at which we can sell those meals. That maybe something like $8.00 a meal. And over here we have \( P^*M \), the profit-maximizing price for the monopolist to charge.
Other Market Models

Monopolies

**Determining the Monopolist’s Profit-Maximizing Output and Price**

So here’s the summary. The question we’ve tried to answer is, “What is the profit-maximizing output and price for a monopolist”, say your restaurant at the airport. The answer is keep producing output, keep adding to your sales, keep lowering the price and increasing the number of customers you serve until the marginal revenue is equal to the marginal cost. At that point, your profit is as big as you can get it. When you find that point, it tells you the quantity that you’ll want to produce as a monopolist and, at that given quantity, go up to the demand curve and over to find the price that you’ll be charging at that point. Notice that the price you’re charging is greater than the marginal cost. It’s greater than the marginal revenue. That’s because for any firm with market power marginal revenue is always less than price, because you had to lower the price to get to that point, which means you had to give up money you could have earned on restaurant meals that you would have sold at a higher price. Marginal revenue is less than price, but equal to marginal cost when a monopolist is maximizing its profits.

So next we’ll take another look at the diagram and see whether you’re making a positive economic profit or not.
Calculating a Monopolist's Profit and Loss

We've been talking about the special problem that faces a firm that has market power. The problem is this: if you charge a higher price, you'll sell fewer units, but if you want to sell more units, you have to lower the price. It's a kind of dilemma. So, how do you balance these competing forces? Now we're ready to answer the question. How does a profit maximizing the monopolist choose the price at which it sells its product and the quantity that it sells? We're now ready to put together what we've discussed about the marginal revenue, of the monopolist together with the cost curves that we analyzed in earlier lectures to identify the point at which the monopoly maximizes profits.

Let's start by a quick review of marginal revenue. If we draw the demand curve that a monopolist faces, the downward sloping demand curve, were showing that as the price of the good rises, the monopolist will be able to attract fewer customers. If the monopolist charges a lower price, he or she will be able to sell more units. So imagine again, we're in the airport and you're running your restaurant. If you charge a higher price for your meals, you'll attract fewer customers, but if you want to have your restaurant full, you'll have to charge a lower price for all of the meals that you serve. This dilemma is summarized in the marginal revenue curve. The marginal revenue curve tells you how much revenue you add to your total by selling an extra unit of output, and remember, marginal revenue is always less than the price you're charging for the good. That is, the marginal revenue curve always lies below the demand curve. The reason is this, in order to sell an extra unit of output, you have to lower the price that you would have charged on all of the units that you could have sold at a higher price. If you are selling ten units at a price of $1.00 apiece and you want to sell an eleventh unit, you've got to lower your price. Perhaps down to 80 cents per unit, that means 20 cents that you are giving up on the ten units you could have sold at the price of $1.00. This is why marginal revenue is always less than the price of the product, and why the marginal revenue curve always lies below the demand curve. So, let me go ahead and label the curve that I've drawn here in purple, MR, for marginal revenue.

Now, let's suppose that we're a monopolist, and we produce a product that has no cost. Suppose we have spring water that bubbles freely out of the ground and that there's no cost at all for us to sell this to customers who bring their own buckets and scoop it up. What price would be appropriate to charge for a good that is freely produced? Think for a moment, if we were in a perfectly competitive market and water bubbles freely out of the ground, what would the price be for a bucket of water? The answer is, the price would be zero. If the water is truly free, then no one could charge a price for the water that's above zero or they would be undercut by competition. Someone else would go uncork a spring somewhere else and offer the water at a lower price, until finally, in equilibrium, the price of the water would fall to be equal to its marginal cost of production. In this case, by assumption equal to zero. So, in a perfectly competitive market, a good that is freely available will sell at a price of zero. However, if we are in a monopolized market, suppose you are the only person who has a spring, you're the only person who can provide water to people and people cannot look elsewhere for substitutes, then you have market power, and in the case of market power, you can maximize your profits, not by charging a price equal to marginal cost, but rather by charging a price equal to what the market will bear. Now, there's a phrase for you, the price that the market will bear. What are people willing to pay for a bucket of water?

We'll consider that we have this diagram working and if you are looking to maximize your profits, you're looking for an area that is going to give you the maximum total revenue, price times quantity. Pick a point on the demand curve, any point that you like, and if you pick a point on the demand curve, let's say this one right here with a price equal to $P_0$, and a quantity equal to $Q_0$; and here for a moment, I'm considering only the red curve. I'm imagining that the purple curve doesn't even exist. Price times quantity, the area of this black rectangle is the total revenue that you earn by selling your water at a price of $P_0$. Now in this case, the total revenue is not maximized; reason being, that if you lower the price that attracted more customers, you would actually increase the size of this rectangle. That is, you would maximize your profit, not by charging price of $Q_0$, but by charging whatever price makes this rectangle the largest, and it so happens that the rectangle is the largest at the point where the marginal revenue curve cuts through the horizontal axis. At this point right here, where the purple curve cuts through the horizontal axis at a price of, we'll call this P-max, and down here we can call this point of intersection Q-max, at this particular point with the price of P-max and the quantity of Q-max, this rectangle is the largest that it can possibly get. Any firm maximizes its revenue by pricing at the point where marginal revenue equals zero. As long as you can continue to sell buckets of water and add to your total revenue, keep selling them, keep lowering the price and adding to your sales until finally, you've made your revenue as large as it can possibly get. The rule for maximizing your revenue is pricing the product where marginal revenue is equal to zero. At that point, you've made your particular chunk of revenue as big as it can get.

Now, you don't have to necessarily want to maximize your revenue as a firm, you may in fact, want to maximize your profits, but in our particular case of the water that bubbles freely from the ground, revenue and profit are the same,
Calculating a Monopolist’s Profit and Loss

that’s because there are no costs. Let’s look now at the problem from a different angle. Suppose that instead of simply wanting to maximize your revenue, you want to maximize profit. When in fact, production is costly. Well, that’s going to require a different diagram. If your product has a cost of production, you need to put the cost curves in the diagram along with the demand curve and marginal revenue curve. Let’s go ahead and put in this picture, the marginal cost curve, marginal cost curve as usual, is an upward slope in curve that tells us how much it costs our firm to produce an extra unit of output. Now, what will the rule be for maximizing profit for a monopolist? Well remember, the definition of profit is this: profit, and I’ll use here the Greek letter π, for profit, is equal to total revenue minus total cost. So, we want to keep increasing output until the change in profit is zero. We want to keep increasing output until our marginal profit, the amount of profit we make from selling one addition unit, has dropped to zero, and at that point, we can’t push our profits any further. So, if we keep producing output until the change in profit, that results from a change in output, is equal to zero, that’s going to require that we keep producing output until the change in total revenue, that results from increasing output, minus the change in total cost, that results from adding output, is equal to zero.

And over on the board, I will summarize this equation by plugging in the definitions that we’ve introduced earlier. The change in total revenue that results from a change in output has a name, what is it? It’s called the marginal revenue that’s represented here by the purple curve. The change in total cost that results from the change in output also has a name, what is it? It’s the marginal cost, so here we have marginal revenue minus marginal cost equaled to zero. That is when we’re at the point of maximizing profits, marginal revenue minus marginal cost equals zero, or to put it another way, marginal revenue equals marginal cost. The extra revenue that we add from producing one more restaurant meal is exactly equal to the cost that it adds to our production. At the margin, we’ve reached the point to where the extra revenue is equal to the extra cost.
Monopolies

Other Market Models

**Graphing the Relationship between Marginal Revenue and Elasticity**

In this lesson, we’re going to take a behind-the-scenes look at marginal revenue, and one of the things that we’ll discover is the relationship between marginal revenue and a familiar concept, the elasticity of demand.

Let’s start with our downward-sloping demand curve. Any firm that has market power is dealing with a downward-sloping demand curve. That is, they can either sell a larger quantity at a lower price or a smaller quantity at a higher price. A downward-sloping demand curve is what it means to have market power. If you as an individual firm, if you as a restaurant operator at the airport, have to be concerned about the quantity that you sell or you might depress your price, you’ve got market power and you’re dealing with a downward-sloping demand curve. That means you have a special problem. That is, do you want higher prices and fewer customers or lower prices and more customers?

Let’s look at how you sort that decision out. Let’s start with a particular price quantity combination on our demand curve. Let’s suppose that the price is initially a high price – $c_0$, maybe $10 per meal – and the quantity is initially a small quantity – we’ll call this quantity $q_0$ – maybe it’s twenty meals a day served. If the price is $p_0$ and the quantity is $q_0$, what will your total revenue be? Can you identify total revenue in this picture? The answer is: Total revenue is always price times quantity – it is the area of this rectangle.

What we’re concerned with is how total revenue changes when the price and quantity change along the demand curve. Suppose you decide that you want to serve more customers at a lower price. If so, you may drop your price down to $p_1$, and the quantity that you are going to be selling will then increase accordingly. So let me put the new quantity in here. It will be $q_1$ at a price of $p_1$. So the total revenue that you will be earning with the new lower price and larger quantity is the area of this rectangle here – below $p_1$ and up to $q_1$. The change in price results in a change in quantity. A lower price means more customers served, and the total revenue is different in the two cases.

Let’s look at how total revenue has changed, first geometrically and then intuitively. You'll notice that our two total revenue areas have in common this box in the corner, but our original situation involved this additional rectangle on top, which we lose when we drop the price. The new total revenue box adds this revenue in addition. We’re going to call this little rectangle right here Area One. Area One is added when we lower the price and the quantity increases. The intuitive explanation of Area One is that when the price is lower, the firm is going to add new sales or new customers. Area One represents the additional revenue that comes when the firm lowers its price and adds additional sales. Area One is an increase in total revenue that results from a lower price and larger quantity. So let me write over here on the side the formula for Area One. So Area One, which I’ll write as a rectangle one, this Area One is going to be equal to the new quantity minus the old quantity, or the change in quantity, multiplied by the new price. So Area One = $P_1 \times (q_1 - q_0)$.

The intuitive explanation for Area One is the additional revenue that comes from new sales at this new lower price. However, in order to get those additional customers, you had to reduce the price, and that resulted in a loss of revenue. This little box up here was lost when the firm went from $p_0$ to $p_1$. Let’s call this Area Two. Rectangle Two, or Area Two, then, is the lost revenue that is associated with lowering the price. Here’s the intuition: See, you could have sold this many restaurant meals – say ten meals a day – you could have sold them at the higher price; but when you lowered the price, you gave up revenue that you would have earned on those meals anyway. You’re now selling the same meals at a lower price, and that’s a reduction in your revenue. Even though you added all of these new sales, your old sales now are less lucrative for you because you’ve cut the price from say $10 per meal down to $8 per meal. You gave up $2 on each of those meals that you would have sold anyway at the higher price. Area Two is the reduction in total revenue associated with reducing the price at which you’re selling units that you would have sold anyway. So this area is going to be Area $P_0$ minus $P_1$ multiplied by the quantity – $q_0$ – that you would have sold at the higher price – the lost margin – the lost price increment – on all of these meals you would have sold anyway.

The marginal revenue that you earn whenever you drop your price from $p_0$ to $p_1$, and your quantity increases from $q_0$ to $q_1$, the marginal revenue is going to be the combination of Area One and Area Two. When the price falls, you gain Area One but you lose Area Two. Subtract Area Two from Area One, and you have the firm’s marginal revenue.

Well, this raises a question, and that is is marginal revenue positive or negative? When you lower your price and increase your sales, are you adding to your total revenue or are you subtracting from it? The answer depends on the relative size of Area One and Area Two. If Area One is bigger, then your marginal revenue is greater than zero. If Area Two was bigger, then marginal revenue is negative. Look, what’s going on here? When you lower your price, you’re always going to be losing money on the units you could have sold at the higher price. The question is are you
Monopolies

Graphing the Relationship between Marginal Revenue and Elasticity

adding enough sales to compensate for that lost revenue? Are you adding enough sales so that Area One – the increment that comes from increasing your customer base – are you adding enough sales so that Area One, the extra revenue from adding more units sold, offsets the loss from giving up money that you would have earned on these units that you would have sold anyway?

Well, what are we talking about here? We're talking about how many extra sales you make when you lower your price a little bit. Does this concept sound familiar? We have a name for it. Do you remember what it is? The answer is we're talking about the elasticity of demand. Elasticity of demand refers to the responsiveness of quantity demanded to a change in the price of a product. If your demand is very elastic, a small change in price results in a big increase in your sales, and Area One is greater than Area Two. If, on the other hand, demand is inelastic, then the given change in price will add only a small increase in quantity demanded, and Area One will be smaller than Area Two. This is what happens when demand is inelastic. So if Area One is bigger than Area Two, then demand is elastic and a decrease in price results in an increase in total revenue. When demand is elastic, marginal revenue is positive. However, if demand is inelastic, Area One is smaller than Area Two, and in that case marginal revenue is negative. Selling extra units shrinks your total revenue.

What I'll do now is show you the relationship between Areas One and Two and the concept of elasticity. It will involve a few lines of math, but I think the payoff makes it worth it. Here I've reproduced the formulas for Area One and Area Two. Area One is the price times the change in quantity. Area Two is the change in price times the original quantity.

What I want to do is convert this into a familiar measure, elasticity of demand, but in order to do that, I'm going to have to use a couple of math tricks. The first one is I want to talk about small changes in price and quantity. In the picture that we drew earlier, we had a large change in price and quantity just to make the diagram easier to read. But if we have a small change in price and quantity, we could write the difference between \( q_1 \) and \( q_0 \) as \( \Delta q \). This difference – \( p_0 - p_1 \) – we can write that as \( \Delta q \). The interesting thing about a small change is that the subscripts are no longer that important. We don't have a big difference like $10 for Price Zero and $8 for Price One. Instead, we have very small differences – $10 versus $9.99. So when the prices are that close, we don't have to be meticulous about keeping the subscripts, we can let them go because \( p_0 \) and \( p_1 \) are pretty close.

The next thing to notice is that whenever we write our deltas, we want to be careful that we have the new price minus the old one, or the new quantity minus the old one. Here I have the old price minus the new price, so when I write this as a change in price, I have to put a negative sign out in front of it to indicate that it's going the opposite direction from the change in quantity. Keeping these two things in mind, if you're careful in measuring the direction of your change and that the change is very, very small, we can write the difference between \( q_1 \) and \( q_0 \) as \( \Delta q \). The interesting thing about a small change is that the subscripts are no longer that important. We don't have a big difference like $10 for Price Zero and $8 for Price One. Instead, we have very small differences – $10 versus $9.99. So when the prices are that close, we don't have to be meticulous about keeping the subscripts, we can let them go because \( p_0 \) and \( p_1 \) are pretty close.

Now let's see if we can simplify this expression. The first thing that I will do is I'll break the fraction into two parts and put the price terms in their own fraction multiplied by the quantity terms. So no big change yet, I just broke the fraction kind of into two separate parts. The next thing that I'm going to do is I'm going to use this price term, and instead of taking it this way, I'm going to write it as its own reciprocal. I'm going to write it as \( \frac{1}{\Delta q} \) of this fraction. So what I get next is \( \frac{1}{\Delta q} \times \frac{\Delta q}{q} \). So all I did between this step and this step was flip this fraction over and put it in the reciprocal form.
Other Market Models

Monopolies

**Graphing the Relationship between Marginal Revenue and Elasticity**

\[ \frac{\Delta q}{q} \frac{1}{\Delta p} \]

denominator of another fraction. I took the reciprocal of it. Well, now what I have is this: I have \( \frac{\Delta q}{q} \) – that is, the percentage change in quantity demanded divided by the percentage change in price. One more thing – I’ve still got a negative sign that I have to put out in front.

Well, what have I got? What I’ve got is the percentage change in quantity divided by the percentage in price; and since these two usually move in opposite directions, with a negative sign out in front, I’ve got the absolute value of that fraction. Ah, hah! That’s familiar. We now have a familiar measure of price responsiveness. This term is exactly the elasticity of demand. So if Area One is greater than Area Two, then the value of this fraction is greater than one. That is, the quantity changes a lot in response to a given change in price, and that’s what we call elastic demand. If Area One is smaller than Area Two, then this fraction is less than one, and that’s what we call inelastic demand. If Area One is equal to Area Two, this fraction is equal to one and we have unit elastic demand.

So, if a firm lowers the price if its product and it sells a lot of the goods, if Area One is greater than Area Two, that’s because customers are relatively responsive to a change in price. Demand is elastic, and the marginal revenue is positive. That is, a reduction in price and an increase in quantity increase total revenue. If, on the other hand, a firm lowers its price and customers don’t respond, Area One is small relative to Area Two. This elasticity measure is less than one, the demand is inelastic and marginal revenue is negative. Anytime demand is inelastic, marginal revenue is negative. Anytime demand is elastic, marginal revenue is positive.

This goes back to a lesson that we discussed earlier, and that is that an increase in quantity sold in a particular market will result in an increase in total revenue if the demand for the product is elastic. On the other hand, if the demand is inelastic, then a restriction of quantity and a raising of the price will increase the total revenue. I’ve gone to the trouble of all of this math just to make very clear to you that marginal revenue is intimately related to the concept of elasticity of demand.
Other Market Models

The Social Cost of Monopoly

Determining the Social Cost of Monopoly

Let's suppose that you are running a restaurant at the airport. The only one, and you are a monopolist. Should you expect to be unpopular? Will you be surprised that the people in the airport complain about the high prices of your meals. Or are you by surveying the customers, going to be making friends? Here's the question, why are monopolists so unpopular? Perhaps it has something to do with social cost of monopoly. That by increasing their own profits, monopolists are actually imposing costs on society, and thereby shrinking the overall pie of economic value. That's what we are going to be looking at in this lesson, the social cost of monopoly.

But first, let me answer a question that some of you are probably asking. We've got a demand curve here and a marginal revenue curve in this diagram, but where's the supply curve? Where's the supply curve that we're used to from our standard supply and demand diagrams? The answer is this, there is no supply curve for a monopolist. The reason is because of the definition of the supply curve. Remember, supply curve tells us the quantity of output that suppliers are willing and able to sell as the price varies in the market. That is, anytime we draw a supply curve, we are assuming passive response on the part of sellers. We're assuming that sellers take the price as given and passively respond by choosing the quantity they want to sell. Now that's a great assumption for a competitive market; but in the case of monopoly, price is not taken as given. Monopolists are not price takers, they are price setters. And because monopolies set prices, rather than take them as given, there is no supply curve for a monopolist. A monopolist has a marginal cost curve like any firm, but because the monopolist gets to choose to produce the output where marginal costs equals marginal revenue, there is no passive response to price. The monopolists sets price, therefore, monopolists have no supply curves. They have marginal revenue curves and marginal cost curves, but the supply decision for a monopolist is more complex than it is for a competitive firm. Therefore, one last time, there is no supply curve for a monopolist.

Now suppose this monopolist were forced to behave competitively. Suppose this were a competitive market where maximizing profit meant that you had to equate marginal cost to the price of the product. That is, if firms could freely enter this market and could compete with one another on price and drive the price down, firms would continue to enter and keep producing output until the price fell to the point in which it was equal to marginal cost. That is, the point in this diagram that we would get if firms behaved competitively would be the point where the marginal cost curve intersects the demand curve. That's the point in which the price would be equal to marginal cost and no further entry would occur. Firms wouldn't want to produce any more of the product. Well, what we would get in that case is price equals marginal cost, so I'll put that in this diagram. We'll call this the competitive price; the price in competitive equilibrium and the quantity that the firm would produce or that all firms together would produce in a competitive market would be Qc. So, in this case, price equals marginal cost when the competitive quantity becomes a kind of benchmark for us. Why are we so interested in the price that would occur in a competitive market and the quantity that would be produced? Why is this benchmark interesting? The answer is this, you recall that economists like competitive markets because competitive markets under certain circumstances maximize economic value. Look at this picture, the red curve, the demand curve, is a list of the reservation prices of the different customers in our market. And as long as there are no wealth effects and no externalities, this demand curve represents social benefits. Each point on the demand curve tells us about the benefit that some customer derives from getting a unit of the product. So think about your restaurant for a moment. If you had the demand curve for meals at your restaurant, what you have is a record of the reservation prices of all of the customers that might choose to come into your store and buy a meal. A meal at your restaurant is worth $11.00 to someone, $10.50 to someone else, $9.50 to another customer, $8.50, $7.00 and so forth on down the demand curve. The demand curve summarizes the benefit that customers in the market derive from enjoying your product. The marginal cost curve, if there are no wealth effects and no externalities, the marginal cost curve tells us about the cost to society of providing another meal at your restaurant. The marginal cost of the first meal might be a dollar, the second meal might have a marginal cost of $1.75, and so forth. Just to jog your memory, you'll recall that economic value is created anytime we produce and trade another unit of the product, for which the benefit is greater than the cost. Anytime the marginal social benefit is greater than the marginal social cost, we are adding economic value when we increase production. In this case, in the case of your restaurant, the competitive price and competitive quantity tell us about that amount of restaurant service that maximizes economic value. All of the meals from zero up until this point, perhaps 40 meals served a day at your restaurant, give us benefits greater than costs. That is, you can look at the red line and the green line and see that this area at the economic value that can be created, the maximum economic value that could be created at your restaurant.

Now here's the question, does that particular choice of price and quantity maximize your profits as a monopoly restaurant owner? And the answer, unfortunately for society, is it does not. If you look here, you find that profit is
Other Market Models

The Social Cost of Monopoly

Determining the Social Cost of Monopoly

maximized where marginal revenue equals marginal cost. We’ve discussed this before. The monopolist maximizes his or her profits by choosing this particular combination. You maximize your profits at your airport restaurant by restricting the quantity of meals that you serve. Instead of serving 40 restaurant meals a day, you may serve only 30 choosing the monopoly quantity and charging this higher price, the higher price on the demand curve, the price that the market will bear. Now whether you have this price in quantity combination when you restrict output and raise the price, you are imposing a cost on society. Let’s see that cost. Look at all of these units that are not being produced and sold. Look at all of these restaurant meals between Q_m and Q_c that you have decided not to offer. In order to maximize your restaurant profits, you have restricted output below the competitive level and you have raised the price above the competitive level. The reason this is bad for society is as follows.

All of these restaurant meals that you do not produce in trade have a higher benefit to society than they cost to produce. There are people in here who are willing to pay more for a restaurant meal than it costs you to produce it, but you are not serving those customers. You are not serving those customers because by raising your price, you are able to charge more to the customers who stick with you than you would earn by serving those additional customers. By restricting your output and charging a higher price, you maximize your profits. You put more money in your pocket. In fact, the money you put in your pocket can be seen in this diagram. By raising the price, from the competitive P_c to the monopoly profit maximizing P_m, you put this additional amount of money into your pocket. All of the customers that you continue to serve, you’re making extra revenue off of those customers by charging them a higher price. What you lose on revenue is the extra revenue that you would have earned by serving these customers here between Q_m and Q_c. By knocking those customers out of the market out of the price range, you no longer get the revenue from them. However, the cost of serving those customers was higher than the cost of serving these other customers, because of the increasing marginal cost. So, the profits that you lose on these customers is small compared to the extra revenue that you earn by charging a higher price to customers who are willing to pay more. Now, here’s the problem.

The problem, from society’s point of view, is all of these people who are not being served, all of these people whose benefits exceed the cost of serving them and they are simply not getting service because the monopolist does not find it privately profitable to serve them. In order to maximize his or her own profits, the monopolist raises the price and charges more to the customers who will pay more, thereby pushing out of the market these marginal customers who are willing to pay enough to cover the cost of their meals but not willing to pay enough to continue to provide profits to the monopolist, to make it worth his or her while to serve them. If I wanted to identify the loss to society from having a monopoly instead of having a competitive market, I would identify the costs of all of these customers here who don’t get service. I’m shading in an area showing you the loss that occurs from society’s point of view when output is restricted from 40 meals to 30 meals a day and the loss is as follows: all of the benefits that the customers would earn, less the cost of serving them. Benefit minus cost is economic value and here’s economic value lost when trade is restricted. We call that a deadweight loss.

This is the deadweight loss associated with monopoly. When the monopolist restricts trade and raises his or her profits, the monopolist creates a deadweight loss relative to a competitive market. A competitive market would serve more customers; in fact, it would serve all of those customers for whom the benefit is as great as the cost, all of those customers that add economic value for society. But the monopolist finds it privately profitable to restrict trade, charge a higher price, push some customers out of the market, and impose a deadweight loss on society. This is why monopoly is viewed as bad for society. Monopoly is viewed as bad for society because it usually, in fact, almost always entails a restriction of trade relative to competition. And the cost of restricted trade is that some people who benefit more than the cost of providing the good, do not get service. Now, in the next lesson we’ll consider why we allow monopolists to hang around even when they are imposing a deadweight loss on society.
Other Market Models

The Social Cost of Monopoly

Calculating Deadweight Loss

In this lesson, I’ll give a simple introduction to the cost of monopoly. I want to go back to the supply and the demand curves that we talked about earlier when we discussed normative economics and the concept of economic value and we are going to look at them now in the context of another different story. Let’s suppose that you have a restaurant at the airport and you are the only provider of services, food service at the airport. You have a monopoly, and we would like to compare what happens, from society’s perspective, in the case of monopolized food service, versus the case of competitive food service. You running a monopoly restaurant, and how much economic value are you creating as you try to maximize your profits. Let’s look then at some numbers that represent the cost and benefits in this situation.

The blue dots along the bottom of this diagram represent marginal costs, the costs to you and to society of serving an extra customer in the restaurant. Here we have $3.00 for one meal, $3.00 for the second meal for a total of $6.00 to serve two people, $3.00 again for the third meal, and so forth. Each of these dots represents $3.00 worth of cost to putting in another meal on the table. I’m assuming for the sake of my example here that you have constant marginal cost in the food industry. That is, every customer you want to serve in the restaurant adds as much cost to production as the customer before, constant marginal cost. We will also assume that there are no externalities and no wealth effects so that these marginal costs that are faced by your firm or by any other firm are equal to the marginal social costs of providing food service. So, here we rule out the possibility of externalities and wealth effects so that private costs are equal to social costs. And we are making the story very simple by assuming that the marginal cost of production is constant.

The red dots have represented the demand curve in our previous examples and they also represent the reservation prices of buyers in the market. If we assume that there are no externalities and no wealth effects, then these red dots also represent the marginal social benefit of serving an extra customer in the restaurant. So for example, someone has a reservation price of $9.00. Whenever we price our product at $9.00 per meal, we serve one customer. If we want to serve a second customer, we have to lower the price down to her reservation price of $7.00 per meal and $7.00 per meal will attract two customers and so forth. So the red dots represent the reservation prices of customers and with no wealth effects and no externalities, they also represent the marginal social benefit of serving an extra customer in the restaurant.

Let’s start with a story that we’re familiar with. Suppose food service in the airport is provided competitively. Suppose there are a lot of providers that are competing for business. In a competitive market, what will the price of restaurant service be? How much will people have to pay to get a meal in a competitive market? And how many restaurant meals will be served? You figure it out. Well, remember, in a competitive market price is equal to marginal cost. The price of a restaurant meal will be bid down to $3.00 per meal and that means there will be one, two, three, four, five customers that will be coming in to get food service and the price will be $3.00 per meal. So it’s where the supply curve would intersect with the demand curve, the equilibrium price and quantity. Now, how much profit are the individual restaurants making in this case? The answer is, they are making no profit. The price is equal to the cost, in this case, marginal cost and average cost because the marginal cost is constant, $3.00 per meal and with price equal to average cost, there is no profit being made in this market. There is no producer surplus. The firm’s are making nothing beyond covering their costs. Does that mean that the market is not creating economic value? Absolutely not. On the contrary, the market is creating lots of economic value, but it’s all coming in the form of consumer surplus. Look at this customer right here. He’s willing to pay $9.00 for a meal. The price of a meal in this example is $3.00; therefore, he gets $6.00 worth of consumer surplus. Next, we have a customer willing to pay $7.00 only paying $3.00 for $4.00 worth of consumer surplus. Four and six is a total of $10.00 consumer surplus on these first two meals. The third customer was willing to pay $5.00. She got her meal for $3.00 also for two additional dollars worth of consumer surplus. And finally, we have a customer who is willing to pay $4.00, got his meal for $3.00 for an extra dollar worth of consumer surplus. The fifth customer paid just enough to cover the cost of the meal, so there is no extra economic value created on this marginal meal. But, all of these customers within the margin, all four of these customers, paid much less than what the meal was worth to them. So they all got consumer surplus. Remember, consumer surplus is part of the economic value that is generated in the market and the total amount of economic value in this case is equal to the total amount of consumer surplus. Thirteen dollars was at economic value, which would be the area between the red curve and the blue curve in this case. In the case, in the case of the competitive market, with a constant marginal cost, all of that economic value goes to the customers.

Now, what happens if we take this competitive market for restaurant meals and turned it into a monopoly and we give you that monopoly franchise? How much are you going to charge customers for their meals and how many customers will you end up serving? Well, your motive, of course, is to maximize your private profits. You are
Calculating Deadweight Loss

The Social Cost of Monopoly

The Social Cost of Monopoly

Calculating Deadweight Loss

concerned about economic value, but your only concerned about the part that winds up in your pocket. Therefore, you're going to calculate the price and quantity that maximizes producer surplus for you, and in this case, producer surplus is the same thing as monopoly profit. Remember, producer surplus is only part of economic value. The other part is the consumer surplus. So, if producers and consumers don't have the same interests at stake, if you want to maximize your share of the pie, you may do it by minimizing the part that consumers get, and in the end, you may actually reduce the total amount of economic value just to put a bigger chunk of it in your pocket. Let's see how this works. Now, remember you're a one-price monopolist so you have to charge all of your customers the same price. It means you can't charge this customer $9.00 and this one $7.00. You can't charge everyone the reservation prices, you have to choose one price and stick with it. So, let's look at your options.

The first option is pricing the meals at $9.00 apiece, if you do that then you'll sell one meal, you'll served one customer and your total revenue would be $9.00. Your total cost would be $3.00 times one meal and your profit would be nine minus three or $6.00. If you decide you want to serve a second customer, you have to lower your price to $7.00. Your total revenue in this case is $7.00 times two or $14.00 and your total cost is $3.00 times two for a total of $6.00. Thirteen minus six is $8.00 worth of profit. If you reduce your price further to $5.00 in order to attract the third customer, then you're going to have a total revenue of five times three of $15.00. Your total cost in this case is three times three or $9.00 and your profit is $6.00. Well, notice your profit has just been reduced, you are making $8.00 worth of profit when you served two customers and now you're making less profit when you serve three customers, and it just gets worse from here. If you lower your price down to $4.00, you serve four customers for a total revenue of $16.00 minus total cost of three times four is $12.00 and your total profits are $4.00. If you lower your price in order to bring the fifth customer in, notice in that case, your making no profit at all. Your price is equal to your average cost and in that case, you're breaking even, you are making zero economic profit. So in order to maximize your profits, you choose a price equal to $7.00. The monopoly price that maximizes profit is $7.00 and the quantity of restaurant meals that you provide is equal to two. The purple box here then represents your total revenue, $7.00 times two meals is $14.00. The cost would be represented here by the blue line, so this area down here is what you have to expend in order to create that revenue, and your profit is $14.00 minus $6.00 for a total of $8.00 worth of profit.

Now, how much profit is created for society? How much economic value is created once we add your profits to the benefit that customers get? See, economic value isn’t just about the monopolist’s profits, it’s also about the benefit that is created for customers and a consumer surplus that comes on top of the economic profit. Well, if you look at our first customer, she was willing to pay $9.00 for a restaurant meal. She only has to pay $7.00 so she’s still getting $2.00 worth of consumer surplus. That $2.00 worth of consumer surplus has to be added on to the $8.00 worth of profit to give us $10.00 worth of economic value so far. Well, that’s where it ends because these other customers don’t get restaurant meals. In order to maximize your private profit as a monopolist you pushed them out of the market. You raised the price so high that they were no longer willing or able to purchase restaurant meals. That means, that this five minus three or $2.00 worth of economic value here, which would have been producer surplus or consumer surplus, it’s simply not created, and we call that a deadweight loss. This is economic value that’s there for the taking. There’s a difference here between the value or the benefit that the consumer would get that is, this particular meal is worth $5.00 to some customer and it only costs $3.00 to produce it, there’s $2.00 worth of economic value that’s right there for the picking. But no body gets it because your monopoly operation chooses not to serve that customer, and you choose not to serve that customer so that you can get more revenue off of these two higher paying customers by raising the price you push this customer out of the market. Losing the economic value that could be created perhaps in the form of consumer surplus, perhaps in the form of extra profit. It doesn’t matter, the point is this economic value is lost and we call that a deadweight loss.

The same for this customer, he’s willing to pay $4.00, it only costs $3.00 to serve him, but you don’t serve him at all, so that you can make more money off of these high paying customers. You push him out of the market as well, and therefore, the economic value that he could create is also lost. Well, if you add up the economic value that could be created but isn’t, you get what we call the deadweight loss and the deadweight loss is the cost of operating a monopoly. In order for a monopolist to maximize his or her private profits, in order to maximize monopoly profit, there’s going to be a restriction of trade relative to what we would get in competition. In competition, we would have five meals produced and traded at the price of $3.00 and we would get $13.00 worth of consumer surplus as we saw before. With the monopoly there is a restriction of trade, the price is jacked up, some customers are not served and the total economic value is the sum of the consumer surplus of $2.00 and the monopoly profit or producer surplus of $8.00 for a total of $10.00. The other $3.00 is lost to society. The $13.00 that we could have created has been shrunken to $10.00 and this right here is the deadweight loss, the cost to society of operating a monopoly.
Other Market Models

The Social Cost of Monopoly

Calculating Deadweight Loss

Now, there is one more thing that raises the social cost of monopoly and that’s the possibility of rent seeking. See, you are running the only restaurant in the airport and your making a fat little profit doing so. That means, that everybody else is going to want to get your franchise which means you are going to have to defend your territory against other people who will want to get a share of it, or perhaps to get you out so they can run the monopoly themselves. This leads to a kind of wasteful competition that we have referred to before called rent seeking. The monopolist in order to defend her monopoly will have to hire lawyers will have to lobby Congress will have to go and constantly be expending the effort and resources to defend her monopoly position against other people who would like to take her place. So, in order to keep your share of monopoly profits, in order to keep your monopoly profit intact, you may have to engage in non-price competition, burning up resources to preserve your monopoly position. That is a kind of rent seeking.

So, to summarize, monopoly imposes two costs on society. First, the deadweight loss that comes from restricted competition, and second any kind of rent seeking activities that go on as people compete to become the monopolist. Now we’ll look at the same argument but we’ll look at it in a picture that looks a little more familiar, that is, we’ll look at the continuous demand curve and the smooth marginal cost curve that we discussed earlier and show the deadweight loss in that context.
Other Market Models

The Social Cost of Monopoly

Understanding Monopoly Regulation

In this lesson we’re going to consider how to deal with the problems created by a monopoly. The problem created by a monopoly is this; in order to increase its own profits, the monopoly will restrict the quantity that it sells and raise the price that it charges. The problem with this behavior is that by restricting trade, the monopoly creates a deadweight loss. It eliminates the production and sale of goods that create more benefit for society than they cost society. So, from the point of view of society, monopoly is costly. Therefore, we might, as a society or as a government, decide to take action to avoid this problem and to try to minimize the deadweight loss or the cost.

