PRINCIPLES OF ECONOMICS

For Honors Students

By Roger D Congleton, Truist Professor of Economics, WVU (April 2024)

For Students at WVU only

Not for Distribution

Table of Contents

Chapter 1: Microeconomics is the Science of Markets

When I was an undergraduate many years ago, I believed that economics was the "science of business." But I was wrong about that. It is the science of markets, where a market is a network of voluntary exchange, production, and innovation. Because all of these activities (exchange, production, and innovation) are activities undertaken by individuals, economics is a social science.

Economics has connections to other social sciences such as political science and sociology because both governmental policies and widely shared norms have effects on markets. For the most part, economic theory assumes that public policies and norms are "constant" and so can be ignored as long as they remain constant. It is only when either public policies or norms change that their effects on various parts of the great network of exchange, production, and innovation need to be explicitly taken into account.

We all have some informal experience with markets, because we are (nearly) all part of market networks. We buy things at stores and through the internet. Many of us have held or hold jobs in which we produce things—either tangible things or services that others purchase from us both directly (our employers) and indirectly (their customers). So, we all know a bit about markets from direct experience.

However, very few of us have taken the time to think about how markets work and why they work as well as they do as sources of material comforts, entertainment, and useful technologies. Much of what generates market opportunities is clearly visible, but much is invisible until one thinks carefully about market relationships. At first the more one thinks about market networks the more overwhelming and confusing they seem to be.

There are market connections among billions of individuals, most of which tend to make everyone better off as long ago pointed out by such scholars as Adam Smith in the eighteenth century and Claude Bastiat.in the nineteenth century.

I. An Illustrative Story about the productivity of specialization and market networks from Claude Bastiat:

> Let us take, by way of illustration, a man in the humble walks of life—a village carpenter, for instance—and observe the various services he renders to society and receives from it; we shall not fail to be struck with the enormous disproportion that is apparent. This man employs his day's labor in planning boards and making tables and chests of drawers. **He complains of his condition; yet in truth what does he receive from society in exchange for his work?**

First of all, on getting up in the morning, he dresses himself; and he has himself personally made none of the numerous articles of which his clothing consists of. Now, in order to put at his disposal this clothing, simple as it is, an enormous amount of labor, industry, and locomotion, and many ingenious inventions, must have been employed. Americans must have produced cotton, Indians indigo, Frenchmen wool and flax, Brazilians hides; and all these materials must have been transported to various towns where they have been worked up, spun, woven, dyed, etc.

Then he breakfasts. In order to procure him the bread he eats every morning, land must have been cleared, enclosed, labored, manured, sown; the fruits of the soil must have been preserved with care from pillage, and security must have reigned among an innumerable multitude of people. The wheat must have been cut down, ground into flour, kneaded, and prepared; iron, steel, wood, stone, must have been converted by industry into instruments of labor; some men must have employed animal force, others waterpower, etc.; all matters of which each, taken singly, presupposes a mass of labor, whether we have regard to space or time, of incalculable amount.

In the course of the day this man will have occasion to use sugar, oil, and various other materials and utensils. He sends his son to school, there to receive an education, which, although limited, nevertheless implies anterior study and research, and an extent of knowledge that startles the imagination.

He goes out. He finds the street paved and lighted. A neighbor sues him. He finds advocates to plead his case, judges to maintain his rights, officers of justice to put the sentence in execution; all which implies acquired knowledge, and, consequently, intelligence and means of subsistence.

He goes to church. It is a stupendous monument, and the book he carries thither is a monument, perhaps still more stupendous, of human intelligence. He is taught morals, he has his mind enlightened, his soul elevated; and in order to do this we must suppose that another man had previously frequented schools and libraries, consulted all the sources of human learning, and while so employed had been able to live without occupying himself directly with the wants of the body.

If our artisan undertakes a journey, he finds that, in order to save him time and exertion, other men have removed and leveled the soil, filled up valleys, hewed down mountains, united the banks of rivers, diminished friction, placed wheeled carriages on blocks of sandstone or bands of iron, and brought the force of animals and the power of steam into subjection to human wants.

It is impossible not to be struck with the measureless disproportion between the enjoyments which this man derives from society and what he could obtain by his own unassisted exertions. **I venture to say that in a single day he consumes more than he could himself produce in ten centuries.**

What renders the phenomenon still more strange is that **all other men are in the same situation.** Every individual member of society has absorbed millions of times more than he could himself produce; yet there is no mutual robbery.

And, if we regard things more nearly, we perceive that the carpenter has paid, in services, for all the services others have rendered to him.

If we bring the matter to a strict reckoning, we shall be convinced that he has received nothing he has not paid for by means of his modest industry; and that everyone who, at whatever interval of time or space, has been employed in his service, has received, or will receive, his remuneration.

The social mechanism, then, must be very ingenious and very powerful, since it leads to this singular result, that **each man, even he whose lot is cast in the humblest condition, has more enjoyment in one day than he could himself produce in many ages.** (*Harmonies of Political Economy*, pp. 452–54)

II. Understanding Market Networks

Bastiat's description on the productivity of market networks with their specialization, economies of scale, and relatively high innovation rates remains even more true today.

The main puzzle that principles of microeconomics addresses is how do such networks operate, given that there is no central planner? Millions and billions of individuals make independent decisions about what to buy, what to produce, and what to invent and yet the result is not chaos but rather a systematic pattern of life and production that generates material comforts, entertainment, and necessities for essentially all of the persons that participate in those networks. How is this possible? What is it that induces folks to make the choices that lead to productive activities—where productive is not simply producing things, but producing the things that people want and in the quantities that they are willing to pay for?

Anyone can build sandcastles or daisy chains, but not everyone can build a house, undertake a surgery, maintain a car, program a computer, fly an airplane, or invent a smart watch. All of which are much more highly valued by most individuals than sandcastles or daisy chains.

The basic answer developed by micro-economics is that prices coordinate the decision of consumers, producers, and innovators so that the most valuable things get invented and produced—where "most valuable" is determined by the preferences and income of consumers Producing things that are more desired by consumers tends to be more profitable for individual producers and organizations devoted to production of goods to sale (referred to as "firms" in most economic textbooks). And, for anyone that prefers more income to less income, such goods will attract more efforts to produce them than goods that are less profitable to produce and sell.

Similarly, innovators may invent lots of things—but again insofar as most prefer more income to less income, they will tend to focus on the innovations that are likely to be most profitable—e.g. that consumers are willing to pay the most for. Thus, market driven innovation tends to occur with the interests (or expected interests) of consumers in mind. In this way expected profits may be said to determine what both is invented and what is produced, and what is sold. The more profitable a product tends to be the more innovators and producers try to produce it.

But what determines profits? This is partly determined by the interests of consumers and their willingness to use their income to purchase both things (often called "goods" by economists) or services. The more consumers are willing to pay the higher potential profits are—other things being equal. However, the more firms compete for consumer expenditures, the lower prices tend to be and the lower the profits actually realized tends to be. Thus, the typical prices at which goods and services are sold are an important determinant of profits. The higher the selling price, the higher are profits. The lower the selling price the lower are profits. This is true for both producers and innovators.

Similarly, the higher the price of a good is, the less interested a consumer (buyer) is in purchasing it—and the lower the price of a good is, the more interested a consumer (buyer) in purchasing it. Thus, prices also affect consumer decisions.

When prices rise, firms tend to produce more, and consumers tend to purchase fewer units of the goods produced for sale. When prices fall, firms tend to produce less, and consumers tend to purchase more of the goods produced for sale. These opposite effects of prices on firms and consumers sound like a recipe for chaos—but it turns out that for the goods that are routinely produced in sold in markets, there is normally a price at which the purchases of consumers (quantity demanded) equal the quantities produced by firms (their profit maximizing quantities.

Such prices are referred to by economists as "equilibrium prices" and theories of "price determination" provide the core principles of microeconomics.

To understand how prices emerge and coordinate the decision of producers, consumers, and innovators, a series of "models" were gradually worked out by economists over the past two and half centuries. The models abstract from a variety of idiosyncratic features of the decision of firms, innovators, and consumers in order to better understand the main factors that determine equilibrium prices and their ability to coordinate a broad range of decision in market networks.

Those same models can be used to predict how prices, sales, and output levels change when the factors held constant (or abstracted from) change. The first half of the course focuses on the logic of price determination and the second half focuses on the logic of "comparative statics," which is how equilibrium prices change as public policies such as regulation and taxes change or when better or worse than usual weather occurs or when major innovations take place.

III. Some Useful Historical Background

Microeconomics emerged as a field of study in the nineteenth century as a relatively small group of people (many, but not all, of them college professors) began to analyze the new, broader scope of commerce in ordinary life.

Prior to 1700, essentially everyone was a farmer, farm hand, or servant. Some owned land, but most worked on land owned by others. Many of the larger farms were largely self-sufficient. They made their own tools (mostly from wood), wove much of their own clothing, and grew most of their own food. w There was always some market activity in towns and between towns, but most farmers, farm hands and maids journeyed to town infrequently. Farm hands and maidens were largely paid with room and board and so had little money to spend on the things that might be purchased in town. [Trade without money is called **barter**.]

Commerce took place, but for most people it was a "side dish" rather than the "main course" of their lived. There was always some "commerce" in towns and at local market fairs, but only a fairly small fraction of the population lived in towns and depended entirely on commerce for their living.

In northern Europe this pattern of life began to change in the eighteenth century. Commerce began to expand for a number of reasons, ideas about a good life changed. Farming and other technologies improved, and the organization of production and innovation also improved. These innovations tended to make commerce a more efficient and reliable method of meeting consumer wants than it had been in the past. Suffrage expanded which tended to make government policies more determined by middle class voters than noble and royal families. Ideas about a good life changed and norms against commerce diminished in strength.

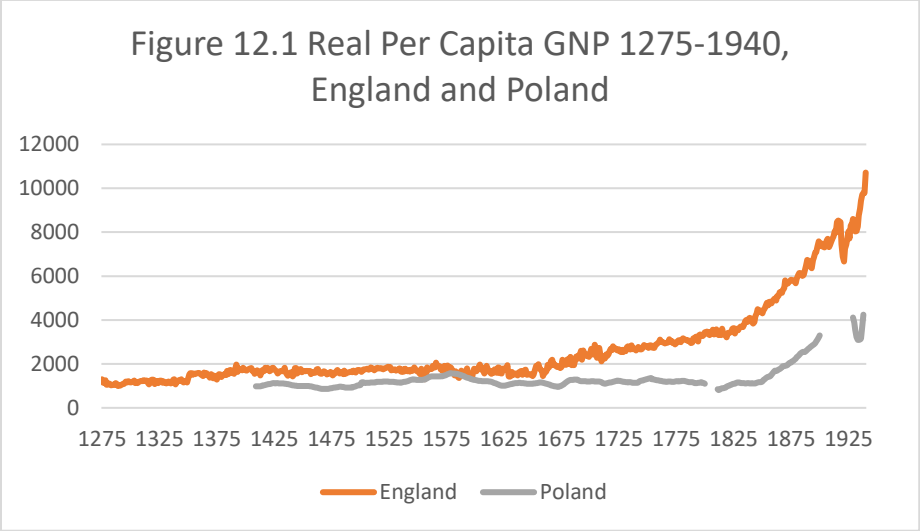
As a consequence, measured average (real) income (per capita RGNP) rapidly increased in the 18th and19th centuries. An entirely new pattern of life emerged, which I refer to as the commercial society. At the beginning of the nineteenth century in the West, 90-95% were farmers, by the end of the century only around 10% were. Today only about 2% are.

Commercial organizations (firms) grew larger and more numerous(because of specialization). Many were much larger than any previous private organizations, because they needed to be large to take full advantage of economies of scale. Many of the new production methods were very capital and knowledge intensive. Many production methods and products of the nineteenth and twentieth centuries were completely new: steam engines, tractors, light bulbs, automobiles, airplanes, computers etc.

All these and other factors—including broader mass education--increased specialization and allowed new economies of scale to be realized. A great new range of occupations emerged, and more and more persons "earned their living" not as farmers but by hiring themselves out to commercial organizations (firms) in exchange for wages. Urban centers in the West expanded rapidly as families left rural farmlands for jobs in the cities.

And, with money wages being paid, new "retail" organizations emerged to serve the rapidly expanding middle class.

Life was very much different in 1900 than it had been in 1800 for most persons in the West. This highly specialized, interdependent form of life was new, and had become the norm. Most of the rest of the world gradually adopted this system in the twentieth century and more or less world-wide market networks developed for a wide variety of products--far more products than had been produced for sale in earlier times.



And so, a **broad range of people, billions of them, became linked together through market trading networks** of many kinds. As a consequence, market activities--economics--is central to most people's life in most parts of the world.

IV. Principles of Economics

The first principles of economics books were written in the late eighteenth century. All such books and the courses based on them attempt to explain how the new commercial systems operated. Their aim was similar to that of Physics during that period. They attempted to develop and apply a small number of principles that could provide systematic explanations of how the new commercial societies and their broad networks of exchange operate, and also how they are affected by factors beyond the networks of exchange, production, and innovation that characterize contemporary commercial societies.

New textbooks are introduced every year during the twentieth century, but most are still based on the topics developed by Alfred Marshall's textbooks of the late 1890s and early 1900s.

This course is not based on a textbook but on web notes developed for the course. It is based on such textbooks and much other reading undertaken over the course of my career as an economist. There are readings from a mainstream textbook and a couple of other books, but they are not the "main dish" of

this course. Instead, the main dish is the class website and the class lectures which are based upon the class web site.

Its main goal is the same as other principles of economics courses: to provide the tools that students need to understand markets in a general way and also the models and ideas that you'll need in upper-level economics courses, should you go on in economics or dabble a bit in it.

As an honors course, the course developed for you places a bit more emphasis on what might be called "neoclassical models," which were worked out in the late nineteenth century and during the first half of the twentieth century. Understanding these models will sharpen your intuitions about how markets work and also for economics as a field of study.

The first half of the course develops the models that most principles of economic courses cover. The second part of the course covers more advanced topics that are often left out of principles of economics courses, but which should be both useful to you and which will deepen your understanding of economic outcomes and choice settings.

The aim of all principles of economics course is not to persuade you to favor specific policies or induce you to memorize particular facts, but to provide a variety of logically consistent models that will help you to think more systematically and carefully about modern market-based relationships and their associated societies.

The remainder of this chapter develops a few ideas that ground the rest of the course.

V. Positive and Normative Economics

For a lot of purposes in economics, it is useful to distinguish between **positive issues**—issues concerning how the world works—and **normative issues**—issues concerning whether the world works in a good or bad way. This is especially true of areas of economics that dealing with the consequences of public policies on markets and incomes. It is possible for people to agree about the future consequences (positive conclusions), yet still disagree about whether a policy is good or bad the ideal policy (normative conclusions). Most of the analysis of the effects of public policies undertaken in this course is positive, although there will be a bit of normative analysis as well.

DEFINITIONS:

A **Positive Statement** is a statement about **what is**, has been or will be. It is a statement about the world.

A **Normative Statement** attempts to evaluate the **desirability** of alternative states of the world.

Normative conclusions are not always based on "gut reactions" or intuitions, but they may follow from logically consistent theories. Examples include cost-benefit analysis, the Pareto Criteria, and related ideas from utilitarian and contractarian theories of the good or just societies.

Positive conclusions are often based on accepted scientific theories (which may be changed in the future) or one "gut intuitions" such as "that color is really green rather than blue." In either case, a positive statement is about the world as it is, has been, or will be rather than whether it is desirable or not, or good or bad, or better or worse.

Many positive statements can be assessed to determine whether they are "true" or "false." Such statements are said to be "operational." Not every positive statement is true. Indeed, science often makes progress by challenging or extending the "positive claims" of earlier theories previously thought to be true.

Generally, normative statements conclude that a particular policy or state of the world is good or bad, is Pareto optimal or not, should be undertaken or not, etc. because some state of the world is "better" than another. Some of these statements are arguably "true" if one accepts a particular normative theory. Policy A may be said to be better than policy B after a careful cost-benefit analysis that can be appraised to determine whether it (the cost benefit analysis) yields the results claimed. (Net benefits can be estimated using data, statistical tools, and theory.)

Confusion can occur about whether a statement is positive or normative, in part, because many reasoned normative statements are "consequentialist" and so use positive statements to support their conclusions. For example, policy X may be said to be a bad policy because it increases unemployment. That policy X increases unemployment is a positive statement. The conclusion that X is a bad policy depends on whether you believe unemployment should always be avoided or not—which is to say the use of increase or decreases in unemployment as a normative measure or indicator.

Positive statements are often confused with operational statements. *Operational statements* are statements that can at least conceptually be tested to determine whether they are true or false. Not all positive statements are testable, and moreover, give a coherent normative theory, some normative statements are testable!

Some Puzzles and Examples:

- The moon is made of green cheese. (p, but false)
- Minimum wage laws always increase unemployment. (p, probably true)
- Tariffs are a bad policy because they reduce consumer welfare. (n, probably true according to cost-benefit analysis)
- Mass transit reduces unemployment. (p, probably true)Mass transit should be subsidized because it reduces air pollution. (n, possibly true)
- Global warming can only be reduced with a high carbon tax. (p. probably false—other policies may also do so)

VI. The Use of Models by Economists

The course develops the geometric tools that economists use for modeling choices by consumers and firms. These can be used to make a series of predictions about markets, prices, incomes, and innovation.

They can also be used to analyze the effects of various public policies and natural shocks. It is interesting to note that the models we cover in class were sufficient for publishing in the top economic research journals until the mid-nineteen seventies, or so. So, although that is no longer true, they provide very general explanations of most economic phenomena.

The models used in this course are **partial equilibrium** ones that assume that much of the world beyond our models is stationary--non changing. This is a way of reducing the complexity of the world down to levels that can be analyzed. (In economics this is often called the "ceteris paribus" assumption.)

It is easier to analyze one person or one market at a time than the whole market or universe at once. It allows us to undertake what economists call comparative statics. If "something" changes, then a change in a market's equilibrium will occur. , holding other things constant. Among the things that are often held constant are (i) the laws that define the "rules of the game," (ii) the distribution of wealth, (iii) technology. Of course, in the real world everything is not held constant. However, understanding can often be increased by focusing on a subset of important factors and abstracting from other factors that are deemed less important. Most sciences use this method.

For the most part, we'll focus on the implications of rational choice and scarcity for market activities and outcomes: prices, production, profits, income, innovation, etc. The main dish of principles of microeconomics is "pure theory," e.g. "principles" that explain much of the course of prices and their ability to coordinate the activities undertaken in market networks. However, we'll also be developing lots of applications for that theory to show how it can be used and why it is used.

Whether the most "important factors" have been identified or not is at least partly an empirical issue—do those factors in fact account for the phenomena that they purport to explain or not? The results of many empirical studies suggest that economic models do in fact capture much that is important. The examples help to demonstrate that the models "work," they can account for much about the way prices and markets behave through time and in response to changes in circumstances—new policies or worse than average weather etc.

VII. The Grounding Ideas of Micro-Economic Models

Microeconomics studies the implications of rational choice in a setting of scarcity. People are not always rational and some goods, such as air, are not always scarce, but the goods traded in market networks are all scarce (otherwise people would not be willing to pay for them) and insofar as individuals and organizations are forward looking, they tend to be rational in the sense that word is used by economists (which is a bit different than its ordinary English usage).

Rational choice can be represented in a variety of ways, but the simplest is that a choice is rational if it is forward looking and attempts to advance some end or move towards a goal. The goal may or may not be reasonable, but if a person or organization acts to systematically advance that goal, then it is "rational" in the sense that economists use the term. For example, a person that attempts to maximize

grades is rational. A person that maximizes wealth is rational. A person that tries to care for his or her friends is rational. A person or organization that systematically attempts to maximizes profits is rational.

A good is scarce whenever too little of it exists to fully satisfy all possible goals or wants by persons currently living.

- Personal time (human time) is scarce.
- Attention is scare.
- Wealth is scarce.
- Labor is scarce (since time and people are scarce).
- Knowledge is scarce.
- Most natural resources are scarce.
- Most manufactured goods are scarce.