Let’s look then at what we might do. The first thing is that the government might decide that the best way to deal with a monopoly is to break it up; that is, to break a large company into a number of small companies and to force them to compete with one another. That’s pretty much the strategy that the government took with the AT&T of 1984, when they broke the large telephone company into a number of smaller companies that were supposed to compete with one another.

The problem with breaking up a monopoly, the problem with certain kind of anti-trust actions is they may actually cost society by limiting economies of scale. A very large company may be good for society, if its average cost of production gets so low that it can provide the good to society at very, very low costs. This is the argument of natural monopoly. And I’ve shown, in this picture, an example of what a natural monopoly might look like. Suppose we have a company that provides its products at a constant marginal cost. You might think of this as a pharmaceutical company that can keep churning out additional pills at a fixed and very small cost. However, if you include the fixed costs of producing these drugs, that includes all of the research, and development, and marketing, and other things that go into creating a market for a pharmaceutical, the average cost is going to be very high for small scale operations. And the average cost is going to continually fall as the operation gets larger and larger. This is usually what we call a natural monopoly. A natural monopoly is one where the price of the grid would get smaller and smaller, if we priced at cost as the operation got larger and larger. It is an operation where the point of efficient scale involves a lot of production relative to the amount of the good that the market wants to buy or, in this case, the point of efficient scale would be infinity, if we have constant marginal costs and a fixed cost. And the average cost is always going to be declining. It’s not possible for a pharmaceutical to get too big according to our usual analysis of efficient scale. The efficient scale will be infinity.

So, breaking a company up into a lot of small competitors is going to be no good for society, if there are large economies of scale in that industry. All you’ll do is create a bunch of inefficient firms, each of which has to cover its own fixed costs. So, breaking up a monopoly is not always the quick and easy solution to the problem of deadweight loss. In some cases, it will be, but in other cases it won’t. And the jury is still out on whether breaking up AT&T was really the right approach to solving the problems that were imagined to come out of that monopoly situation. Notice what’s happened; all of the Baby Bells that were created are now in the process of rejoining one another to create large companies that can enjoy economies of scale. So, there’s always a tension, when you’re thinking about breaking up a monopoly, between enjoying scale economies and having smaller competitive firms.

The second approach that can be taken to solving the problem of deadweight loss is to restrict a monopolist’s ability to exercise its market power. This is what we call monopoly regulation; telling a monopolist what price they can charge. Left to its own devices, the monopolist, in this case, would choose a price where marginal revenue equals marginal cost at this point right here, and the quantity the monopolist would produce would be \( Q_N \) and the price the monopolist would charge would be up here on the demand curve. In this case, we have a deadweight loss that comes out of this operation and the deadweight loss is the difference between what a competitive market would produce; that is, where the demand curve crosses the marginal cost curve, and what the monopolist chooses to produce, where marginal revenue equals marginal cost. If you take the difference between these two outputs, the competitive output and the monopoly output, and look at the difference between the demand curve, that is, the benefit to society, and the marginal cost curve, that is, the cost to society, you see all of these drugs that don’t get produced and traded, so that the monopolist can raise his or her profits.

Now, suppose we decide that we want to restrict the monopoly’s ability to do this and thereby shrink the amount of the deadweight loss. Well, to start with, in a market that’s completely unregulated, the deadweight loss is going to be the area of this triangle right here, between the demand curve and the cost curve, between the competitive output and the monopoly output. This area is the cost to society of the monopoly’s restriction of trade.
Other Market Models

The Social Cost of Monopoly

Understanding Monopoly Regulation

Well, let me propose two ways of trying to solve this problem. The first is the government can use what is called average cost pricing. Average cost pricing is telling the monopoly to charge a price that just allows it to cover its average cost, when the market buys all it wants to buy at that price. So, suppose the regulator sets a price of a car that’s $P&C$, the average cost price. If the monopoly can’t manipulate the price, then it has no incentive but to sell all that the market will buy at that price; that is, go over to the demand curve and find the quantity, and then drop the quantity down into the horizontal axis and find out how much the monopoly sells at that particular price. So, if the government says the monopoly has to charge a price of $PAC$, the monopolist will choose this quantity of output and $QAC$ is the amount that ends up being produced in trading. The interesting thing about $QAC$ and $PAC$ is it’s the point where the demand curve and the average cost curve intersect. This is a price and quantity that guarantees that the monopolist breaks even. The government doesn’t usually have too much trouble figuring out what this price is; that is, if it has good information about the monopolist’s costs. Well, that means there are a few problems associated with average cost pricing. The first is that the monopolist may be tempted to misstate its costs, to overstate its costs, to inflate or pad its costs, so that the government will allow it to charge a higher price and, perhaps, it can make some secret profits. That’s one problem.

The second problem with average cost pricing is that once the monopolist is guaranteed that will break even or make a reasonable return on its investment, then the monopolist has no incentive to engage in any kind of activities that might lower its costs. It won’t do any kind of innovation, nothing at all. Once the monopolist is regulated, and especially if the government will lower the price the monopolist can charge anytime his or her costs fall, well, then you’re being punished for innovating. You’re being punished for lowering your costs. So, not only might the monopolist misstate her costs, but she also will have incentive to control her costs, or to innovate, or to do anything that would result in costs falling, because she doesn’t get the benefit from it. In fact, the government would take it away by then lowering the regulated price.

So, average cost pricing, even though notice it does shrink the deadweight loss than the big triangle to this little old triangle, it has some problems in the implementation. It may not be practical because of the incentives it gives the monopolist to misstate costs or to inflate or not keep their costs under control.

Well, the other possibility then is to try to eliminate the deadweight loss completely by having the price equal to the marginal cost. This is what would happen in competition. Having the price equal the marginal cost is what we call marginal cost pricing. If marginal cost pricing is in effect, then the monopolist will then produce the competitive quantity. Ah, the nice thing about this is we completely eliminate the deadweight loss. Well, the problem with this is it’s very, very hard to implement. First of all, marginal cost is much more difficult to estimate and report than average cost is. For average cost, you just take your total costs, divide by the amount of output and you’ve got average cost. With marginal cost, on the other hand, you’ve got to be able to estimate or otherwise figure out what the cost is of increasing your output by one unit. And that’s not something that a lot of firms have at ready access to that information. So, first of all, it may be impractical to get the information that you need. Well, in the case of the drug story we were just telling, it shouldn’t really take too much effort to figure out how much it cost to turn the crank and produce some extra pills. However, there are still other problems. Even if you can get the information reliably, it still may be difficult to implement marginal cost pricing. The main problem is this; if you force the firm to price, where price equals marginal cost, it may be that you’re imposing a loss on the firm. Look, if you go up to the average cost curve, you can see that the cost of producing this quantity of drugs is greater than the price. You’re imposing a loss on the firm. Now, it makes sense where this loss comes from. You’re not giving the firm enough revenue to cover their fixed costs. You’re only allowing them to cover the cost of producing an extra pill. That doesn’t include all the money they spent on research, and development and marketing. All that would be lost under marginal cost pricing. So, if a monopolist is forced by the government to price at marginal cost, the monopolist would have to go out of business, or else get a subsidy from the government equal to the amount of its fixed costs. It would need some kind of additional assistance.

So, marginal cost pricing eliminates the deadweight loss, but it also imposes a loss on the monopolist, and that requires some extra help. Average cost pricing doesn’t completely eliminate the deadweight loss. It is a little easier to implement, but it imposes bad incentives on the monopolist. And finally, what the monopolist wants to do imposes a very large deadweight loss on society.

None of these proposals is without its drawbacks. But there’s been a new proposal that’s been made recently by some economists that has some of the advantages without all of the drawbacks, and that is the proposal that the
Other Market Models

The Social Cost of Monopoly

**Understanding Monopoly Regulation**

government might actually buy the patent from a drug company that had come up with a successful drug. Buy the patent from the company and then allow any firm that wants to, to produce the drug competitively. That would give us marginal cost pricing, because of competition, and by buying the patent from the drug company, it would allow the drug company to cover some of the fixed costs. Of course, there are all kinds of issues, figuring out exactly what to pay for the patent, whether there should be an auction or not, all of those things. But it’s good because it combines some of the best of our regulatory strategies; eliminating the deadweight loss, still allowing the company to make a profit, and using the market, that is, if we have an auction for the patents, to figure out what the value of them is.

Monopoly regulation is a strategy that’s used to try to reduce the social cost of monopoly. Shrinking the deadweight loss. And even though none of the strategies is perfect, all of them are probably some improvement over a pure, unfettered monopoly pricing.
Other Market Models

Oligopoly

Introducing Oligopoly and the Prisoner's Dilemma

We've considered a market that's dominated by a single seller. We called it a monopoly. We've also studied markets that have so many sellers that they each take the price of the product as given. We call that a competitive market. In our last lectures, we looked at another case, a market that has a small number of sellers, but each of them has some market power. That was called monopolistic competition and it allowed us to study the creation of variety. However, in some ways, monopolistic competition is a special case. What we really are interested in is knowing how a small number of sellers behave strategically. We're interested in a model, in which 3, 4, 5, some small number of sellers makes decisions, considering the effect their decision has on other players, where players are considering each other's moves and making strategic decisions, trying to second-guess their rivals as each tries to maximize his or her respective profit. What we're looking for here is a model of what economists call oligopoly. The term oligopoly refers to a market where there is a small number of sellers and where the sellers interact strategically with one another; that is, where each player tries to guess what other players are going to do, because the choices of all the players affect the outcome of the game for each player. My profits depend not only on what I do, but what each of my competitors does as well.

There is, in economics, no generally agreed upon single model of oligopoly. There are several different models of oligopoly, each of which has its own strengths and weaknesses. But one thing economists tend to agree upon today is that any description of strategic behavior among a small number of firms is probably best described using what we call a game theory. In this lesson, I'll introduce you to game theory, giving you an exposure to the most basic game of all, the prisoner's dilemma. It turns out that the prisoner's dilemma has a lot of applications in economics and it applies quite well to the behavior among a small number of firms that are considering how to price their products or choosing the quantity of output to produce. So let's introduce the prisoner's dilemma and I'll introduce it with a story that really doesn't have much to do, at first glance, with economics.

Let's suppose we have two prisoners that have been apprehended at the scene of a crime. And we'll call our prisoners Bonnie and Clyde. So let's suppose that Bonnie and Clyde have been apprehended at the scene of a crime and the district attorney has put them in separate jail cells. And the district attorney comes in and says to Bonnie, "All right, Bonnie, here's the deal. We think that you robbed the bank with Clyde and we're going to give you two choices: you can either confess to the crime or you cannot confess. If you confess to the crime and Clyde confesses, we will give you both 10 years in jail. Now, if neither one of you confesses, we can't convict you on the bank robbery, but we're going to trump up some charge and we're going to convict you each for one year for loitering. Now, Bonnie, if you confess and Clyde doesn't, we will let you be the State's witness and you can go free and Clyde will go to jail for 10 years for the bank robbery and an extra year for perjury, lying to the jury. However, if you don’t confess and Clyde does, Clyde gets to be the State’s witness and he goes free and you go to jail for 11 years."

Well, there's the game. Now, think about it. Bonnie realizes that her freedom or jail time depends not only on what she does, but also on what Clyde does. So let's suppose that Bonnie and Clyde have been apprehended at the scene of a crime and the district attorney has put them in separate jail cells. And the district attorney comes in and says to Bonnie, "All right, Bonnie, here's the deal. We think that you robbed the bank with Clyde and we're going to give you two choices: you can either confess to the crime or you cannot confess. If you confess to the crime and Clyde confesses, we will give you both 10 years in jail. Now, if neither one of you confesses, we can't convict you on the bank robbery, but we're going to trump up some charge and we're going to convict you each for one year for loitering. Now, Bonnie, if you confess and Clyde doesn't, we will let you be the State’s witness and you can go free and Clyde will go to jail for 10 years for the bank robbery and an extra year for perjury, lying to the jury. However, if you don’t confess and Clyde does, Clyde gets to be the State’s witness and he goes free and you go to jail for 11 years."

Now, in a game, your strategies are the choices that are available to you. And, in this case, Bonnie’s strategies are she can confess or not confess. And Clyde’s strategies are the same; he can either confess or not confess. So the strategies are the choices that are available to each player.

Once we know the players and their strategies, we can draw a graph that represents the game. We call this particular graph a matrix. It's an array of numbers, and I will write this matrix out carefully here on the page. The dimensions of the matrix represent the number of players. So Bonnie will be over here and Clyde will be up here. So when I'm looking at the box from the side, I'm looking at Bonnie’s choices and, when I'm looking at the box from the top, I'm looking at Clyde’s choices.

So Bonnie gets to choose which row of the matrix we wind up on. And the row depends on her choice. Remember her choices are she can either confess, which I’ll represent with a C, or she cannot confess, which I’ll represent with NC. So Bonnie can either confess or not confess. She chooses the row on which we end up.

Clyde chooses the column in which we end up. He can either confess or not confess. Those are his two strategies. And by choosing which one to do, he puts us in the first column or the second column of the matrix.
Other Market Models

Oligopoly

Introducing Oligopoly and the Prisoner’s Dilemma

Now, this box represents the outcome where Clyde confesses and Bonnie confesses. This box represents the outcome where Clyde doesn’t and Bonnie does confess, and so forth. Each box in the matrix represents an outcome that depends on the choice of Bonnie and the choice of Clyde.

Now, we need one more thing to make this diagram complete. We need to represent the outcomes for each player in each of the combinations of strategies. So, for instance, if Bonnie confesses and Clyde confesses, each of them goes to jail for 10 years, and here’s how I write that. I write Bonnie’s 10-year jail term first, and then I put a comma and I write Clyde’s 10-year jail term. So within any box in the matrix, I write the payoffs for the players. The first number represents the payoff for Bonnie and the second number represents the payoff for Clyde. Now, if neither of them confesses, then they both go to jail for one year. So there’s Bonnie’s one year in jail and here’s Clyde’s one year in jail. However, if Bonnie confesses and Clyde doesn’t, Bonnie goes free, because she becomes the State’s witness, and Clyde goes to jail for 11 years. Same is true if Clyde confesses and Bonnie doesn’t. Then Clyde becomes the State’s witness and he goes free and Bonnie goes to jail for 11 years.

Well, now I’ve completely represented the game. This matrix represents everything about the game that we know: the players, their strategies and the payoffs in each case. Now we’re ready to ask the question, “How will this game turn out? What’s the solution to this game?” The solution to a game is the outcome when each player is doing the best they can, given what their rivals are doing. We have a name for an outcome like this, we call it an equilibrium. In particular, we call it a Nash equilibrium after the game theorist John Nash, who described it.

So a Nash equilibrium is defined as follows: a Nash equilibrium is an outcome of the game, where each player is doing the best they can, given what all other players are doing. When each player is giving a best response to the other player’s best responses, then you have a stable situation. No one wants to change his or her choice. Let’s look at how this works.

Bonnie’s best response to Clyde’s decision to confess, if Clyde were to decide to confess, would be what? Which outcome is going to minimize Bonnie’s jail time? The answer is Bonnie minimizes her jail time by confessing anytime Clyde confesses.

Suppose now Clyde chooses not to confess. What’s Bonnie’s best response? Which response would give Bonnie the minimum jail time? The answer is Bonnie would choose to confess. Now notice, whether Clyde confesses or doesn’t confess, Bonnie can minimize her jail time by confessing. Bonnie’s best response is the same, whether Clyde confesses or doesn’t confess. She minimizes her jail time with that particular choice. So we call that Bonnie’s dominant strategy. In game theory, a dominant strategy is a strategy choice for a player that maximizes that player’s satisfaction no matter what the rivals are doing. So no matter what Clyde does, Bonnie minimizes her jail time or maximizes her satisfaction by confessing. Confessing is a dominant strategy for Bonnie.

Now, by the same token, we’re going to get the same results when we look at Clyde’s best responses. If Bonnie confesses, Clyde’s best response is going to be to do what? Which of these two columns, which Clyde gets to choose from, gives Clyde the minimum jail time? The answer is Clyde minimizes his jail time by confessing. If Bonnie chooses to put us on the bottom row by remaining quiet, which of these two columns is going to maximize Clyde’s satisfaction? Once again, Clyde maximizes his satisfaction by confessing, even if Bonnie doesn’t confess. So Clyde wants to confess whether Bonnie confesses or doesn’t confess; that is, Clyde has a dominant strategy also, and that is to confess.

Now, a Nash equilibrium is a place where Bonnie is giving her best response to Clyde at the same time that Clyde is giving his best response to Bonnie. In a diagram like this, you can see the Nash equilibrium is a box where there are two circles sitting together; that is, if Bonnie is confessing and Clyde is confessing, then nobody wants to switch their decision, given the play of the rival. Bonnie says, “Well, if Clyde is confessing, I certainly want to confess. Otherwise, I’ll be in jail for 11 years.” So Bonnie considers, but doesn’t want to change her choice. Given that Bonnie is confessing, Clyde is certainly happy to confess, too, because if he kept quiet, then he would go to jail for 11 years. So given what Bonnie is doing, confessing is Clyde’s best response. And since Bonnie is confessing, then she’s also giving her best response to Clyde’s confession. Neither one of them can improve his or her position by changing their choice. This is a stable outcome, a Nash equilibrium. And it’s also what we could call a self-enforcing agreement; that is, once Bonnie and Clyde are both confessing, there is no tendency for either of them to change. There’s no temptation for either of them to switch to the other strategy. So the outcome here is self-enforcing.
Other Market Models

Oligopoly

Introducing Oligopoly and the Prisoner’s Dilemma

Well, there’s one more thing to notice about this. Just because it’s an equilibrium doesn’t mean that it’s a good outcome. It certainly doesn’t mean that it’s efficient. After all, which combination of choices would minimize jail time for Bonnie and Clyde? The answer is if Clyde were to be quiet and not confess, and Bonnie were also not to confess, they would wind up in this little box down here, with Bonnie only in jail a year, Clyde only in jail a year, and minimizing the combined jail time of the two players. That would be the best choice, because both of them would only have to spend a year in jail. We can put a little star in this box, because, in an important sense, this is the box that Bonnie and Clyde would prefer to wind up in. This is the one that minimizes jail time.

However, even though this is the best box to wind up in, it’s not a Nash equilibrium. What’s Bonnie going to do is Clyde is committed to not confessing? Well, if Clyde has decided that he’s not confessing, Bonnie says, “Now, hold on. If I confess, then I can go free and leave Clyde in jail for 11 years.” She may get no satisfaction, knowing that Clyde is stuck in jail, but, according to these numbers, she’s certainly more satisfied having no jail time rather than one year. So Bonnie is tempted to change her move and reduce her jail term from one year to zero years. Because of that temptation, Bonnie may defect and then we are in a different outcome. Once Bonnie is confessing, then Clyde doesn’t want to keep his mouth shut. He wants to confess, too, and we wind up back at our Nash equilibrium. Even though this box is the best one for Bonnie and Clyde, this one is the self-enforcing agreement.

Now, this is a pretty grim view of the way people behave and, in fact, one of the things that happens in business and real life is we spend a lot of time forming relationships with people that we know we can trust, so that whenever we have a situation like this, where there’s a temptation to cheat, we can be reasonably sure that the other person won’t cheat. One of the ways that we enforce cooperative behavior in real life is through repeated play. The fact that Bonnie and Clyde may play this game over and over again every day gives them an opportunity to reward each other for cooperative behavior and to punish each other for opportunistic behavior. If Clyde were to defect and try to stick Bonnie with the 11-year jail term so he could go free, Bonnie would get him back next time. So Bonnie and Clyde, if they play often enough and know each other well enough, would be able to enforce cooperation. But in the one-time play, if these numbers tell the whole story, there’s only one outcome from rational choice, and that is Bonnie and Clyde both choose to confess.

Now, I always feel weird telling people that this outcome is rational, because we all have programmed so deeply into us the good outcome; that is, for many reasons, we are inclined to cooperate and to trust people who we want to believe are trustworthy. On the other hand, it’s important to realize that, in many cases, there are opportunities for defection, so sometimes we need more than just the general trust to keep cooperation enforced. We may need third party enforcement. Bonnie and Clyde may want to hire a hit man to knock off anybody who finks. Therefore, the costs are changed for them and it’s individually optimal to keep their mouths shut. We might have third party enforcement of contracts that make people cooperate. But, more likely, people form reputations through repeated play that incline them to be cooperative rather than take advantage of their fellow players by cheating.

There are all kinds of prisoner’s dilemmas floating around us in the world. Pollution – am I going to clean up my own mess, or am I going to dump it on other people? There’s a good example. It’s cheaper for you to dump your stuff in the river, but if everybody does that, we wind up with a bad outcome. Therefore we have both rules and customs that incline us to clean up our own messes. The prisoner’s dilemma then creates a problem. The best outcome may not be individually rational and it may require repeated play, reputations or contracts to make cooperation more likely.

In the next lesson, we’ll see how an oligopoly is an example of the prisoner’s dilemma, how, if the firms cooperate, they can get the monopoly outcome among them; that is, they can act as a monopolist and split a very big pie among the players. However, once everyone else is holding the price up, you may be tempted to cheat and increase your output so that you can bring in more revenue. But if everybody cheats, then you’re once again in a competitive situation with very low profits.

So in the next lesson, we’ll develop more clearly the example of the prisoner’s dilemma in an oligopoly.
Other Market Models

Oligopoly

**Understanding a Cartel as a Prisoner's Dilemma**

Let's apply gain theory to the case of a cartel. A cartel is an oligopoly where the individual members try to act as a monopolist. Remember, the monopoly always maximizes profits in a given market. So, if we have a number of firms that can get together and act as one, that is, set price and quantity together, they can choose the monopoly price, the monopoly quantity and earn the monopoly profits, which they can then divide among themselves. A cartel arrangement like this always maximizes profits for the members of the oligopoly.

However, holding together a cartel may prove difficult, because of the prisoner's dilemma. After all, if your two rivals are showing restraint, that is, they're restricting the quantity that they produce so as to help keep the price high, then you're tempted to try to sell a larger quantity to take advantage of that higher price. It's individually rational for a single firm to take advantage of its rival's restraints by increasing the quantity that it sells. However, if all of the firms subject to the same temptations try to increase their outputs simultaneously, then the price falls back towards the competitive levels. The firms compete, output increases, price falls and the cartel, for all practical purposes, doesn't exist. The firm sells more output than a monopoly would at a lower price and earns a much smaller profit.

What we're going to do, in this lesson, is show how the cartel's problem is a special case of the prisoner's dilemma. Let's look at two countries that are part of a want to be cartel; they are Nigeria and Venezuela. These two countries would like to cooperate in the production and sale of oil so that the price of oil stays high and they are able to act as a monopoly jointly and maximize their profits. Then, Nigeria and Venezuela each have two strategies. The first is to cooperate with the cartel agreement. In which case, you produce your share of the monopoly output. The monopoly output of oil is determined and Venezuela produces its half, Nigeria produces its half and in that case the monopoly profit is produced and shared between the two countries.

Alternatively, you can cheat on the agreement. You can let your rival hold the price high by restricting his output and then you can try to sell a lot of output to take advantage of the high price. This would be called cheating on the cartel agreement and what it really means is that Venezuela would be producing a larger quantity of oil than the cartel prescribed for it. So, instead of restricting itself to one-half of the monopoly output, Venezuela might produce three-quarters of the monopoly output or some larger amount of oil. That then has a bad effect on your rivals who have restricted their quantity, because now you're pushing down the price and you’re leaving them in trouble.

So, let's fill in the table then, with the numbers. Let me make an assumption first of all. The assumption is going to be that if both firms cooperate with the cartel arrangement, if the firms act jointly as a monopoly, that they will make a total of $10,000,000, and that if Venezuela produces half of the output and Nigeria produces half of the output, each of them gets half of that monopoly profit. So, if they both cooperate and the cartel is successful, then each of them will get $5,000,000 worth of profit, 5,000,000 for Nigeria, 5,000,000 for Venezuela. Now, notice here that the numbers in the matrix represent profits from oil sales and since profits are good, each country would like to have as large a number as possible. This contrasts with our previous example where the numbers represented jail terms.
Other Market Models

Oligopoly

Understanding the Kinked-Demand Curve Model

We’ve seen that monopoly faces a downward-sloping demand curve. And a perfectly competitive firm has a perfectly elastic demand curve. What does the demand curve look like for a firm that’s in an oligopoly? Remember an oligopoly is a small number of firms watching each other’s every move, interacting strategically. Like in the airline industry – what does the demand curve for air travel look like for one particular airline?

Consider here we are with Airline A, and Airline B, C, and D are watching its every move; they’re all trying to second-guess what each other might do next, with respect to the price they charge for their tickets, and the quantity of tickets they sell each week. Suppose the initial situation is that Airline A is charging a price of $200.00 a ticket, and selling about 10,000 tickets a week. What does the demand curve look like for Airline A’s tickets? Well, if Airline A were to decide to raise its price to try to get more revenue from each customer, probably Airline B, C, and D would not follow suit, because if they hang back, and keep their prices constant at around $200.00, they can get a bigger market share as price-sensitive customers defect from Airline A and jump to B, C, or D instead.

This was the idea that Paul Sweezy had whenever he came up with his kinked-demand curve, that if a firm starts with an initial situation and raises its price, they’re likely to lose a lot of customers, because rival firms will hang back, keep their prices low, and tempt customers away. On the other hand, Sweezy says, if Firm A were to decide to lower the price of its airline tickets, probably all of the other firms would follow suit. There would be a big decline in price, and Firm A would not add nearly as many customers as it might expect, because of retaliation from its rival.

This kinked-demand curve is a graph to describe this strategic interaction. When you see a kinked-demand curve, it’s a way of saying that a firm can expect that rivals will not follow it when it takes its price upwards, but rivals will retaliate and follow suit if it should cut prices. The kink occurs at the point where we begin – the initial situation, the original price and quantity. What does the marginal revenue curve look like for a firm that has a kinked-demand curve? Well, the marginal revenue curve, you’ll recall, lies below and is steeper than the demand curve, because as you lower your price, you’re losing revenue on all of the inframarginal units – the tickets you could have sold at the higher price.

So the marginal revenue curve is going to look like this down to the point of the kink, at which point it immediately changes slope and is going to have a discontinuity and start up again further down the page. This marginal revenue curve tells the change in revenue that this airline can expect if it adds an extra passenger, sells an additional ticket. So if we put the marginal cost curve in here, we’re going to be able to predict where Firm A is going to choose to operate – where marginal revenue equals marginal cost – which will usually occur somewhere in this discontinuity.

Well, now this is all starting to seem kind of technical, and you might very well ask the question, "What determined this original point anyway? What determined the location of this kink in the demand curve? How did we get to this original price and this original quantity?" That’s a great, because the model doesn’t tell that, and that’s one of the criticisms of it. It’s really a story about how firms in an oligopoly respond to each other’s moves, but it doesn’t do an especially effective job at telling us how the kink gets established to begin with, how we arrive at the original situation. It also doesn’t tend to match what we actually observe; empirical data doesn’t confirm this theory. The theory, like a lot of theories, is a good story; it’s logically consistent, it even has kind of a compelling, sensible element to it. But it’s not nearly sophisticated enough to really explain the kind of strategic interaction we get in oligopolies; so we’ll have to consider some alternative explanations.
We've just completed a series of lessons on monopoly, a market that is dominated by a single firm. The interesting thing about monopoly is that the firm has market power, so it maximizes profits by pricing where marginal revenue equals marginal cost. This happens any time a firm has market power. Now before monopoly, we studied another form of market structure, perfect competition, where free entry drove economic profits down to zero. Firms entered and exited until, in the long run, the firms in the market were breaking even, price equals average cost.

Now, we're going to combine these two market structures to come up with something new. We'll take the idea from monopoly of price being set, so that marginal revenue equals marginal cost. That is, we'll take the notion of market power from monopoly and combine that with the notion of free entry from perfect competition. What we have then, is a market in which firms compete to create little monopolies. Firms compete with each other to get a share of the market, each firm keeping a downward sloping demand curve by convincing its customers that its product is in some respect distinct from the competitors.

With monopolistic competition, we'll see that we get results, both from monopoly and from competition. From monopoly, we'll get marginal revenue equals marginal cost. We'll get the result of market power in the profit maximizing decisions of the firm. From perfect competition and free entry, we'll get price equals average cost, that is, firms that are breaking even. The interesting thing about monopolistic competition is that it gives us a model for how variety arises in the market. So, we'll begin by studying product differentiation, what it means and how firms create market power. Then, in the next lecture, we'll use some technical tools to show what the equilibrium looks like when firms compete to create monopolies.

So, let's begin. Look at these two products. Can you tell them apart and if you can, how did you learn to tell them apart? As I study them, I see both have buns with sesame seeds on top. The buns have slightly different shapes. Both of them have meat patties. Both of them have cheese and other condiments on them, but to an untrained eye, they both look like hamburgers. Both of them are products of large companies that have spent a lot of advertising revenue to create the impression among customers that in fact, the products are not perfect substitutes. And if you happen to be the fan of one or the other, you may fight to the death over which one is better. In fact, you may not be willing to switch to the competitors product or only be willing to do so, when the price of it is much lower.

So, the fact that these two both look like burgers to me, just means that I'm not watching enough television, because then I'd clearly know the difference between one and the other. Or, let's take another example.

Here's a product where a lot of revenue has been spent by the company to create the impression of distinctiveness. Now, both of these products are bubbly, brown liquids packaged in aluminum cans. And yet, I have had vigorous arguments with friends who tell me that in fact there is no way that you could consider these two products substitutes. That you either want one and must have it and wouldn't consider the other or you're a die-hard fan of the other and would never consider the first. Here we've got two products, both of them brown, both of the bubbly, and as far as I'm concerned, they look pretty much alike. I can't smell the difference. In fact, they smell identical to me. I can't even taste the difference. Now, what does that tell you about me?

It tells you that I haven't had enough experience to be able to distinguish these two. In fact, I would fail any taste test that these two companies put me through. As far as I know, these products are perfect substitutes, they're practically identical. But, I bet, you don't think so, at least if you're a fan of one of the products. And yet the companies spend a lot of money trying to persuade me that the other product is not an acceptable substitute. This money that's spent creates in our minds an impression of variety. The impression that in fact there really are two soft drinks where the untrained eye can see only one. But there really are two different kinds of hamburgers, where the untrained pallet detects only one product.

The question is this. Why do firms spend money buying advertising services to persuade me that their product is distinct? And then answer is, so that they can create market power. This product is able to raise its price without losing all of its customers, simply because it has persuaded its customers that it is a superior product to the rival. This product is able to keep its customers even when its rival lowers its price, because there are people who simply would not switch away from the product that they prefer. The impression of distinction is created from small things, advertising perhaps, maybe a subtle difference in the formula, more bubbles, a different temperature, but really are the products distinct and in whose mind? But all the company needs to do is to persuade you that they're different, to
Monopolistic Competition

**Defining Monopolistic Competition**

allow you to stick with it even as it raises its price, or not to defect to the competitor just because the competitor lowers its price a little bit.

These two products have created little monopolies. Little efforts to create market power on their behalf have paid off by earning them captive customers or at least customers who are loyal because they believe that one product is better than the other. What we’ll do in the next lesson is see how this creation of market power leads to a different kind of equilibrium than we get either with pure monopoly or with perfect competition. And I’ll let you in on the ending a little bit early. The answer is, we’re going to wind up with a little more inefficiency than we have in a competitive market, but the payoff is we get more variety, or at least the perception of it.
Other Market Models

Understanding Pricing and Output Under Monopolistic Competition

In this lesson, we're going to explain how monopolistic competition works and I'm going to do it using some graphs that are familiar, but I'm going to use them in an unfamiliar way. Let's look at the situation that faces a single fast food restaurant selling hamburgers. This fast food restaurant has been able to achieve a downward sloping demand curve by differentiating its products from the hamburgers of competitor restaurants. Maybe they have more sesame seeds on their bun, maybe they advertise harder, maybe they have some kind of special promotion that make their restaurants look fun, something like that. But, for whatever reason, you believe as a customer that this product is not a perfect substitute for other hamburgers. Therefore, you're willing to stay with this firm, even when they raise their price or if competitors lower their price, you won't defect and buy the other product.

Product differentiation gives this firm a kind of monopoly. They are the only providers of their particular brand of fast food hamburger. Because they have a downward sloping demand curve, they also have a marginal revenue curve, like any firm with market power. And remember, marginal revenue is always less than price, because in order to sell another hamburger, you have to lower the price that you could have charged for a smaller number of burgers and earned more money on them. So marginal revenue is declining and is below the demand curve.

Finally, notice, I've put in the cost curves for producing hamburgers. This fast food restaurant experiences increasing marginal cost as it produces more hamburgers, due to diminishing labor productivity. And the average cost curve lets us determine whether the firm is making a profit or not. So, looking at this diagram, we can describe the firms behavior the way we would describe the behavior of any monopolist. This hamburger restaurant in going to produce the quantity of hamburgers that makes marginal cost equal to marginal revenue. So, here's the quantity of hamburgers produced right here. If we follow up to the demand curve, we see the price that this firm would be charging for its hamburgers and since price is greater than average cost, this firm is making a profit. And the amount of profit would be price minus average cost multiplied by the number of hamburgers produced.

You've seen all of this before and we've given you a little reminder over on the board to help you remember how a monopoly is described, the choice of the quantity, the choice of the price and the profits that are earned. Now, here's where the model gets new and interesting. Instead of assuming that there is no entry into this market, let's imagine that there is free entry. The cost of entering the market is the cost of differentiating your product from that of competitors. We could have anybody start up their own fast food chain, as long as they can persuade customers that the hamburger that they sell is somehow different than the other burgers that are out there on offer.

Once you've got a new burger that you've defined in the minds of customers as somehow unique, then you've got your own downward sloping demand curve, you've got your own little monopoly, a monopoly in a market that you've created by product differentiation, either through advertising or other small changes in the nature of the products. Well, think about this. One chain has its little chunk of the market, another chain has its little chunk of the market, a third chain has its little chunk of the market and if you add up all the demands for all of these different kinds of hamburgers you get the total demand for hamburgers in the market.

See what's happening is, out there in the market, there is a demand for hamburgers and firms are satisfying that by producing hamburgers and selling them, but in order to get more profit the firms are differentiating their products from each other. And when they do that, they have a downward sloping demand curve for their particular variety of hamburgers. However, it's not that we could get an infinite number of hamburgers. There's not any way that the market would support 80 or 90 different particular kinds of fast food burger. That's because there's not enough demand in the market to support that many different chains.

However, if you look at the diagram that I've drawn here, there's certainly room for at least one more chain. How do you know, how do you know by looking at this diagram that another chain could enter with a differentiated hamburger and still make a profit. You know because, in this situation, the firms that are already in the market, this being one of them, are making extra profits. They're making economic profits. So, firms will enter to try to take some of the market share and when they enter, they'll be creating a new differentiated product.