Perhaps surprisingly, these two characteristics—rationality and scarcity—can be used to construct models that can shed light on a wide variety of private choices, choice in markets, choice in politics, and many other choice settings. This is not to say that every human activity can be completely understood using economic reasoning, but it is a rare activity that cannot be better understood using micro economic ideas, geometry, and logic.

Chapter 2: Rational Choice and the Demand and Supply Model of Price Determination

I. The Geometry of Net Benefit Maximizing Choice

There are **two widely used models of rational choice**: (1) the net benefit model and (2) the utility maximizing model. Both models can be developed geometrically, and both can be used to create surprisingly well integrated general theories of market activity (and many other activities). These models, surprisingly, allow one **to deduce** a broad range of properties about markets--most of which have been verified with statistical tests of one kind or another.

The models do not work perfectly, but it is amazing how much can be explained about a wide range of market, individual, interest group, and political activities from so few assumptions.

We'll focus on the net-benefit maximizing model in this course, because it's a bit easier, and is in some ways it is more important for undergraduate applied micro courses. (The utility-maximizing model is developed in intermediate microeconomics courses and will not be used in this one.)

Nearly all partial equilibrium economic models can be developed from the assumption that individuals and firms maximize their private net benefits.

- Definition: Net benefits are total benefits less total cost.

Economists have given special names to the net benefits of consumers and firms.

- For example, Consumers are assumed to maximize consumer surplus: the difference between what a thing or collection of things is worth to them and what they have to pay for it. $CS(Q) = TB(Q) - TC(Q)$
- Firms maximize their profit, which is the difference in what they receive when they sell their products and what it costs to produce them. $\Pi = TR(Q) - TC(Q)$

However, the geometry and logic of net benefits is essentially the same whether one deals with consumers, firms, other roles, or many other types of activities.

Margins and the Maximization of Net Benefits

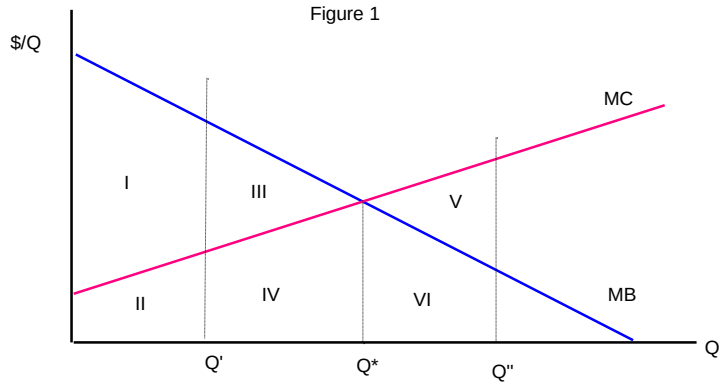
It turns out that to maximize net benefits, one normally chooses the quantity at which one's marginal benefits equals one's marginal costs. We'll demonstrate why in a few moments.

- "Marginal" is an adjective, and refer to how much another unit of a good or service produces additional benefits or costs, etc.

- **Definition**: Marginal "X" is the change in Total "X" caused by a one unit change in quantity. It is the slope of the Total "X" curve. "X" ∈ {cost, benefit, profit, product, utility, revenue, etc.}
- *Important Geometric Property*: Total "X" can be calculated from a Marginal "X" curve by finding the area under the Marginal "X" curve over the range of interest (usually from 0 to some particular quantity, Q).
- This "**area property**" allows us to determine consumer surplus and/or profit from a diagram of marginal cost and marginal benefit or revenue curves. This property will be used over and over again in this course to help us understand the implications of the diagrams that we'll be using to model the way prices are determined in competitive markets.
- For example, consider the MB of ice cream cones. The first ice cream cone on a hot day might provide a lot of marginal benefits, the second one less so, and the third one even less so, because one gradually becomes satiated. On a cold day, the marginal benefits of each ice cream cone tend to be than on a hot day for most people.
- A MB curve represents the highest price that an individual is willing to pay for successive units of a good. An individual's willingness to pay for successive ice cream cones would diminish in both cases—but he or she would be willing to pay more for them on a hot day than on a cold day. Thus, the typical individual's MB curve for ice cream cones is a downward sloping curve and the MB curve for ice cream cones on a hot day is above the MB curve for ice cream cones on a cold day.

Figure 1 illustrates typical slopes of a marginal benefit (MB) curve and a marginal cost (MC) curve for an individual that produces something for his or her own consumption. MB curves are typically "downward sloping" and MC curves for production are typically "upward sloping." The Roman numerals label areas of the diagram that are important for calculating the total benefits and total costs of three quantities labeled along the horizontal axis, Q', Q*, and Q".

Figure 1



The marginal cost and marginal benefit curves of Figure 1 allow us to calculate the total cost of Q' and the total benefit of Q' . To do so we **use the area rule mentioned above**.

To make just a bit less abstract, we'll imagine that the individual of interest is named Al (which is short for Allen or Alice), and that the MC curve is his or her assessment of the MC of gathering apples for his or her own use and baking a pie, The MB curve is the marginal benefit associated with different sized apple pies after all the work is done..

- Al's total cost of producing Q' units is the area under the MC curve from zero to Q'. In the diagram, that area is labeled are II.. TC(Q') = II
- Al's total benefit associated with consuming Q' units is determined by the area under the MB curve from zero to Q', which is composed of two sub areas, area I and area II. TB(Q') = I + II

Al's net benefit from producing Q' units of this good (an apple pie of size Q'), and then consuming it is TB(Q) – TC(Q') = (I+II) – (II) = I. Thus, **AL's net benefits from Q' units of this good is area I.** This area, might for example, represents the net benefit Al realizes by spending the early morning gathering apples, then spending the late morning baking an apple pie, and finally enjoying the pie after it comes out of the oven in the early afternoon.

Exactly the same process can be used to characterize the total cost of producing any quantity of a good. If you know the MC of production and the marginal benefit of consuming the good, you can calculate the net benefits of producing and consuming the good.

Next we'll use Figure 1 to determine the areas that correspond to the total benefit, total cost, and net benefit associated with the other two labeled outputs: Q* and Q''.

A. Results:

TB(Q*) = I + II + III + IV , TC(Q*) = II + IV , and thus, NB(Q*) = I + III
TB(Q') = I + II + III + IV + VI , TC (Q'') = II + IV + V + VI , and thus, NB(Q'') = I + III - V

If you compare the areas that represent Al's net benefits at the three quantities, it should be clear that a net benefit maximizing individual will prefer Q* over both Q' and Q'', because Q* generates the highest net benefit of the three labeled quantities.

- **Al's net benefits are larger at Q*** than at any quantity less than Q* because I+III > I, and it generates a higher net benefit than any quantity greater than Q*, because I+III > I +III - V.

Note also that **Q* is the quantity at which AL's marginal benefits equal equals his or her marginal costs**. In this case and most others of interest in this course, Al's net benefits are maximized at the quantity where MC = MB. Such quantities are normally labeled Q* in this class.

- Our calculations imply that any person or organization that attempts to maximize net benefits will do so by producing and consuming the quantity of a good that sets his or

her marginal cost equals marginal benefit. The same logic applies to cases in which a person purchases a good rather than produces it.

- **This rule of thumb** reflects the graph we just analyzed, and its basic logic applies to most cases in which Q is very "divisible" and MC intersects MB at just one point with marginal benefits initially being greater than MC up to that intersection (up to Q*) and lower after that.
- There are a few exceptions and extensions that are useful to understand, but **most of the results and theories worked out in the entire class follow from this one property**.
- This characterization of net benefit maximizing decisions is **very general** and can be used to model the behavior of both firms and consumers in a wide variety of circumstances.

Net-benefit maximizing decision makers tend to choose activity levels where **their own marginal costs equal their own marginal benefits**--not because they care about "margins" but because this is how one maximizes net benefits.

II. Deriving a Consumer's Demand Curve Using the Net Benefit Maximizing Model

If consumers are rational net benefit maximizers, it turns out that both individual and market demand curves are downward sloping. To demonstrate this, we have to determine what net benefit maximizing behavior implies about an individual 's demand curves.

A demand curve is a curve that describes how many units of a good or service an individual will purchase at a given price, other things being equal. In other words, if "you" tell "me" a price, I can use the demand curve to tell "you" how many units an individual will purchase other things being equal.

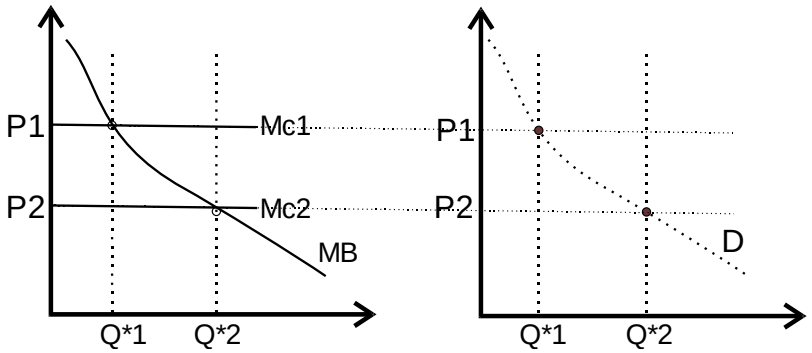
(Other things being equal means that other factors that might influence a person's decision to purchase a good such as his or her preferences, personal income, and the prices of other goods are held constant when one uses a particular demand curve).

To derive a demand curve from a consumer's MB curve:

(i) Pick a price, P
(ii) Find the associated MC curve. If the individual can purchase as much as he or she wants at the prevailing market price, then his or her MC curve is a horizontal line equal to price.
(iii) Find the NB maximizing quantity, Q (the amount bought by a rational consumer). This will generally be the quantity that sets MB equal to MC.
(iv) Plot P and Q on a separate diagram.
(v) Repeat with another price, until you have traced out the individual's demand curve.

Figure 2.2 illustrates the geometry of this process for 2 prices. It begins with the consumer's marginal benefit curve (MB), then uses two prices, P1 and P2, to find two points on this person's demand curve.

Figure 2.2 Deriving a Demand Curve



The derivation process has been left abstract, so that you can see that it applies to any MB curve that is generally downward sloping and is initially above the MC curve of interest. (The marginal cost lines are the dashed lines.)

In principle, the process of plotting out an individual's demand curve continues forever. One can keep choosing prices, finding the associated CS-max quantities, and plotting the price and quantity. However, in most cases three or four points are enough to get the basic geometry of a consumer's demand curve.

The **highest price** that an individual is willing to pay occurs at a price that just touches the top of the individual's MB curve. At prices higher than this one, the MC curve is always above his or her MB curve, and the individual will **purchase none** of the good because negative net benefits are realized in by purchasing any of the good in such circumstances. Such losses can be avoided by purchasing none of the good. Some prices that are just "too high" to make it worthwhile to purchase a good or service.

Such cases are an exception to the rule of thumb that we generated using the first diagram. There are also other exceptions, but that is the main one, and one that is of some practical interest. The most common quantity that an individual purchase in a grocery store is zero. Out of the thousands of products in a typical large grocery store, a typical consumer purchases only a few dozen.

Similarities and differences between an individual's MB and demand curves

Note that whenever the marginal benefit curve is downward sloping over its whole range, the consumer's demand curve **goes through exactly the same points** as the MB curve and is also downward sloping. In such cases, one can use estimates of individual demand curves as estimates of MB curves. However, although they go through the same points, marginal benefits and demand are not the same function, because they **have different meanings,** and the **functions go in opposite directions**.

MB functions go from Q into $/units, whereas demand functions go from P into Q. They are inverse functions of each other.[1]

Implied Properties of Demand Curves

Notice that we **already have some predictions** about behavior that follows simply from the net-benefit maximizing model of consumer choice:

- Demand curves generally slope downward.
- As price rise, consumers tend to purchase fewer goods.
- A change in MB will cause demand curves to shift up or down according to the shift in MB.
- Consumer net benefits tend to fall as prices rise, other things being equal.

Keep in mind, that the point of the diagrams is not intended to induce you to memorize a long series of diagrams, but to induce you to understand the logic of the choices being modeled. The diagrams will all be a bit different, but their underlying logic is nearly always similar to that used in the first two sets of diagrams.

It is the logic that is general, rather than the particular diagrams. The rational person or organization will do X in circumstance Y, because X maximizes his or her net benefits in those circumstances.

III. Deriving a Firm's Supply Curve from MC Curves

Firms are also net benefit maximizers. Thus, the derivation of a firm's supply curve is very similar to that used to derive a consumer's demand curve. The main difference is that the selling price is the marginal benefit realized by selling good, and the marginal cost of producing the good reflects the firm's particular production process and the firm's costs for inputs. Thus, derivation of a firm's supply curve begins with a firm's MC curve, rather than its MB curve.

Another difference is the name given to a firm's net benefits—namely **profits**, and the name given to its marginal benefits, namely **marginal revenue**. Although the names have changed, the basic geometry of maximizing net benefits remains the same. Firms will produce outputs that maximize its profits, and those outputs (usually) require the firm to produce where its marginal benefits (e.g. its marginal revenue) equals its marginal production costs.

---

[1] In some odd cases, as when the MB curve has a bump or two in it, the demand curves may include only a subset of the point on an individual's MB curve, but the points in common are always from the downward sloping portions of the MB curve. Thus, when a demand curve is derived in this way from MB curves, it turns out that **every** individual demand curve slopes downward.

1

Given this, it should not be surprising that the method used to derive a firm's supply curve is very similar to that used to derive a consumer's demand curve. A firm's supply curve characterizes its output as a function of price. So, if "you" tell "me" a price, "I" can use the supply curve to provide "you" with the answer.

To derive a firm's supply curve,

B. (i) Choose a price.

C. (ii) Find the associated marginal revenue curve (MR). If it sells as much as it wants at the market price, then this will be a horizontal line equal to that price.

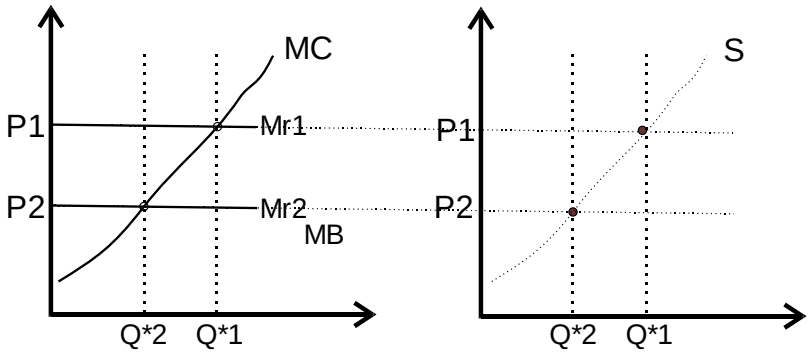
D. (iii) Find the profit maximizing quantity of the good or service, given that MR curve and its MC curve. This will normally be the quantity that sets its marginal revenue equal to its marginal cost.

E. (iv) Plot the price and the profit-maximizing quantity on another diagram.

F. (v) Repeat with several other prices to trace out a supply curve.

The figure below illustrates these steps for a particular firm's marginal cost curve (MC) and two prices, P_1 and P_2.

Figure 2.3 Deriving a Supply Curve



For price P1, the quantity that maximizes profit is found by going over to the MC curve and then down to the quantity axis. That quantity is labeled Q_1. This output maximizes profits given price P1. This combination of price and output is plotted on the Supply diagram on the right as the point (Q1, P1). (At price P_1, the firm will produce and sell quantity Q_1.)

Next we try another price, P2. A horizontal line through P2 is the firm's marginal revenue curve, since its revenue now increases by that P2 dollars every time another unit of the good is sold. Given that new MR curve, this firm chooses the quantity that maximizes profits, which is again found where MC=MR and is labeled Q2. (The marginal revenue lines are the dashed lines). This price output combination is plotted on the Supply diagram as the point (Q2, P2). (At price P2, the firm will produce and sell quantity Q2 if the price is P2.)

In principle this process is continued forever: choosing new prices, finding the associated profit-maximizing quantity, and plotting the price and quantity. But in most cases a few points are enough to get the basic geometry of a firm's supply curve, although more will be needed when the MC curve has a complex shape.

Note that when the MC is upward sloping, the firm's supply curve goes through **exactly the same points** as its MC curve. This implies that one can use estimates of a firm's supply curve as estimates of its MC curve.

(As in the case of an individual's demand curve, a firm's supply curve is the inverse function of its MC curve. Supply goes from prices ($/unit) into quantities; whereas, MC goes from quantity into $/unit measures. The functions go in opposite directions.[2]

Again, deduction generates a series of predictions about a firm's behavior with respect to the relationship between selling price and output.

- Firm supply curves tend to be upward sloping.
- As prices rise, other things being equal, firms produce and sell more units of their products.
- Any change in market conditions that affects a firm's marginal costs will shift the firm's supply curve. For example, an increase in input prices tends to cause the marginal cost of production to increase.
- If MC rises, the supply curve shifts **back to the left**.
- If MC rises, profits will fall, other things being equal.
- If market prices rise, a firm's MR shifts upward, and both its output and profits will increase.

[2] However, if MC has a more complex shape, only the points on MC that can represent profit maximizing outputs are on both the MC and firm supply curve. Since it turns out that those points in common are all from the upward sloping parts of the firm's MC curve, **a firm's supply curve slopes upward.**

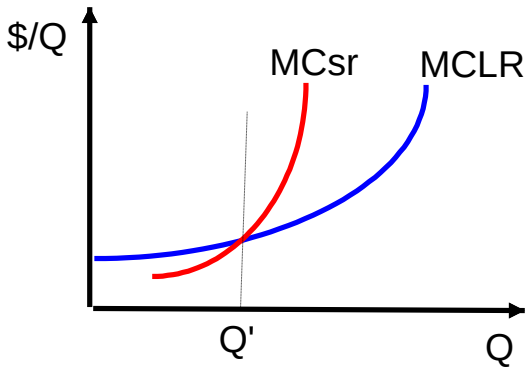
Exactly the same method can be used to derive a firm's Long Run Supply curve given its Long Run Marginal Cost Curve. The only difference is that one should use a firm's long run MC curve rather than its short run marginal cost curve. And the same conclusions will follow.

IV. A Digression on Fixed and Variable Costs in the Short and Long Run

The area(s) under the marginal cost curve is technically, its total variable cost, rather than total cost. It is the part of a firm's production costs that vary with its output. In some cases—in the short run—a firm may have costs that are necessary in order to produce its product(s), but which do not vary without put.

- Total cost is Total Variable Cost plus Total Fixed Cost. TC(Q) = TVC(Q) + TFC
- Total variable costs vary as output changes. Fixed costs do not, and so they are not "picked up" by the area under the MC curve.
- **For most purposes in this class, and in many other economics classes, fixed costs are ignored**, because although they affect profits, but they do not affect output decisions.
- **In the long run all costs are variable, because the firm can alter the size of its plant, change locations, and even change what products it makes. Since, we are often mainly interested in long run supply, we can ignore fixed costs because they are irrelevant in the long run.**

At the level of a firm, the difference between its short run and long run supply is that short run marginal costs tend to rise faster than long run marginal costs, because there are fewer adjustments that a firm can make to keep costs down in the short run than in the long run.



For each level of fixed factors that a firm might have, there is a different short run marginal cost curve (MCsr curve). In the diagram above Q' is the output for which the mix of fixed and variable inputs is ideal (e.g. they minimize the total cost of producing output Q'.) MC_{SR} and MC_{LR} are equal only when the optimal mix of all factors is used to produce the output of interest.

To derive a firm's short run supply curve, use the firm's short run MC curve.

To derive a firm's long run supply curve, use the firm's long run supply curve.

(In a diagram, it is usually enough just to changes label on the relevant curves from the SR to LR. MC_{SR} to MC_{LR} and S_{SR} to S_{LR} , however if the point is to illustrate differences you'll need draw both curves.

V. A Degression on the Relationship between Average and Marginal Costs

For the most part, this course will focus on marginal costs, marginal benefits, and marginal revenues, because these most fully characterize a rational consumer or firm's behavior. However, many textbooks spend quite a bit of time talking about average costs and average revenues, partly because in many real-world settings these are easier to measure than marginal costs and marginal revenues.