So, what happens then, when a competitor enters this market? Is the share that's left for our particular hamburger shrinks? Once a new burger is created, some people will be persuaded that it is the best burger. It will become their burger of choice and they'll switch to it. What happens then, is the market share of our little firm shrinks. The market share of our particular fast food restaurant shrinks. The demand curve moves inward and with it the marginal revenue
Other Market Models

Monopolistic Competition

Understanding Pricing and Output Under Monopolistic Competition

curve and as that changes, as demand shifts inward and marginal revenue comes with it, the firm adjusts its profit maximizing output.

With a smaller market share, now, the firm is going to produce a smaller quantity of burgers and change the price that it charges and perhaps make a different amount of profit than before. Again, what caused the curves to shift? The curve for your particular fast food chain shifts inwards when a new chain enters the market and claims a share. It says, "Hey, there’s enough to go around." And it appears that there is, because your firm is still making enough money to cover its average cost.

However, I shifted the curves so that we have arrived at a new equilibrium. Notice, the equilibrium has the firm producing where price equals—where marginal revenue equals marginal cost and price is on the demand curve. But here’s something that’s true, in this case, that wasn’t true before. Price is now equal to average cost and when price is equal to average cost the firm is breaking even. There’s no room for any more hamburger chains in this market. Every hamburger chain has entered, distinguished itself and claimed a chunk of the market.

Now the firms are just breaking even as they would in a competitive equilibrium. Ah-ha. That’s why we call this monopolistic competition. Monopoly, because of the downward sloping demand curve, the market power, the marginal revenue, marginal revenue equals marginal cost, all that’s familiar from our model of monopoly. But we call it competition, because free entry occurs. That is, firms continue to enter this market with distinguished products until price is equal to average cost, until the demand curve is tangent to the average cost curve. At that point, price equals average cost, firms are breaking even and there will be no further entry. We now have an equilibrium in a monopolistically competitive market.

So, is that good or is it bad? Well, it depends. It depends on what you like. The advantage of monopolistic competition is we get a whole lot of variety, all different kinds of hamburgers, at least if you can recognize the variety, it’s there. One the other hand, the cost to customers is, instead of arriving at this point of minimum average cost, the way that we do in a competitive market, we wind up on a downward sloping portion of the average cost curve. Average cost never falls all the way down to its lowest level, the way it would in a competitive market. Costs are higher in a monopolistically competitive industry. Costs are higher and there’s still a bit of a deadweight loss, like in any monopoly.

But that turns out to be the price we pay for the variety that monopolistic competition delivers. If you like variety, more than a little extra efficiency, monopolistic competition is a great bargain. On the other hand, you’ve got to recognize the deadweight loss is still there and the firm never gets to minimum average cost. So, a quick review. Monopolistic competition has firms behaving as monopolies, but the additional element is free entry. Firms can enter, creating their own new little chunk of monopoly until the market doesn’t have room for any more of them, until the market share of an individual firm is just enough to cover the average cost of producing the burgers they sell.
Other Market Models
Monopolistic Competition

Understanding Monopolistic Competition as a Prisoner’s Dilemma

Here’s another example of how gains theory can help you describe strategic interaction between competing firms. In our last example, we looked at how the quantity choice could become a strategy. In this example, we’re going to look at how advertising expenditure can become a strategy. Let’s suppose we have two rivals, Coke and Pepsi, and they’re both competing for a share of the soft drink market. Let’s suppose for the sake of easy mathematics that the soft drink market has $10 million worth of sales every year. So, either company can go after that money, and if they split the market, each of them will get five million. Let’s suppose that if neither of the companies advertises they will split the market right down the middle. Coke will get five million, Pepsi will get five million.

Let’s suppose, however, that either one of them can launch an ad campaign and the ad campaign costs three million dollars. If you launch an ad campaign and your rival does not, you get all the customers. However, if you launch and ad campaign and you rival launches an ad campaign, you continue to split the market down the middle, just as before. Let’s look at what kind of gain we have when Coke and Pepsi are trying to make decisions about advertising. If neither company advertises, they split the market down the middle. So, ten million dollars divided by two is five million for Coke and five million for Pepsi. So, Pepsi has five million, and Coke has sales of five million, and since neither of them is advertising, that’s what they get.

Let’s suppose now, that Pepsi decides to advertise and Coke doesn’t. That means Pepsi gets all the customers, all ten million and we subtract three million for advertising revenues. In that case, Pepsi winds up with seven million dollars total and Coke gets nothing. Let’s suppose now that Coke tries to advertise when Pepsi doesn’t. In that case, Coke will wind up with the seven million dollars profit and Pepsi winds up with nothing. Finally, if both of them advertise then they split the market in half and when they split the market in half, each of them gets five million, but now each of them has a three million dollar ad campaign, which reduces what’s left over down to two million per company.

Now, we’ve translated our story into a matrix. That’s the first step in solving a gain. The second step is finding the best responses. If Pepsi decides to advertise, Coke’s best response is to do what? Well, of course, Coke is going to make nothing if they don’t advertise and they’re going to make two million if they do advertise. So, Coke’s best response is to advertise. If Pepsi decides not to advertise Coke’s best response is? To advertise, of course, because then they get the whole market and seven million. So Coke has a dominant strategy and that’s to have an advertising campaign. What about Pepsi?

Well, if Coke decides that it wants to advertise, Pepsi’s best response is? Check again, these two numbers and you find that Pepsi has a higher profit if they choose the first row. That is, if Coke has us in the first column, Pepsi certainly wants to be on the first row and get two instead of zero. If Coke puts us in the second column, by not advertising, well once again, Pepsi finds that it can make a higher profit by advertising then not advertising. Pepsi has a dominant strategy, just like Coke and it’s to advertise.

Well, by golly, we’ve got a prisoner’s dilemma on our hands. That is, the best outcome for Coke and Pepsi will be, neither one of them to advertise and share the customers between themselves. When they do, they have a very large pot of money to share and no money spent on advertising. What happens in this case, however, is that we don’t have a self-enforcing agreement. Although Coke and Pepsi might like to be in this box down here, if neither of them is advertising, then both of them are inclined to cheat. That is, if Coke says to Pepsi, “Let’s don’t advertise this year.” Or Pepsi says to Coke, “Let’s don’t advertise this year.” Well, once your rival isn’t advertising, then you can make the big money by having an ad campaign.

So, we don’t have a self-enforcing agreement if neither of them are advertising. If neither of the companies advertises, then both of them are going to find that it’s individually rational to launch an ad campaign. However, when both of them launch an ad campaign, then we wind up in this box up here, where the profits are very low. In fact, the combined profits of the two companies are lowest in this box of all three. This is the worst outcome for the two companies. How did they wind up here? They wound up here because of the prisoner’s dilemma, because the good outcome was not a self-enforcing agreement, because this is the only box where both players are giving their best response to their rivals. This is the only stable outcome.

Well, if Coke and Pepsi decided they didn’t want to advertise and every year they kind of looked at what each other did and they said, “Come on, come on, let’s just share the market, forget the advertising.” In that case, we might be able to enforce this outcome through repeated play. Or if the government declared that advertising soft drinks was
Other Market Models
  Monopolistic Competition

  **Understanding Monopolistic Competition as a Prisoner’s Dilemma**

  illegal, then Coke and Pepsi would be down here in this box and they would like that law, in this particular story, because it would make them more profitable. It would protect them from the prisoner’s dilemma that puts them in this unprofitable box up here at the top.

  Imagine what would happen one year if one of the major soft drink advertisers, instead of running an expensive Superbowl ad campaign, came on the Superbowl commercial and you saw a black screen with white letters that said, instead an expensive ad campaign we’re donating all the money we would have spent on Madison Avenue to charity. Now, how would that effect your behavior? Would you be inclined to buy that soft drink? Maybe that’s the way that a soft drink company might try to resolve this prisoner’s dilemma, by using some kind of judo on the advertising process. But until something amazing like that happens, we’re kind of stuck in a dilemma. Advertising is like an arms race. Both companies would prefer not to do it, but because of the possibility of your rival doing it, you have to keep up and in the end, everyone’s profits are reduced.
Resource Markets
The Derived Demand for Labor

Deriving the Factor Demand Curve

We've seen how a profit-maximizing firm continues to increase its output until the marginal cost is equal to the price of the good. Or, in the case of a firm with market power, the firm continues increasing production until the marginal cost is equal to the marginal revenue. In the short run, remember that a firm can only increase its output by hiring more of the variable input – for instance, labor. Everything else is fixed, so if you want to make more stuff, you've got to get more workers. This means that in the short run, there's an equally good way to describe the condition of profit maximization, and that is that the firm keeps hiring workers until the last worker adds just enough revenue to the firm to cover the cost of that worker.

What we're going to do now is consider this alternative way of describing profit maximization. And the payoff will be that we come up with a demand curve for labor, a derived demand – that is, firms hire labor because of the things that you want to buy from the companies that use labor to make them.

Let's start by defining the concept of marginal revenue product. The marginal revenue product is equal to the change in the firm's total revenue that results when it hires an extra worker. The change is what's added to the firm's revenue as a result of the decision to bring another worker into the factory. The marginal revenue product has two components. The first component is the marginal revenue – that is, how much extra revenue is added when the firm produces another unit of output. And the second component is the marginal product – that is, how much output is added when the firm hires an extra worker.

If we flip these over and look at the definitions of the marginal revenue and the marginal product, you can see clearly the relationship between the marginal revenue product and these two concepts that we've learned elsewhere. That is, when a new worker comes into the firm, that worker adds a certain amount of output; and that certain amount of output is going to have a direct effect on the total revenue.

Now consider how this output is going to influence the total revenue. If the firm has market power, then the firm has to lower its price to sell this extra output, and we become very interested in the elasticity of demand and the marginal revenue. But to make this problem easier, let's imagine for the time being that we're dealing with a competitive firm – that's a firm that's able to sell as much output as it likes without influencing the market price of its good. If that's the case, then the change in total revenue that results from a change in output – the extra revenue that's added whenever you sell an extra unit of the good – is simply equal to the price of the good. So we can replace marginal revenue with the goods price. It's always true for a competitive firm that price is equal to marginal revenue.

So now we have a simpler version of the marginal revenue product, and this simpler version has its own name – that is the "value" of the marginal product. The value of the marginal product is defined as the price of the good times the extra output that's produced by the last worker hired. So if you want to know how much a worker adds to the revenue of a competitive firm, the value of the marginal product answers that question; it's simply the physical output of the worker multiplied by the price at which that good sells in the market.

So let's go ahead and use this concept on a firm that we've looked at before, and then see how that firm might use the information in choosing how much labor to hire whenever it's after profit maximization. So here we go to a set of data that we've considered before – the production possibilities of this firm in the short run. We can hire one worker, two workers, three workers, four, and so forth. And here we have the physical output of any given number of workers. Remember we're in the short run with certain fixed inputs, like the size of our factory and the tools the workers have to work with. We're varying only a single input – labor. So one worker can make two television sets a week, two workers can make ten television sets a week, and so forth.

The next thing we want to do as we're trying to figure out how much labor to hire is look at the physical productivity of each additional worker; and we've calculated this information before – it's called the marginal product of labor. So the first worker takes us from zero television sets produced to two television sets produced. The marginal product of that first unit of labor is two television sets.

The second worker brings the total production up to ten units. So ten, minus those two units that we got when we had only a single worker, gives us a marginal product of eight units for the second worker, and so forth. In each case, look at the number of television sets that can be produced when another worker is added, and subtract what was produced without that worker, and you get the marginal product of that last worker. The third worker has a marginal product of 20, the fourth worker has a marginal product of 10, and so forth.
Resource Markets

The Derived Demand for Labor

Deriving the Factor Demand Curve

Now, let’s suppose that the price of a television set is $100.00. That means that this competitive firm can sell all the television sets it wants to at a price of $100.00 apiece; and that, because the firm has no market power, they don’t have to worry about saturating the market. But they will be concerned with how much revenue is added by each additional worker they put in their factory.

So to calculate now the value of the marginal product in this case, we’re going to multiply the physical output of each additional worker by the market price of the television – $100.00 – and that’s going to give us the contribution of each additional worker to the revenue of the firm, an amount measured in money. And we call that amount the value of the marginal product.

So now I’m ready to add another column to my table, and I add that column by taking the physical output of each worker, multiplying it by the price of a television set, and recording it, and that becomes the value of the marginal product. So the first worker produces two television sets that weren’t there before. If we multiply two television sets by the price of $100.00, we get the value of the marginal product of that first worker is going to be $200.00. The third worker adds 20 television sets, so the value of the marginal product of that worker is 20 television sets at $100.00 apiece. So value of the marginal product for the next worker, worker number four, is $1,000.00. The value of the marginal product for the next worker is $500.00. In each case it’s as simple as multiplying the actual physical output of the worker by the price of the television sets. This last worker, by the way, has negative marginal product; the firm would never hire that worker anyway, so we don’t even have to calculate that.

So what do we do with this information? What does the firm do with this information? The firm asks itself: “How many workers should we hire?” This is essentially the same question as: “How many television sets should we produce?” because, in the short run, the only way you can change your output is by hiring more or less of the variable input. By choosing the amount of labor to have on the factory floor, the firm is, in the short run, choosing how many TV sets to produce. The choice of having five workers is the choice to produce 45 television sets a week.

So how does the firm make that decision? Suppose the firm is considering going from four workers to five workers. You can see that that fifth worker adds $500.00 to the revenue of the firm. Is that enough? Well, it depends completely on how much you have to pay that worker. If the worker’s wage is $300.00, by all means, hire that worker, because the worker is adding $500.00 to your revenue, but only $300.00 to your costs. On the other hand, if that worker costs $600.00, it wouldn’t pay to make that extra hire, because the $500.00 in extra revenue wouldn’t cover the costs of bringing that worker into your factory.

So now you can see how the firm is making its decision. Keep hiring workers up to the point at which that last worker add just enough to your revenue to cover the cost of bringing that worker onboard. This is the decision to continue hiring until the wage is equal to the value of the marginal product. Keep hiring your workers until congestion of your fixed inputs pushes productivity low enough that that last worker adds just enough output to allow you to cover the wage of hiring that worker.

So now we’re ready to draw a demand curve for labor. The demand curve for labor is a graphical representation of this rule – that the firm continues to hire workers until the wage is equal to the value of the marginal product. If I graph the numbers that I just derived, the value of the marginal product turns into this nice green curve. The value of the marginal product is increasing when the marginal product is increasing, and then is decreasing when the marginal product is decreasing. That means this region over here is the region where teamwork and specialization are dominating the production process and this region over here is the region where the fixed inputs are becoming increasingly congested.

Now the firm that wants to maximize its profits is going to keep hiring labor as long as the value of the marginal product is greater than the wage. That is, as long as each additional worker is adding more to the firm’s revenues than it’s adding to the firm’s costs.

You wouldn’t stop back here with less labor than L*, because those extra workers are adding more to the firm’s revenue than they’re adding to the cost; the VMP is greater than the wage. You wouldn’t proceed to a labor output that’s greater than L*, because the workers are still costing the same wage, but they’re adding now not enough to the...
Resource Markets

The Derived Demand for Labor

Deriving the Factor Demand Curve

A firm’s revenue to justify the expense. This is the point at which the firm has maximized its profits with respect to labor hires. Keep hiring labor until the last worker that you hire has just added enough to the firm’s revenue to cover the wage.

This is a derived demand. "Derived" means that it’s demand that comes from another demand. It is because customers want to buy the output of this firm that this firm has a demand for labor. And anything that changes – either the technology of production or the demand for the final good – will shift this green curve. Consider, for example, what would happen if this good were to get very popular, and its price to rise. In that case, the green curve would shift upwards, as a higher price for each unit of output would mean that each additional unit is adding more to the firm’s revenue. So an increase in price would shift the VMP curve upwards.

Technology could also do the same thing. If labor became more productive – perhaps because the firm acquires more capital in the long run – or if labor just learns how to do things better, then the value of the marginal product curve would shift upwards, because each additional worker hired adds more output, and, therefore, more to the firm’s bottom line.

We’ve been calling this the value of the marginal product because we’ve maintained the assumption that the price was constant. That is, the firm was competitive, it had no market power. But you get exactly the same shape of a curve if the firm has market power; you just have to be careful about calculating marginal revenue, which depends on the elasticity of demand. In the end though, you get a very similar looking curve called the marginal revenue product curve. It has the same properties of the curve that we’ve been studying.

So let me conclude by making one observation then about this new rule for profit maximization of firms in the short run, this new rule that wage is equal to the value of the marginal product, or a wage is equal to the marginal revenue product. And that is that if you look closely at this rule, it’s exactly the same as the old rule that we use for profit maximization, which was price equals marginal cost.

Let me go through this argument carefully with you. The rule that we’ve just derived says that the firm will keep hiring labor until the wage is equal to the value of the marginal product. Now the value of the marginal product is going to be the price of the good multiplied by the marginal product of the variable input. So price times marginal product is equal to the wage for the profit-maximizing firm. Now I’m going to move these symbols around just to get a little bit more room on the board. But watch what happens when we carefully consider what the marginal product is. The marginal product of labor is the change in output that you get when you hire an additional worker.

Well, the reciprocal of this marginal product is going to be the amount of labor that you need to hire to produce one additional unit of output. Follow my logic – if one additional worker can produce four television sets, then you need only 1/4th of a worker, or 1/4th of a worker’s time, to produce an additional television set. So if we multiplied both sides of this equation by the reciprocal - that is, the labor requirement, how much labor do you need to produce an extra unit of output – we get a very familiar equation.

Look over here on this side. Here we have the changing labor necessary to produce another unit of output. Multiply by the wage – that is, what you have to pay to get that extra worker. Well, that’s nothing more than the marginal cost of producing another television set. The labor you need multiplied by the wage it takes to hire that labor. So let’s go ahead and re-label that expression “marginal cost.” Now look at the other side of the equation. Here we have change in labor, change in labor, change in output, change in output. This term simply cancels; multiplication leaves us something equal to 1.

And here you have it, a familiar rule. The firm will continue adding workers, continue expanding output in search of higher profits, until marginal cost is equal to price. And that’s the familiar rule for profit maximization. So whether you look at it as a matter of choosing your output, or whether you look at the firm’s demand for labor, wage equals value of the marginal product. It all boils down to the same intuition, the same decision.
The Derived Demand for Labor

Deriving the Least-Cost Rule

When we’ve talked about the firm’s profit-maximizing decision before, we’ve usually considered the case as a short run, where some inputs were fixed and one input, usually labor, was variable. But that rules out a bunch of interesting questions, such as: How does the firm choose the combination of labor and capital to use to produce a given amount of output? That’s a question for the long run, when all factors can be varied; in particular, when the ratio in which capital and labor are employed can be chosen by the firm to minimize the cost of producing any given level of output.

Let’s consider that question now. How will the firm combine labor and capital in the long run when it’s free to vary all of its inputs? What rule will it use for deciding the mix of capital and labor to employ to produce its output? The rule is called the “least cost rule.” That is, the firm will use whatever combination of labor and capital minimizes the cost of producing its target output level. Let’s see how that works. Suppose a firm has only two inputs – labor and capital. “Capital” refers to the tools that are used, and “labor” to the people who show up and work there. And suppose that each of these factors has its own price, and let’s suppose for the time being that that price is given in competitive markets. Let’s suppose that the wage, or the price of labor, is equal to $10.00 per unit, and the price of capital is equal to $5.00 per unit. By the way, in economics, we always use the letter “K” for capital, even though the English word “capital” starts with “C” – I think that’s in order to distinguish from consumption, which is often represented with a “C.” Anyway, I find that confusing sometimes, but P_K refers to the price of capital.

So in what ratio is this firm that wants to minimize the cost of production going to employ labor and capital? Should we hire an extra worker, or should we get an extra tool? Or have we found the combination of labor and capital that’s somehow un-improvable, that maximizes our profit, and contains our costs. Well, here’s the rule it turns out to work: Keep hiring labor and capital up to the point at which the marginal product of each factor, divided by the price of that factor, is equal. For instance, the profit-maximizing condition requires that we keep hiring labor until the marginal product of an extra worker, divided by the price that we would have to pay to get that worker, is equal to the marginal product of an extra unit of capital, divided by the price that we’d have to pay to get that capital.

Here’s the intuition: Suppose that the marginal product of labor is equal to six. That means for $10.00 more you can add six more television sets to your output by hiring another worker. And suppose the marginal product of capital in this same story is equal to two. That is, if you spend $5.00 on another tool, you can add two more television sets to your output. Well, which of these is a better deal? If you spend $10.00 on tools in this case, you would employ two tools, with a marginal product of two television sets per tool; that’s a total of four additional TV sets for $10.00 spent.

On the other hand, if you spent $10.00 on an extra worker, you could get six additional TV sets, which is a much, much better deal. Since this ratio is greater than this ratio, the firm would do better by increasing its labor employment. In fact, it should move some of its budget away from capital expenditure towards hiring more labor. Since you’re getting more bang for your buck, literally you’re getting more television sets per dollar spent on labor than you are per dollar spent on capital.

Eventually, if the firm scales back its employment of capital, the marginal product of capital will rise. And suppose we came to the point at which the marginal product of capital were equal to three – that is, each additional tool employed produces three more television sets. In this case, the ratios are equal – $10.00 spent on capital gives you six additional television sets; $10.00 spent on labor gives you six additional television sets. The ratio of marginal products to price of factors, are identical across the factors of production. There is no way for you to reallocate your production budget from capital to labor, or from labor to capital and increase your output per the amount of money that you’re spending. This is the way the least-cost rule works: keep employing factors of production until these ratios are equal, and when they are, you’ve found a cost-minimizing method of producing the output that you’re producing.
Resource Markets

The Derived Demand for Labor

Analyzing the Labor Market

What causes the market wage to change? How do economists explain the equilibrium wage? Well, if you think about it, the wage is just the price of labor, and we certainly know how to explain prices – they’re set by the interaction of supply and demand. So to understand the wage, we need to think about tools we already have for explaining other prices. Let’s look at the demand for labor, the supply for labor, and the way they interact, to determine the wage, which the labor market clears.

First, we’ll do a quick review of the two sides of the labor market – the demand for labor, and then the supply of labor. The demand for labor is determined by the behavior of profit-maximizing firms. Let’s look first at the individual firms’ demand for labor, and it’s simply that firm’s marginal revenue product curve – that is, the curve that describes the amount of revenue added by each additional worker that the firm hires. The profit-maximizing rule is that a firm will continue to hire workers as long as the marginal revenue product exceeds the wage – that is, as long as each additional worker adds more to the firm’s revenue than it adds to the firm’s costs.

Anytime you’re looking at a graph like this in economics, it pays to ask three questions, and if you can answer those three questions, then you really understand the story that’s in front of you. The first question is, what is the story or the relationship that’s being described in this picture? In this picture, it’s a relationship between the wage in the market and the amount of labor that an individual firm finds it profitable to hire. The second question you should ask is, what accounts for the slope of the curve? That is, in this particular case, why is the demand for labor, or the marginal revenue product curve, downward sloping? And the answer depends on the marginal product of labor. Remember each additional worker that you hire eventually adds less output than the previous worker. The marginal product of labor diminishes as you hire more and more workers, and that means each additional worker adds less to the firm’s output, and therefore less to the firm’s revenue, than the worker previously hired.

You can also think about the elasticity of the demand for labor, which influences the slope of the curve. Whenever labor demand is more inelastic, then a given change in the wage has a smaller impact on the firm’s labor hiring than when the demand curve for labor is more elastic. For example, when the marginal product of labor is changing very rapidly, a small change in the wage will cause a firm to make a very large change in the amount of labor that it hires. But if the demand for labor is more inelastic, then the same change in the wage results in a small change in the firm’s quantity of labor demanded.

The third question you want to ask about any graph in economics is, what would cause the curve to shift? So, for instance, if we start with a given marginal revenue product curve describing the demand for labor by some particular firm, we want to know what would cause the firm now to want to hire more workers at every wage. That is, what would cause the marginal revenue product curve to shift to this new position, where at any given wage now this firm wants to hire more labor? Well, remember this curve represents the amount of revenue that each additional worker adds to the firm’s revenue. So there are two things that would cause the marginal revenue product curve to shift. The first is if the price of the good or service that this firm is selling were to increase. If the price increases, then workers producing a given amount of output are actually adding more revenue to the bottom line of the firm, because the price at which this output can be sold is higher.

The second thing that would cause the curve to shift is an increase in labor productivity – if either technology improves so that workers know how to make more output from the given amount of input, or if the workers get more capital to work with, or anything that increases productivity is going to shift outwards the marginal revenue product curve, which means that at any given wage, the firm finds it profitable to hire more labor than before. So here’s a complete description then of the demand for labor on the part of an individual firm. Let’s move now from the case of an individual firm to the market for labor as a whole, and look at how the overall demand for labor is calculated. We use a method called “horizontal summation.” We find the total demand for labor by adding up the labor demands of the individual firms in the market. Suppose we’re in the market where firms are producing television sets, and they’re hiring workers to assemble those television sets. Here is the labor demanded by Firm A, here’s the labor demanded by Firm B. And in this graph, we’re going to calculate the market demand for labor by summing the labor demands of the individual firms.

So suppose we took an initial wage – W0 – maybe this is $10.00 an hour – and at $10.00 an hour, Firm A demands this quantity of labor right here. So I’ll measure the amount of labor that’s demanded at $10.00 an hour, and I’ll go over to the market demand curve and make a little mark indicating that Firm A wants this quantity of labor at that wage. Then I’ll look at Firm B and find that at $10.00 an hour, Firm B wants this quantity of labor. So I go over here,
Resource Markets

The Derived Demand for Labor

Analyzing the Labor Market

take Firm A’s quantity of labor demanded, and add on Firm B’s quantity of labor demanded, and now I’ve got a point on my market labor demand curve. At a wage of $10.00 an hour, the two firms together want this quantity of labor. I found that quantity by summing horizontally their demand curves.

Now suppose that the wage rises from $10.00 an hour to say $20.00 an hour, and I’ll call this point W₁. I could simply find the new point on Firm A’s labor demand curve and the new point on Firm B’s labor demand curve and add them horizontally to get a new point on the labor demand curve. But it turns out that that method is less than thorough. Why? Because at this higher wage, both firms are going to find it profitable to hire less labor, which means that fewer television sets are going to be assembled, and if these are the only firms making television sets, television sets are going to be scarce, and their prices are going to rise. Remember, higher price increase the marginal revenue product for the firm. The output of each additional worker can now be sold at a higher price. So I have to represent the rise in price of television sets by shifting the marginal revenue product curve for each of the firms involved. So, with new higher prices, or scarcer television sets, the marginal revenue product curve at Firm A is going to shift upwards. So I’ve now got the marginal revenue product for Firm A, and we’ll call this Curve-1, representing the new higher prices after the change. And the same is going to be true for Firm B – marginal revenue product for any quantity of labor hired would be larger, because the price of TV sets have gone up.

Well, with the higher price of television sets, we now look at the interaction between the new higher wage and the new shifted marginal revenue product, and I’m going to measure the amount of labor that’s going to be hired at that particular wage – this new higher wage – and go over here and mark it in my market labor demand graph. Then I’m going to do the same thing for Firm B – look at this new wage of $20.00 an hour and find out how much labor Firm B wants to hire – it’s going to be a smaller amount than before – and I record it in my graph. With the new higher prices and the new shifted-outward curves, I find a new point on my labor demand curve, and this new point takes into account not only the productivity of the firms, but also the higher prices of television sets when overall production falls.

So here’s my market demand curve for labor. My overall market demand curve for labor includes the change in the price of television sets when all firms hire fewer workers, as well as the underlying productivity of labor. So there you have it – the demand curve for labor in the market.

Now the next step in this discussion is going to be a consideration of the supply curve of labor. And the supply curve of labor, of course, represents the behavior of individual households – people trying to decide how much labor they want to supply in response to changes in the wage. People make a decision about how much of their time they want to enjoy as leisure, and how much of their time to devote to work so they can make money to buy stuff. Well, think about this. It makes sense to imagine that as the wage rate rises, households will supply a larger quantity of labor, because it’s more rewarding to spend your time at work so you can then buy stuff. Each additional hour buys you more stuff as the wage rate rises.

After some point, however, it may be that the supply curve of labor actually bends backwards. That is, higher wages actually lead households to work less. This is because the income effect of the wage is to lead you to feel richer. If you can pay your bills with fewer hours at work, why not actually reduce the quantity of labor that you supply, and enjoy a given standard of living with more leisure? On the other hand, it may be that the substitution effect continues to dominate, and as the wage rate rises, households will supply a larger quantity of labor, because it’s more rewarding to spend your time at work so you can then buy stuff. Each additional hour buys you more stuff as the wage rate rises.

So whether the supply curve of labor actually bends backwards at some point, whether the income effect of higher wages actually begins to dominate, may not be so important. We know that at least over some important range, the supply curve of labor is upwards sloping – for individual households and for the market as a whole. As the wage rate rises, households find it profitable to spend a larger quantity of their time at work, and a smaller quantity of their time in leisure. So ask the three questions: What’s the relationship? The relationship is between the market wage and the amount of time that households are going to supply the factories in the form of labor.

Second question: Why does the curve have the upwards slope? Because, as the wage rate rises, households are willing to supply a larger quantity of their time to labor, because labor has a higher reward – each additional hour that you spend on the job gives you a bigger reward in terms of stuff you can buy. And finally, what would cause the curve to shift? Well, if households decided that they were willing to work more hours at any given wage rate, the curve would shift outwards. That is, at any given wage rate now you would have a larger quantity of labor supplied by households.
Resource Markets

The Derived Demand for Labor

Analyzing the Labor Market

as a whole. This could be if income taxes were reduced so that any given wage now translated into more disposable income than before. It could also be if households are suddenly presented with bills that they didn’t have before, and they suddenly needed to work longer hours, or just find it profitable, or their preferences shift to work in favor of leisure. There are several things that could account for this shift; but anything that would lead households to want to work longer hours at any given wage rate would shift the curve outwards.

Well, now we’ve got the demand curve, and we’ve got the market supply curve. Our last step is to put the two together to calculate equilibrium, and that’s what we’re going to do now. So we go to a fresh graph, in which we’re telling a story about a relationship between the wage rate and the quantity of labor, and we’re going to put two stories in this picture at once.

The first story is the demand curve for labor – that is, the marginal revenue product of the individual firms added up, accounting for the effect of changes in labor hires and the prices of the product they produce. And the supply curve for labor, which describes the behavior of households – how much of their time are they willing to devote to factories and the production of goods and services, in exchange for wages that they can spend on goods and services that please them. So if we’ve got a demand curve and a supply curve, well, you know how to go from there. You find the place where the curves cross, and this is the point at which the quantity supplied is equal to the quantity demanded. This is the point at which there is no further tendency to change, and an equilibrium wage is established.

Let’s see how that works. Suppose we have a wage that is below the point of crossing – that is, an equilibrium wage is up here, this wage is below equilibrium. At that point, firms are really interested in hiring workers, because it’s profitable to hire a whole bunch of workers because workers are inexpensive and product a whole lot of output. So the quantity of labor demanded at this point is quite large. On the other hand, households are uninterested in working in factories and other employment opportunities, because the wage is relatively low. If this is the market for television workers, then assembling televisions doesn’t pay as well as the opportunity cost of most people, and they’re not going to devote their time to that industry. Well, what have we got? The quantity of labor demanded exceeds the quantity of labor supplied, and we’ve got excess demand for labor. What’s going to happen? Because factories aren’t able to employ workers – workers that would be profitable for them to have on the assembly lines – factories are going to begin to bid against one another, and the wage will be pushed up by the bidding mechanism. And as the wage rises, some firms find it unprofitable to operate, and they simply don’t hire as many workers as before.

On the other side of the market, as the wage rises, more workers will enter this market, drawn away from competing opportunities.

Finally, as the quantity demanded shrinks, and the quantity supplied increases, eventually an equilibrium wage is established at which the quantity of labor supplied equals the quantity of labor demanded, and we’ll call that wage $W^*$. And the quantity of labor in which the market clears we’ll call $L^*$. You can take it from here – you can do all kinds of comparative statics exercises. Suppose the price of this product increases – the demand curve for labor would shift out, creating excess demand at the original wage, pushing the equilibrium wage upward. Suppose income taxes are cut so that more workers want to work at any given wage rate? That would drive down the equilibrium wage in this market. Suppose technological progress increases the marginal product of labor? Shift out the demand curve and tell the story that follows.

So the market for labor is really just like the market for any other good or service, and the price is the wage – the equilibrium wage is set where the quantity demanded is equal to the quantity supplied. So what this picture does is it takes the behavior of firms that are pursuing profits, the behavior of households that are pursuing maximum satisfaction in the use of their time, and combines them so that the forces balance to give us a quantity and a price at which the market for labor clears.
Resource Markets

Monopsony

Understanding Labor Market Power and Marginal Factor Cost

We've discussed the labor market, and the firm's demand for labor; and we've maintained the assumption that the wage is constant, or given. This will be the case if the labor market is competitive, and a firm can hire all the labor it wants to at the going wage. But in some cases, firms don't have that luxury. Suppose you're the major employer in an area. If you want to hire more workers, you've got to coax more people into the labor market by raising the wage you're willing to pay, and that means you're going to have to pay that new, higher wage to everyone, including the people who are already working for you. If this is the case, you're going to think twice about adding an extra worker, because it's not just a matter of paying the wage; it's a matter of increasing your labor cost when you have to pay everyone a new, higher wage.

This is the situation anytime a firm is a monopsonist – the only buyer of a particular good, service, or, in this case, factor of production – like labor. If you want to buy more of something, you're going to wind up raising the price to get more. And that new higher price applies to all units, including those units you could have hired at a lower price.

Suppose now we consider the case of a single firm that is in a monopsonist position in the labor market – you're the only major employer of a particular kind of labor in a particular market. You can hire any number of workers, but to hire additional workers, you've got to raise the wage that you're willing to pay in order to coax extra workers into the labor pool.

Suppose now that you want to hire one worker. Well, to get one worker, you have to meet his reservation wage of $10.00 an hour. That means, to get one worker, your total wage bill is going to be $10.00. If you want to get a second worker now, you've got to raise the wage to be high enough to attract him into the labor pool, and it looks like that wage is $15.00. However, if you're going to raise your going wage to $15.00, not only do you have to give that $15.00 to the second worker, you've got to give it to the first worker as well. So your total wage bill has now increased from $10.00 with one worker to 2 times 15 – or $30.00 – to have two workers. That means the marginal factor cost – the extra money you have to pay to move from one worker to two workers – is $30.00 minus $10.00, or $20.00. So the marginal factor cost is actually greater than the wage that you pay the second worker, because in order to attract the second worker into the labor force, you've got to raise the wage that you pay to the first worker as well.

Suppose now you want to get a third worker. Well, the third worker is only going to work if the wage rises to $20.00 an hour. But remember, if you pay her $20.00 an hour, you've got to pay workers number one and two $20.00 an hour as well. Three workers now involves a total labor cost of 3 times 20, or $60.00. Before you were getting your workers for a total of $30.00, so $60.00 minus $30.00 means that your labor costs have increased by $30.00 as you brought the third worker into the firm, and so forth.