- **Definition**: Average "X" is Total "X" divided by quantity, TX(Q)/Q where "X" can be total cost, variable cost, fixed cost, total revenue, total product, etc.
- **Useful Mathematical Property**: Given a value for **average X** at some specific quantity Q', one **can calculate total X** by multiplying average X by Q'. that is AX(Q')*Q' = TX
- **Useful Geometric Implication:** Given an average X curve, one can calculate total X at Q' by finding the area of the rectangle formed using the point AX(Q') as the righthand upper corner and (0,0) as its lower lefthand corner. This rectangle is Q' wide and AX(Q') high, so its area is just AX(Q')*Q' = TX(Q') (see the above).
- This geometry allows one to use average revenue and cost curves to calculate profit -- although not to directly determine the profit maximizing output.

However, for this class, we will mostly be using the areas under marginal curves as developed above.

VIII. Illustrative Calculation of Average and Marginal Costs from a Total Cost Function.

Table Illustrating the Calculation of Average and Marginal Cost from a (Discrete) Total Cost Function

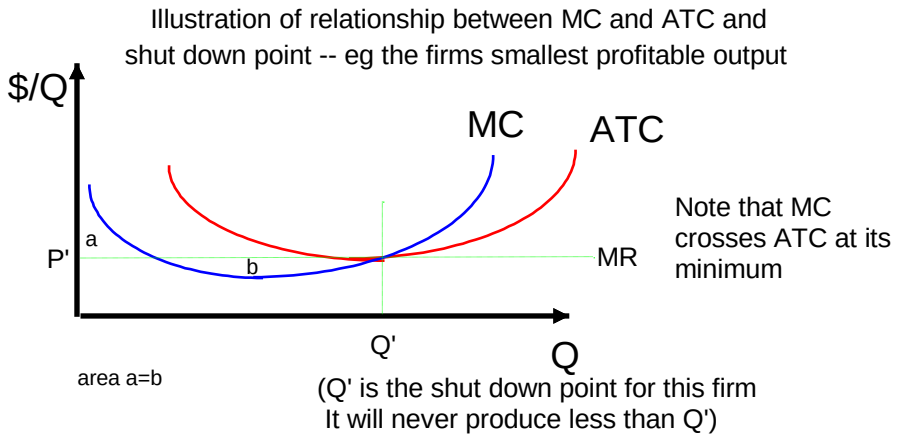
Quantity of Output	Total Cost	Average Cost =TC(Q)/Q	Marginal Cost = DTC
Total Cost Function		*Calculations*	
1	100	100	100
2	180	90	80
3	250	83.33	70
4	340	85	90
5	440	88	100
6	550	91.67	110

It turns out that marginal X and average X also have a variety of mathematical relationships. Among the most useful in undergraduate economics classes are the following:

- If marginal X is above Average X, then the Average X curve will be rising.
- If marginal X is below Average X, then the Average X curve will be falling.
- If marginal X = Average X, then the Average X curve will be neither rising nor falling.

All of these relationships apply to your grade point average, if you have a 3.0 average and get an A, then your average will be pulled up, because in this case your marginal grade is higher than your average grade. If instead you get a C, then your average will be pulled down (in this case your marginal grade is below your average grade). If you get a B, nothing will happen to your average (in this case your marginal grade equals your average grade).

Illustration of the relationship between MC and ATC



Notice that you can use either MC or ATC (actually ATVC) to characterize a firm's shut down point.

- At price P', the firm earns zero profits at it profit maximizing output, Q'
- (Profits can be calculated using the average or marginal curves.)
- At prices below P', there is no way for the firm to earn a profit (even ignoring fixed costs) and so it will shut down.

VI. Market Demand and Supply Curves

In **competitive markets**, both firms and consumers are "**price takers**," they take the prevailing market price as "given" and adjust their production and consumption choices to the prevailing market price.

Thus, in competitive markets, the market price is every firm's marginal revenue curve, which is just a horizontal line equal to the prevailing market price. Each firms gets "P" new dollars of revenue every time it sells an additional unit of its product or service at that price. So, the first MR curve is a horizontal

line through P1. Similarly, in competitive markets, the prevailing market price is every consumer's MC curve.

Equilibrium in markets occur at the price the set market demand equal to market supply. To characterize that equilibrium, we'll need to derive both a market demand curve and a market supply curve. It turns out that market demand curves are simply the "horizontal sum" of individual consumer demand curves and that the market supply curve is the "horizontal sum" of individual firm supply curves. The derivations below should make that clear.

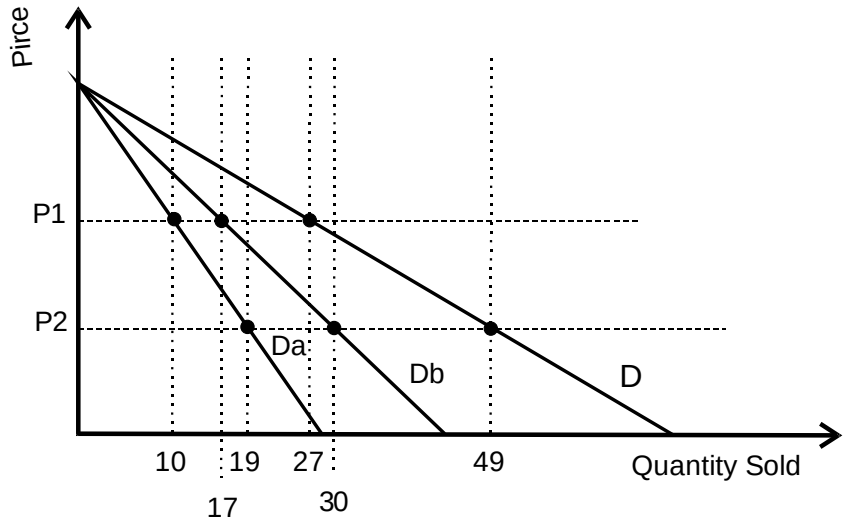
Deriving a Market Demand Curve

Market Demand characterizes the total amounts that consumers will purchase buying at given prices.

Similarly, Market Supply (in the short and medium run) characterizes the total amount that all firms in an industry (market) industry will produce at given prices.

A market demand curve can thus be derived by choosing a price and plotting the total purchases of all consumers (in the market) at that price, and plotting the price chosen and the total purchase. Repeat with a series of prices to trace out a market demand curve.

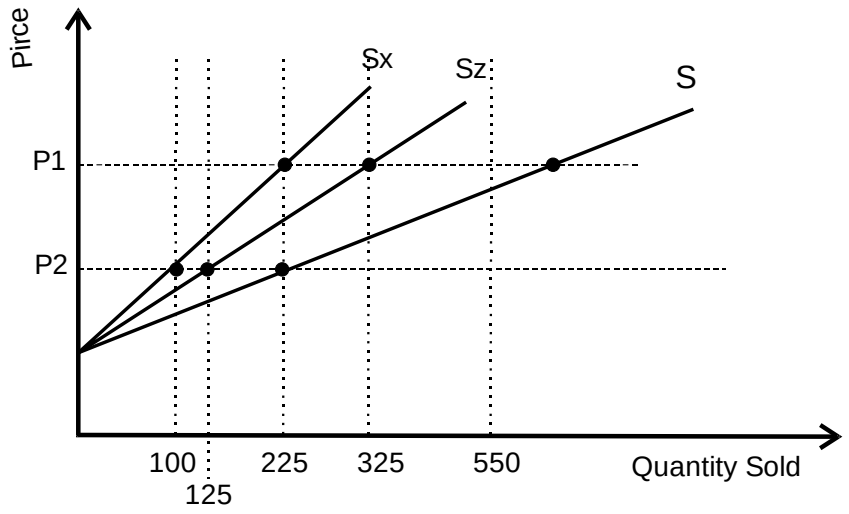
Figure 2.4: Derriving a Market Demand Curve



Deriving a Market Supply Curve

Similarly, a market supply curve can be derived by choosing a price, finding the amounts that will be supplied by every firm in the market, adding up those quantities, and plotting the chosen price and the total output of the industry. Repeat with a series of prices to trace out a market supply curve.

Figure 2.5: Derriving a Market Supply Curve



G.

H. **Market Demand** curves for ordinary private goods are "horizontal" sums of individual demand curves.

I. **Market Supply** curves for ordinary private goods are "horizontal" sums of individual firm supply curves.

Having derived the individual market demand and supply curves using the net benefit maximizing model, it should now be clear that:

- A market demand curve is (approximately) the horizontal sum of the marginal benefit curves of the individual consumers. So, a market demand curve can be used to estimate the aggregate MB curve for all the consumers in the market of interest.
- A market supply curve (short run) is approximately the horizontal sum of the marginal cost curves of individual firms in the market. So, a supply curve can be used to estimate the industry's MC curve.
- [An **exception** is the case in which Marshallian assumptions are used, in which case, S_{LR} is the industry's ATC curve. The distinction between Marshallian and Ricardian long run supply is discussed below and revisited in an appendix to this chapter.]

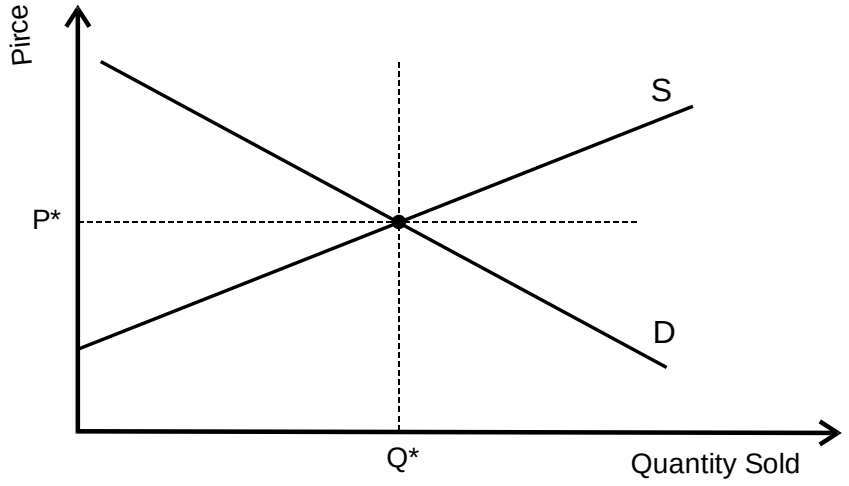
IX. Equilibrium Prices in Competitive Final Goods Markets

Economists use market supply and demand curves to model the determination of market prices in "competitive markets." A competitive market exists when a reasonably "large" number of producers (firms) sell very similar products to relatively large numbers of consumers. Their large numbers tend to make both sides of the market "price takers" as has been assumed when we derived the individual consumer demand curves and the individual firm supply curves. Large numbers make it very difficult, if not impossible, for firms or consumers to have a significant impact on prevailing market prices.

A market is said to "clear" when the quantity supplied by all firms selling a particular type of good or service provide the total amount of those goods or services demanded by all the purchasers (consumers) of those goods or services. A market clearing price, P*, is one that induces the firms in a particular market to supply quantity of a good or service equal to the amount that consumer's demand at that price in that market.

Market prices tend to move to levels where the total quantity supplied by all firms equals the total amount demanded by consumers. (This defines P*, and Q*)

Figure 2.6: A Market Clearing Price



The "market clearing" model of price determination predicts that output and prices tend to move toward P* and Q* as prices adjust to "ration" the quantities produced for sale to consumers.

Equilibrium prices cause all markets to "clear" at which market supply = market demand.

Note that this adjustment process is, in principle, an entirely decentralized process requiring governments to do nothing more than enforce property rights and contracts.

- If there is a shortage, prices tend to rise, causing consumers to buy less and firms to produce more, until a price emerges that sets demand equal to supply.
- If there is a surplus, prices tend to fall, causing consumers to buy more and firms to produce less until a price emerges that sets demand equal to supply.

X. Normative Analysis Using Supply and Demand

These properties allow us to use some of our results from deriving supply and demand curves to analyze gains from trade, the efficiency of markets, and problems that might be associated with trade barriers or externalities. (The latter are taken up in chapters 5 and 10.)

Market demand curves can be used to approximate the marginal benefits received by all consumers. So, the areas under demand curves can be used to approximate the total benefits received by consumers.

Market supply curve can be used to approximate the marginal cost of suppliers. So, the areas under a short run supply curve can be used to approximate the short run total variable cost of the industry, and the areas under a long run supply curve can be used to approximate the total long run total costs of the industry of interest.

Consequently, the demand and supply curves can be used to estimate the net benefits realized by all firms and consumers in an industry (profits and consume surplus).

We can use these properties of demand and supply curves to show that competitive markets tend to be "efficient" or "Pareto optimal," because that they induce production levels that maximize social net benefits.

In the absence of externalities or monopoly power (both covered later in the course), markets tend to produce social net benefit maximizing output levels of goods and services.

The Geometry of Social Net Benefits

The supply curve can be used to represents the industry's marginal cost of production, which is approximately the **social marginal opportunity cost** of the resources used to produce the good being analyzed. Marginal opportunity cost represents the value lost from other uses of the resource (labor, energy, materials etc.) when it is shifted from other markets (uses) to producing the good of interest.

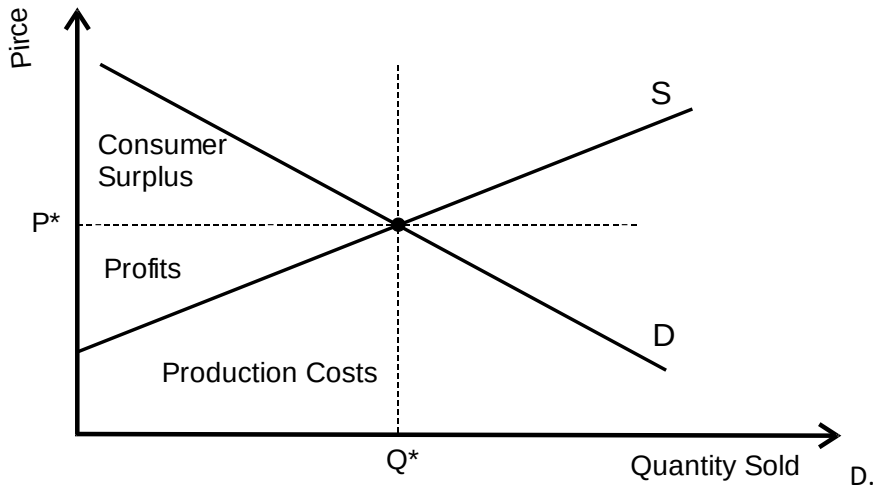
The demand curve can be used to represents the marginal benefits realized by all consumers that purchase the good of interest. So, in the absence of externalities, the **demand curve can be used to approximate the social marginal benefits** associated with producing and consuming the good or service of interest.

The areas under the demand and supply curves correspond to society's total benefits and total costs of production, respectively.

(Again, we are assuming that there are no externalities and little or no monopoly power.)

Figure 2.7 uses this geometry to determine the social net benefits of production, sale, and use of a product that is produced and sold in a competitive market. Notice that the social net benefits are divided up between firms and consumers. The firms share of the net benefits called profits and the consumers share of net benefits called consumer surplus.

Figure 2.7: Net Benefits at a Market Equilibrium



The geometry of "market clearing" price implies that markets to produce the output levels that set **marginal social benefits (demand) equal to marginal social costs** (supply).

- Consequently, competitive markets tend to produce the social net benefit maximizing level of output, the output that maximizes the sum of consumer surplus and profits.
- This conclusion is one very widely used normative argument in support of competitive markets as an "efficient" welfare maximizing form of social organization.
- It is sometimes called the first theorem of welfare economics: competitive markets produce (Pareto) efficient outcome.
- They do this in part, because firms have incentives to minimize the resources used to produce their products.
- They also do it because firms have incentives have incentives to produce what consumers want, what generates relatively high marginal benefits for them.

I. Appendix : On the difference between Ricardian and Marshallian long run supply.

A. In the short and medium run, the number of firms in the industry can be taken as fixed.

As a first approximation, this is also often true in the long run.

B. On this point, however, the Ricardian and Marshallian conceptions of long run supply are quite different and imply different things about the nature of the long run supply curve and long run profits.

- Long run supply in the Ricardian framework reflects “entry and exit” of firms with different cost functions, with higher cost firms entering last. The low-cost producers enter first, other things being equal, and the high-cost firms last. (The LR supply curve already includes these effects.)

- **Ricardian long run supply is thus always "upward sloping**" and reflect the **marginal production costs** of firms in the industry. In general, relatively efficient firms (those with better technologies, locations, or natural resources) will continue to earn profits greater than zero in the long run, because they have lower costs than "marginal" firms.

- Long run supply in the "**Marshallian**" long run reflects **entry and exit of identical** efficiently sized firms into and out of the industry. This process takes place until profits fall to zero. As long as input prices do not change, **Marshallian long run supply curves are always horizontal** and reflect the industries average cost (the minimum average cost for each identical firm).

- In the long run, Marshallian firms always produce at efficient scale (at the bottom of their ATC curves) and earn zero profits.

C. Note the derivation of long run supply in the Ricardian sense, is exactly like that used above for both short and medium run analysis. The number of potential suppliers is essentially fixed (determined by production costs), although the individual firms will alter their production as prices vary.

D. Long run supply in the Marshallian model takes place entirely through the entry and exit of identical firms each producing at the bottom of their LR average total cost curves and earning zero profits.

It bears noting that the Marshallian model of long run supply works well for some industries (say auto repair shops), and the Ricardian model works well for others (oil, minerals, sports, farming, etc.).

E. We will mostly use the Ricardian model of long run supply in this class, but it should be kept in mind that the Marshallian entry and exit may also be important in some markets.

Chapter 3: Comparative Statics: Why Prices Change

I. Introduction: A Recap and Overview

Recap

The previous chapter provides the basic logical foundations of microeconomics. It showed how **rational decision makers**—in the sense of being net benefit maximizers—would behave in their roles as consumers and producers. That behavior, in turn, implied that when consumers and producers are price takers, equilibrium price emerge that are equal to the marginal benefits of consumers and the marginal costs of firms.

Consumers and firms tend to be price takers when there are many consumers and firms in the market of interest. In such cases, no single individual or economic organization has significant control over the average price at which goods and services are sold in the market of interest. It was "**price taking**" behavior that allowed us to derive both individual demand and supply curves and demand and supply curves for local, regional, or international markets for particular products.

Prices in such settings tend to move toward the levels that "clear the market" of interest. Such price set market demand at that price (P*) equal to market supply at that price (P*). If prices are below P*, too little is provided to satisfy demand and consumers compete with each other for the product on sale and, in effect, bid the price up until supply equals demand. If prices are above that level, then supply exceeds demand and firms, in effect, compete with each other for consumers by lowering the price of the goods on offer until supply equals demand. Because such activities are undertaken by many consumers or firms at the same time, they jointly induce prices to adjust to equilibrium levels—the price characterized by the intersection of the demand and supply curves.

Chapter 2 also showed why trade takes place. **Both firms and consumers benefit each time a trade is transacted**. Consumers realize consumer surplus because their total benefits from purchasing the good of interest exceed their total costs. Otherwise, they would not purchase any of the good. (Keep in mind, that **individuals buy none of most goods** on offer exactly because their expected benefits from many purchases are less than the expected cost of most goods.)

On the other side of the market, **firms produce and sell goods and services because they can profit** from doing so—which is to say, because the revenues realized by selling the good or service of interest is higher than the cost of producing it. Firm owners seek profits, because that is one of the sources of income that allows them to purchase goods in their roles as consumers, to generate funds for "a rainy day," and to make other investments to improve prospects for future consumption.

Trade is always voluntary, and unless one or the other side is fooled into making a mistake either when purchasing or when selling a good, both parties have to benefit—otherwise the exchange would not take place.

Specialization occurs because some entrepreneurs (firm owners or firm organizers) are better at some forms of production or at some forms of organization than they are at others. This allows them to profitably sell their products or services at a lower price than other less able producers. (Input sellers also tend to specialize in providing the services, equipment, or intermediate goods that maximize their income. However, analysis of input markets is taken up in the next chapter and so are neglected in this one.)