By now the point is clear, in order to bring the fourth worker in, you've got to raise the wage to $25.00 an hour, and to get the fifth worker in, you'd have to raise the wage to $30.00 an hour. In each case, you've got to raise the wage for all of the workers, and that extra cost is going to guarantee that the marginal factor cost is always higher than the wage. We can label this curve MFC – the marginal factor cost – or the change in your total labor cost that results from hiring an extra worker. Marginal factor cost is greater than the wage, because not only do you have to pay the new higher wage to the marginal worker, you've got to pay it to all of the inframarginal workers as well. The only way to get more labor is to keep raising the wage.

So you can see there's a big difference between being a firm who can take the wages given – that is, assume the labor market is competitive – and a firm that is influencing the wage by its decision to hire labor, a firm that acts as a monopsonist, for instance.

Well, once we've got this marginal factor cost curve developed, we can put back in the marginal revenue product curve from other discussions. The marginal revenue product tells you how much revenue is added when an additional worker is hired. A firm that has to worry about rising wages as it hires more workers is going to keep employing labor up to the point at which the marginal factor cost is equal to the marginal revenue product. Keep hiring workers as long as those workers are adding more to your revenue than they're adding to your costs. Finally, at the point where the green curve and the red curve cross, we've got profit maximization. If a firm hired any more workers, it would be actually reducing its profits, because the output that it's producing has got the market saturated to the point that the price is falling; the workers are becoming less productive as they congest fixed inputs in the short run.
Resource Markets

Monopsony

Understanding Labor Market Power and Marginal Factor Cost

And finally, the new innovation in this lecture is, to get those extra workers, you've got to bid the wage up, and that higher wage applies to all of the other workers as well. So, in general, if you're a firm with market power in the factor markets, and the goods markets, when you seek to maximize your profits, you've got to be worried about three things. Thing number one: As you increase your output, labor congests the fixed inputs in the short run; as congestion sets in, marginal productivity diminishes, marginal costs rise. Thing number two: As you produce more of your good, the price in the output market falls; you've got to be concerned about saturating the market. And number three: In the factor markets, as you hire more labor, the wage rises, increasing your total labor costs. So the right balance among these forces: Increasing your revenue, managing your costs, managing productivity is what leads to profit maximization.
Resource Markets
Capital Markets

Analyzing Capital Markets

Capital refers to any good that’s used to produce other goods. When we talk about a firm’s capital stock, we’re talking about the tools that are available to that firm to produce its final goods and services that it sells to customers. So, one example of capital is the factory that the firm produces output in. Another example would be the equipment that’s in that factory – drill presses, lathes, automatic sanders, painting machines – any tools that are used to produce goods and services.

When a firm is trying to decide how much capital to employ, it uses a logic that’s very similar to the logic it uses in its labor decision – that is, should we hire an extra worker? And the answer depends on the revenue that that worker adds and the cost of getting that worker to work in your factory. A firm continues to employ labor until the extra revenue is equal to the extra cost, or the marginal revenue product is equal to the wage.

The decision that the firm uses in employing capital is quite similar. Capital has a special property, and that is, it is durable. If you buy a factory, you can use it year after year; if you put a machine to work, it’s going to be producing output for you for many periods to come. Now, of course, a certain amount of that capital is going to wear out each year. We call that depreciation. And because of depreciation, you’ve got to repair your factory from time to time, you’ve got to spend money to service your machines, and so forth. But, for the most part, capital is defined by its durability. It is a tool that lasts; it is an investment; it is an asset that creates profits for the firm into the future. So when we consider the firm’s decision about how much capital to put to work, we’re thinking about an investment decision. How much should we pay for a particular asset, given the returns that we expect to get from it in the future?

Let me lay out this problem systematically. If our firm under consideration – suppose a television manufacturer – is considering buying another assembly plant, what kind of logic will it work through to decide whether or not to buy the factory? Well, the factory’s got a price tag sticking on it, and the firm can pay the price or it can let this factory go. If it pays the price, it has gained access to an opportunity to earn profits. Well, how much profit? The firm is going to sit down, look at the capital, consider different strategies for using that to make television sets, decide how much labor it would have to hire to man the factory floor, and it’s going to come up with a calculation for the number of television sets that it can produce in this factory – say 1,000 television sets a year. One thousand television sets a year multiplied by the price of $100.00 per television set is $100,000.00 worth of revenue that this factory will be creating.

Well, we’ve got to subtract from that revenue the cost of making those television sets in the factory, and that’s going to depend on how much we have to pay the workers who are going to be doing that task. Let’s suppose we add up wages times workers employed, and we find that the total labor bill for cranking out those 1,000 television sets is $90,000.00. That means $100,000.00 worth of revenue minus $90,000.00 worth of variable costs equals $10,000.00 worth of profit each year from this asset.

Well, $10,000.00 a year worth of profit is a stream that’s going to persist into the future. Maybe that also includes the cost of offsetting the depreciation of the capital asset if we had to throw in a little money to fix windows and grease the wheels from time to time. But any way you look at it, you’ve got an income stream now that you expect if you buy this asset. And here I’m writing down this income stream – benefit minus cost, or the revenue you get from producing and selling those television sets minus the money that you have to spend on variable inputs and depreciation to make those television sets come out the other side. What you do is you take each year’s profit and you discount it by 1 plus the interest rate.

So what you’re doing here is calculating the present discounted value of the stream of profits that you expect if you own this factory and employ it in the production of television sets. So the first year you discount profits by 1 plus the interest rate, the second year you discount the profits by 1 plus the interest rate squared, and so forth, all the way up to the Nth year – which could be way out towards infinity – in which you discount the profits earned by 1 plus the interest rate, raised to the power of N – this is the present value of the profit stream.

Now, if you know the interest rate, then you can plug the interest rate in here and calculate present value. Alternatively, you can think about this problem from a different angle. Suppose what we know is not the interest rate, but we know the price that we have to pay to acquire the asset. If we write here the price of that factory – and suppose the price of that factory is $200,000.00 – then $200,000.00 is the price. We take annual profits from this factory of $10,000.00, and we can solve this complicated equation in terms of R. That is, what is the interest rate that would make $200,000.00 – the price of the factory – equal to the present discounted value of the profit streams that the factory is going to generate whenever we stock it with workers and make TV sets?
Resource Markets

Analyzing Capital Markets

The interest rate that solves this problem is called the “internal rate of return” on this factory. That is, if this factory were a bank account, and you put $200,000.00 into it, what would the interest rate have to be that would allow you to take out $10,000.00 each year, the same way that this factory allows you to earn $10,000.00 in profits? The answer, if this problem stretches on out towards infinity, is about 5%. So at an interest rate of 5%, an annual profit of $10,000.00 has a present discounted value of $200,000.00. So, if the actual interest rate in the market – that is, if the opportunity cost of capital, the rate of return that you could earn by taking that same $200,000.00 and using it to purchase a bond, or to put it right into your bank account – if the interest rate is 7%, then you would do better to take the money that you were going to spend on that factory and invest it somewhere else. Put it in the bank, or buy a bond, earn 7%; don’t put it in this factory and get a rate of return of only 5%.

On the other hand, if the interest rate in the market is only 3%, then this factory looks awfully good by comparison; the internal rate of return is high relative to the opportunity cost of capital. In fact, it even pays you as an entrepreneur to go out and borrow money at 3%, buy this factory for $200,000.00, and get an annual rate of return of 5%; borrow money at 3%, earn 5% on it, and there’s your profit on the capital.

So the rule for a firm is: Borrow money as long as the market’s interest rate is less than the internal rate of return on the projects you’re considering. If the internal rate of return is lower than the market interest rate, then you would do better to forego the investment opportunity and invest any surplus funds you may have in the bank or in other bonds. By all means, don’t borrow money at an interest rate that’s higher than the internal rate of return on the project you’re considering.

So how can we turn this logic into a demand curve for capital? Well, here’s what we can do. We write the opportunity cost of capital on the vertical axis – this is the interest rate – and on the horizontal axis, we write the amount of capital that the firm is going to acquire in a given period of time. Now, capital is kind of a hard thing to measure. I mean, what is the unit in which you measure capital? I mean, some capital is factory building, some of it is tools, some of it may be patents, and other intellectual properties.

So the way we’re going to measure capital is in total dollars that the firm spends on capital – that is, the firm’s capital stock will be measured in terms of dollars invested. So let’s suppose that a firm can earn an internal rate of return of say 10% on a project that costs $200,000.00. So at 10% interest rate, the firm would be willing to invest in that $200,000.00 project. When the rate of return drops down to 8% – that is, when the market rate is 8% – the firm finds it profitable to invest in a project that has an internal rate of return at 8%. And let’s suppose that project costs $300,000.00. So add $300,000.00 to the $200,000.00 that you’d spend on the 10% project to the total amount of investment spending of $500,000.00. Then keep going down. Look at projects with internal rates of return that are lower. Here’s one with an internal rate of return of 7% that costs $100,000.00. Here’s one with an internal rate of 6% that costs $400,000.00. And at each internal rate of return, add on those projects that have that internal rate of return.

So here we go – eventually we can connect the dots, and this gives us the firm’s demand curve for capital – how much money the firm wants to spend on investment projects as a function of the market interest rate. When the market interest rate is very high, there aren’t very many projects that are worth undertaking, because the internal rate of return on most projects is going to be lower than the market price of capital.

So when the market price of capital is high, the total amount of investment spending that the firm wants to do, the total amount of capital that it wants to acquire, is low. On the other hand, at a low interest rate, there are lots of projects whose internal rate of return exceeds the opportunity cost of capital. And, therefore, the firm is going to be doing a lot of investment spending, it’s going to be acquiring a lot of capital.

So here you have it – the demand curve for capital. Firms make a comparison between the internal rate of return on an investment project, like buying a factory, and the amount of interest they would have to pay on the money they borrowed to get into that project. When the interest rate is high, few projects are profitable. When the interest rate is low, lots of projects have internal rates of return that are above the opportunity cost of capital, and firms acquire more.
Market Failures

Overview of Market Failures

Understanding Market Failures

By now, you've got the impression that economists really like free markets. Economists like free markets because of the belief that free markets deliver the maximum economic value with minimum administrative costs. Remember, this is the picture we used when we talked about the ability of the market to deliver an efficient outcome. Here's the quantity of the good traded, here's the price, at which the good is traded, and we're ready now to draw a supply and demand diagram to represent a market.

Here's the downward sloping demand curve. In normative economics, the downward sloping demand curve indicates the marginal social benefit of producing an extra unit of the good. This is what someone is willing to pay to get the next loaf of bread, or the next gallon of gasoline, or the next paperback novel. The height of the demand curve then represents the marginal social benefit created by an extra unit of the good produced. And the area under the demand curve represents the total social benefit of a certain quantity of that good provided to consumers.

The supply curve represents the cost of providing that good; that is, these reservation prices for sellers can be interpreted as the marginal social cost of bringing those goods to market. This is the cost of putting on the market an extra loaf of bread, an extra gallon of gasoline or an extra paperback novel. As we move up the supply curve, the increasing opportunity cost is reflected by the higher reservation prices. But the marginal social cost is shown all along this blue curve and, if we add up the cost of providing a certain quantity of output, we'll get an area under the supply curve, up to that quantity.

So, if we subtract cost from benefit, what we're left with is economic value. And the maximum economic value we can get out of the market is this point right here, where the margin of social benefit is finally equal to the marginal social cost. All of these units back here on this side of the graph, from zero up to the quantity where the curves cross, all of these units add economic value, because the benefit is greater than the cost. Every additional loaf of bread that we produce in trade is benefitting consumers more than it cost producers. And because of that, we are creating economic value. Remember, economic value is the difference between the benefit people derive from the good and the cost of providing it. The difference between benefit and cost is value. It's a kind of surplus or profit. So, we maximum economic value by producing all of those units, for which benefit is greater than cost, and that will be all of those units up to the point, at which the demand curve finally dips below the supply curve.

So, the free market takes us to this point. The bidding mechanism takes us to the point where the quantity demanded equals the quantity supplied. And we like to call that Q* and P* because it's the equilibrium point, but also because in a free market, if the demand represents social benefit and supply represents social cost, this is also the point, at which economic value is maximized. None of these additional units over here should be produced, because the cost exceeds the benefit, but all of these units, where benefit exceeds the cost, should be produced. And economic value is maximized with this particular quantity produced. And with this particular price in the free market, all of these buyers come forward to buy the good, and all of these sellers produce and sell the good. We've got the right quantity of the right good going into the right hands and produced by the right sellers. You can't ask for a better outcome. And this is the story of the efficiency of the free market. The free market maximizes economic value. However, in order to get to this story, we have to make a lot of assumptions. If any of those assumptions are violated, it can really mess the story up completely.

What I'm going to do now is give you a list of the factors that can lead to what economists call market failure. A market failure is a situation, in which free trade among agents fails to deliver maximum economic value. A market failure is a situation, in which the free unregulated market gives an outcome that is, in fact, improvable, an outcome that could be improved by something besides the individual choices of all the individual players. A market failure is a situation where supply and demand crossing, giving price and quantity, is not the same thing as maximizing economic value. It is not the same thing as an un-improvable, good outcome.

Well, let's then think about the things that could mess this story up. The first thing that could mess this story up is if supply and demand don't actually represent marginal benefit and marginal cost. What if the demand curve is not a reliable measure of benefit? What if the supply curve is not a reliable measure of cost? In that case, our story no longer makes sense. Just because we're at a point where supply and demand cross doesn't mean we're at a point where we'd extracted all the economic value from the market.

I can come up with a couple of reasons why this might not be the case. The first is if we have wealth effects. With wealth effects, remember, people's willingness to pay for the product is not necessarily equal to the social benefit that
Market Failures

Overview of Market Failures

Understanding Market Failures

that product generates. Poor people, who have limited income, may not be able to register in the market the true social value of them buying a loaf of bread or paying college tuition. If wealth effects are present, income limitations are driving reservation prices more than how much the good is really worth to someone. If the demand curve is not reliably measuring social benefits because of wealth effects, then we can’t rely on this diagram to tell us that wonderful, free market efficiency story that we were telling before. The same thing is true with the supply curve. If the supply curve is affected by wealth effects; if poor people’s willingness to accept payment is somehow distorted by income limitation, then the supply curve can’t be relied upon as a measure of social cost. So, wealth effects is the first thing that can lead to what we call market failure, or a situation where the outcome is not as good as it could be.

The second way of getting a market failure is if the demand and supply curve don’t represent benefiting costs because of externalities. Now, remember, externalities are situations, in which your choice imposes costs on other people that aren’t part of the trade. If I get a flu shot, I’m creating benefits people who aren’t present when I make a decision to get the shot. They aren’t present when I decide whether to buy the shot or not. If I produce bread, using the smokestacks that soil the laundry of the people next door, well, they’re not present when I make the decision to produce and sell the bread, but I’m still imposing a cost on them. That’s an externality. With externalities, the social benefits are not equal to the private benefits. My private benefit of getting a flu shot doesn’t include the benefit that I’m creating for you, but the social benefit does. You have to add my private benefit to the benefit that I’m creating for other people, the external benefit, in order to get the true social benefit. In that case, people’s willingness to pay may not include benefits that they’re creating for other people, and the demand curve is not a reliable measure of social benefit. The same thing with costs. The supply curve reflects the private cost of the baker. It may not include the costs associated with the pollution that he creates. In that case, the supply curve is understating the true social costs of the baker’s bread. So, just because the baker is willing to provide the bread doesn’t mean that the bread is socially profitable. It could be that the reason I’m selling this bread at such a low price is my ability to dump the pollution, the extra costs, on other people. If I had to clean up the pollution myself or live with its consequences, then I might reduce my production of bread. So, the supply curve, in the case of external costs, is not reliable measure of the true social cost of production. The supply curve and the demand curve always reflect private benefits and private costs, but they don’t always reflect social benefits and social costs. In the case of external benefits or external costs, the demand curve and the supply curve are unreliable measures of social benefit and social cost. And, in that case, the place where the curves cross is not necessarily the best place for society.

The third thing that can lead to problems in the market is if there is noncompetitive behavior; that is, if there are agents in the market that have what we call market power. Remember, when we drew this diagram, we assumed that all buyers and all sellers were price takers; that is, they accepted the price as given and responded passively to it. No one believes that they could influence the price by changing the quantity that they were willing to buy or sell. The price was believed to be given and agents imagined themselves small relative to the market, only a small part of the total economic activity. They were price takers, competitive behavior. However, in some cases, it may be that the product is sold by a single company, in the case of monopoly. A monopoly is a great example of what we call market power, the ability to influence the price of the good. The monopolist influences the price of the good by restricting the quantity that she sells and driving up the price. Now, when the monopolist does this, she may be able to make more profit by making her good artificially scarce. There may be units that are socially profitable; that is, the cost of those units may be less than what people are willing to pay to get them. But the monopolist restricts trade, cuts back the quantity that she sells, so she can charge a higher price to people who are willing to pay more. And then she prices people out of the market who would pay a price that covers the cost of producing the good. Then she’s imposing a cost on society. She’s restricting trade, creating a deadweight loss and, in that case, messing up the market. Monopolists make private profit by artificially restricting trade. And when you artificially restrict trade, you’re destroying economic value. People who would be willing to pay a price that covers the cost, but they’re pushed out of the market, so that the monopolist can make more money off of people who are willing to pay higher prices. When market power exists, then we can’t be sure that the free, unregulated market will produce the right quantity of trade. With monopoly and the incentive to restrict trade and jack up prices, when people have the power to manipulate the price, then we no longer have competitive price taking behavior. This diagram doesn’t even really apply anymore. The assumptions behind it have been violated and you usually will get an efficiency, a social loss, associated with the restriction of trade. So, the third source of market failure is market power, monopolists, or anyone else who can affect the price of their product, because they’re looking for private profit, not necessarily profit for society.

Finally, the fourth source of market failure is problems with information distribution. This is what we call asymmetric information, when one party to a trade knows more about the quality of the good being traded than the other party does. A great example of this is the market for used cars. The person who’s selling the car knows more about the
Market Failures

Overview of Market Failures

Understanding Market Failures

quality of the car being traded than the person who is considering buying it. In this case, the buyer’s suspicious that the seller may be trying to sell a bad car, and therefore the buyer lowers the price that he’s willing to pay. This lower price guarantees that people who have good cars are going to bring them off the market. They’re going to pull them off the market because a good car is worth more than the price the buyer is willing to pay, because the buyer’s scared of being taken advantage of. When the buyer knows less about the trade than the seller does, sometimes entire markets can be wiped out. Buyers and sellers can’t get together and agree. Even though the buyer would like to have the high quality product, he is unable to buy it because the risk of getting a low quality product causes him to lower his price, lower the price he’s willing to pay, and that causes the sellers of high quality products to bring them off the market. The market for used cars turns into the market for lemons, the market for low quality cars, because scared buyers won’t pay a high enough price to bring the good cars onto the market. Asymmetric information can cause an entire market to go away. Another example of asymmetric information is why do we not have a market for insurance against traffic tickets? Because once you were insured, you would engage in risky behavior; that is, you would drive faster because there’s no cost to you of getting a ticket. This is what we call moral hazard. It’s another problem of information. Because insurers couldn’t be certain that you wouldn’t break the law after you have insurance, they can’t afford to give you that kind of insurance. So, the market for insurance for traffic tickets simply doesn’t exist. Sometimes entire markets break down because one party knows more about what’s actually being traded than the other party does, either because the quality is uncertain, or because there are choices that will be made that won’t be monitored by both sides.

So, a quick review. The free market is a wonderful thing, under certain circumstances. Under certain assumptions, the free market delivers an outcome that maximizes economic value. But in four cases, there are four assumptions that have to be satisfied to bring us to this point and, if any of them are violated, we cannot trust the free market to give us an un-improvable outcome. And those cases are: wealth effects, when individuals don’t have the wealth to register social benefit or social cost; second, externalities, when private costs and private benefits don’t necessarily equal social costs and social benefits; third, the case of market power, when agents can manipulate prices and create artificial scarcities that give them profits; and fourth, asymmetric information, when the quality of the product being traded is not equally well understood by all parties and people are afraid of being taken advantage of.

Well, these four cases are probably pretty common in the economic world, so can we just then throw out the idea that free markets are good? No, we can’t do that! Free markets are an ideal and they’re an ideal that are more closely approximated in some real world cases than they are in others. I mean, it’s probably the case that, in some financial markets, the free market is a pretty good description of the way things work, or in markets that really are competitive, when there are a lot of sellers and a lot of buyers and the product is very well understood and easily inspected and, in fact, there are low external benefits or costs. In those cases, we use a great description of the way the world works. But when you’re dealing with a situation that is characterized by externalities, or the wealth effects are really important, or a situation where there is a monopolist, or where asymmetric information is central to the whole story, then you want to be careful jumping to the conclusion that the free market gives the best outcome. In these cases, there may be a market failure.

Well, what do we do about a market failure? We’re going to talk about economists’ response to market failure, when we consider these individual cases. We will have individual lectures on all four of these market failures and, in each case, I’ll be discussing how economists propose dealing with them, in order to create some kind of improvement.
Market Failures
Public Goods and Public Choice

Defining Public Goods

Economists have faith that flea markets provide the right allocation of goods and services. That is, the bidding mechanism, prices, supply and demand, leads to outcomes where the right people produce the right things, the right people get them, the outcome is efficient, or un-improvable – value maximized – except in cases of market failure.

Let's consider one case of market failure – the case of public goods. Public goods are goods that, once provided, people cannot be excluded from. Now usually the market excludes us from goods by only giving them to the people who pay for them; if you don’t pay for it, you don’t get it. And that way the market directs goods to the people who want them the most. But if a good is nonexclusive, if it is a public good, we can have some peculiar problems arise.

Consider the case of public television. Once the signal is out there on the airwaves, anyone who has a set can tune it in, which means that you can enjoy it, whether you pay for it or not. If you choose not to pay for it, if you don’t send your check in to public television, public television doesn’t collect enough money to pay for its programs, and before too long, it’s off the air. This is a problem with public goods. We call it the “free rider problem” – if people can get something for free, they won’t pay, and if they don’t pay, then we don’t get the good, or we don’t get enough of it.

To understand public goods better, consider the following strategy for categorizing goods and services. We can start by distinguishing between goods that are exclusive and goods that are nonexclusive. These are the so-called public goods – once the good is provided, no one can be excluded from it. Exclusive goods, on the other hand, might be called private goods – goods that you can only have if you pay. Another distinction we can make between goods is goods that are rival in their use and goods that are non-rival. A rival good is one that, if you’re using it, it’s not available for me to enjoy. A non-rival good, on the other hand, is a good that we can both use at once, a good that if you’re using, I can use also, with no diminishing of the benefits.

Let’s consider some examples in each of these cases. A good that is private or exclusive, and also rival, is the hamburger you ate for lunch yesterday. It was rival because you enjoyed it and I couldn’t because you were eating it; it was exclusive because you couldn’t have the hamburger unless you paid for it. An example of a good that’s exclusive and non-rival is a software package. It’s exclusive because you can’t use it, it’s not licensed to you unless you pay for it. It’s non-rival because your use of that software package doesn’t in any way decrease my benefit from using the same software if I go buy my own copy.

A good that exclusive might be called a “private good,” and these are goods that allocated well by markets. On the other hand, public goods, or nonexclusive goods, are not necessarily allocated well by markets; in fact, markets usually fail whenever a good cannot exclude people from its use. Let’s consider, for instance, the case of a good that’s nonexclusive and rival. One instance would be a freeway. Anybody who wants to can drive on the freeway once the government has provided it.

However, the freeway is rival because if your car is on the freeway, that means there’s less room for my car, and the more people pile on the freeway, the slower traffic moves, and eventually the whole thing is spoiled by overuse. This is what economists call the “tragedy of the commons.” The tragedy of the commons results when it is in individually rational for people to overuse a public good. Because people can’t be excluded from the road, traffic keeps entering the freeway until nobody else wants to get on, and they don’t want to get on because the traffic isn’t moving anymore. The value of the road has been reduced to zero – people would rather take the side streets or stay home.

Another instance of a nonexclusive rival good is a park. Once the park is built, unless there’s a gate and a fee, anyone who wants to can use it. But it’s rival, because the more people who show up at the park, the less room there is for you to play Frisbee, the less there is peace and quiet for you to enjoy under a tree. The tragedy of the commons applies also to the park; once the park is open, people will keep crowding in until it’s no longer desirable to go there, and the park, in that sense, becomes spoiled – it’s no better than any of the other alternatives around.

One way of solving the tragedy of the commons is to privatize, or enclose, the commons. That is, instead of making the road a nonexclusive good, make it a toll road, and exclude people who don’t pay. Put a gate on the park, and charge people an admission fee, so that you can make the good exclusive. So one solution to the problem of the tragedy of the commons is to create a market where one didn’t previously exist. And to do that, you have to be able to turn the good from a public good into a private good.
Market Failures

Public Goods and Public Choice

Defining Public Goods

Another example is the case of the nonexclusive/non-rival good – like public television, or like a streetlight. Once the streetlight has been set up in your neighborhood, anyone who walks on the sidewalk can enjoy its light. The problem here with a non-rival/nonexclusive good is the free rider problem. Who’s going to pay for that streetlight? If no one can be excluded from it, then it may be individually rational for people to behave opportunistically and to not send in their check to public television, or not send in their fee to the people who provide streetlights. What happens then, of course, is we get fewer streetlights; there isn’t enough good programming on public radio because a lot of people listen and don’t pay.

So the free rider problem can be solved if we can somehow make the good exclusive, and then eventually information technology will be such that satellites that are beaming you radio signals will also be extracting money from your account using wireless technology as you’re listening to the radio station. And at that point, radio will no longer be a nonexclusive public good, it will become an exclusive good, just like your subscription to a magazine.

So that’s the problem of public goods. Public goods or nonexclusive goods are subject to the tragedy of the commons and the free rider problem. That way, we tend not to get enough public goods, and the public goods that we do tend to have tend to be overused to the point of being spoiled.

Let’s look now at how an economist would describe the problem of public goods, and the solution. We’ve seen how economists add up the demands of individuals to form the market demand curve whenever we’re dealing with an exclusive good. We call this the process of horizontal summation. We take a price, and we find out how much each individual in the market wants to purchase at that price. For instance, Person A, at a price of $1.00 a taco, may want three tacos a week, and Person B, at a price of $1.00 a taco, may want five tacos a week. Three plus five equals eight, so at a price of $1.00 per taco, we have a total demand of eight tacos per week in this market; that’s the way things work when you’re dealing with goods that are exclusive. Person A’s tacos and Person B’s tacos have to be added together because they’re going to be consumed separately, because we’re dealing with an exclusive good.

On the other hand, if we’re dealing with a public good, we don’t add up quantities. Rather, we add up prices, because once a given quantity has been provided, each individual in the market will be able to enjoy it, and what’s important to the economist is discerning how much benefit in total is created by this public good. So, for instance, suppose we provide a streetlight on the street where A and B live. How much value – how much benefit – does that streetlight create for the society that consists of Person A and Person B? Well, since we’ve got one streetlight, and the same quantity is going to be enjoyed by each person, we don’t want to be adding up quantities at a given price. Rather, we’re going to take the quantity of streetlights as given – one streetlight – and we’re going to add up the benefits that these individuals receive, respectively, from the streetlight. So instead of doing a horizontal summation, we’re going to be doing a vertical summation, and it’s going to look something like this.

To derive the market demand curve in this case, we’re going to fix the quantity and look at how the prices, or values, vary between the members of the market. Suppose we have one streetlight being provided to the society that consists of A and B. Person A values one streetlight at $10.00 – that’s what he would be willing to pay to get it. Person B values the same streetlight at $5.00 – that’s what she would be willing to pay to get it. The total social benefit derived from this one streetlight, then, is the sum of A’s benefit plus B’s benefit – $10.00 plus $5.00 equals $15.00.

We can do the same vertical summation for any given quantity of streetlights – two streetlights, three streetlights, half a streetlight – and in so doing, we will be adding up vertically to construct the market benefit curve for streetlights. One streetlight has a market benefit, or social benefit, of $15.00 – the vertical sum of B’s benefit and A’s benefit. So that’s what this kinked red curve is all about. It’s saying for any given quantity of streetlights that we provide, what’s the total benefit that our society derives from that quantity of public good.

Now, what’s the value-maximizing quantity of public good to provide? How many streetlights should we actually build on this street? Well, that depends on the marginal cost of an extra streetlight. This dotted blue line now represents the marginal cost of streetlights – how much it costs to put one in. And let’s suppose that it costs $15.00 to install a new streetlight. Well, in this case, the marginal cost curve crosses the demand curve, or marginal benefit curve, at a quantity of one streetlight. One streetlight costs society $15.00, and it creates a benefit to society of $15.00 also. So this is the socially optimal quantity of streetlights – one streetlight with a benefit and a cost at the margin of $15.00.
Market Failures
   Public Goods and Public Choice

Defining Public Goods

The last question is who’s going to pay for the streetlight? And if we were doing things ideally, we would assess a tax, or a charge, on Person B of $5.00; we would charge her the benefit she’s getting from the streetlight; and likewise, we would charge $10.00 to Person A, the benefit he is getting from the streetlight. The total benefit, then, is collected in the form of fees, which go to cover the cost of the streetlight. This is how things work in economics.

Now we are stuck with the problem that, how do we get Person A to tell us the truth about how much this streetlight is worth to him? How do we get Person B to tell us the truth? How do we actually go and get them to fork over the money that it takes to build this streetlight? At this point, we probably are going to have to resort to some kind of second-best scheme, like a sales tax or a property tax or some other way of collecting money from these two members of society to provide their streetlights. But in the ideal world, what we would do is charge each person the marginal benefit that they get; equate that with the marginal cost, and it would determine the quantity of streetlights and how much each person paid for them.

See the provision of public goods is a relatively complex business, because people get different amounts of satisfaction out of a good that everyone is enjoying at once. So people, in effect, should be paying different amounts for the streetlight. But that’s kind of a difficult scheme to implement, so we resort to other schemes, like coercive taxes, collecting pools of money, out of which we pay for national defense and streetlights and roads and so forth.

To summarize, then: Whenever you’re thinking about public goods – goods that are nonexclusive – to find the social benefit, rather than summing horizontally, the way we usually create a demand curve in a market with exclusive goods, you sum vertically to find the total sum of the benefits that people derive from goods that they are all enjoying at once. And then you have to have some kind of scheme for collecting the money that’s needed to provide the goods. Usually that’s going to be done through taxes, which will roughly approximate the ideal provision scheme. However, public goods can in general create a market failure. Because they are nonexclusive, they can’t be provided the way hamburgers and software and other goods and services are. And if the goods rivalrous, you’re going to be subject to the tragedy of the commons. And if they are non-rivalrous, you still have the free rider problem. And this is why economists believe that whenever we’re dealing with public goods we need something besides the usual free market to give us the right outcome. It’s why we need some kind of more cooperative or collective resource allocation mechanism – the government’s involvement, a private club to provide these things – something that takes account of the fact that the good that’s being dealt with is fundamentally nonexclusive, and therefore different from most goods that are traded in most markets.
Market Failures
Public Goods and Public Choice
Analyzing the Tax System

If a government wants to provide public goods for its citizens, it has to somehow collect their money to pay for them, and that usually means taxes. Let's consider now taxes from an economic perspective. The first question is how do we raise taxes; on what base are taxes levied? That is, what is it that we're taxing to collect the money to provide for streetlights and public television? The predominant tax base for the federal government of the United States is income; that is, people pay a proportion of their income each year to the government as federal income tax. Alternatively, the government could tax wealth; that is, it could require each citizen to pay some proportion or fraction or percentage of his or her accumulated wealth to the government each year for providing public goods.

Another idea that has been in the press a lot in the last few years is taxes on consumption; that is, rather than paying income taxes, you would pay taxes on money as you spend it. You would take your annual income, deduct the amount that you’ve saved by putting it in the bank or buying stocks, bonds, or other investments. The remainder would be considered your consumption, and you would pay a fraction of your total consumption expenditures to the government in the form of a consumption tax. Now, does it matter whether we tax wealth, consumption, or income? It does, but before we talk about the implications of each of those taxes, let’s develop a few more tools.

Next, a matter of definition: What’s the difference between a proportional tax, a progressive tax, and a regressive tax? A proportional tax is a tax where the revenue a consumer pays to the government increases proportionally with the tax base. That is, if there were a proportional income tax at a rate of say 25%, you would pay 25% of each additional dollar in income that you get to the government no matter what your income was. The tax amount that you pay to the government – your total tax bill – would increase in proportion with your income.

A progressive tax, on the other hand, is one where people pay a larger percentage of each additional dollar that they earn in income as they get wealthier. So, in a progressive tax system, people earning incomes of $30,000 a year might pay tax at a rate of 25% of each additional dollar of income they earned, where people earning $200,000 a year might pay income tax at a rate of 40% of each additional dollar of income earned. A progressive tax is one where richer people pay more to the government in the form of taxes, because each additional dollar is taxed at a higher rate as your income gets larger and larger. And this is what is meant by the idea of tax brackets. If you’re earning $30,000, you’re in one tax bracket with a lower marginal tax rate; if you’re earning $200,000 a year, you’re in a higher tax bracket, where each additional dollar has to be paid to the – a larger percentage of tax. That is, the marginal tax rate is higher on the $200,000 income.

What about a regressive tax? A regressive tax is one where poor people actually pay a larger percentage of their income than richer people do, or a tax that bites harder on lower incomes. Now, very few people would propose that we actually have a regressive income tax reflected in tax brackets. That is, no one would say that someone making $20,000 a year should pay 50% of their income in taxes while someone who is making $200,000 a year ought to get by with a tax rate of 5%. But some taxes wind up being regressive, because they fall especially heavily on poor people. For instance, taxes on food are regressive, because poor households spend a larger percentage of their income on food, and therefore are paying a larger percentage of their income in taxes if they’re being raised by excise taxes on groceries. Whereas rich people, who don’t really buy that many more groceries than poor people do, would be paying a much smaller fraction of their income through that particular tax.

So, proportional tax is one that increases proportionally with income; a progressive tax is one where the tax rate rises with income; and a regressive tax is one where the tax rate falls with income.

The next distinction we want to make is between the marginal tax rate and the average tax rate. I’ve already alluded to this; let me see if I can make it very clear. The marginal tax rate is the fraction of each additional dollar of income that someone pays in taxes. The average tax rate is the ratio of the total tax bill to your total income. I think I can make this clearer with a numerical example. Let’s suppose that the tax formula for our income tax is given by the following equation: that people are going to pay taxes on all of the income – “Y” – that they earn beyond a certain amount – let’s say $20,000. So every dollar that you earn in addition to $20,000 a year, you have to pay, let’s say 25% of it in taxes. So our formula for total taxes that you owe as a household is your income minus $20,000 times 25%.