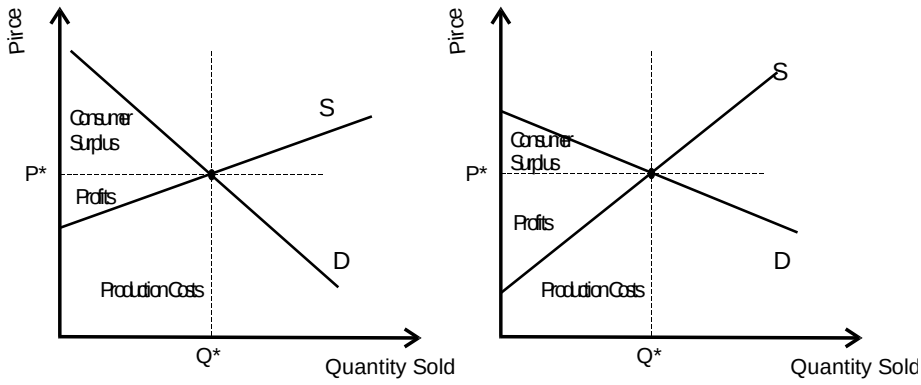
Most producers, thus, sort themselves out (in the pursuit of profits) into markets where their skills produce the most efficient (least cost) means of bring products to market. This process enhances the gains from trade realized by consumers by reducing production costs, which tends to reduce prevailing market prices at the same time that they generate profits for entrepreneurs and firm owners.

All this is not to say that no one ever makes mistakes, but rather it is to say that, on average, prices gravitate to levels that set demand equal to supply in every market. As a consequence, surpluses (excess supplies) and shortages (excess demand) tend to be small and on average are approximately zero for all goods in well-functioning markets.

The last graph in chapter 2 showed how the gains from trade are divided up in a market where both sides are price takers rather than price makers. That division was approximately equal, which is not the only possibility but varies with the slopes of the demand and supply curves in the market of interest. Figure 3.1 show how the division of gains to trade between sellers (firms) and purchasers (consumers) varies with the slope of the supply and demand curves.

In general, the flatter (more price sensitive) the demand or supply curve is, the smaller the fraction of the net benefits of exchange are realized by consumers. In the first diagram, the demand curve is steeper (less price sensitive), and consumers realize most of the net benefits of exchange. In the second diagram, the supply curve is steeper and more of the net benefits are realized by suppliers.

Figure 3.1 On the Slopes of the Demand and Supply Curves
and the Distribution of the Net Benefits at a Market Equilibrium



It is not that firms are "greedier" in the second market than in the first. Rather it is because production costs rise faster in the second than the first and so there are more units of the good or service that are sold with "high margins" (e.g. with relatively high profits on those units). Remember that supply curves simply reflect the marginal cost curves of the sellers in the market and their numbers. The flatter these MC curves are and the more firms there are competing for consumers, the flatter the market supply curve tends to be (and the lower supply profits).

Likewise, it is not that consumers are dumber in the second market than in the first, but rather than the marginal benefits from the good or service for the average consumer is relatively high and diminishes rapidly as additional units of the good or service are purchased. It is the nature of the marginal benefit curves of each consumer and the number of consumers that determines how "flat" a market demand curve is, and thereby the extent to which consumers realize the lion's share of the gains to trade or somewhat smaller shares.

However, it is clear that each side of the market always benefits because otherwise trade would not take place! All market transactions (except those mandated by governments) are voluntary—although consumers would always like a lower price and firms would like a higher one. It is ultimately economic pressures (Supply and Demand) that determine both market prices and the manner in which the net benefits from exchange are realized.

Focus of Chapter 3: Why Prices Change

Chapter 2 provides a general explanation for the prices that we observe when we shop for goods and services. Chapter 3 focuses on factors that cause price to change and provides a general theory of the factors that cause prices to rise and those that cause prices to fall.

Generally, the factors focused on in this chapter are in a sense "natural" one—they are factors associated with nature and human nature, rather than governmental policies—which can also induce prices to change in various ways as developed later in this course.

In general, any change in circumstances, preferences, or expectations that affects the marginal benefits of a good or service (for most consumers in the market of interest), will affect the location of the market demand curve. And any change in circumstances that affects the marginal cost of producing and selling a good or service (for most firms in a market) will affect the location of the supply curve.

The geometry of supply and demand imply that a shift in either the market demand curve or the market supply curve tends to change the prevailing market price for the good or service sold in the market of interest. In general, an increase in demand tends to increase prices whereas an increase in supply tends to reduce them—other things being equal.

II. Factors that Tend to Shift the Demand Curve

Demand curves reflect the effect of prices on the quantity demanded by everyone in a particular market. So, a change in the price of a good (almost) never shifts a demand curve. Instead, it causes movement along the curve. Shifts in a market demand curve occur when the marginal benefits of the good or service of interest change for some reason.

Sometimes the marginal benefit of a good varies with the weather—other things being equal. For example, consider the demand for umbrellas. If one lives in a climate where it never rains, there is little benefit to owning an umbrella, and so the MB of umbrellas tend to be low, which implies that an average individual's demand for umbrellas is low in such regions. On the other hand, if one lives in a place where it rains 4 days a week, the marginal benefit associated with an umbrella or two (or three) tends to be higher and so an average individual's demand for umbrellas tends to be higher in such reasons than in drier ones. The average person in the wet region will own more umbrellas than the average person living in dry regions if the price of umbrellas in both places are about the same.

Other examples of **weather dependent** (or conditioned) marginal benefits (and therefore demands) include barbeque grills, picnic tables, waterproof coats, sandals, fans, air conditioners, convertibles, golf clubs, windows, swimsuits, cold drinks, bicycles, swimming pools, irrigation systems, boots, gloves, furnaces, heat pumps, warm coats, skis, ice skates, hot drinks, and snow mobiles. All these goods have marginal benefits curves that are partly determined by weather. A change in weather or location will, thus, change the demand for such goods. (Many folks from farther south purchase their first winter coats during their first winter at WVU.)

This is true at the individual level and therefore at the market level of demand. If a change in weather increases the MB associated with owning a good or purchasing a particular service, then then the market demand curve tends to increase in response to such weather.

Some marginal benefits vary with **expectations**. If one expects rain to occur, the MB of bringing an umbrella on a trip is higher than when one does not expect rain. Similarly, if new information makes one expect greater marginal and total benefits from a particular good than one had before the information was received and digested, then one's marginal benefit curve shifts up and one's demand for the associated product increases. For example, during the covid year prior to the vaccine, many news sources said that masking would lower one's probability of contracting covid which increased the expected marginal benefits associated with masks, so individual and market demands for masks increased dramatically. Similarly, if information about the health benefits of a particular lifestyle or type of food suggests that "it" is better for one's health than previously thought, then the MBs of that lifestyle or food increase and the demand for that lifestyle or food increases.

Changes in income also affects the MBs associated for many goods, because one's marginal benefit associated with an additional unit of a good is one's willingness to pay for that unit of the good. (Remember that marginal benefits are in terms of the dollar amounts (or other currencies) that one is willing to pay for another unit of the good. The higher is one's income, the more money one has, and (generally) the more money one is willing to pay for an addition unit of most goods. So, as consumer

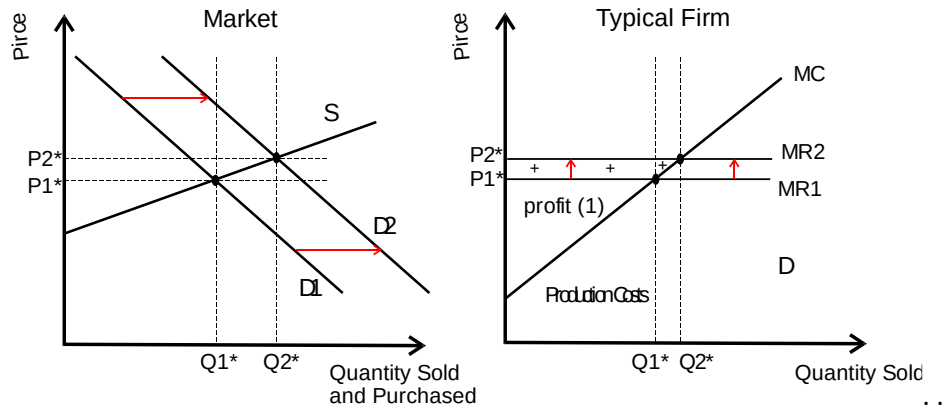
income increases, demand for all **normal (and superior) goods** increase. (There are a few goods where, as one's income increases, one is less likely to purchase fewer of such goods—such as good purchased at thrift shops. Economists call such goods "**inferior**" goods.)

The prevailing **price of substitutes or complements** also affect the marginal benefits (reservation price) for particular goods. If one is wearing a raincoat, the MB of an umbrella is less. Raincoats are substitutes for umbrellas, although not perfect ones. So, if the price of raincoats decreases, the MB of umbrellas tend to decrease as well. A substitute (raincoat) can be used instead of the good of interest (umbrellas). A complement is something that you tend to use along with the good of interest. A beach towel is used with a swimsuit. A daypack is used with hiking shoes. A textbook or a particular set of webnotes (such as these) is used along with a college course. As the price of complements increase, the marginal benefits of the good of interest tends to decrease. **So, an increase in the price of a complement causes demand to diminish (shift to the left), whereas an increase in the price of a substitute tends to cause demand to increase.**

There are a huge range of changes in circumstances and expectation that can affect the marginal benefits of a good. What may be a bit surprising is that any change that increases the marginal benefits or expected marginal benefits of a good or service has the same effect on an individual's demand. It increases demand. If the same change affects most people in a market, then the market demand also increases.

And, when market demand increases—and the supply curve does not change—then price tend to rise, as indicated by figure 3.2 below.

Figure 3.2 The Direct Effects of an Increase in Consumer Demand



.Notice that the effect of an increase in demand on price is to increase the quantity of the good produced for sale by sellers, because an increase in price increases the profits that firms realize by additional production at the margin.

As prices increase, the marginal revenue of each firm increases, which induces firms to produce additional units of output. (Note that the figure to the right depicts the effect of a higher equilibrium

price on a typical seller-producer.) The increase in profits by a typical firm is the area denoted with the "+" signs. This increase in production and sales implies that prices do not rise as much as would have happened without the adjustment to higher prices by firms.