So if your income is $20,000 in this case, plug $20,000 into the formula, minus $20,000, you don’t owe any taxes at all. Suppose now that your income is $50,000. Well, each additional dollar of income now that you’re earning above $20,000, you’re having to pay 25% of it in taxes. So your marginal tax rate is going to be 25%. In fact, your marginal
Market Failures

Public Goods and Public Choice

Analyzing the Tax System

tax rate is going to be 25% whether you’re making $50,000, $100,000, or $1,000,000, because the change in your tax bill that results from an increase in your income once you’re past $20,000 is equal to 25%.

Your average tax rate, on the other hand, is your total tax bill divided by your total income, which is going to be $Y minus $20,000 times 25%, over your income, which is equal to – let’s see – 25% of $20,000 is $5,000, and $Y over $Y -- if I do my algebra here, I get this – 25% minus $5,000 over $Y.

So if my income is $50,000, I can plug $50,000 into this formula – $5,000 divided by $50,000 is 10%, 10% subtracted from 25% – my average tax rate is 15%. That is, my total tax bill is going to be $50,000 minus $20,000, or $30,000 times 1/4th, which is $7,500. And $7,500 is 15% of my total income of $50,000.

So your average tax rate is the ratio of the taxes you pay to your income. And it’s going to change, because, as our income increases, we’re going to be paying a different percentage of our total income in taxes because of this $20,000 exemption, because we don’t actually start paying taxes until we reach an income of $20,000. So let’s say instead that my income is $500,000. An income of $500,000 – of course, my marginal tax rate is still 25% – each additional dollar of income, I’ve got to pay a quarter to the government. But if I plug $500,000 into this formula – $5,000 divided by $500,000 is 1%; then 25% is 24%. The average tax rate for someone who is making $500,000 a year by this calculation is 24%. Notice, the income tax rate is rising, from zero for someone with an income of $20,000 or under, to an average tax rate of 15% for someone with an income of $50,000, to an average tax rate of 24% for someone with an income of $500,000. So this tax system that’s described by my formula here is a progressive tax system.
Public Goods and Public Choice

Understanding Public Choice

We've seen how the free market fails to provide the efficient amount of public goods, and how government – collective action, cooperative action – can give an improved outcome. So now you might wonder: How does the government work to deliver the efficient amount of public goods? Do governments work according to some different set of rules? Well, some people believe so. But economists are going to be slow to give up their belief that people are always behaving rationally, in self-interest. And if you change the rules of the game, they're just going to learn to play by a different set of rules.

Now the government may work differently than the market, but we can still imagine that with this different set of rules, the people involved – the voters, the politicians – they're all going to be pursuing self-interest in whatever environment they find themselves in. And this helps us to understand what economists call “public choice theory” – the tools of economics applied to understanding how governments work, and how public goods get provided, and whether or not the outcomes are in some sense efficient, or good.

Let's look at some instances of public choice theory applied to understanding how voting works. Now voting is a peculiar thing. Voting is different than the market. In the market, if you want something and you want it really bad, you can pay more money for it. That is, if you want the cheeseburger and I want the cheeseburger, the person who bids the highest price will get it. However, in politics everyone has one vote; therefore, you're limited in your ability to express the intensity of your preferences, and that leads sometimes to some weird outcomes. Let's look at some instances of voting rules and how they play out in public choice situations.

First, let's look at how majority voting can lead to an inefficient outcome. Suppose we are in a city where people are voting on whether or not to build a new park by the highway, and let's suppose there are three people who live in the city – Person A, Person B, and Person C – and they're going to have an election to decide whether or not they are going to pay the taxes that it would take to build this park. And let's suppose the total cost of this park is $300.00, and that if this election is successful – that is, if the park is approved – each person will have to pay an equal share of the cost of the park. So the taxes that are going to be imposed in this case will be one-third of the cost of the park on each of the three voters. So the cost to each voter is going to be $100.00. The total cost of the park is $300.00, and the cost per voter is $100.00.

Now let's suppose that we know something about each of these three voters. What we know is the satisfaction that they would get from being able to play in this park, and let's suppose it's different for each person. Let's suppose the guy on top really likes parks a lot, and if and when he goes to the park, he's going to get a satisfaction of $110.00 worth of pleasure, being able to visit this park daily. Let's suppose that he would pay $110.00 rather than do without the park. The next guy is also a big park fan, but he doesn't enjoy it quite as much. Maybe he lives further from the site of the park, and his benefit is only $105.00. And let's suppose our last citizen is not so crazy about the park; it's worth $75.00 to her – not as much as it is to the other two citizens.

First, let's look at how majority voting can lead to an inefficient outcome. Suppose we are in a city where people are voting on whether or not to build a new park by the highway, and let's suppose there are three people who live in the city – Person A, Person B, and Person C – and they're going to have an election to decide whether or not they are going to pay the taxes that it would take to build this park. And let's suppose the total cost of this park is $300.00, and that if this election is successful – that is, if the park is approved – each person will have to pay an equal share of the cost of the park. So the taxes that are going to be imposed in this case will be one-third of the cost of the park on each of the three voters. So the cost to each voter is going to be $100.00. The total cost of the park is $300.00, and the cost per voter is $100.00.

Now, first question: Is this park going to be approved in an election? Well, look, our Voter A finds that the benefit is greater than the cost – that is, the satisfaction is more than enough to compensate for the extra taxes. He is going to vote in favor of the park – “yes.” The next guy gets a benefit of $105.00 – greater than the taxes he’s going to incur; he likes this idea, he votes “yes” for the park. And our third voter finds the park is not going to be worth enough to her to compensate for the extra taxes; she votes “no.” Well, the park is going to be approved by majority vote. But here’s the rub: The park should not be approved according to economic analysis because the benefits are not enough to compensate for the costs. The total cost of providing this park for this city is $100, $100, $100 – $300.00; the total amount that we had to raise in taxes covered the cost of the park. But how much benefit is provided in total by the existence of this park? – $110.00 plus $105.00 is $215.00, plus $75.00 is only $290.00 worth of benefit. As you can see, the benefit is less than the cost, and by the economic way of thinking, this park would be an inefficient use of society’s resources. But majority vote approves it anyway simply because enough voters find it individually profitable to have the park, while socially the park is unprofitable.

So there you have it, an instance of an inefficient election outcome; majority voting can lead to an inefficient result. Now we're going to see how majority voting may fail to yield any meaningful result at all – what we call the “impossibility theorem of voting.” The implications of the impossibility theorem are best seen in a story. Suppose we have a society where we have a million dollars to spend, and we're considering three possible public goods projects -- a streetlight, we'll call that Project A; a public television station, we'll call that Project B; and a park, we'll call that Project C. Now, which of these three projects does our society really want?
Understanding Public Choice

Now hold on, does that statement really even make sense – “projects that our society wants”? I mean, does society “want” anything? Individuals want things; individuals have clearly-defined preferences, but can we clearly define the preferences of society? This is what the impossibility theorem is all about. It says that there is no consistent way of taking individual preferences and adding them up into the preferences of society that could then be used to guide public choice.

Let me show you how it works. Now we’ve got three voters in our society. Here’s our first guy, and let’s suppose that he expresses his preferences thusly: He thinks that the park is the best use of the money, he likes the streetlight next, and he thinks the public television state is third. Our second voter has got a different set of preferences. He likes the public television station the best, the park second, and the streetlight last. And our third voter thinks that the streetlight is the most important thing, the public television station is second, and the park is third. Now these preferences are clear, right? The individual preferences are clearly stated, but can we add them up now to get the will of the people? Well, it turns out that we can’t; there’s no consistent rule for adding these up, and it’s easy to see the problem that we get by majority vote in trying to choose which of these projects is the most desired by society.

Suppose we have an election, and all three of these projects are on the ballot at once, and we ask each voter to tell us how we should spend the million dollars. Well, this guy’s going to say that the park is the priority, and he’s going to vote for it; this guy will vote for the TV station; and this woman will vote for the streetlight. We’re deadlocked; society has told us nothing – there is no clear will of the people. Suppose instead of putting all three of these projects on the ballot at once, we do what people usually do in an election; that is, we let people make a choice between two alternatives. That is, we have pairwise comparisons. So we take one thing off the ballot, and let people vote on the remaining two items.

Well, let’s see what happens in that case. Suppose we take the streetlight off the ballot. Which of the remaining two projects would win the election? Well, this guy prefers the park, this guy prefers the TV station, and she prefers the TV station; so the TV station would beat the park in an election. Suppose now that we take the TV station out of the picture and make a pairwise comparison between the streetlight and the park. He likes the park better, he likes the park better, and she likes the streetlight better; so the park would beat the streetlight in a pairwise election. Suppose now we take the park out of the picture and make a comparison between the streetlight and the TV station. Well, he prefers the streetlight, he prefers the TV station, and she prefers the streetlight; so the streetlight beats the TV station in a pairwise comparison.

Now hold on, this is beginning to sound like that game of scissors and paper and rock, where the rock can beat the scissors but the scissors can beat the paper and the paper beats the rock; things seem to be going around in circles. Well, that in fact is the case, and if you’re having a set of pairwise comparisons – that is, you’re having one election and then the winner goes up against the remaining candidates, you can wind up with any outcome simply by choosing the order of the elections carefully. Let me show you what I mean.

Suppose what you really want to wind up with is the TV station – Option B – and you’re a clever politician and you’re going to stage a set of elections so that people will actually vote for this TV station. The way to get this TV station in is to make sure that it’s running against something that it can beat. Now you know that the TV station can beat the park; however, you know that the streetlight will also beat the TV station if it goes up against it directly. So you’ve got to arrange this election so that the streetlight is out of contention before the TV station enters the picture. So the first election that you want to see held is between the streetlight and the park. If the streetlight and the park are running against each other, he votes for the park, he votes for the park, she votes for the streetlight, and the streetlight is now out of the picture – people have voted for this park. Then in the second election, you run the park against the TV station, and see what happens. With the streetlight out of contention now, you’ve got two voters – right here and right here – who prefer the TV station, and the park loses, and the TV station is how we’re going to spend public money.

By now it’s clear that any of these three projects could wind up winning a series of elections depending on the order in which the elections are held, and that is a kind of weird outcome. I mean, does the society really prefer the TV station or the park or the streetlight? The answer is, we don’t know, because there’s no consistent clear way of adding up individual preferences to get to society’s preference. There is no way for us to come up with a society preference ordering that guides us in how to spend public money, because we can use any set of rules we want to, to create this social preference function; and, depending on the rules we use, we’ll wind up with a different recommendation. That’s
Public Goods and Public Choice

Understanding Public Choice

the impossibility theorem at work, and its implication is that we should be cautious in the faith that we place in majority voting systems to express the will of the people when it comes to how to provide public goods.

Now we can also use public choice theory to explain the behavior of politicians, and this leads to one more model – this is the model of the “median voter.” Suppose we are in a society where people hold a variety of political views that range from those on the far left to those on the far right. And suppose, just for the sake of our example, that people’s political views are distributed evenly along the political spectrum; that means any little chunk that I take out of the political spectrum here has the same number of people in it – people are uniformly distributed along this line. Now that’s probably not the way things are in society; in fact, people’s politics are probably distributed according to a bell curve, with the majority of people’s opinions clustered toward the center. I could do this example with a bell curve, but it would be a bit more complicated visually, and give us no really useful additional information. So let’s assume for the sake of simplicity that people’s views are distributed evenly along the political spectrum.

Now let’s see what’s going to happen. People are, of course, going to vote for whichever politician is closest to their point on the political spectrum – whichever politician is closest to their particular views. So let’s suppose that politician red locates himself down here at the far left end of the spectrum, and politician green locates herself down here at the far right end. Well, all of these voters down here are going to vote green, and all of these voters over here are going to vote red, and the voters in the middle are going to look both ways to see which politician is closest to their particular position.

So if we take this line and find the point that’s halfway between green and red, we’re going to divide the voters between the two parties. Now red is a strong left politician, and green has strong right views. But if green is pragmatic, green will realize that she can get more votes by moving closer to red, because as green moves this direction towards the left, all of these voters that are down here in the far right into the spectrum still find green the most attractive alternative. And yet now the dividing line is moving further and further to the left; that is, more voters now find that green is closer to their preferred position. So green is going to keep moving, keep moving, keep moving to the left until green is right up against red. See, now green has a big share of the market over here; lots of voters now find green closer to their preferred position, and red is stuck down here with just a little niche on the far left.

So what’s red’s best move going to be? Red’s best move is going to be jump over green to get between his rival and the largest share of the voters. Well, two can play at that game, and green is going to jump over red, and the two are going to jockey back and forth until they find themselves practically on top of each other at the center of the market. The competition for voters leads these two politicians to press themselves towards the center.

Now this is something that we talk about all the time – politicians clustering towards the center; there’s really no difference between the candidates, they wind up saying the same things. But that’s not always going to be the case. What if it’s true that voters in the end, if they don’t find you close enough to their position, simply drop out of the market, because they’re disappointed, and they won’t vote for anyone at all? If it’s the case that moving too far away from the extreme positions causes you to lose votes, then the voters will start to – the politicians will start to spread out, so as to keep the extreme voters in the market. If there’s a chance that you could lose people on a particular issue, then you’ll find yourself spreading yourselves out, moving back towards the ends, in order to bring voters back in. And this happens frequently when a third-party candidate enters the market and threatens to take away the extremes. You’ll see that politicians tend to differ more on issues that people hold extreme and firm opinions about. So “hot button” issues tend to divide politicians, whereas issues that are less controversial and not make-or-break issues for voters tend to lead towards more clustering towards the center.

So the median voter model is really asking us where is the center? Where are most of the voters clustered? If the voters are clustered towards the left, we’ll find both politicians clustered towards the left. If the voters are clustered towards the center, that’s where the politicians will be. Studying where the politicians are clustering together, that is telling us where the median voter is. Where’s the person in the middle, where’s the person that the politicians are trying to get on top of. When the government provides public goods, it collects tax money from constituents and uses that money then to pay for roads, and, bridges and streetlights, and public television, and so forth.

But public choice theory alerts us to the fact that once a big pot of money has been gathered through taxes, or through any method, for that matter, people are going to begin to swarm around it, wanting a piece. That is, people who are fans of public television are going to lobby the government for more public television, and people who are
Market Failures

Public Goods and Public Choice

Understanding Public Choice

Fans of streetlights are going to lobby the government for more streetlights. Special interests form around public goods, as well as transfer payments. Other people will show up just simply wanting a chunk of tax money for their own particular interests. And before you know it, there are a lot of people swarming around the central government’s tax collections, spending their time and driving around organizing to try to influence the machinery of government to divert more of this tax money in the direction of their special interests.

Now, this seems like good, clean fun, doesn’t it? It’s just another form of competition – competition for the government’s largess. However, to an economist, this kind of competition is not benign; that is, it is wasteful – wasteful competition, in the form of what we call “rent seeking.” Rent seeking refers to people’s willingness to burn up their own or someone else’s resources in an activity that is purely redistributed. That is, by showing up in Washington and lobbying congress and organizing voters and trying to get a bigger share of the pie for yourself, you’re not actually creating anything – you’re just kind of mud-wrestling over a limited pot of money that’s been taken from other people. Nothing is being created; it’s just a fight over the stuff that’s already here.

So the time that people are spending in this activity is time that they could be spending doing something else, creating something like pizza, that’s valuable to society. And the resources that people are burning up in this activity are resources that would have other uses, other places – he could be delivering Meals on Wheels, or helping a friend move his furniture. So rent seeking refers to a form of wasteful competition, when people burn up resources or impose opportunity costs on society, or otherwise shrink the overall pie just to try to get a bigger chunk of the smaller pie for themselves. The cost of rent seeking is that things that society would otherwise have are no longer available to us; it’s just as if people were burning up resources – that is, making these things no longer available. It’s as if they’re simply vanishing so that people can get a bigger chunk of the tax money for themselves. And this is the cost to society of all of that activity – lost resources.

So, rent seeking is socially costly; it creates no net benefit, but it does impose resource costs on society. And public choice theory alerts us to different aspects of government involvement in the economy that are likely to provoke or promote rent seeking.

Next time you’re engaged in some kind of activity, ask yourself, “Am I actually creating value here? Am I writing a song? Am I accumulating human capital in the form of additional knowledge? Am I baking bread? Am I making someone happy?” Or “Am I engaged in a rent seeking activity? Am I simply trying to get a bigger share of an already created pie by burning up my own time, my own energy, my own resources. Creative activities are ones that add to the overall value of society – productive employment of resources. And managers are always looking for ways to promote productive competition – competition that makes people work harder, or faster, or create more beautiful things. Rent seeking competition, or wasteful competition, on the other hand, actually shrinks the pie. And one of the complaints about government is that the tax pots that are collected by the government encourage rent seeking – lobbyists, litigation, and other wasteful efforts to redistribute income in the direction of special interests.
Market Failures

Understanding Expected Value, Risk, and Uncertainty

In this lesson, we'll introduce how economists describe people's behavior under uncertainty or the way people make choices in the face of risk. Let's suppose I offer you a gamble. I'm going to flip this coin and if it comes up heads, I'll pay you $2.00. If it comes up tails, on the other hand, I'll pay you nothing. How much would you be willing to pay to play this game with me? Well, let's figure out, first of all, what the expected value of this gamble is. The expected value is the average return if you play the game over and over and over again. The expected value is calculated as follows.

First of all, take the odds of each outcome and multiply those odds by the value of that outcome. So, for instance, the odds of the coin coming up heads are 50 percent or one-half. The value in that case would be $2.00, the amount of money that I would pay you. The odds of the coin coming up tails are also one-half and your payment, in that case, would be zero. You get nothing according to the rules of this game. So, if you multiply the odds one-half times the outcome, $2.00 and add that to the odds one-half multiplied by zero dollars, you get the expected value of this gamble, which in this case is equal to $1.00. Think of this $1.00 this way. If you played this game with me over and over and over and over again, your average payoff would approach $1.00.

Now, that's the expected value of the gamble. How much would you be willing to pay me to play this game? Well, some of you might be willing to pay $1.00 to play the game. In that case, you are what we call risk neutral. A risk neutral agent values a gamble at its expected value. You're the people who would pay me a dollar to play the coin toss game. Some of you might be willing to pay me even more than $1.00, maybe a $1.05 or $1.25 to play this game. If your willingness to pay is greater than the expected value, we say that you are risk inclined or risk loving. These are people who are willing to pay more than the expected value to take a risk. These are people who are excited about the prospect of winning and value that more than the possibility of the loss.

Now, many of you out there would not be willing to pay me a dollar. Your maximum willingness to pay for this gamble may be only 75 cents, or 50 cents, or 25 cents. This is what we call risk aversion. People who are not willing to pay the expected value of a gamble are described as risk averse. For a risk averse agent, the risk of losing is valued more heavily than the possibility of winning and therefore the value of the gamble to that person is less than the average return, less than the expected value. Now, it's kind of interesting why it is that people will buy insurance and also go to Las Vegas and gamble on a vacation.

Sometimes we act as though we are risk averse. Whenever you buy insurance on your house, you're paying someone to take the risk away from you and usually what you are paying is greater than the expected value of your loss. There's a very small probability that your house will actually burn down. But, if you multiply that probability by the value of your loss, in the case of a fire, you'll get the expected value of the loss to you. The fact that you're willing to pay more than that, that you're willing to give up more money than you expect to lose in order to be protected from the risk is an indication of your risk aversion.

On the other hand, if you buy a lottery ticket and go and gamble in Las Vegas, you are paying more to play that game than you are expecting to win. That's because you're focused on the possibility of a big return is enticing you into the game. You're behaving as a risk lover. If there is a million to one chance of winning the lottery and the return on winning is a million dollars, then the expected value of a lottery ticket is $1.00. If you're willing to pay $2.00 for a lottery ticket, it's because the jackpot excites you and the risk of losing your $2.00 is not as important to you. You're behaving as a risk lover.

So, sometimes we'll behave as a risk lover, in one aspect of our lives, but as risk averse in other aspects. Most of us are risk averse when it comes to the possibility of losing our income or losing our house or losing something else that is of great value. On the other hand, we're likely to gamble with small amounts, because we enjoy the prospect of winning a big jackpot. So, the way economists describe risk attitudes is with three concepts. People who are excited about risk would be called risk lovers or people who are risk inclined. These are people who are willing to pay more than the expected value in order to take a gamble. People who are neutral about risk, who value a gamble at its expected value are called risk neutral and finally, most of us, with at least large amounts of value, are risk averse. We would be willing to pay more than the expected value of a loss to be protected from the loss and less than the expected value of a gain in order to have the chance of winning. We're the people who value the coin toss game at 50 cents or 75 cents, but never at $1.25.
Market Failures
Uncertainty

**Understanding Expected Value, Risk, and Uncertainty**

This is the notion then of how economists think about risk behavior. This idea of risk behavior is the basis for a theory about how socks are priced, based on their expected value, how insurance is priced and people’s willingness to pay to avoid risk becomes the price of insurance. Also, it has to do with how much extra you would have to pay one of your employees if you put her on a sales commission. Once her income is uncertain, she’ll need some kind of compensation to make her willing to take that risk. So the theory of risk behavior is used frequently in Economics to describe financial markets, to describe a sales plan, to describe insurance markets. It’s a useful concept.

So, for a quick review. To calculate the expected value of a gamble, take the probability of each outcome and multiply it by the value of that outcome, add them up and you get the expected value. Now, to consider how much an agent is willing to pay to take that gamble, you need to know about the agent’s risk attitude. If the agent is risk inclined, he’ll pay more than the expected value. If the agent is risk neutral, he’ll pay exactly the expected value and if the agent is risk averse, he’s willing to pay some number that’s less than the expected value.
Why is it so difficult to buy a good used car? Have you wondered about that? This is a great example of a problem economists call asymmetric information. Asymmetric information means that one party in a transaction has more information or knowledge about the quality of the good being traded than the other party does. And asymmetric information can really mess up a market.

In a previous lecture, we talked about risk. Risk is when an agent looks at the odds and takes a gamble. You decide whether to buy insurance on your house or not. You decide whether to play roulette or not. You decide whether to buy a lottery ticket or not. You decide ahead of time whether you want to subject yourself to the odds, whether you think your chance of winning is enough to warrant the cost that you are incurring to take the gamble.

Asymmetric information is a different kind of problem. This is a problem that stems from the distribution of information. What different people know. The fact that one person knows more about a trade than the other. And one of the things that we learn, when we study information, is that sometimes a little bit of information can make the situation much worse than no one having information at all.

Let’s look at this example, as we consider what economists call the problem of lemons. The market for lemons. Now, a little bit of vocabulary here. For any of you who are non-native speakers of English, a lemon refers to a car that is so much trouble that it isn’t worth it. You have to spend so much time and money repairing it that you don’t get enough value to warrant the expense. We will contrast a lemon with a gem. A gem is a good car that you buy that gives you a lot of satisfaction and very small repair bills.

Now, let’s suppose that we have a situation where there is complete knowledge about the quality of cars. This would be a market, in which there is perfect information. We’ll consider the first the case of perfect information. And let’s suppose that in this market the price that a buyer is willing to pay for a gem is equal to $10,000. So, the price a buyer is willing to pay for a gem is $10,000. And let’s suppose also that the price that a seller will accept for a gem, the reservation price of sellers, is equal to $8,000. Now, in this case, you can see very clearly that it’s great for buyers and sellers to get together; that is, buyers are willing to pay a price that’s higher than the price seller’s are willing to accept; that there is $2,000 worth of economic value created by getting the buyer and the seller together. So, if there’s a gem, the trade should certainly take place.

On the other hand, let’s suppose that with lemons - there is no value for a lemon to anybody. Buyers wouldn’t pay any money to get one and sellers would be willing to give the lemon away for free. So, we’ve got gems and lemons, and we’ve got the reservation price that buyers have and sellers have for each of these two kinds of cars.

Now, if we had perfect information, our first case, what’s going to happen? If buyers and sellers all know whether any given car is a gem or a lemon, what is going to happen in that case? The answer is the sellers will sell the car to the buyer at a price somewhere between $8,000 and $10,000. What do economists think about this outcome? They think it’s great, right? Because the car goes from the hand of the seller, who values it less, to the hands of the buyer, who values it more. Great. With perfect information, no problem. Trades take place and economic value is created.

What about the lemons? Well, if you’re sure that a car is a lemon, then you’re not going to pay anything for it. But the seller would just as soon give it to you for free, and the economist doesn’t care anyway, because lemons aren’t creating any economic value for anybody. So, no matter who gets the lemon, it’s just a wash.

Let’s consider now a second case. In the second case, let’s suppose we have uncertainty on both sides, that any given car could be a gem or a lemon, but the buyer doesn’t know which is which and the seller doesn’t know which is which. You could think of this as the market for new cars that roll off the factory line and the seller doesn’t know if it’s a gem or a lemon and neither does the buyer. There is the same uncertainty n both sides of the story.

Let’s suppose, however, that there is one other piece of information that both the buyers and sellers have; and that is that 50% of the cars that roll off the factory line are gems and 50% are lemons. Now, given that you’re buying a car and you don’t know whether it’s a gem or a lemon, but you do know the odds, the market for cars becomes a kind of lottery. When you buy a car, you’re buying lottery ticket. You’re buying something that half the time is going to be worth $10,000 to you, if you’re a buyer, and half the time it’s going to be worth nothing to you, because it’s a lemon. On the other hand, if you’re selling these, you’re also selling lottery tickets. Fifty percent of the time you’re selling something that’s worth $8,000 to a seller and 50% of the time you’re selling something that’s worthless.
Understanding Asymmetric Information as an Economic Problem

So, let’s suppose, for the sake of this story, that all of our agents are risk neutral; that is, they are equally excited about gains as they are concerned about losses; that is, the value of any gamble to a risk neutral person is the expected payoff from the gamble, the expected value. Well, in that case, for the buyers, the expected value of a car is 50%, or one-half, times 10,000, that’s the chance of getting a good car, plus 50% times zero, the chance of getting a lemon, multiplied by the value of a lemon. And that equals $5,000. Five thousand dollars is the expected value when you buy this new car. And that’s what a buyer would pay, if the buyer is risk neutral.

Now, how much would a seller want to sell the car for? Remember, 50% of the time it will be a gem, 50% of the time it will be a lemon. You calculate. What’s the value of which a seller would sell a car under this situation of uncertainty? The answer is the seller would sell for one-half times $8,000 plus one-half times zero, for an expected value of $4,000.

Now look, the buyer’s expected value is 5,000, the seller’s value is 4,000. All the trades are going to take place. The sellers will be willing to part with this car for anything, any amount that’s greater than $4,000. The buyers will be willing to pay any amount that’s less than $5,000. So, there’s room here for agreement and, once again, all the cars move from the hands of the seller to the hands of the buyer. A good outcome, as far as the economist is concerned, because, once again, the buyers get all the gems. And since the gems are worth more to the buyers than to the sellers, the economist says value is created by moving them from the seller’s hands to the buyer’s hands.

Well, what about the lemons? All the lemons move, too, don’t they, because they buyers buy cars and half the time they’re getting lemons. But, once again, the economist doesn’t care whether a buyer gets a lemon or whether the seller keeps the lemon. Since lemons aren’t creating any value, there’s no concern about who gets the lemons. Some of the buyers will wake up the next morning, turn the key in their new car and be unpleasantly surprised, but that’s just a redistribution. That’s just like losing some money. It’s not like value hasn’t been created. Value would be lost if some of the good cars stayed in the hands of people who valued them less.

Now, let’s consider the final case. The final case is the case of asymmetric information and this is what describes the market for used cars. The seller knows whether the car is a gem or a lemon and the buyer doesn’t. The buyer knows that 50% of the cars that come off the factory line are lemons, but the buyer doesn’t know whether the particular car he’s looking at from this particular salesman is a lemon or a gem. But the salesman knows. The salesman knows. The salesman has private information about the quality of the car. What happens in this case?

Suppose you’re a buyer and you go into a used car lot, and a salesman offers to sell you a car for $5,000. What do you know? You know for sure that that car is not a gem. Why? Because if it were a gem, the salesman wouldn’t part with it for less than $8,000. That’s what a good car is worth to the seller in our story. So, if the seller is willing to sell you a car for under $8,000, you know it’s a lemon. To have any chance at all of getting a gem, you’ve got to pay a price in excess of $8,000. However, if you pay $8,000 or more, what are you going to get? Well, you’re going to get a 50/50 gamble, because all the cars will be on the market at $8,000 and some of them will be gems and some of them will be lemons. In our case, 50% will be each. So, what’s it worth to you to get that 50/50 gamble? Well, we already calculated it. The 50/50 gamble is only worth $5,000. It’s expected value to the buyer. So, why pay $8,000 for a lottery ticket that’s only worth $5,000? You wouldn’t do it. The problem is when you lower your price below $8,000, then you know for sure you’re going to get a lemon. And that’s why this market falls apart. The market falls apart because buyers are not willing to pay a price that’s high enough to draw the gems into the market. They’re not willing to pay that high price because of the risk of getting a lemon. This is what we call, in economics, adverse selection.

Adverse selection exists from the presence of low quality. Leads buyers to behave in such a way that it drives the high quality out of the market. The presence of these lemons causes buyers to try to protect themselves by paying a lower price, which leads the owners of the good cars to take them off the market. No one will sell a good car at a price that buyers are willing to pay, and the price buyers are willing to pay is low because of the risk of getting a lemon. This is the problem of adverse selection. Adverse selection is also why there’s no market for scratched-off lottery tickets. If someone wants to sell you one, you can be sure it’s not a winner. It’s also the reason why the market for health insurance sometimes has problems. The people who are inclined to buy health insurance are people who have higher risks, and this raises the price of insurance for everyone, which means people with low risks decide not to buy, and the market then has problems. Only the high risks are in the market. The price keeps going higher, pushing out the low risks, cooling the dynamic of adverse selection.
So, asymmetric information then is a problem that can really mess up a market. The solution would be to pay extra money to try to eliminate the asymmetric information. There's a service called Lemon Busters that will come and examine your car before you buy it. You have to pay them a fee, but the fee then helps you sort out gems from lemons. So, there can be some kind of certification process, or maybe fellows will be willing to offer you a 90-day guarantee. You can bring the car back, if it happens to be a lemon. There are ways of dealing with adverse selection, but if it's left untreated, it can cause a market to vanish.
In the previous lesson we introduced the concept of asymmetric information. Asymmetric information exists when one party to a deal knows more about the quality of the good being traded than the other party does. And we looked at time at a particular kind of asymmetric information, when one party knows about the traits, some unobservable trait of the good, or some unobservable trait of themselves. I'm a seller and I have a car for sale. I know whether it's a gem or a lemon and you, the buyer, don't. I'm applying for health insurance and I know what my health risks are and you don't. You're the seller of insurance. Or I'm a student who wants to join the study group. Perhaps I know more about my aptitudes and skills than you do, my other people in the study group that I'm trying to get to invite me. Whenever one party has special knowledge about some kind of characteristic, then we get the problem of adverse selection. The presence of low quality drives high quality out of the market. The presence of lemons causes people who have good cars to take them out of the used car market. The presence of people with high health risks in the market for insurance drives up the price of insurance and causes people with low risks to drop out of the market. The presence of people who are not especially hardworking students or devoted to serving their study group causes people to be very careful about who they invite to their study group and, once people get into the study group that are hardworking and are contributing, then it's always the smartest students who seem to leave first, because they could do better on their own. This is the problem of adverse selection.

Now, there's another problem that stems from asymmetric information, and that's called the problem of moral hazard. Moral hazard arises when one of the parties to the deal takes an unobservable action after the deal is done. This is the example of someone who gets insurance and then decides to behave in a risky fashion. For instance, when I went to apply for insurance on my house, I was very happy not to have to worry about things being stolen from my house and therefore I was more inclined to leave the windows open when I went out for a walk, or maybe even leave the back door unlocked whenever I went to work. After all, carrying around keys is a bit of effort and I like the windows open during the day because the house then smells fresh when the breeze blows through. However, whenever I take that choice, when I make that decision, I'm imposing a cost on the company that insured me, because I make it more likely then that my house will be ripped off, things will be stolen and I'll file an insurance claim, and they'll have to compensate me for a lost television set, or a lost answering machine or stolen clothing. This is the problem of moral hazard. When an insured person then begins to take extra risks and these extra risks impose a cost on the person who offers insurance.

Let's suppose I'm a salesman and I'm working on commission, but I don't like the risk. Some days are good, some days are bad. I'd rather have a nice, safe, reliable income. So, I go to my boss and I say, "Look, I want to get rid of the commission and go on a fixed salary. The fixed salary then protects me from risk and I feel better." And the boss, perhaps, decides that he's persuaded by my case and he puts me on a fixed salary. Well, what's the temptation then? The temptation is I don't work as hard to make the extra sales because I'm no longer subject to risk. I'm not insured and being insured, I may be inclined to make decisions that impose a cost on my boss. Not working as hard to sell clothing then, or sell shoes, or cars or whatever then causes my boss and the company to lose revenue. But I'm only doing that because I'm not scared anymore. I'm only doing that because I'm insured against risk.

Moral hazard is a difficult problem. It often causes insurance markets to break down. People aren't going to insure you if, by insuring you, they are going to give you incentive to take risks, which then cause the insurance not to be a profitable business. If we insure people with life insurance they all decide then they want to go skydiving because they don't have to worry about anyone who may be hurt, if the have a skydiving accident, besides themselves, of course, then the life insurance become a very difficult product to offer. The price of life insurance goes way, way up because the people you insure then are behaving in a risky way. If you offer health insurance and people decide, as a result of that, that they want to have a diet or other habits that increase the risk that they'll make a health insurance claim, then the price of health insurance has to go up to insure only risk taking people. Moral hazard can cause a market to break down and the solutions to moral hazard are to return some of the risk to the person who's insured. That's why insurance companies give you a deductible. If your car is hit, then you have to pay the first $1,000 worth of damages yourself. That way you still face some of the risk. Other things that can happen are when you go to a bank and you take out a loan, they may ask you to post collateral. This gives you an incentive, so that if the business that you're starting goes bankrupt, you're still going to lose something, perhaps you car, perhaps some bonds or something. In any event, with some of the risk restored to you, you're more inclined to invest in safe and profitable business ventures rather than going out and doing something highly speculative or risky.
Market Failures

Uncertainty

Understanding Moral Hazards in Markets

A second thing that happens is that there is coinsurance; that is, an insurance company will only insure you for 50% of your loss or 80% of the loss. That way, you still bear a chunk of the risk. You still have to pay 20% of the damages, and that may be enough to incline you to make safe decisions and to protect the insurer from a big loss.

Finally, a third thing that can happen that can reduce the problem of moral hazard is monitoring; that is, a company may decide that if you want to have insurance on your house, you have to be subject to an inspection every so often to make sure that you are locking your doors and windows. Or there may be a provision in your insurance policy that says, “If your house is ripped off and you left the windows or doors open, then the insurance company doesn’t pay as much.” This, again, is a kind of monitoring and the way that the insurance company enforces this is they require you to call the police any time there’s an accident or any time there’s a crime. This makes sure that the police come, investigate and report that, in fact, yes, the doors were locked. This person was imposing extra risk through his choices.

So, moral hazard then arises when one of the agents makes an unobservable choice. You can’t tell directly whether I’m leaving my doors open or not. You can’t tell directly whether I’m going out and eating fatty foods or smoking cigarettes, but you can, however, you can make it a provision that either I be a part of the risk through a deductible or through coinsurance, or you can incur the extra expense of monitoring my behavior, so as to make sure that I’m not imposing extra cost after I’m insured.

Moral hazard is different from adverse selection, and we can distinguish the two with a story. It is part of our modern legend and it’s confirmed by statistics that people who drive Volvos get more traffic tickets for moving violations than people who drive other cars. Why is it? Why are people in Volvos so often cited for running red lights, running stop signs and driving too fast? Well, there are two answers that we could give. The first is based on adverse selection; that is, people who know they’re bad drives and they’re scared that they might get in an accident, they go and they buy a Volvo to protect themselves from the risk of the accident. That way, when somebody shows up at a Volvo dealership, the first question you want to ask them is, “You know, are you a bad driver? Is that why you’re hear? Are you scared of getting in a wreck?”

There’s also another explanation for why Volvos are in so many traffic incidents, and that is moral hazard. Once you own a Volvo, you’re not scared of other people any more because you’re driving a tank, and so you can drive fast, and run red lights and things like that without any fear. The Volvo protects you from risks and therefore creates a kind of moral hazard. Not being afraid may lead you to take more risks or engage in riskier behavior.

So, how do we tell whether this particular story is an example of adverse selection or moral hazard? Maybe we inspect people’s driving records ahead of time. When people show up to buy a Volvo, if we find out that they already have lots of citations for traffic accidents or traffic violations, then they’re coming to get a Volvo because they’re already risky. That would be adverse selection. Moral hazard, on the other hand, would be people who drive safely before they get a Volvo and then once they get a Volvo, they start driving like a maniac because they’re protected.

In summary then, asymmetric information causes problems for markets. If people have unobservable characteristics before the deal, then the problem is adverse selection. Low quality drives high quality out of the market. On the other hand, if people make unobservable decisions after the deal, then the problem is moral hazard, the presence of protection from risk, the presence of insurance leads people to engage in risky behaviors that impose costs on someone else.
Externalities

Defining Externalities

I think you’d be hard pressed to find a more obnoxious bumper sticker than this one and yet, apparently somebody paid two bucks for it. I’ve seen it on people’s cars, which lets you know that the people are getting at least $2.00 worth of satisfaction from sharing this message with you. Now, I myself, will be willing to pay at least 25 cents to the person to keep them from putting the sticker on their car, from being spared having to encounter this message in traffic. And I can’t be alone, because if there are 40 of us who feel that way, then we will be willing to take up a collection of $10.00 to pay the guy not to put it on his car. Right? He’s getting $2.00 worth of satisfaction from it and we’re getting $10.00 worth of dissatisfaction, collectively. So, why is the sticker on his car? It’s because we’re dealing with a particular kind of market failure called an externality.

An externality is a cost or a benefit that you can pass on to other people. Let’s take the case of an external cost. An external cost is the cost of your transaction that is borne by someone who is not a party to your transaction. For instance, if I’m a box maker and I sell you a box, but I’m polluting the lake when I produce the box, the pollution is an external cost. It’s a cost that’s passed on to swimmers and fishermen and other people who aren’t part of your and my transaction and yet they bear a cost. They’re affected by our trade.

Suppose you buy a package of cigarettes from a vending machine. The vending machine sells you the cigarettes. You buy them because of the satisfaction, but when you light up in a crowded room, other people bear the cost of your decision. You have externalized consequences. You have externalized cost of your action onto other people who are not part of the decision that creates the cost. That’s an externality. Other people outside the transaction are affected.

There are also external benefits. Let’s consider the case of a flu shot. Suppose I go to the doctor and get a flu shot. Well, I’m protected against the flu, but so are you. If you’re students in my class or co-workers or my friends, now you don’t have worry about me getting the flu and coughing and sneezing on you and perhaps subjecting you to the same kind of torment. You don’t have to worry about the flu because you get a benefit from my flu shot. Now, chances are I didn’t think about you when I went to the doctor and made the choice to get the shot, but you happen to get a benefit simply because I’m vaccinated. The benefit is external to my decision to get the shot. It’s external to me. I made a private calculation. A private calculation of the benefit from the shot and the cost of getting it and if the private numbers worked out okay, if I made a profit off this deal, then I got a flu shot, if the cost was less than the benefit to me.

However, everybody else, all of my students and co-workers, they got a benefit too, from reduced risk of exposure. And that’s an external benefit, because it’s external to my choice to get a flu shot. They weren’t part of the decision. Now, here’s the problem. When there are externalities present, when there are external costs or external benefits, we cannot rely on the free market. We cannot rely on decentralized private choice to give us an outcome that maximizes economic value. In fact, in many cases a factory will see that’s it’s profitable to make their good and sell it to the market, because the revenue they bring in is greater than the costs they incur.

However, the revenue will only be greater than the cost, because they are able to dump part of the cost on other people, dump the pollution in the lake or in the air. If you added the true cost of the pollution onto the private cost of making boxes, the cost might be greater than the benefit. It might be socially unprofitable to make those boxes. Those boxes may actually be shrinking the pie, because the true benefit is less than the true social cost. However, the boxes get made anyway, because some of the costs can be externalized and they look privately profitable to the company.

The same might be true of my flu shot. When I think about the money I have to pay the doctor for this flu shot and the sore arm that I have to live with for three days, it may not be worth it to me, privately, to protect myself from the flu. However, the only reason I don’t get the flu shot is because I am ignoring all of the benefits that are created for everyone else. If in fact, I somehow were experiencing or internalizing the external benefits, if I were getting a nickel for every person that I’m protecting against the flu by getting the shot, it might turn out to be a profitable venture for me to go to the doctor and get it anyway.

So, when there are external costs or external benefits, there are choices that get made that don’t add economic value and there are choice that should be made, but aren’t that would add economic value. In the next lecture we’ll look at some specific examples, involving numbers that show how, in the presence of externalities, a market can fail. And, we’ll suggest some remedial measures that might help increase economic value.
Market Failures
Externalities

Explaining How to Internalize External Costs

In this lesson we’ll show how externalities can lead to market failure, and we’ll consider some actions that can be taken to remedy the problem.

Let’s start with the case of an external cost, and we’ll use an example of pollution. Let’s suppose that we have a market, in which boxes are produced and traded. People buy boxes to ship things in and box factories produce boxes and sell them. But the production of boxes involves generating waste chemicals that might be dumped into the water somewhere, and that can be an external cost. So, let’s consider this story carefully. Let’s take the first box that’s produced in this market, and let’s suppose that that box generates $5 worth of benefits for some customer somewhere. Now, if there are no wealth effects and no externalities, this $5 represents the marginal benefit to society of the product of that box. Let’s suppose that the box maker is willing to produce and trade that box at a price of 50 cents. So, here’s a point on the supply curve. We get one box offered for sale at a price of 50 cents per box. Now, if there are no wealth effects and no externalities, then the difference between the $5 benefit and the 50 cent cost would be economic value that’s created by this trade. However, if there are external costs of making the box; that is, pollution that’s created that’s dumped in the water, we have to add the external cost onto the private cost to get the true social cost of producing the box. See, the 50 cent reservation price is the box maker’s private cost of making the box. These are the costs that he has to cover, in order to make a profit or to break even. These include the cost of any cardboard products that he brings into his factory, the cost of his labor and so forth. But it doesn’t include costs that he is able to dump on other people. It doesn’t include any free disposal, as far as he’s concerned, of pollution products in the lake. And in order to find the true social cost of producing the box, we have to add the external cost onto his private cost. If we do that, it may be that the true cost to society of producing this box is not 50 cents, but rather $2. The $2 social cost is the sum of the 50 cent private cost and $1.50 in external costs.

Now, right now you probably have a question. Well, how do we calculate external cost? After all, calculating private cost is very easy. We know that if the private cost isn’t covered, the box maker won’t offer the box for sale. The reservation price of the box maker depends on the cost of his labor, and inputs and so forth. But how do you put a value on the pollution that’s dumped in the water? Think for a moment about some different ways that we might value pollution or the cost of pollution. One might be the cost to the box maker of cleaning up the pollution himself or preventing the pollution from being dumped in the water to begin with. He might impose some kind of control system on his factory that uses strainers or things like that, and that would be costly. Or he might dump the pollution in the water and then hire a crew to go out and clean it up. It may also be that he could simply pay the people who swim in the water or drink it, to accept the pollution, or pay to cover their doctor bills to solve the problems they develop from contact with the pollution. All of these are possible ways of valuing the pollution. What you as the swimmer would be willing to pay to have the water clean instead of dirty. The cost of your doctor bills, the cost of the factor of preventing the pollution from being dumped in the water, or the cost of cleaning it up once it is dumped in the water. Now, which of these different methods will the economist use? Which of these different methods would the economist say is the appropriate way of calculating the cost of the pollution, the external cost? The answer is whichever one is the least expensive. That’s what the economist would say. If it’s cheaper to clean the pollution up after it’s in the water than it is from going in the water to begin with, the economist would say that lower cost of returning things to the original state, that lower cost of dealing with the pollution, is the appropriate one for this story. In fact, if you go to an extreme example, the economist would say that if it’s cheaper to pay people’s doctor bills from getting sick from the pollution to restore them to health, well, then that’s the appropriate cost we should use for the external cost of the pollution. So, whether it’s cheaper to clean it up, prevent it from happening or pay to cover the consequences, whichever one of those is least expensive, that’s what the economist would use for the external cost of the pollution.

Now, the box maker doesn’t consider the cost of the pollution. The box maker externalizes that cost. So, will this trade take place? And the answer is yes, it will take place because the willingness to pay of the buyer is greater than the private costs that determine the seller’s willingness to sell. So, as long as the $5 is greater than the 50 cents, then the trade takes place. Should this trade take place? From the point of view of normative economics, is this trade adding value to society? Yes, it is, as long as $5 is greater than $2, as long as the social benefit of the box is greater than the social cost. So, in this case, even though there’s an externality involved, we don’t have a failure of the market. The box should be produced and traded, and it is.

Let’s look now at a second box. Let’s suppose that this second box now has a value to the next customer of $3.50. And let’s suppose that the private cost for the box maker of producing this box has gone up to $2.50, because of diminishing marginal product labor, or something like that. So, increasing opportunity cost is raising the seller’s reservation price. Now, should this second box be produced and traded? Well, we’ve only go part of the costs here.
Market Failures

Externalities

Explaining How to Internalize External Costs

If we add the pollution cost onto the private cost of the seller, we get this; $1.50 on top of $2.50 makes the true social cost of this box equal to $4. Whoa, now we’ve got a problem. Because the buyer’s willingness to pay is greater than the seller’s willingness to accept payment, we know the trade will take place. The buyer’s private benefit is greater than the seller’s private cost, this trade will take place. However, this trade is not good for society because the true cost of this box is greater than the benefit it creates. In fact, the production and trade of this box generates 50 cents worth of economic loss for society. We have deadweight loss here that’s created because the box maker is able to externalize the cost and, because he doesn’t experience the true cost, he doesn’t have to deal with the true cost of the box he creates and he can dump the cost on someone else. We get a box produced and traded that’s actually less valuable to society, it generates less benefit, than it costs. It generates negative value. This trade should not take place, but it does in a free market, where people are able to externalize the costs. The costs are borne by somebody else.

Now, how do we solve this problem? How do we solve the problem? In economics, we have a principle that applies generally to situations of market failure. It’s called the principle of the second best. The second best refers to how to clean up a mess, the second best. The best situation would be if there were no externality and everything that was produced and traded was privately beneficial and privately costly. It would be great, from the point of view of the economist, if there were no externality, if there was a complete set of markets here for trading all the goods. But because this pollution is external to the market it’s external to the trade, things get messed up here. The externality creates a mess. The principle of the second best says this: when you have a market failure, the best policy for dealing with it is the one that most precisely corrects for the original problem. Whatever policy that the principle of the second best recommends will be that policy that most precisely offsets the original distortion. In this case, what’s the original problem? The original problem is an externality, people’s ability to pass costs on to someone else, who doesn’t have a [????] in the matter. So, we correct that problem by making the decision makers internalize the externality. Internalize the externality, so that they now have to pay all the costs that are associated with their choices. Rather than being able to dump the cost onto someone else, we’re going to make them pay to cover the cost, whether in the form of imposing the control technology on their factory and keeping the pollution from happening, or whether it’s cleaning up the water, or compensating the people who suffer from the pollution. We’re going to make the box maker now pay the true social costs of his choices. In that case, what happens is the private costs become the same as the social costs. They slide up to coincide with the social costs. Now that he has to pay not only the private costs of the cardboard and the labor, but also the cost of cleaning up the pollution, the blue triangles move up to lie on top of the green triangles. We have made the seller’s private costs the same as society’s costs. Now, the way we could do this would be imposing a tax on the production of boxes. And if every box generates the same amount of pollution and all pollution has the same cost to society, then we impose a tax of $1.50 on the production of each box. If the seller has to pay that tax, then that hire tax of $1.50 shifts his private costs up, so that they’re the same as society’s costs. Now, his incentives, the box maker’s incentives, are the same as society’s incentives. And, in that case, the first box will be produced and the second will not. By taxing the box producer, we cause him to internalize the externality, to experience the cost that he was previously able to externalize, pollution dumped in the lake, now he experiences that as part of his cost of production. And the boxes that are not profitable for society will be not profitable, not privately profitable for him to produce. So, with this tax, we get a change in behavior. Boxes that should be produced and traded will continue to be produced and traded, but boxes that are not socially profitable will no longer be produced and traded. The externality is corrected for. The principle of the second best applies in this case. If the problem is an externality, you solve it by imposing some kind of tax that forces the person who was externalizing to internalize the cost of their actions.

Next, we’ll look at an example that involves external benefits.
Market Failures

Externalities

Explaining How to Internalize External Benefits

Let’s look at a market failure in the case of an external benefit. The good that we’ll consider is the flu shot. Let’s look at a particular agent who’s trying to decide whether to get flu shot or not and she says, “Um, you know, the benefit I get from prevent the flu this season, I’ll value that a $3.00.” and the cost of getting the flu shot at the doctors is $2.00. Well $3.00 is greater than $2.00, so we know that she’ll get a flu shot. What she doesn’t know or may not have thought about is the true benefit that she generates for society is not just $3.00, but also the benefits that she generates from all of the other people who don’t have to get sick because she doesn’t spread the flu to them.

If that external benefit is $2.00, the $2.00 plus her private benefit gives us a social benefit for that flu shot of $5.00. Because the social benefit is way above the social cost, then it’s very clear that we’ve generated a lot economic value here. Five minus two is $3.00 worth of economic value for this flu shot. Part of it comes in the form of her private benefit, part of it comes in the form of the external benefits and, of course, the cost is the cost. Well, nothing to worry about here, the flu shot takes place and it’s just a nice, happy accident that it generates a lot of extra benefits for people who weren’t involved in the decision.

Let’s look now at a second candidate for a flu shot and let’s suppose that the flu shot is going to cost $2.00 to provide and that her private benefit from the flu shot is only a $1.50. Maybe her arm gets sore or maybe she’s less prone to get the flu anyway and therefore her private benefit is going to be less than the private cost. However, the social benefit would still be the same. If she got the flu, she’d be spreading it to other people. So, in keeping with our previous example, we’ll add on a social benefit. We’ll add on an external benefit of $2.00 to give us a total social benefit of $3.50. Here’s the external benefits; the difference between the green and the red and here’s the private benefit and here’s the cost.

Well, what’s the problem in this case? The problem is, she’s not going to get a shot because her private benefit is less than the cost. It’s not worth it to her to go to the doctor and get the shot. However, the shot would add value for society, because the overall benefit of the shot, the external benefit plus the private benefit adds up to a social benefit of $3.50, which is greater than the cost of the shot. So, this shot should be given. However, our candidate does not want the shot, because her private benefit is less than it costs to get the shot.

So, even though this is privately unprofitable for her, it would be socially profitable if she got the vaccine. So, what do we do in a case like this? Notice, the second shot doesn’t take place and that’s a deadweight loss, a regret, a lost economic value of a $1.50 that could be created, buts not created because of externalities. Let’s suppose then, in this case, we help our agent internalize the external benefits that she creates by giving her a subsidy. Let’s suppose that everyone who comes in for a flu shot gets $1.50 or a coupon for $1.50 worth of something or juice and cookies worth a $1.50 or something that raise their private benefit up to be equal to the social benefit.

If we give a subsidy, then the red diamond lie on top of the green one and private benefits are equal to the social benefits. Once the private benefits and the social benefits are equal, then we get the right number of flu shots. The original candidate gets the flu shot, just like before and he gets the juice and cookies too, and our second candidate decides now that she gets the juice and cookies, it’s worth it for her to come in and get the flu shot. Now that she’s been subsidized, her private benefits are the same as the social benefits and she makes the decision that in the market maximize economic value.

So, once again, we’ve got the principle of the second best at work. The original problem was the external benefit, the people who benefit that are not part of the transaction. So the way that we bring them into the picture, the way we help our decision makers internalized these externalities is we subsidize them. We give them an amount of money that reflects the benefits that they are creating for other people besides themselves and once the subsidy is in place, private benefits and social benefits are once again the same. The subsidy is equal to the amount of the external benefit. Once we subsidize you for an amount equal to the external benefit, we cause you to internalize the externality. We make the externality part of your private calculation, so that your choices reflect social benefits, not just private benefits.

With the subsidy, in this case, the social benefits and the private benefits are the same and the correct number of flu shots is given. We correct for the market failure with a carefully targeted policy, in this case a subsidy to help agents internalize external benefits.
Market Failures

Solutions to Externalities

Finding a Market Solution to External Costs

We've talked today about economic costs and economic benefits and how they can mess up a market. Let's take the insight that we got from the simple examples now and apply it to the supply and demand diagrams that we are familiar with. These supply and demand diagrams, they're useful for explaining how externalities can lead to market failure. If the red curve represents marginal social benefit and the blue curve represents marginal social cost, then economic value in the market is maximized where the red curve and the blue curve intersect. All of these units have social benefit greater than the social cost and they should all be produced and traded. All of the units beyond four in my example have marginal cost greater than marginal benefit, they are not economically valuable and they should not be produced and traded. So, the reason economists are excited about competitive free markets, when there are no externalities and no wealth effects, is because they lead to the quantity produced that maximizes economic value. This whole little wedge right here can be thought of as economic pie, benefits, less cost or economic value. And all of these units up to number four are up to units where the benefit is greater than the cost, and so they are economically profitable to produce. They're good for society.

Now, the problem with externalities is that they cause private costs and social costs to diverge. They cost the private willingness to pay of the box factory not to be equal to the social cost of making problem. They cause his reservation price to understate the true cost of boxes, because part of the cost is dumped in the lake and passed on to other people. So, in order to explain an externality in a diagram like this one, we have to add the external cost onto the private cost. And when we add the external cost onto the private cost, we get a new curve that represents marginal social cost. Marginal social cost would be greater than the original supply curve by the amount of the externality. So, what I've done here is I've added the pollution cost of box making onto the private costs of the box maker and gotten a new curve that we'll call MSC, marginal social cost. Marginal social cost is equal to the private cost of the box maker plus the external cost of the pollution. Add those together for every box and you get the marginal social cost curve.

Now, clearly, economic value is maximized where benefit and cost are equal at the margin; that is, we should stop producing boxes back here at three boxes instead of four, because after three boxes, the true cost of making a box, including cardboard, labor, all of that stuff plus the pollution, is greater than the benefit that people get from those boxes. Because the social cost at the margin is higher than the social benefit, those boxes are not profitable to society. And the only reason they are produced is because the box maker can pass some of the costs on in the form of the pollution externality. If he had to internalize those costs, he would stop at this point. So, what we wind up getting is a kind of deadweight loss. We get a deadweight loss that is equal to this area here; that is, if we're producing four boxes, the true cost of that fourth box is going to be way up here on the green curve. The true benefit is down here on the red curve, so the cost exceeds the benefit. That's a loss to society. And the same is true for all of these fractions of boxes in between three and four. All of these fractions have a higher cost than they have benefit. And if I shade in this area, I get a deadweight loss, giving me a loss that society incurs that it would not incur, if it were not for the externality. If the box maker had to internalize these costs, these boxes would not be produced. All of these boxes have costs greater than benefit, and that's a loss to society. Think of this little chunk right here as a deadweight loss. It's bad pie.

Now, how can we correct this problem? We can correct it by making the box maker internalize the externalities. That's the principle of the second best. If the problem is an externality, something that you're able to dump on other people, make you keep it at home, make you have to pay the cost. Then you'll make decisions that are consistent with the interests of society, that are consistent with maximizing economic value all around. So, now we impose a tax on the box maker per box produced, and it shifts up the supply curve. It increases the private costs of the box maker to the point, at which the marginal social cost curve and the supply curve are equal. Once we've imposed a tax, marginal social cost and the private decisions of the box maker coincide; that is, the box maker will stop at three boxes because once has to pay for the pollution, the price of producing those boxes is not going to be enough to cover his costs. He will stop at three boxes, produced and traded, rather than going on and making those socially unprofitable boxes. Notice how this works. It works because of the principle of the second best. We figured the problem was the box maker was able to dump his costs on society. But if we make him pay the true costs himself, then, in that case, the blue curve will shift up to lie on top of the green curve. The private costs of the box maker will be the same as society's costs, and the box maker will have his incentives aligned with those of society. He will stop at three boxes, this bad pie will not be created at all and we will maximize economic value for society.

Now, let me take a moment and talk about this notion of using taxes and subsidies to correct for an externality. If you're taxing the production of boxes, well, you're telling box makers that they should produce fewer of them. You're telling them that the cost, the true cost of making a box is higher. The principle of the second best says find out what
Market Failures

Solutions to Externalities

Finding a Market Solution to External Costs

the problem is and design a policy that, as precisely as possible, offsets the original problem. Don’t introduce any new problems. Don’t add anything extra. Find the original problem, target it and design a policy that offsets it, that corrects it, that compensates for it. Well, the policy that I’ve been describing is a tax on the production of boxes. It turns out that that’s kind of a crude policy, because the problem here is not that boxes are produced; the problem is pollution. So, if we really wanted to take the principle of the second best seriously, we wouldn’t charge box makers a tax for making boxes. We would charge them a tax for pollution. We would charge them a fee for every gallon of pollution that they dumped into the river. That’s the source of the problem. That’s the externality, not the boxes themselves. So, one thing that we could do then is charge them a tax per unit of pollution they dumped in the water. If that happens to be constant amount per box produced, then the box tax and the pollution tax are the same thing. But if different technologies for producing boxes produced different amounts of pollution, or if the first box produces more pollution than the tenth box at the margin, then we want to try to separate the box tax from the pollution tax. Forget the box tax. Make your focus on the pollution itself. Tax only the thing that creates the external cost.

Well, that then raises the question how much should you charge the firm to pollute the water? What kind of tax should you impose per gallon of pollution? Where do we get the information about the right amount for that tax? And the answer is, well, we kind of have to guess or use scientific studies, unless we try to return our situation all the way to the market; that is, why not create a market for the right to pollute? Create a market where the businesses can buy certificates that entitle them to dump pollutants. And they can trade them among each other, buying and selling them, to determine a market price. Clean firms that use relatively little pollution will drive up the price of these certificates by being able to pay more for them and still be profitable. Firms that use dirty technology will find that it’s too expensive to buy the certificates that they need and therefore they’ll switch to cleaner technologies, and so forth. Plus, people who like to swim and fish can let their organization, their environmental groups or swimming clubs, bid in the market and buy up the pollution rights, so that the lake stays clean; that is, if its really worth it to them to keep the lake clean.

Ah! See what we’ve done? We’ve taken a situation that was previously outside the market, external, pollution dumped in the water, and we’ve recreated a market. We’ve brought all of this back into the logic of the market, so that the information that different players have gets mixed together in determining the price of pollution. In this case, the market can allocate pollution in the same old effective way that it allocates ball bearings, or loaves of bread or anything else. See, really, the problem with an externality is there is a missing market. Something that should be valuable to people, or invaluable to people, is not being traded, because it’s just being dumped in lakes. It’s being pushed outside of the market. The principle of the second best says see, the problem isn’t just an externality. That’s just a shorthand way of talking about a missing market. There should be a market for something and there isn’t. So, let’s recreate the market and restore efficiency the way economists think about it.

So, there’s been a lot covered in this little segment. Let me give a quick review. The quick review is that when private costs are different from social costs, we can get overproduction of goods with external costs; that is, the supply curve is below the marginal social cost curve by the amount of the cost of pollution. In this case, we have a deadweight loss from overproduction of boxes, boxes that aren’t really socially profitable, that get produced because of external costs. With the right kind of tax, you can make the private costs of the box maker equal to the costs of society and restore an efficient outcome. However, it may turn out that if you’re really trying to take the principle of the second best seriously and really trying to focus on the problem at its source, the problem is not solved by taxing boxes, but rather by creating a market for pollution rights, so that people can register the costs and benefits of pollution with the same kind of efficiency that they register information in other markets. Creating a market that’s missing is usually a much better way of solving a problem than using some kind of crude approximation, like a tax on box production. So, this is the way that we deal with externalities, either by the cruder methods of taxing to cause people to internalize the costs, or the more efficient solution of recreating missing markets.

In the next lesson, we’ll look at how government action may not even be necessary to resolve some problems created by externalities.
Market Failures
Solutions to Externalities
Finding a Negotiated Settlement to an External Cost

In the previous lecture, we discussed how taxes and subsidies can correct the market failure created by externalities. In this lesson, we’re going to consider the possibility that government action isn’t needed to correct the market failure. That under certain circumstances, bargaining among the people involved can solve the problems created by externalities. This lesson introduces an economic concept called the Coase Theorem. The best way to explain this theorem is using a story. Let me introduce the players.

Suppose we have a lake that’s situated out here in the country, and this lake is used by two producers. One producer uses it as a sink. This is a box manufacturer, who dumps the waste products of the manufacture of his boxes into the lake rather than having to bottle them up and ship them off at cost to himself somewhere else. So, we’ve got a boxer here, who is running a factory that produces boxes, and let’s suppose the following things are true about his operations: First, the price at which he can sell a box on the market is equal to 50 cents. And this 50 cents represents the benefit to society of having an extra box. So the price of boxes is equal to 50 cents apiece. Let’s suppose now that whenever he produces a box, that he dumps pollution into the lake that; every box that’s produced by his operation generates one gallon of pollution that goes into the lake. Let” suppose that he could prevent the pollution from going into the lake by incurring a cost of 25 cents per box. So, if he dumps the pollution into the lake, then it’s free for him to produce the boxes. But, if instead he decides to bottle the pollution and haul it off for disposal somewhere else; that is, not to pollute the lake, then it costs him 25 cents per box to prevent the pollution.

All right, here’s some important numbers for you to keep in mind. Now, over here on the other side of the lake, we’ve got a brew operation, a beer maker. So the beer maker is producing beer, and here’s the beer maker, and the beer maker can sell each brew that she produces for $1. So, the social value of having an extra beer, the social benefit is $1 per beer. Let’s suppose that the beer brewer uses the lake as an input. The water comes from the lake that goes into the beer and, if the water is clean when it goes into the factory to make beer, then there’s no cost to produce the beer. However, if the water is dirty, the brewer has to clean the water up to begin with before the beer can be produced, and let’s suppose that the cost to the brewer of cleaning up the water is 50 cents for every gallon of pollution that she has to siphon out of the water. So, let’s suppose that she’s producing a beer, if she can produce it for free and sell it for $1, then she’s going to make a dollar’s worth of profits. But if to get that dollar’s worth of profits, she has to incur 50 cents worth of cost to clean up the water because it’s dirty, then her profit is only going to be 50 cents.

Let’s consider now two possibilities. In the first case, we’re going to have the brewer and the boxer operating independently of each other. Each of them will be maximizing respective profits without considering their effect on the other guy’s operations. So, in that case, what’s the boxer going to do? Is the boxer going to clean up his pollution, or is he going to dump it into the lake? Well, which one would give him maximum profits? You calculate and it’s not hard. Of course, he’s going to choose, if he thinks only about his own profits, he’s going to choose to dump the pollution from his boxes into the lake. That way each box that he produces earns him revenue at 50 cents and at a cost of zero. On the other hand, if he were clean up the pollution at a cost of 25 cents per box, then he would only be making 25 cents worth of profit. So, for the boxer, the individually rational choice made in isolation would be to use the lake as a sink.

Let’s consider the brewer over here on the other side. The brewer says, “Hmm. You know, I’ve got this water that I’d like to use for beer, but it’s kind of dirty. I’m going to have to incur the cost of cleaning it up.” Remember, the cost to her of cleaning the gallon of pollution out of the water is 50 cents when she wants to make a beer. So, it costs her 50 cents to use the dirty water, clean it up and then she sells the beer that she makes for $1. Her profit per beer on this operation is going to be $1 minus the 50 cent cleanup cost, or 50 cents profit.

All right, the total profit in this case is 50 cents, earned by the boxer, plus 50 cents earned by the brewer. This is the outcome in the case of independent decisions. With these independent decisions, both of them are doing the best they can without thinking of the other one. And the joint profits of the two operations, the combined profits, are 50 cents from the box and 50 cents from the beer, for a total of $1.

Let’s suppose now that we introduce a second way of doing business here. Let’s suppose that instead of the boxer and the brewer operating in isolation, let’s suppose that they get together and try to make a joint decision. What kind of arrangement would the two work out? Well, first of all, the brewer would convey her costs and her revenue information to the boxer and say, “Look, when the water is dirty, I have to pay 50 cents to clean your box pollution out of the water, and that lowers my profits.” She says, “Why don’t you clean up the pollution? After all, as you told me, it
Solutions to Externalities

Finding a Negotiated Settlement to an External Cost

only costs you 25 cents to block the pollution. It costs me 50 cents to clean it up once it's in the lake. So it makes sense for me to compensate you for the extra costs that you incur.” That is, even if the brewer has to bribe the boxer to go along with this arrangement, there's still profits for them to be made once the boxer starts to clean up the pollution himself. Suppose, for a moment that the boxer owns the rights to the lake and he’s entitled to pollute, if he wants to. In that case, the brewer is going to have to make it worth his while to clean up the pollution. What she does is this: she says, “Look, if you’ll keep the water clean, I can earn $1 off the sale of my beer. That $1 leaves me enough money to pay you the 25 cents that it costs you to clean up the pollution and still have some extra money left over that we might share. A dollar, minus the 25 cents that I pay you to clean up the pollution, leaves me 75 cents worth of profit on my beer, which is greater than the 50 cents I was making before. That extra 25 cents, 75 cents minus 50, that extra 25 cents worth of profit will be available for you and to split once we’re cooperating.”

Next, we think about things from the boxer’s point of view. He says, “Hmm. You know, if I have to go to the trouble of cleaning up the pollution that adds 25 cents to the cost of producing a box. However, the brewer is going to give me that 25 cents plus some of her extra profit. So, I do better, in this case, by cleaning up the pollution and cooperating with her.” Now what happens is this: the boxer makes the choice that he is going to be responsible for cleaning up the pollution. He changes his operation, at cost to himself, and now, when he produces a box, he incurs the extra 25 cent cost of blocking the pollution within his factory and proper disposal of it. Meanwhile, the brewer gets water into her factory to produce beer that has no pollution in it. The pollution has been blocked in the box factory. Now, she’s able to earn $1 worth of profit, pay the boxer to cover the cost of controlling the pollution, and still have money left over to share. In this case, the joint profits of the operations are $1, earned by the brew factory, and 25 cents, earned by the box factory after he’s cleaned up the pollution, for a total of $1.25. The total economic value, the joint profits of the two operations are higher when the box maker cleans up the pollution himself. The brewer, on the other hand, will pay the box maker to do this, because that also increases her private profits. Once there’s a possibility of making the pie bigger, the two can get together and figure out how to split the costs of moving from the status quo, the present situation, to this new situation that involves more total profit.

Now, here’s an interesting thing to note. It doesn’t matter whether the boxer owns the rights to the lake or whether the brewer owns the rights to the lake. If the boxer owns the rights to the lake, the brewer will have to pay the boxer to do the extra cleanup. If the boxer owns the rights to the lake, if the boxer does not have the right to pollute, the brewer will simply say that he has to stop polluting, and still we get the same outcome. We get the same outcome whether the brewer or the boxer owns the lake. In either case, the boxer is going to start cleaning up the pollution himself. If the boxer owns the rights to the lake, then he’s entitled to compensation from the brewer. He will have to pay him the 25 cents and maybe a little bit extra to go along. If the brewer owns the rights to the lake, then the boxer may not get compensation, he may simply wind up having to clean up the pollution, if he wants to keep his business in operation. The point that I’m making is this: the ownership of the property rights to the lake has no effect on the final outcome. It does have an effect on the distribution of the extra profits between the brewer and the boxer. It does influence that. But it doesn’t influence the economic activity. It doesn’t influence the fact that the boxer is going to end up having to clean up the pollution. Either he’ll be paid to do so, or he’ll be ordered to do so.

The Coase Theorem says the following: When two parties can get together with relatively low transactions cost; that is, it doesn’t cost too much for them to get together and bargain with each other, then the result of that bargain is going to be the result that maximizes the economic value in the situation. That is, remove from $1 total profits to $1.25. The Coase Theorem adds an extra insight, and that is if there are no wealth effects, then it doesn’t matter who owns the property rights to the lake. You will still get the outcome that maximizes profit. The bargaining between the two factory owners will result in the economic outcome that maximizes profits. The only difference that the property rights make is this: The property rights determine how the value is divided between the boxer and the brewer. But the property rights do not influence the fact that the brewer is going to wind up getting clean water, and the boxer is going to wind up taking care of the pollution himself.

So, a quick summary here. Notice we’ve got a situation of an externality. We have a situation where there is a resource, a lake that is used by two different agents in the economy. We have a boxer, who can impose an external cost on the brewer; that is, he can externalize the cost of his pollution, which raises the cost of making beer for the guy across the lake. What we discovered, in this case, is that if the two can get together with relatively low cost and reach a bargain, they will agree on the outcome that maximizes joint profits of their operation. The ownership of the lake, the ownership of the property rights has no effect on the outcome they will agree upon. It influences only the division of the economic profits between the two parties.
Market Failures
Solutions to Externalities

Applying the Coase Theorem

Let's do another example of the Coase Theorem that will help us see when government actions may be necessary.

Let's change the numbers a little bit. Suppose the box maker can still sell his boxes for 50 cents a piece and let's suppose the brewer can still sell her beer at $1.00 per brew. But let's change the cost of pollution now. Let's suppose that the cost for cleaning up the pollution for the box maker is 50 cents. That is, every box he produces creates pollution and if he has to stop the pollution from going into the lake, it's going to cost him 50 cents per box.

Let's suppose the cost to the brewer of cleaning up the pollution, once she takes the water in to her brew factory that it only costs her 25 cents to clean it up and make it fit for beer. So, we've changed the costs of dealing with the pollution. The box maker now has a lower cost of dealing with the pollution. The outcome in this case is going to be different. If we have independent operations, the brewer is still going to have to deal with polluted water, because the boxer is going to choose to dump the pollution into the lake, rather than incur the 50 cent cost of cleaning it up. So, his profits from making a box are now 50 cents, because he's dumping the pollution in the lake.

The brewer now, in order to clean up the water, is going to have to incur a cost of 25 cents. So her profits from making a brew, after she incurs the 25 cent cost of cleaning up the water is $1.00 minus 25 cents or 75 cents. The total profit of the operation now is 50 cents for the box, 75 cents for the beer or a total of $1.25. Now, if the boxer and the brewer get together and want to try to improve upon this situation, the brewer is only willing to offer the boxer 25 cents to stop polluting the lake. Well, 25 cents is not going to cover his cost of having to bottle the pollution, remember, that's 50 cents. So, the brewer and boxer will be unable to reach an agreement that would result in the lake staying clean.

Now, this is an important point. The value in this particular case is maximized when the brewer cleans up the water, because her cost of cleaning it up are lower than the boxer's cost of cleaning it up. In this case, the outcome that maximizes profits for the two parties involved will be for the boxer to pollute the lake and for the brewer, then, to clean up the lake before she makes the beer. So, the lake gets used as a sink before it gets used as a drink. Now, if you don't believe me, consider how much profit would be created if in fact the boxer had to go to the trouble of protecting the lake. If he had to go to the trouble of blocking the pollution himself, in this case, he'd sell the box for 50 cents, it costs him 50 cents to make it, you get zero profit out of the box operations. The brew, now, is made with clean water and with the clean water there's no cleaning costs, so the brewer makes $1.00 per beer.

The total economic profit from one box and one beer is now only $1.00, which is less than $1.25. You get less total economic profit from these two operations when the boxer cleans up than when the brewer cleans up. So, the Coase Theorem, once again, results in the outcome that maximizes value. However, we're only considering value for the two players in the story. Let's suppose I introduce another player.

Suppose, here's a swimmer who likes to swim in the lake. And let's suppose that not only one, but hundreds of people come to swim in this lake on a regular basis. If the water gets polluted, all of these swimmers may decide that it's not worth swimming. It's not worth getting ear infections. It's not worth getting nauseated. Or the water may be so dirty it's not even enjoyable to get in. In that case, the value of the lake to the swimmers may be destroyed. So, will the swimmers get together with the boxer and the brewer and reach an agreement about the pollution? Well, they could, but if there are a whole lot of swimmers, the transaction cost of getting everybody to the bargaining table may be so high that it wouldn't pay them to go to the trouble of trying to protect the lake.

Rather they would just let the lake become polluted, they would complain, they would be unhappy and they would go look for someplace else to swim or stop swimming altogether. When the transaction's cost of organizing an agreement like this become very, very high, it may be less expensive for society to have a government come into the picture and declare that the lake must remain clean. If the government represents the interests of all of the swimmers, the government can enter the arrangement on behalf of the swimmers and declare that, in fact, it's better for society if the boxer has to clean up the pollution.

Even though the boxer and the brewer might agree that it's better for the lake to be polluted, that it's better for the brewer to clean it up than the boxer to prevent the pollution, if we introduce a lot of other interests parties, who have a very high transactions cost of making their wishes known, then it may be easier to have a government step in and impose a solution on the problem, to impose a rule rather than simply allowing these two parties to bargain. Again, we go back to the rules of the Coase Theorem. The Coase Theorem assumes that the cost of getting the parties to
Market Failures
   Solutions to Externalities

**Applying the Coase Theorem**

the table to bargain is relatively low. If that transactions; cost is low, then we can count on bargaining among the parties to maximize economic value, without any kind of extra government involvement.

However, if transactions' cost are high, it may be less expensive to have the government step in and help to solve the problem. So, there you have it. The Coase Theorem challenges the notion that the government is needed in every case to deal with externalities. Some externalities can be solved by private bargaining, or private negotiations between the parties involved. We can expect that if those parties can get together with low transactions cost, they will agree on an outcome that maximizes the size of the pie, that maximizes economic value. Moreover, if there are no wealth effects, then the property rights, the ownership of the lake doesn't influence the outcome. The water will either be clean or dirty and it doesn't matter who owns the lake. The only difference the property rights make is how the profit is divided between the two parties. However, if there are other interested parties and their transactions cost of participating in a bargain are very high, then, it may be less expensive to society to have some kind of government actions, some kind of law or regulation to help manage the externality.
International Trade

The Basics of Open Economies

Determining the Difference Between a Closed Economy and an Open Economy

The last time we talked about the market we put supply and the demand together and came up with the equilibrium price and quantity traded. That was the case where the market was closed to trade. That is, all the supply and all the demand were within the economy.

This time we're going to study international trade. We're going to use the supply in the demand diagram with one addition. In the case of international trade, we're going to consider a market or a country that can bring more of the good in from the outside. That is, it can import. Alternatively, this same market or country can sell more of its good to the outside world, that is, it can export. In this case, rather than determining an equilibrium price and quantity, we'll be determining an equilibrium volume of international trade—imports or exports.

Now we're going to be using the same tools that we used before with one modification. Let's look at it. The demand curve this time represents the demand of all of those people who live within the market or the country. To make this more concrete, let's suppose we're talking about a country like France and a product like this one. This is a pommelo, the world’s largest citrus fruit. Now, there are not a lot of these that are actually grown in France, because the climate isn't suitable, and it will be very expensive, that is, a large opportunity cost for the French to give up the other goods that they would have to give up to make these pomeles. So the pomelo is important.

On this curve we're representing the demand of people in France for pomelos. At each price there is a quantity that the people in France will want to buy of the pomelo, given their income, the price of substitutes, the price of compliments, expectations, and all of those other factors that influence demand.

Let's look now at the supply curve. This is the supply curve for pomelos in France, that is, the domestic supply curve. This curve tells you at each price for pomelos how many of them are actually produced by farmers in France itself. This supply curve refers to the supply of pomelos domestically, so I'll go ahead and put a little \( d \) in my diagram to remind me that this is the supply of domestic farmers. The demand curve is the demand for pomelos domestically, so I'll put a little \( d \) here to remind me that these are the people in the market itself, in France, who want to buy different quantities of pomelos at different prices.

Now, if this were a closed economy the price of the pomelo would be determined where the supply curve and the demand curve meet. That is, there's only one price at which the quantity supplied and the quantity demanded are equal. That is, there is one price at which we have an equilibrium. And the quantity demanded of pomelos is on the quantity axis below that intersection.

But this is going to be a different story. Rather than considering the closed economy, we're going to consider an open economy, and we represent the open economy by inserting a world price line. The world price line reminds us that as far as pomelos are concerned, France is a small country. That means the farmers in France can produce as many Pomelos as they're capable of without producing so many that they actually affect the world price. People in France can buy all of the pomelos that they might be inclined to buy also without influencing the world price. Most of the pomelos in the world are grown outside of France, and most of them are consumed outside of France. So as far as this particular fruit is concerned, France is a small country, a country without influence on the world price.

So we can put the world price in our diagram, and it's going to be a constant. No matter how many pomelos are supplied and demanded in France, there will be no effect on the world price. Let's put the line in. Suppose the world price for pomelos is right here, \( P_w \). This is the world price, and we might imagine that it's $1.00 per pomelo. That is, you can buy or sell all the pomelos you want to on the world market for $1.00 apiece. So I'm going to let this line go across the diagram. It represents the opportunity for international trade. It represents the opportunity to buy and sell outside of France, that is, in the rest of the world, at a price of $1.00 per pomelo.

Now, with the world price given at $1.00, people in France are always going to have the opportunity to buy pomelos for $1.00 apiece from someone outside of France. They might import them from Spain or Africa or somewhere else. Any farmer who produces pomelos in France has to compete with these imports. That is, they have to compete with farmers other places who are able to produce and sell pomelos for $1.00. That is, no one will pay more than $1.00 because of the alternative to import the good. And no farmer can charge more than $1.00 because otherwise he or she will be undercut by competition.
International Trade

The Basics of Open Economies

Determining the Difference Between a Closed Economy and an Open Economy

So what happens in this market? The world price is fixed. France is a small country, and it takes this world price as given. What then do the people of France do? Well, let’s look at our supply and demand curves, which represent the behavior of people in this market. On the blue curve we can see what the French farmers are going to do when the price of pomelos is set at $1.00 per fruit. That is, they can cover the opportunity cost of producing only a few pomelos before the opportunity cost rises to a point where it’s no longer profitable for them to make the fruit. The quantity supplied domestically will be given by this intersection. This intersection tells us the quantity that the French farmers can afford to supply, so we’ll use a QSD to represent the domestic quantity supplied.

Going over to the demand curve, we see how many pomelos French customers want to buy at a price of $1.00 each. This is the quantity that people in France are willing and able to purchase at the price $P_w$, maybe $1.00 per pomelo. Notice that the quantity demanded is a much larger quantity than the quantity supplied. The quantity demanded domestically is very large, because French people find the pomelos to be a bargain and are very happy to buy a lot of them. Now, here’s where international trade is different than the closed economy. In the closed economy, you couldn’t have a situation where the quantity demanded is greater than the quantity supplied. Remember, we call that excess demand. And when excess demand is present, the bidding mechanism pushes up the price of the good, and so we return to equilibrium.

In this case, however, we imagine that because international trade is possible, the quantity demanded domestically can be larger than the quantity supplied domestically. It’s possible for people in France to eat more pomelos than French farmers produce. How do they do it? They import the difference between the domestic demand and the domestic supply. The difference between these two points represents the quantity of pomelos that are imported in equilibrium.

In the international trade equilibrium, we don’t determine the price, the price is given from outside. What we do determine is the difference between the quantity supplied domestically, and the quantity demanded domestically. That is, the volume of imports. And the volume of imports will depend on domestic supply, domestic demand, and the world price. So to summarize, with international trade, instead of finding the place where the curves intersect, you put in the world price line and find the difference between the quantity supplied domestically and the quantity demanded domestically. This is our equilibrium.
International Trade

The Basics of Open Economies

Understanding Exports in an Open Economy

We’re back on location talking some more about international trade. That’s why I’m wearing my special shirt. In the last lecture we showed how you could use the supply and demand diagram to represent an equilibrium when the world price is given. Remember, we were talking about France, and this product, the pomelo.

Let’s look back at that diagram that we drew before. The red curve represents domestic demand, and the blue curve represents domestic supply. The world price is given and the difference between domestic demand and domestic supply at that price is the volume of imports. Now, notice something. If France were a closed economy and had to make its own pomelos, the price would be up here at $P^*$ where the curves intersect. Because the world price is below this price, the price that would occur in a closed economy, we say that France does not have a comparative advantage in the production of pomelos. That’s why France imports them. Because, if French farmers had to make pomelos themselves, their costs would very quickly rise above the world price. That is, the blue curve rises very quickly above $P_w$. It’s cheaper for France to import its pomelos rather than to make them at home.

Now we’re going to talk about another case. In this second case we’ll look at another product where France does have a comparative advantage. That is, a situation where the world price is greater than the domestic equilibrium price. The product that we’ll choose to examine is wine. Let’s imagine now that France is, for the sake of our story, a small country when it comes to the production of wine. That means that France can buy all of the wine that it wants and make all of the wine that it wants without influencing the world price for wine. We have to make this assumption of a small country, even though it’s unrealistic. We have to make it in order to be able to use the model that we’ve been using.

So let’s look at the picture. In this case, the red demand curve represents the domestic demand for wine. It tells us how much wine people in France would like to consume at different prices, given their incomes, price of substitutes, compliments, and all of those other things that influence demand. The blue curve represents the domestic supply of wine. As the price rises, producers in France will make and offer for sale more and more wine as they are able to cover the increasing opportunity cost of making more and more wine.

If France were a closed economy, the price of wine would be determined where the demand curve and the supply curve intersect. This would be the equilibrium price, and here you’d have the equilibrium quantity. Now, let’s suppose, for the sake of our story, that the world price for wine is greater than this equilibrium for a closed economy. Let’s put the world price for wine way up here, maybe $8.00 a bottle. At this world price for wine we can draw in a dashed line that touches the demand curve and the supply curve and we’ll see how the French economy would respond to the opportunity to trade internationally.

Well, notice first, at this high price for wine, French consumers want to purchase this quantity determined by the intersection of $P_w$ with the demand curve. The quantity of wine demanded domestically we can write down here on the axis as QDD. The quantity of wine supplied domestically can be found by going over to the domestic supply curve. At the price of $8.00 per bottle, the domestic producers of wine want to produce this larger quantity, and we can mark it down here on the axis as QSD, the domestic quantity supplied.

Now notice, the quantity supplied domestically is greater than the quantity demanded domestically at this higher price. If we were in a closed economy, of course, this could not be an equilibrium. The bidding mechanism would push the price of wine downwards until we reached the intersection of the two curves. But because we have an opportunity trade internationally, this higher price for wine does not give us an excess supply, but rather gives us a lot of extra wine that can be exported. The difference between domestic consumption of wine and domestic production, at this high price, becomes France’s export volume, that is, the difference between the two is the quantity of wine exported. The equilibrium in this case is not a price level. The price level is given to us from outside the model, outside the story. The equilibrium is a volume of wine exported, a difference between the quantity supplied and the quantity demanded in France at this particular world price.
I’m still wearing my special shirt and that can mean only one thing, we have yet another lecture on international trade. This time we’re going to look at how changes in the economic environment influence the equilibrium.

Let’s go back to the story we were telling before, the case of the pomelo, France’s imported good. You’ll recall from our diagram that we showed that at the world price of $P_w$, there was a larger quantity of pomelos demanded domestically than there is supply, so the extra pomelos are imported. Now we’re going to do some comparative statics exercises, that is, we’re going to show how the equilibrium changes when we make a change in one of the variables that we had previously held constant.

Let me give you an opportunity then, to walk through this comparative statics exercise with me, and see if you can guess where it’s going. Here’s the change that we’ll be analyzing. How will the equilibrium be affected if French consumers become wealthier, that is, if there is an increase in income in France? Now let’s go through those steps that we go through when we do a comparative statics exercise.

Step number one—identify who in the market cares about this change. That is, whose behavior is affected directly. Will it be the buyers or the sellers? You pick. The answer is, the buyers care directly. As income increases, that is going to influence the quantity demanded at every price. If we suppose that pomelos are a normal good, that means as incomes increase, that French consumers will respond in a predictable way.

Now we’re ready for step two. Which curve will shift and which way will it shift? The supply curve, or the demand curve, inward or outward? You decide. Well, since pomelos are a normal good and since income is increasing, the demand curve is going to shift, representing a change in the behavior of the buyers. Since pomelos are a normal good and French consumers are richer, the quantity demanded will be larger at every price than it was before. That is, French consumers will be buying more pomelos at every price than they were buying when their incomes were smaller. So we can show the shift as the demand curve moving outwards to its new position.

Now, the final question. What happens to the equilibrium? Does the equilibrium price go up or down? Ah, I tricked you, because this is international trade, remember? The price does not change. The price is given by the world. France is assumed to be a small country, so the price is always going to be $P_w$, the price at which pomelos trade on the world market. What happened as the demand curve shifts outwards is we move outwards to a new equilibrium. The quantity of pomelos imported actually increases as the demand moves outwards. That is, our new equilibrium occurs at the same price as before, only now, because of the increased demand associated with the increased income in France, the volume of import has increased.

So with international trade, your three questions will be—who is affected directly by the change, which curve shifts and which way, and finally, what happens to the pattern of trade? Do imports increase or decrease? Or do exports increase or decrease? You’re now ready to try some more problems on your own.
We’re going to do some normative economics now. Normative economics, you’ll remember, answers questions like, is this outcome good? Can it be improved? The question that we’ll be asking at first is this question, how much value is created by the existence of a particular market? What is this market worth to society? Now, the question of value means measuring the benefits that come from an activity and comparing those benefits with the costs. Value is defined as the difference between the benefits and the costs of a particular activity. So whenever we look at the value created by a market, we’ll be looking at the benefit created by the trades that occur in that market and comparing them with the costs that are created when those trades occur. Benefits and costs and value.

The first question that we’ll need to answer is, how do we measure the benefits that come from trades? How do we measure the benefits that are created in a market? Let me propose this answer to the question. That the benefit that you get from buying something is equal to the maximum price you would have paid to have that good. The maximum price that a consumer would pay to have a particular good is called the reservation price for that consumer for that particular good. Suppose we’re considering a good like bread, and you’re hungry, and the maximum price that you would be willing to pay for a loaf of bread, rather than do without it, is $5.00. In that case, $5.00 is your reservation price for that loaf of bread. The question is this—is $5.00 a good measure of the benefit you derive from eating that loaf of bread? Is $5.00 really what that loaf of bread is worth to you and is worth it to society that you enjoy?

Now, let’s break that question down into two parts. If your reservation price for a loaf of bread is $5.00, that is, that’s the maximum price you would be willing to pay for that bread, does that $5.00 represent the benefit that you get from that bread? Well, suppose you’re a poor person. That is, suppose you only have $5.00, but you’re very, very hungry, and if you had more money, say, $10.00 or $20.00, you would pay all of it to get that loaf of bread. In that case, the $5.00 isn’t telling us so much about how much benefit you get from the bread. It’s telling us how much money you have in your pocket. If we gave you more, you’d be willing to pay more.

In this particular case we say that you are subject to wealth effects. Wealth effects means that your income influences how much you would be willing to pay for something, and this becomes a problem when we’re talking about necessities like food. Surely you’d be willing to pay more for food rather than do without it. But you’re not able to pay more, because your income is restricted. If there are wealth effects, then a person’s reservation price is not a reliable measure of the benefit that he gets from consuming that good.

So if we want to make an equation between reservation price and benefit, we have to assume that there are no wealth effects. That is, this is not a situation where poverty is important, or poverty is having a big influence on what people are willing and able to pay.

The second concern that we have concerns the question of the social benefit of the bread. Sure, the bread may be worth $5.00 to you, but does it have extra value for some other people who aren’t eating it? Maybe children that you’re able to take care of when you feel stronger from eating the bread. Or to take another example besides bread, what about a flu shot? The flu shot may be worth $5.00 to you, but if you get a flu shot, then you’re less likely to sneeze on me and give me the flu later. So there are other people who are affected by your choice. Your choice affects other people besides just you. There are not only the private benefits, but there are also what we call external benefits. Your consumption creates benefits for other people.

Now, when I go to get a flu shot I may not think about all of the other people that benefit from my health, and because of that, my willingness to pay for a flu shot is lower than the benefits than I’m creating for society as a whole. That is, my price—the price that I’m willing to pay, reflects my private benefits, not benefits that are created for others.

So if we want to make an equation between social benefit and reservation price, we’re going to want to assume that we’re dealing with a product that does not have a lot of external benefits. That is, a product whose benefits are confined to the person who buys it and uses it.

So a quick summary. Reservation price might be a good measure of social benefit. That is, the benefit that a good creates for society. But it will only be a good measure if two conditions are satisfied. First, there should be no wealth effects and second, there should be no external benefits.
Evaluating Market Outcomes

Normative Economics

Using the Demand Curve as a Measure of Benefit

We’re talking about normative economics. We’ve discussed that the demand curve may be a way to register the benefits that trade generates for society. The demand curve can be viewed from two perspectives. The way we’ve looked at it before when we were doing supply and demand diagrams is as follows. We took a price, say a price like $2.10 for a loaf of bread, and we followed that price horizontally over to the demand curve and then down to the axis to find the quantity of bread that is demanded at that price. That is, the quantity that people in the market are willing and able to buy when bread is selling for $2.10 a loaf.

The way we’ve usually read the demand curve is as follows: people take the $2.10 prices given and they respond by purchasing five loaves of bread. When we do normative economics, we read the demand curve in a different way. Assuming that there are no wealth effects and no external benefits, the demand curve tells us the benefit that’s created for society when an additional loaf of bread is purchased.

Suppose we’re talking about a market where four loaves of bread have already been traded, and we want to know how much value is created for society when that fifth loaf of bread is consumed by someone. Well, go to number five, the fifth loaf, and define the benefit that it creates for society. Follow the number five up to the demand curve. Go from this point on the demand curve over to the vertical axis, and you’ll find that this loaf will be traded when the price of bread is $2.10. That is, only when the price of bread falls to $2.10 a loaf will someone in this market purchase the fifth loaf of bread. What does that tell you? It tells you that $2.10 is somebody’s reservation price. Only when the price reaches that threshold will someone purchase that fifth loaf. Because $2.10 is somebody’s reservation price, we imagine then that that fifth loaf is worth $2.10 to that person, that is, if there are no wealth effects. And since bread is considered a private good, that is, all the benefit, we will say, is enjoyed by the person who bought that loaf, there are no external benefits, then the personal benefit that that consumer gets is also society’s benefit, since that person is a member of society. And society’s benefit is a sum of all of the individual benefits from the people who enjoy bread in this market.

So let me repeat. When we read the demand curve the usual way, we pick a price, go over to the demand curve, and find the quantity demanded. When we do normative economics, we look at the demand curve from the other perspective. We take the extra loaf of bread—in this case, loaf number five—go up to the demand curve and over to find the reservation price of the customer who buys that fifth loaf. And the reservation price, if there are no external benefits and no wealth effects, will be equal to the benefit that that loaf creates for society.
Evaluating Market Outcomes
Calculating Total Economic Value

Quantifying Benefit

If you'll look over in the content box, you'll see a demand schedule. You may recognize it as the demand schedule we used earlier in the course when we were deriving the demand curve. Now we're going to look at those same numbers from a different perspective.

The first entry in the demand schedule says one loaf of bread is demanded when the price of bread is $5.00 per loaf. What that means is this—that someone somewhere in the economy will buy that first loaf of bread if, and only if, the price falls as low as $5.00 per loaf. That is, the reservation price of someone, for that first loaf of bread, is $5.00. Let's look at the diagram.

I can put my dot right here to represent a point on the demand curve. Only now I'm thinking about this point as a normative economist. This point represents the benefit that a consumer derives from consuming that loaf of bread, that is, if there are no wealth effects, and no external benefits.

The second loaf of bread will be purchased, perhaps by a different consumer, when the price drops down to $4.00 a loaf. At that point, we have a second loaf of bread purchased, and a different amount of benefit because the reservation price is different. Two loaves of bread generates how much benefit, in total, for society? Why don't you make a guess? The answer is, the two loaves together will create a total benefit for society of $9.00, $5.00 worth of benefit from the first loaf, plus $4.00 worth of benefit from the second loaf. What I'm doing is adding up the reservation prices.

The third loaf now has a reservation price of $3.00, so the total social benefit from three loaves of bread is $5.00 plus $4.00 plus $3.00 for a total of $12.00. The fourth loaf of bread is going to sell when the price of bread drops down to $2.50 a loaf, so I can put this in here with a price of $2.50 and a quantity of four loaves. And the fifth loaf of bread will sell when the price drops as low as $2.10 per loaf.

Now, another question for you. What is the total benefit for consumers when five loaves of bread are produced and traded? The answer is $5.00 plus $4.00 plus $3.00 plus $2.50 plus $2.10. The total amount of benefit generated--$16.60 to society. If you want to see that in the picture, you can shade in vertically the areas under each one of these loaves. The total benefit generated by five loaves of bread produced and traded is this area underneath the demand curve. If we were to connect these dots with the demand curve, which would look sort of like this, the total benefit generated by five loaves of bread being provided to customers, would be the area underneath the demand curve. That is, the sum of each one of these vertical segments adds up to the area underneath the demand curve for all the loaves of bread that are provided to consumers.
Evaluating Market Outcomes

Calculating Total Economic Value

Quantifying Cost

Remember, we're doing normative economics and our goal is to answer the question: how much value is created in a market? We've looked at one side of the question of values, that is, the benefit side, and we saw that benefits could be measured by looking at the demand curve. Now we're going to concern ourselves with measuring the costs of economic activity. And the costs we're going to gauge by looking at the supply curve.

Ask yourself this question—what is the cost to society of producing a loaf of bread? Well, the first thing that we'd want to do is go to the baker and ask him, for what price would you be willing to provide us with a loaf of bread? The minimum price that a baker would accept to produce a loaf of bread is called the baker's reservation price. Now, is the reservation price a reliable measure of the cost of that bread to society? The baker probably is not going to accept a price that's less than the cost that he incurs to make the bread. He's got to cover his utility bills, he has to pay his workers, and he has to pay for resources like flour and water that he uses to produce the bread.

However, if this baker were poor and desperate, he may be able to make and accept a price that's lower than it's true cost. So the first thing that we want to assume whenever we try to equate the baker's reservation price with the cost of making bread is that there are no wealth effects. The second thing we want to assume is that the baker bears all of the costs himself. That is, the baker is not able to push some of the cost off onto others, perhaps by running an oven that pollutes his neighborhood and dirties the clothes at a laundry that's next door. We want to make sure that all of the costs to society are private costs of the baker. That is, the baker bears all of the cost, so that any price the baker is willing to accept reflects all of the costs he has to cover, and that all of the costs associated with the bread are private costs of the baker. If there are no wealth effects and no external costs, then the baker's reservation price will be equal to the cost to society for the social cost of that loaf of bread.

Now, we can show this on a supply curve just like we showed benefits on a demand curve. Let's take a look at the graph. Before, whenever we had a supply curve, we picked a price to find out what quantity that producers will be willing to supply. Suppose we pick a price like $2.10 a loaf of bread. At a price of $2.10 a loaf of bread, we went over to the supply curve to find out how many producers could cover their opportunity costs at that price, and then we went down to the axis to find out how many loaves of bread would be supplied when the price was $2.10. When we did usual supply and demand diagram, positive economics, we went from the price to the curve and down to the axis to find the quantity supplied.

When we do normative economics, on the other hand, we look at things from the other perspective. We ask ourselves, what is the cost to society of providing that fifth loaf of bread. We go the number five on the quantity axis, and then this time we go up to the supply curve to find the price at which some producer is willing and able profitably to offer than fifth loaf of bread for sale. We read the curve from the other perspective, or in the opposite direction. What we want to know is, what is the cost to society of the fifth loaf of bread. When we discover that some baker is willing to offer that bread, the fifth loaf, at a price of $2.10, that is, his reservation price, then we conclude that if there are no wealth effects and no externalities, that $2.10 is the cost to society of that fifth loaf. This is what we call the marginal economic cost, or the marginal cost to society of that particular loaf.
Evaluating Market Outcomes
Calculating Total Economic Value
Determining Total Social Cost

Remember, we're doing normative economics and we're trying to find the cost to society of producing bread for a market. You'll look over in the content box and see a supply schedule. The supply schedule that we've used before when we derived the supply curve. We'll use the same numbers now with a different interpretation.

If you'll look at the supply schedule you'll notice that the first loaf of bread is supplied when the price is as high as 40 cents a loaf. We'll put a dot here on the supply schedule to represent that point. We give it the interpretation that the cost to society of the first loaf of bread will be 40 cents. We can only use that interpretation, however, if there are no wealth effects for the producer, and no external costs.

The second loaf of bread will be produced when the price rises as high as 60 cents a loaf. Put another dot here to represent the opportunity cost to some producer of producing that second loaf of bread. The third loaf of bread will come onto the market when the price goes as high as $1.00 a loaf. Each dot represents the marginal cost to society of producing the extra loaf of bread.

Now, a question for you. What's the total cost to society of producing three loaves of bread? The answer is 40 cents, plus 60 cents, plus one dollar, for a total of $2.00. Let's add two more loaves. If we add the fourth loaf of bread, the opportunity cost of producing it for some producer to society is going to be $1.50, and the fifth loaf of bread is produced only when the price rises as high as $2.10 a loaf.

Now, another question. What’s the total cost to society of producing five loaves of bread? The answer is found by adding up the points and getting a total of $5.60. How could we see this in the graph? It's the area underneath each one of these dots added up. If you add up these vertical amounts, you get the total cost to society of producing five loaves of bread. Or if you connect the dots and consider this part of the supply curve, these first five loaves of bread have a cost to society equal to the area underneath the supply curve for five loaves. If you look at the area that I've shaded in, it represents the total cost to society of making five loaves of bread. That is, if there are no wealth effects and no external costs.
Evaluating Market Outcomes

Calculating Total Economic Value

Understanding Economic Value

Now that we have a measure of economic benefit, and a measure of economic cost, we can combine them to get a measure of economic value, and answer the question that we started with—what is the economic value created by the existence of the market for bread. Let's go back to our picture and look. We took the demand schedule and put the dots along this axis to represent the benefit created by each additional loaf of bread consumed in this market. We took information from the supply schedule representing it with blue dots. In this case, each blue dot tells us the marginal cost of producing an additional loaf of bread.

Let's see now how much value is created by each loaf of bread that's produced and traded. Suppose one loaf of bread is produced by a baker and given to the consumer to enjoy. This first loaf of bread is worth $5.00 to some consumer. It costs 40 cents to produce. How much economic value is created by this first loaf of bread? The answer is $5.00 worth of benefit minus 40 cents of cost for a total of $4.60 of economic value.

How much economic value is created by the second loaf? The answer is, $4.00 is the benefit to a customer, 60 cents is the cost to a baker, so the economic value will be $3.40. We can continue in this way, adding up the economic value of each additional loaf by finding the marginal benefit created for a consumer when that loaf is produced, and subtracting the marginal cost that the baker incurs to make it. In each case we'll get a measure of the marginal value or the economic value created by the extra loaf.

If we keep going past the fourth loaf, we come to the fifth loaf of bread. This fifth loaf of bread is interesting. It's the one where the marginal economic benefit of $2.10 is exactly equal to the marginal economic cost of $2.10. This loaf of bread is a break-even for society. That is, the fifth loaf of bread generates no economic value. The benefit and the cost are equal. This loaf is interesting because if we push past it and go to the sixth loaf of bread, we wind up with negative social profits, or a social loss. You can see here that the sixth loaf of bread is worth $1.80 to some consumer, but it cost the baker $2.50 to produce. The economic loss incurred is the excess of cost over benefit, in this case $2.50 minus $1.80 for an economic loss of 70 cents. This bread is bad bread as far as society is concerned, because the cost is greater than the benefit. And even more so, if you go on to the seventh loaf, the eighth loaf, and the ninth loaf.

So if you want to calculate the economic value associated with the market, take all of the loaves of bread that are consumed and add up the benefit of each loaf. Then subtract from that total benefit the total cost, and you'll find the total cost by taking the cost of each individual loaf, adding it to the cost of the other loaves to get total economic cost.

If, in this case, we have five loaves of bread that were consumed by these consumers, those consumers who value the bread the most, and they were produced by these bakers, these bakers who have the lowest cost to produce bread. We'd have total benefit, the area under this curve, minus total cost, the area under the blue curve, and the difference would be the area under the red curve minus the area under the blue curve, or the area between the two curves. That is, if we have these five producers and these five consumers, the total economic value generated in this market would be the area between the red and blue dot for all of those loaves of bread that are produce and traded. If we stop at this point right here and produce these five loaves for these five consumers by these five bakers, what would be the total economic value generated in this market? The answer is—$16.60 worth of total benefit, minus $5.60 of total cost for a total economic value of $11.00. If you can get these buyers and these sellers together, you will generate total economic value of $11.00. Well, that's certainly better than nothing, and that's why the economists would say that these trades should take place.

But if we produce any additional bread, even if we take the lowest cost producers and put them together with the consumers that are willing to pay the most, no matter what we do, we wind up reducing that total value. That is, any additional bread produced beyond five loaves will necessarily shrink the pie. And when I say, “shrink the pie,” I mean reduce the total economic value associated with these trades.
Evaluating Market Outcomes

Consumer and Producer Surplus

Understanding Producer and Consumer Surplus

In this segment we're going to see how the economic pie is divided among the players in the market. Recall that when a buyer and a seller get together and make a trade, the buyer and the seller create economic value. Let's go to an example. Think about that first loaf of bread that we've studied so much. There's a buyer out there who's willing to pay $5.00 for that loaf of bread. That's his reservation price, and if there are no wealth effects and no external benefits, then we imagine that $5.00 is social benefit.

There's a seller out there who's willing to make that loaf of bread for 40 cents with no wealth effects and no external costs, that's society's cost for that loaf of bread. So if we put the buyer and the seller together and they make a deal, then they're creating $4.60 worth of economic value. And I'm going to shade that in my diagram with green to represent economic value. The question is, who gets that economic value? Who really gets the gain from trade in this case?

The answer depends on the price at which this buyer and this seller get together. Let's suppose this buyer and this seller get together and make their deal at a price of $2.10 per loaf of bread. So I'm now going to plug in a price equal to $2.10. If the buyer was willing to pay $5.00 and got the loaf of bread for $2.10, he got a good deal. If the seller was willing to produce the loaf of bread for 40 cents and was able to sell it for $2.10 instead, then the seller got a good deal also. Both of them got some gain from trade. Let's consider now two concepts of economic value that relate to who gets the value.

The first concept is the concept of consumer surplus. Consumer surplus is defined as the difference between the consumer's reservation price and the price at which the consumer actually purchases the good. In this case, our consumer was willing to pay $5.00 for a loaf of bread, and he got the bread for $2.10. The consumer surplus would be $5.00 minus $2.10, or a total of $2.90 worth of economic value. And I'm going to shade that in my diagram with red to remind us that it's the consumer's. And I'll use the abbreviation CS to stand for consumer surplus. The consumer surplus in this case is $2.90, $5.00 minus the price you paid of $2.10.

But suppose, on your way out of the store the baker calls out after you and says, “Excuse me, sir. We're having a special today. Instead of paying $5.00, we're going to give you a special deal. We're going to give you back $2.90 and you get to take that out with you. Or you can stay in the store and spend it on jam or bagels or something else that pleases you.” The 2.90 is your extra bargain, it’s your surplus. It's something that you got in addition to the value of the bread. The bread was worth $5.00 to you, you paid the $5.00, you're no better off. But when he gives you the $2.90 back, now you've got something. The $2.90 is your consumer surplus.

If you look at this picture you can see the consumer surplus. It's the difference between the $5.00 you would have paid and the $2.10 that you had to pay. We can shade this area in with red to remind us that it's the consumer's. And I'll use the abbreviation CS to stand for consumer surplus. The consumer surplus in this case is $2.90, $5.00 minus the price you paid of $2.10.

What about the baker? The baker was willing to give you that loaf of bread for 40 cents, but instead, he got $2.10. That's money beyond his costs. This baker got $1.70 more than it cost him to make that loaf of bread, $1.70 more than he needed to make it worth his while to offer you that bread for sale. You can think of this as a kind of a profit. And in this context, the profit has a special name. We call it producer surplus. Producer surplus is defined as the difference between the price the producer received and the price he would have accepted, that is, his cost of making the good. Producer surplus, in this case, is going to be $1.70, which is $2.10 minus the producer’s reservation price of 40 cents.

Now notice one more interesting thing—interesting, but not surprising. If you add $1.70 to $2.90 you'll get $4.60, and that's the total economic value created by this trade, $5.00 minus 40 cents. The total economic value is divided between the part that goes to the consumer—that's the consumer surplus, and the part that goes to the producer, that is, the producer surplus. What determines the total amount of economic value created is the difference between the consumer’s reservation and the producer’s reservation price. What determines the division of that total value into these two parts? Well, that will be the price at which they agree to make the trade. The price at which they agree to make a trade divides that total economic value into the consumer’s chunk and the producer’s chunk.
Evaluating Market Outcomes

Consumer and Producer Surplus

Calculating Total Economic Value

Now we’re going to solve the problem, using mathematical representations, of the supply curve and the demand curve. Over in the content box right now you’ll see formulas for the supply curve and the demand curve. The equation that describes consumer behavior is the demand equation, that is, the formula for the demand curve—price equals 100 minus quantity demanded. That’s the formula we’ll be using for our example. Let’s start by drawing a graphical representation of that formula.

Here in this diagram we go up and find the intercept point, that is, the price if the quantity is equal to zero. In that case, we’ll have a price of 100 when the quantity is set equal to zero. Now the slope of this equation is seen as the coefficient on the quantity term. That is, -1. That means for each one unit change in the price, the quantity demanded increases by one unit. So we can draw that in, and there you have the demand curve. Let’s label it with a D for demand.

The formula for the supply curve is price equals 20 plus the quantity supplied. That is, in order to get larger quantities supplied, producers have to be able to get higher and higher prices. Let’s graph now the supply equation that you can see in the content box. If you set the quantity equal to zero, the intercept will be 20. The slope of the supply curve, once again, is the coefficient on the quantity term. In this case, positive one. So we can draw an upward sloping line with a slope of one.

Let’s label this blue curve S for supply. Now, we know from our study that the equilibrium point is the point where the supply curve and the demand curve intersect. That is, the price at which the quantity supplied and the quantity demanded are equal to each other. That will be this point here, and the equilibrium quantity traded will be the point below that intersection on the quantity axis. We’ll call that Q*. So what we’d like to do now is solve the formulas to find the equilibrium price and the equilibrium quantity. You can follow my work on this page or over in the content box.

Let’s write down the equations. The price on the demand curve is equal to 100 minus the quantity demanded. The price on the supply curve is equal to 20 plus the quantity supplied. Now to find the place where the curves cross, we want to set the prices equal to one another or the quantities equal to one another. Let’s do that now. Setting the prices equal, we get 100 minus QD is equal to 20 plus QS. Since the quantities are equal at this point of intersection, we can use one quantity for both of these terms. Moving the 20 over to this side of the equation and the quantity over to this side of the equation, we get 80 is equal to two times the quantity traded, or the quantity traded will be equal to 40. That is, in equilibrium there will be 40 units of this good that are supplied and demanded. So let’s put that into our graph.

How do we find the price? We’ll place the quantity that we solved for and plug that in to either one of these equations, either the supply curve or the demand curve. Plugging in, you’ll see that the price at equilibrium is equal to 100 minus 40 or 60, which is the same thing as 20 plus 40. The equilibrium price is 60, the equilibrium quantity traded is 40.

Now, we can use this information also to find out what the value of this trade is, that is, what’s the total gain from trade that’s created in this market. Now, the gain from trade, as we’ve talked about before, is divided between two components. That part of the gain that the consumers get, and that part of the gain that the producers get. The gain that the consumers get will be everything under the demand curve and above the market price. This area represents what the consumers would have paid less what they actually have to pay, and that’s our measure of consumer surplus.

To calculate consumer surplus, we want to find the area of this triangle. To find the area of the triangle, you have to multiply the height of the triangle times the base of the triangle times one-half. You can follow my formulas over in the content box. 100 minus 60 is 40. That’s the difference between the maximum price here and the price that the consumers have to pay in equilibrium. There are 40 units traded, so that’s the base of the triangle. Multiply 40 by 40 times one-half to get that the total consumer surplus in this case is $800.00 worth of economic value that goes to the consumers.

We can now use the same method to calculate the total amount of producer surplus. How much value are producers getting in this market? Take the price that producers are actually able to charge for their product—60, and subtract the cost that the producers incur when they make the product. That’s represented by the height of the blue line. The producer surplus then will be the area of this triangle—60 minus 20 times 40 times one-half. In this case, producer surplus is also equal to $800.00 worth of economic value. Add the two together—consumer surplus plus producer surplus.
Evaluating Market Outcomes

Consumer and Producer Surplus

**Calculating Total Economic Value**

surplus to get a total of $1600.00 worth of economic value. That’s the total surplus, or total economic value created in this market.
Evaluating Market Outcomes

Market Interference and Economic Value

Understanding the Effects of Price Control

In this lecture we'll look at another way that a government might interfere with the market. This interference is the price control, and we'll see that it has different consequences from a tax or a tariff. Let's start with a situation in which we have a baker who's willing to bake a loaf of bread and sell it for $1.00. That's his reservation price. Let's take a customer who has a reservation price of $5.00. Now as you know, with a reservation price of $5.00 representing economic benefit, and a cost of $1.00, the reservation price of the seller, there are $4.00 worth of economic value to be created if we can get this buyer and this seller together.

Suppose now that a price control is imposed in this situation, and the price control will be set at $2.00. The government says that no one can charge more than $2.00 for a loaf of bread. Would that have any effect on this market? The answer is, no, it would not. At a price of $2.00 this buyer would certainly buy the bread and get $3.00 worth of consumer surplus. The seller would certainly sell the bread and get only $1.00 worth of producer surplus. But there's no reason not to do the trade because the price control lies squarely between the reservation price of the buyer and the reservation price of the seller.

The situation gets more complicated, however, if the price control were to fall below the seller's reservation price. If the government decreed that bread could not sell for more than 50 cents a loaf, then this trade could not take place. The seller could not cover his opportunity cost with the price control and therefore the trade would be blocked, it would be a deadweight loss. That is, if the seller and the buyer observed the price control. If they got around the price control by doing some kind of deal in an informal market, they may still actually trade at a price between $1.00 and $5.00 per loaf. They would be ignoring the price control, and the cost there would just be whatever effort they had to expend in order to evade the government's controls.

Price controls in markets, however, can have consequences if we have a different kind of situation. Suppose we have a situation where two buyers are competing for the same loaf of bread. One buyer has a reservation price of $5.00 for the loaf, the second buyer has a reservation price of $4.00, and the baker, like before, has a reservation price of $1.00. He's willing to sell the loaf for any price above $1.00. If this were a free market, these two buyers would find themselves competing using the bidding mechanism, and the price would rise until one of the buyers was pushed out of the market. The market price for this loaf of bread would go slightly above $4.00 and this buyer, with the higher reservation price would wind up getting it. That's the way competition works in the market, and the baker would certainly be very happy for this competition because it allows him to charge a higher price and get more producer surplus.

However, suppose now the government intervenes with a price control. The price control blocks the bidding mechanism and keeps these two buyers from competing by bidding. That is, we can no longer use price as a means of competition. If these two buyers both want this loaf of bread, and certainly they would both be happy to buy it at a price of $2.00 per loaf, they're going to have to resort to some kind of competition besides prices, since the price has been fixed, or at least controlled, by the government. The baker's going to get $1.00 worth of surplus, but now these two guys have to wrestle it out to figure out who's going to be able to buy the loaf of bread at this low bargain price. What we will find the two buyers doing is resorting to non-price competition. That is, since price competition has been ruled out by the price control, they're now going to have to do something else to do the bread. Who's willing to stand in line the longest? Who's willing to run faster to get to the bakery early in the morning? Who's willing to hire someone to go to the bakery first and make sure that they buy that loaf of bread on their behalf? That is, there are other ways to compete to make sure you get the loaf of bread. But whereas the baker would prefer price competition, which gives him a bigger surplus, in this case, with a price control, the buyers would have to revert to non-price competition.

Economists don't like non-price competition, because it seems like resources always get wasted. Somebody wastes their time standing in line, someone wastes his time hiring a bread locator who goes out and looks for bread for you, someone wastes his time running fast to get to the bakery. This is what we call in economics, rent seeking behavior. Rent seeking behavior is the willingness to destroy resources, or expend resources to get a bigger share of the economic value for yourself. There's only one loaf of bread, and these two guys are willing to burn up resources wrestling, running, waiting in line, hiring assistants, in order to make sure that they take the bread away from each other. In the end it will be this guy with the high reservation price who is able to spend the most resources, and it becomes kind of like an arms race. He simply outlasts his competition. He burns up more resources than his competitor until finally his competitor is pushed out of the market.
Understanding the Effects of Price Control

This rent seeking is wasteful, because all of the resources that are used in competition, in non-price competition for bread—all the burned up shoe leather, all the burned up time, all the wasted energy, could be used producing other things of value. It could be used hooking rugs, writing children’s books, anything like that—sembling computer boards. But instead, it’s burned up in non-price competition, activity that does not create value, but merely redistributes it.
Evaluating Market Outcomes

Market Interference and Economic Value

Understanding How Price Controls Destroy Economic Value

We are going to look some more at the problem of price controls, but first, let’s do a quick review of how a competitive market works. Suppose we have three customers in the market for bread: one has a reservation price of $5.00, one has a reservation price of $4.00, and one has a reservation price of $3.50. Suppose we also have three suppliers in this market: one with a reservation price of $1.00, one with the reservation price of $3.00, and one with a reservation price of $3.50. What would the equilibrium price be in this market? The equilibrium price will be $3.50. At that price, the quantity supplied equals the quantity demanded. How many loaves of bread will be produced and traded? The answer is three loaves of bread produced and traded at a price of $3.50.

Calculate now the total consumer surplus, the total producer surplus, and the total economic value created in this market. The total consumer surplus will be $3.50 per loaf of bread minus $1.00, that is, $2.50 worth of producer surplus for the first loaf. For the second loaf, 50 cents worth or producer surplus and zero for the third loaf. So total producer surplus will be $3.00. Consumer surplus will be $5.00 minus $3.50, that is, $1.50 plus $4.00 minus $3.50 for 50 cents and zero consumer surplus for the third loaf. The total consumer surplus, then, would be $2.00. Add $2.00 to $3.00 and get the total economic value of $5.00 created in this market.

Now, let’s consider what would happen in this market if we imposed a price control. Suppose the government enters this market and declares that bread must trade at a price of $2.00 per loaf and no more. That is, a price ceiling is imposed. So we can write out here price control, that is, a prohibition on the bidding mechanism, the price cannot go above that. How many loaves of bread will be supplied in this market? That answer is only one loaf. Only one producer can afford to cover his cost at a price of $2.00 per loaf, that’s this baker right here. The other two bakers find that the price control is less than their opportunity cost, and since the opportunity cost exceeds the price they can charge, they’re both out of business. The first thing to notice, then, is there are going to be some blocked trades here. Some loaves of bread that had economic value, that is, created value for society in this market, are no longer produced and traded. The market is going to be restricted.

The next question is, which customer is going to get this one loaf of bread, and how is he going to get it? This customer right here, the customer who has the highest reservation price will definitely get the loaf of bread. He will get it by using non-price competition. He’s going to expend enough resources to knock the bread out of reach of this customer and this one. That is, since all three customers are willing to buy this loaf of bread at $2.00 per loaf, and since there’s only one loaf of bread to buy at that price, non-price competition will set in, and this particular customer is going to be willing to expend enough resources, $2.00 worth of extra resources, to wave the price of bread out of the reach of the $4.00 customer and the $3.50 customer. That is, what’s going to happen is the first customer is going to burn up resources equivalent to $2.00 worth of economic value, whether that is wasting time or hiring the bread locator or spending his own time in shoe leather, tracking down this loaf of bread where ever it is in the city. He’s going to burn up $2.00 worth of economic value in the process of rent seeking in the process of non-price competition. Once the true price of bread, that is, the $2.00 you have to pay plus the $2.00 that you have to spend in non-price competition, that is, in wrestling activities. Once the total price of bread rises above $4.00, then we find that our other two customers drop out of the market, they can’t keep up with the non-price competition of the highest paying customer. How much economic value is now created in this market with the price control? Well, first of all, these two customers are out of the market and these two bakers are out of the market. That means we have some deadweight loss, that is, economic value that was originally created gains from trade that’s no longer available because customers and bakers have dropped out of the market. There’s only one trade that occurs now, and that trade occurs between this baker and this customer. This customer is willing to pay $2.00 for the loaf of bread plus $2.00 worth of economic waste to make sure he gets there first. So his consumer surplus now is restricted to under $1.00. That is, after he pays the cost of fighting off these other two customers, he only gets about a dollar’s worth of extra economic benefit.

Meanwhile, the baker who charged $2.00, the legal price for his loaf of bread, his cost was $1.00, so he gets one dollar’s worth of producer surplus. Well, if you add this dollar’s worth of consumer surplus to this dollar’s worth of producer surplus, you find that the total economic value in this case has shrunk to $2.00. That is, before we had $5.00 worth of economic value created in the market for bread, now we only have $2.00 worth. What happened? Well, what happened was this: The price control led two of our bakers to drop out of the market. When those two bakers dropped out, we lost a valuable trade that could have taken place between this $3.00 baker and one of these customers.

The next thing that happened was we had three consumers, all three of them were willing to pay the $2.00 price to get the one remaining loaf of bread. What they then had to do was engage in non-price competition and the non-price
Evaluating Market Outcomes
Market Interference and Economic Value

*Understanding How Price Controls Destroy Economic Value*

competition ended up raising the true cost of bread so high that two of them dropped out of the market, and that non-price competition from an economist point of view is purely waste. Time, and energy, and resources that could be used doing something else. If you combine the fact that we restricted supply with the price control, with the fact that non-price competition wasted resources, you shrink the total economic value in this market from $5.00 down to $2.00 and that’s the problem with the price control.
In this segment we’re going to see how an excise tax affects the economic value that’s created in the market. We’re also going to look at how the excise tax changes the distribution of that economic value. Let’s look at the graph that we were developing when we showed how the excise tax affects equilibrium. When we put a wedge between the buyers and the sellers, the buyers pay a higher price, the sellers get to keep a lower price and the quantity that’s traded is a smaller quantity than you would get if we were in a free market, with no taxes.

Look at what happened. In the free market we would be at the point where the supply curve and the demand curve cross and the quantity traded would be a larger quantity—Q*. The first thing that economists worry about when an excise tax is imposed is the deadweight loss that comes about when the wedge is put between the buyers and the sellers. Because of this $2.00 excise tax we have all of these trades where the buyer and the seller can’t get together. That is, the buyer’s reservation price is within $2.00 of the seller’s opportunity cost. They can’t get together and make these trades happen because there isn’t enough value to cover the excise tax. That is, the trades are blocked by the tax. A deadweight loss is created.

So an economist, whenever an excise tax is imposed, is going to worry first about the trades that don’t take place. That is a reduction in economic value. And I could shade in this area, this triangle, which represents blocked trade and call it the deadweight loss or the social cost of the excise tax. Now, let’s look at what we actually get in the taxed economy. How much economic value is actually created? Well, to find the total economic value or the total surplus, look at the difference between what buyers are willing to pay, that is, the benefit of the good to the buyers and the cost of producing them on the part of the sellers. That is, what the sellers are going to accept.

To calculate total economic value, subtract the area underneath the blue curve from the area underneath the red curve. That is, everything between the supply and the demand curve is economic value up to the last unit that’s traded. So in this case, with this excise tax, we have kind of a weird shaped trapezoid here. Our weird shaped trapezoid represents our total economic value. Now how is that total economic value divided up between buyers, sellers, and the government? Well remember, buyers get everything below the demand curve and above the price that they pay. This is what the good is worth to the buyers, this is what the actually have to pay and everything between the red curve and the price that buyers pay is consumer surplus, extra bargain for the consumers.

So let me label this. This is the consumer surplus that we get in the case of the excise tax. Well, there’s the value for the producers. Producers are getting a smaller price for the bread that they’re selling, a smaller price than they would get in a free market. Here’s the price that producers get for each loaf of bread that they sell and here’s the cost to producers of making that bread and selling it. So the difference between the price that they are able to charge and the cost of making the bread is producer surplus, or profit. We can put that down here with PS, labeling the producer surplus, and that’s the area of this small triangle down below the producer’s price.

Well, what about this area right here, the area between the buyer’s price and the seller’s price for all of the goods that are traded? Remember, the difference between the buyer’s price and the seller’s price is the amount of the tax. The difference between PD and PS is the $2.00 excise tax. That is, it’s the amount of the tax that’s charged on each loaf of bread that’s traded. If you multiply the tax per loaf times the total number of loaves that are traded, we get the total tax revenue. So we can label this box Tax Revenue. Again, it’s part of the total economic value that’s created, it’s just the part that goes to the government instead of to the buyers and the sellers.

So once again, our total economic value is the consumer surplus, the tax revenue and the producer surplus, the area of the trapezoid. The last thing we want to do is compare this economic value with what we would get in a free market. In a free market we would trade this larger quantity, Q*, and the price in the free market would be found at the intersection of the supply curve and the demand curve. We’ll call this P*. In a free market everything above P* and below the red curve would be consumer surplus, as we saw before. It would be the area of this big triangle. Everything below P* and above the blue curve would be consumer surplus. What happens when the excise tax is imposed?

The quantity traded shrinks and the price rises, shrinking consumer surplus. The price to the sellers falls, shrinking the producer surplus. Consumers use to have all of this area, the area of this small box and this triangle as part of their consumer surplus, but now that’s gone. This rectangle goes to the government in the form of taxes and this triangle becomes part of the deadweight loss, or lost economic value. This area, this rectangle and this triangle are what we call the cost of the tax to the consumers. Down below it we see the bottom of the revenue rectangle and this
Evaluating Market Outcomes

Market Interference and Economic Value

Evaluating the Effects of an Excise Tax

triangle used to be part of producer surplus, but now they're gone. This rectangle has been transferred to the government as part of tax revenue and this triangle is destroyed when trades are blocked. We call this bottom rectangle and the triangle the cost of the tax to producers. So, part of the cost to producers goes to the government, then part is completely destroyed.

To summarize, an excise tax really does effect the market. It effects not only the quantity of trade that takes place and the total amount of economic value created, but it also effects the distribution of economic value among producers, consumers and the government. Let's take one more look at the graph. In a free market the price would be $P^*$, the quantity traded, $Q^*$, the consumer surplus would be the entire area above the price and below the demand curve. The producer surplus would be the entire area below the price and above the supply curve. When the excise tax is imposed, the buyers will pay a higher price than the sellers will receive and the difference will be the amount of the excise tax. Now, let's look at how this tax effects the market.

The first thing that happens is when the buyers pay a higher price, they're no longer willing to buy the product. The sellers, also, when they get a lower price, they're no longer willing to sell the product. That is, there are some trades that are blocked and that becomes the deadweight loss. The difference between the buyer’s price and the seller’s price is the amount of the excise tax. And so, if you multiply that tax by the number of loaves of bread that are actually traded then you get the tax revenue.

Now remember, this part use to be consumer surplus. Part of it became tax revenue and part of it was destroyed. If you put these two areas together you get the amount of consumer surplus that’s lost when the tax is imposed. We’ll call this the cost of the tax to the consumers. Remember, all of this used to be producer surplus. This part, with blocked trade, and this part was lost when the government to the tax revenue. If you put these two together you get the cost of the tax to the producers or the total amount of producer surplus that’s lost when the excise tax is imposed.

Economists are concerned about excise taxes, not because they don’t want the government to take money. Economists are concerned because when you put a wedge between buyers and sellers you block trades that would otherwise take place.
Evaluating Market Outcomes

Market Interference and Economic Value

Assessing the Effect of an Excise Tax on Economic Value

You’ll recall that economists are very excited about free markets. We like them because they maximize economic value without a whole lot of extra information being required. That individuals trading, individual buyers, individual sellers, will reach a point where economic value is maximized, at equilibrium.

We’re going to talk now about a few ways that the government can interfere with the market and the ways in which government interference can be costly. The first example is that of the excise tax. Now an excise tax is a special kind of tax. It’s a tax that’s imposed on each unit of a good that’s traded. For example, on a gallon of gasoline there might be an excise tax of 20 cents. That means for every gallon of gasoline that’s bought and sold, that 20 cents has to be paid to the government. That’s the definition of an excise tax, a per unit tax.

Let’s look now at how an excise tax, or per unit tax, would affect trade. Let’s go back to an example that we considered before, two people who want to trade bread. And let’s take a loaf of bread where a buyer somewhere is willing to pay $5.00. That’s his reservation price. And a seller that we’ve identified will sell that bread for 40 cents. Notice here we’ve got $5.00 minus 40 cents, or $4.60 worth of economic value to be created if we can get these two together to make a trade.

Let’s suppose now that in addition to the trade there’s an excise tax, that the government expects that for every loaf of bread that’s traded, $2.00 has to be paid to the government. Well, who’s going to pay it? We’ll see in a minute how the buyer and the seller decide how to split the tax. But let’s suppose, for the sake of our example, that the buyer agrees to pay $3.50 for the loaf of bread and after the tax, that means that $3.50 minus $2.00 goes to the seller. So the seller will get $1.50. Notice what the tax is doing. It’s making a difference between the price that the buyer has to pay, and the price that the seller gets to keep. The $2.00 here in the middle goes to the government, and that is what we call a wedge, a tax wedge, because it separates what the buyer pays from what the seller gets.

Well, let’s look here at how the tax influences the division of economic value. If the buyer is paying $3.50 for this loaf of bread, and he’s willing to pay up to $5.00, then he’s getting $1.50 worth of consumer surplus. The seller was willing to make the bread for 40 cents. That’s his opportunity cost. But he actually gets $1.50, so he gets $1.10 worth of producer surplus. The remaining $2.00, the difference between what the buyer pays and what the seller gets goes to the government. That’s the excise tax. Well, the excise tax isn’t destroyed, it’s just taken away from these traders and transferred to someone else. So I’ll use my green marker here to remind me that this is still economic value. It’s just been moved somewhere else for someone else to spend. And we’ll call this the tax revenue.

The tax revenue is presumably available to be spent on public good, like road and bridges or anything else that the government may be providing for the citizens. So what we see in this case is the total amount of economic value that’s created is the same as it was before, $4.60. But the division of the economic value is affected by the excise tax. There’s less consumer surplus and producer surplus, and more for the government.

What does an economist say about this outcome? Is it good? Is it bad? Well, the economist doesn’t really care. What economists care about is how much surplus, how much economic value is created. Economists don’t really have an opinion about who gets the economic value, that is, how it’s distributed. Economists as economists don’t have an opinion about that. Economists only get worried or concerned if the amount of economic value that’s created starts to shrink. But in this case we still get the $4.60, it’s just divided differently than it was before.
Let's look at a case now where the excise tax does affect trade and economic value.

Recall that the government requires a $2.00 excise tax on every trade of bread, on every loaf of bread traded. Suppose now we've got a buyer who's willing to pay $3.50 and a seller with an opportunity cost of $2.00. Can they get together and trade? Well, look; if the seller has to pay the excise tax, then he's going to want to charge $4.00 for the bread; that is, $2.00 for the government and $2.00 for him. On the other hand, if this buyer has to pay the excise tax, then the seller would only be able to get $1.50 after the government's share was subtracted. There's no way to fit this $2.00 excise tax wedge in between these two reservation prices. If these two traders have to pay $2.00 to the government, there's no way they can get together and agree on a price for trading the bread. Now that's a problem. It's a problem because, in the absence of the tax, you can see there's economic value here to be had. The buyer's reservation price, or the social benefit, is greater than the seller's reservation price, which is the social cost. So if we could get them together, there would be economic value of $1.50 created by having them make a trade. But because the government wants $2.00 worth of tax for every trade of bread, there's no set of prices that will make these guys get together and make a trade. The seller will either charge a price that's too high or he'll have too little left over after paying the tax to the government. So if a $2.00 excise tax is required on every trade of bread, this trade of bread won't take place. That's $1.50 worth of economic value that's just not created or lost, because of the excise tax. And we put a big mark over this trade to indicate that it doesn't take place, and we call this lost potential value, a deadweight loss.
I'm back in my special shirt on location to talk about international trade. Before we talk about the cost of interfering with international trade, let’s make sure we’re clear on where the benefits from international trade come from. Let’s look back at the graph that we used for closed economy. Suppose we are in France and we’re looking at the market for this product. Do you remember what it’s called? That’s right, it’s a pomelo. We’ve got a supply curve for pomelos in France and a demand curve for pomelos that tell us how many pomelos the French people want to buy at different prices.

Now, if you have a closed market for pomelos, the equilibrium point will be where the supply curve and the demand curve intersect. Here’s the price, \( P^* \), in equilibrium. Here’s the quantity, \( Q^* \), in equilibrium. How much economic value is created in this market? Can you identify the consumer surplus? That’s right, the consumer surplus is the area above \( P^* \) and above the demand curve. Where’s the producer surplus? The producer surplus is the area below \( P^* \) and above the blue supply curve. The total economic value created in a closed economy is the area between the demand curve, which represents economic benefit, and above the supply curve, which represents economic cost. Between these two curves is the difference between benefit and cost or economic value for this entire market.

Suppose now that we introduce international trade, and in the case of the pomelo, we’re going to allow France to import this fruit at a lower price from the rest of the world from outside the market. So, let’s put in the lower price for pomelos, \( P_w \), and at this lower price, people in France can buy as many pomelos as they want and farmers in France can only charge this price if they produce pomelos domestically. With international trade, we think of the red curve as domestic demand, the blue curve as domestic supply or the cost of producing pomelos domestically, and we realize that all trades that take place have to take place at the world price \( P_w \) because no customer is going to pay a higher price when they can import the good at a lower price from somewhere else. That means the farmers in France cannot charge a higher price, they must charge the world price. At this lower world price, how much consumer surplus is created? See if you can identify the area that represents total consumer surplus when people in France can import pomelos at this low world price. In order to identify the total amount of consumer surplus created, start by finding the total quantity of pomelos demanded at the world price. That’s going to be following \( P_w \) over to the demand curve, here’s the quantity demanded domestically. With people buying this large quantity of pomelos at the price \( P_w \), consumer surplus is now the area underneath the demand curve and above \( P_w \), that is, it’s this entire big triangle. How much producer surplus is now available with international trade? To calculate the total amount of producer surplus, follow \( P_w \) over to the supply curve to remind yourself what the quantity supply domestically is. Producers are able to charge \( P_w \) and their costs are represented by the height of the blue line, the supply curve. Producer surplus with international trade will be this small triangle, that is, below the world price and above the cost of making pomelos domestically.

What is the total amount of economic value created in this market when international trade is available? It will be the sum of the consumer surplus, this large triangle and the small triangle of producer surplus. Now, here’s a question. Has the total amount of economic value increased because of international trade or has it not increased? The answer is, it has increased, and let’s see if we can show why. First of all, notice that as the price falls from \( P^* \) to \( P_w \), as the import opportunity lowers the price of pomelos, the amount of producer surplus created in France for the producers of pomelos domestically. That quantity shrinks, and it shrinks by the amount of this rectangle. The lower price means that a lot of producers in France are going to drop out of the market for pomelos, they’re no longer going to produce them. So we have a shrinkage of domestic supply, a reduction in the quantity supply domestically. Also, because the price is lower even those farmers that are still in the market, still able to produce pomelos, are earning less revenue. They’re making less money instead of earning this large price for pomelos; they’re getting a small price. So there’s a smaller amount of surplus for them even on the pomelos they’re still making. So this entire trapezoid, this area right here, is producer surplus that is no longer available. That is, producer surplus that has been destroyed for French farmers by the fact that imports are now replacing their business. However, the amount of consumer surplus that that’s created is this very large trapezoid. With a lower price, consumers are ecstatic because they’re getting a great bargain and a lot of consumers are now going to buy pomelos who couldn’t afford them or chose not to buy them before. Not only are more customers in the market for pomelos, but even those customers that were buying pomelos at the high price, they’re even getting an extra advantage because they’re able to get their pomelos at a discount, that is, the price is falling, adding to their consumer surplus.

The total increase in consumer surplus is this large trapezoid here. Notice, the large trapezoid, which represents the increased consumer surplus, completely offsets the loss of producer surplus. That is, this small trapezoid lies entirely within the big trapezoid. The net gain for the economy in France, the net gain is the area of this triangle, the triangle
Evaluating Market Outcomes
International Trade and Economic Value

Evaluating the Gains from International Trade

that's below the old equilibrium down to the import price. You can think of this triangle as equal to the quantity of import multiplied by the difference between the old price and the new price multiplied by one-half. That is, this rectangle represents the gains from trade, the difference between the additional consumer surplus or the benefits that consumers get from buying low price pomelos and the loss to producers of losing a market. If you subtract the loss producer surplus from the gain in consumer surplus you get the gain from trade. That is what's left over and it's always going to be positive.
Evaluating Market Outcomes

International Trade and Economic Value

Understanding the Effects of Tariffs on Consumer and Producer Surplus

We talked in an earlier lecture about how taxes can reduce gains from trade. Now we’re going to look at the same set of questions in the context of international trade. Suppose we have a country like France, and France has been able to increase economic value by importing a product, pomelos, in which it lacks comparative advantage. Let’s suppose now the pomelo growers of France get together and decide that they want to have their market protected and lobby the government to impose a tariff on the import of pomelos from abroad. What would happen to the market for pomelos if this tariff were imposed, and what would happen to the total amount of economic value if this tariff were, in fact, enacted?

Let’s look at this problem in a diagram. Recall that with international trade we have our domestic demand curve and our domestic supply curve. Instead of going up to the point of intersection, however, we focus on the world price. We look at the quantity that’s demanded domestically, the quantity that’s supplied domestically and the difference between the two of them is the quantity of imports. Now the total amount of economic value created with economic trade is the consumer surplus, which is this large triangle between the demand curve and the world price, and the producer surplus, which is this little triangle between the world price and the supply curve.

Suppose now that in addition to paying the world price everyone in France who buys a pomelo has to also pay a tariff, an import tariff. The import tariff maybe legally levied on the grocery stores that import pomelos but they’re going to pass the price along. They’re going to pass the cost of the tariff along to their customers. So what happens when the price of pomelos in France rises by the amount of this tariff or this extra tax? The tariff is like an excise tax imposed upon each unit of an imported good. So when the pomelos are imported, the price rises by the amount of the tariff. And the price that people in grocery stores in France will have to pay will be the world price plus T, where T is the amount of the tariff, per unit, maybe 10 cents per pomelo.

What happens when the price of pomelos goes up by the amount of the tariff? Well, if we take the line, the dotted line on across our diagram, you’ll see that the tariff changes the behavior in this market. Well, that always makes the economists suspicious. We were very happy with what people were doing before, because it was creating a lot of economic value. Now, when you’ve got a tariff and people change their behavior, we suspect that things can only get worse. Well, let’s see what happens.

The first thing that happens with the higher price is that the quantity of pomelos demanded domestically shrinks. That is, people in France now want fewer pomelos because they are willing and able to buy fewer at this higher price. They buy some kind of substitute instead or they simply do without. The second thing that happens as a result of the tariff and its higher price is that more French farmers get back into the pomelo business. More French farmers decide that they can afford to supply pomelos, they can afford to be sellers because they can cover their opportunity cost at this higher price. So buyers buy fewer pomelos, farmers sell more pomelos.

The quantity of imported pomelos shrinks when this large volume of trade at the world price to a small volume of trade when the tariff is imposed. What happens to the economic value in this market? What are the consequences of this tariff on consumer surplus, producer surplus, and total economic value? Well, let’s look. Remember that consumer surplus before was the entire area above the world price and below the demand price. Can you identify total consumer surplus after the tariff is imposed?

Total consumer surplus is now the area above the new price, distorted upwards by the tariff, that is, people have to pay a higher price, and below the demand curve. Total consumer surplus is now this smaller triangle. The total amount of consumer surplus has shrunk by the area of this trapezoid. This trapezoid, notice, has four parts. Part one, part two, part three, and part four. All four parts use to be part of consumer surplus, but no more. With the higher price consumer surplus has shrunk and knocked off this entire trapezoid.

The next consequence is that at the higher price domestic farmers are getting more producer surplus. That is, at this higher price producer surplus is going to be bigger than it was before. Can you identify the area of the new producer surplus? The new producer surplus is the area below the tariff distorted price and above the supply curve. All of this area now is profit for producers of pomelos in France. The additional producer surplus, that is, the part that’s added over the free trade case is this area right here that we called area one.

So observe that of the lost consumer surplus, part of it is captured by the producers. Part of the lost consumer surplus goes to domestic farmers who are now making pomelos. In addition, this area that we called area three, is the...
revenue that the government raises from the tariffs. This area consists of a height, and the height is the amount of the tariff, which was added on to the world price. So here’s the tariff, 10 cents per pomelo, multiplied by the number of pomelos that are imported. So if you take the total volume of imports, multiply them by the amount of the tariff, we get amount times volume of imports equals tariff revenue.

So area three is captured by the government; a lot like the tax revenue was captured in the excise tax case. So of the four areas that have been lost to consumers, one area was restored by being transferred to producers and another area was restored by being transferred to the government. That leaves these two areas, area two and area four. This is the deadweight loss from the tariff. These areas are the lost economic value that is not regained by any one.

Let’s look at each of these areas in turn. We’ll look first at area four. Area four represents a reduction in domestic consumption. These are pomelos that should have a low price to import and had a greater benefit for customers, but these customers are no longer buying pomelos, because the price has been pushed upwards by the tariff. The reduction in consumption associated with the tariff creates a deadweight loss. Consumers who should be buying pomelos, if they could get them at the true price, but are not buying them at the high price created by the tariff. This area is a deadweight loss. It’s lost value from a reduction in consumption.

Let’s look now at area two. Area two represents lost opportunities to the economy. Look at these French farmers that are producing pomelos. The true cost of growing pomelos is high and getting higher. This is land that’s been taken away from the production of other products, such as wine or apples or some other good that can be grown in France and has more value. The opportunity cost of making pomelos is getting high. Instead France should be importing its pomelos, down here at a low world price. It would be cheaper for the people of France to be getting their pomelos from abroad, but because domestic producers are making those pomelos instead, people in France are doing without other goods that the same resources could produce, such as wine or apples.

This area, which is deadweight loss, represents the loss from inefficient domestic production. That is, the land is being used to make pomelos and it should be used to make something else, something more valuable. Instead, the pomelos should be imported at the low world price. So, to summarize. When a tariff is imposed the volume of imports shrinks. The cost to the economy is a loss of consumer surplus, as consumers have to pay higher prices to get products that they previously imported at lower prices. Some of that lost consumer surplus is merely a transfer. That is, some of it winds up in the hands of domestic producers and some of it winds up in the hands of the government as tariff revenue. But part of the loss from the tariff is never recovered, and that is the deadweight loss. The deadweight loss comes from consumers that are pushed out of the market permanently and from inefficient domestic production, which uses land to produce pomelos that would have a higher value in other uses. The only reason pomelos are being grown is because of the tariff and the extra subsidy that that provides to producers.

If you look carefully you can see that this tariff really has two effects. It’s like a tax on consumers and a subsidy to producers. That’s one of the reasons why economists are suspicious of tariffs. We usually like to pursue policy goals one at a time. Anytime you’ve got a policy that does two things at once, we’re going to be suspicious that it’s not doing either one of them as well as it can. Stay tuned and we’ll slice open the pomelo.