III. Factors that Tend to Shift the Supply Curve

Equilibrium prices can also be affected by changes in supply (e.g. shifts in the supply curve). Such changes are induced by any factor that affects either then number of firms in the market of interest or the marginal cost of production borne by a typical firm in that market or industry.

Weather is, as for demand, a factor that can affect supply. It may do so by changing marginal cost directly or by changing the productivity of the process used to produce and bring the products of interest to market. For example, a late frost may greatly reduce the productivity of farms producing various fruits and vegetables. Indeed, for a subset of farmers (agricultural firms) a late hard frost may eliminate its production of many fruits and vegetables. For example, a few decades ago Florida was a major producer of oranges and orange juice, and a heavy frost could wipe out most of the oranges in a typical orange grove either by damaging the orange blossoms, the fruit on the tree, or in some cases the trees themselves. Bad weather such as a late flood, a drought, or, in some cases, excessive rains, can directly reduce the market supply of a broad range of agricultural products. Such reduction in supply tend to increase prices for the products of interest and thereby indirectly reduce the supplies (by increasing marginal costs of other products such as orange juice, pasta sauce, or bread that are made with the fruits, vegetables, and grains most affected by bad weather.

Any change in circumstances that induces input prices to increase tends to increase the marginal costs of products created with those inputs. Production is often undertaken by a series of firms that make up a "supply chain" of inputs used to produce the goods ultimately sold to consumers. Any change in the inputs or outputs along the supply chain tends to increase the marginal cost of producing the "final" goods and services sold to consumers, and thereby tends to reduce the quantity of goods that can be supplied at a given price.

Inputs include, for example: fruits and vegetables, various mined minerals [copper, iron, lithium, aluminum, etc.], intermediate goods [steel, computer chips, electricity, cardboard, transportation services, etc.], and the many types of labor used to produce the final goods sold to consumers. Increases in any of those "inputs" increases the marginal production cost of the final products ultimately sold to consumers and thereby reduces market supply. Each firm in the final goods market has higher marginal costs and so their supply curves shift back to the left and so does the market supply.

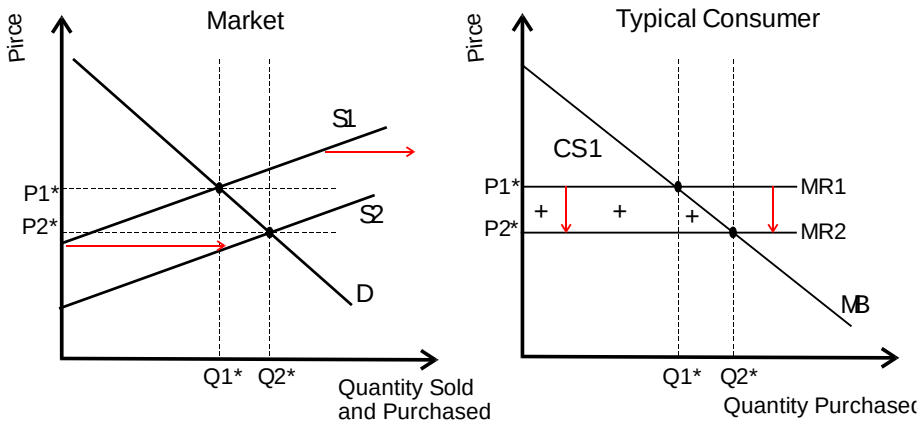
Another factor that can change the marginal cost of a product is technology. Innovations in productive methods generally reduce the marginal cost of production. Innovations that increase cost without improving quality, naturally, are not adopted by firms because such "innovations" increase their marginal costs and reduce their profits. So, only cost reducing innovations tend to be adopted and all

such inventions reduce marginal costs for the typical firm in the supply chain, and thereby tends to reduce the marginal cost of producing goods that use those inputs.

Competition among sellers for consumer "dollars," in turn induces producers to pass on most or all of their cost savings. So, profits tend to increase for firms benefiting from improved production technologies—which is what induces such technologies to be induced (at least in the short run). However, profits increase by less than the amount saved from marginal cost reduction because of competition among sellers for the purchase of consumers of their products.

Any change in circumstances (shock) that tends to reduce the marginal cost of production tends to increase market supply (shift the market supply curve out to the right) which—other things being equal—tends to cause the equilibrium price to fall, which tends to benefit all consumers who realize greater net benefits (consumer surplus) from their purchases of the goods with lower prices. (The area that characterizes a consumer's increase in consumer surplus is marked with "+" signs in figure 3.3.) Increases in the marginal cost of production have opposite effects.

Figure 3.3 The Direct Effects of an Increase in Supply



Many consumer goods are much less expensive today (in inflation adjusted dollars) than they were a decade or two ago, reflecting both the effects of innovation on production methods and the quality of final goods throughout the chain of producers that bring final goods to markets. Exceptions occur in areas where innovation is more difficult or impossible—as with higher education (so far)—whose prices have increased in real terms over the decades as universities compete for talented workers (professors) in the labor market for such persons.

Shifts and Both Supply and Demand

There are many cases in the real world where a change in circumstances affects only market demand curves or market supply curves, but there are others—such as unusually bad or good weather—that affects both sides of the markets at the same time. In such cases, both the demand and supply curves change simultaneously, and predictions about sales (Q*) and prices (P*) may not be as clear as in the cases in which only one side of the market is affected.

In some cases, a shift in supply (upward) may cause prices to fall while an associated shift in demand (upward) and the result will depend on which effect is larger. Nonetheless, the diagrams still help to illuminate what is going on in the product market of interest, and also help explain why the result may be ambiguous as far as theory (economic principles) are concerned.

There are also cases in which the adjustments take place through time and the short run and long run effects of a demand shock or supply shock differ somewhat in the long and short run. In general, the adjustments to sales or output levels tend to be greater in the long run than in the short run—while the price effects are often smaller in the long run than in the short run for reasons developed in the next section.

IV. Long Run and Short Run Adjustments

In addition to factors that directly affect consumer marginal benefits and firm marginal costs, there are also price adjustments that take place through time as consumers and firms adjust their plans, because more factors can be taken into account and adjusted in the "long run" than in the "short run."

Differences between the long and short run occur because many changes in plans or methods requires time. A college degree takes three or four years to finish. Thus, changing from a person with a high school degree and such a persons' employment opportunities to one with a college degree with quite different employment opportunities takes several years to implement. Similarly, change from one location to another takes a bit of time as some research may be necessary and financing and contracts take time to implement—in this case less than earning a college degree but more than grocery shopping.

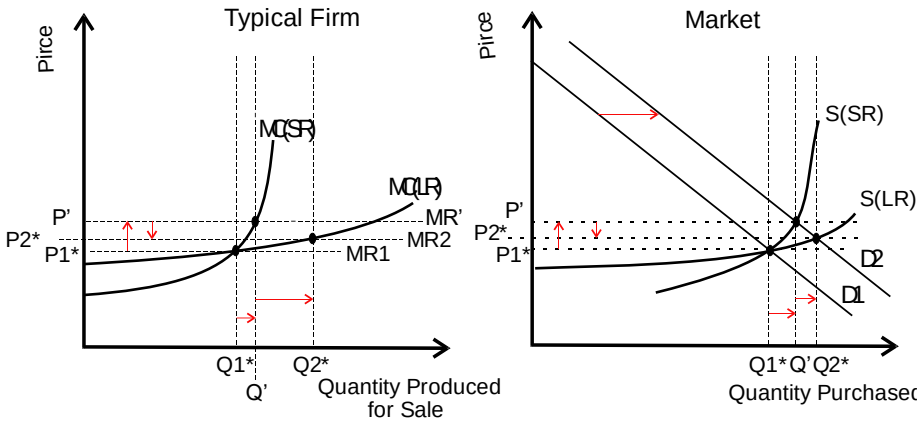
Similarly, many of the adjustments that economic organizations can undertake take time to execute. It takes time to build a new factory, order specialized capital equipment and have it delivered and installed, and to train people to effectively use the new facilities. It also takes time to train (and or identify) relatively talented employees and even more to relocate the organization's productive or sales facilities from their present location to a new location.

Consequently, time matters—and adjustments to new prices, new technologies, or long-term shifts in demand or supply can take many years to occur. Such adjustments in "plans" may also affect both demand and supply as they are implemented. For example, the supply of trained or diligent people increases as more persons seek and obtain useful college degrees. An increase in the supply of such persons may induce firms to reorganize their production methods to take advantage of an increase in the numbers of such persons. A new method of production—better robots, AI-enhanced inventory and quality control, changes in the locations where important inputs are produces, and so forth may induce facilities to be modified and relocated. Such changes generally increase supplies for consumers and change the pattern of demand for inputs (including both many types of labor and capital).

Such long-term adjustments by both consumers and firms to new economic circumstances implies that long-run demand and long-run supply curves are more price-sensitive (e.g. flatter) than short run demand and supply curves whenever price changes induce significant changes in plans for or methods of future consumption or production.

Modeling Long and Short Run Supply

Figure 3.4 Long and Short Run MC and Supply



The most common way that economists have modeled the difference between short and long run decisions by firms it to assume that at least one factor of production (usually capital) cannot be varied in the short run. Since that factor cannot be varied in the short run, it cannot affect short run marginal cost, because it is a "fixed cost" rather than a variable cost. Because one cannot increase capital to expand output over the current level, marginal costs for expansion are higher than they would otherwise have been. Moreover, because one has more fixed capital than ideal for reduced output levels, reductions in output is "cheaper" in the short run than they would have been with the proper amount of capital. (That is, in a sense, fixed costs are higher than they would have been, with the cost-minimizing level of capital.) So short run MC is greater than long run marginal cost for expansions, but lower than long marginal costs for reductions in output. The lefthand diagram of figure 3.4 above illustrates this effect. Thus, every firm's LR marginal cost is "flatter" than its SR marginal cost for a "given" amount of capital—where that "given amount" is the amount that the firm regarded to be optimal for the particular level of production anticipated in the past—namely, its profit maximizing level for the price anticipated at that time.

This relationship has implications for the firm's long run and short run demand, because the same logic used in Chapter 2 applies for both a firm's short run and its long run supply curve. A firm's short run supply curve is constructed from a subset (possibly all) of the points on its short run marginal cost curve. The same firm's long run supply curve is constructed from a subset (possibly all) of a firm's long run marginal cost curve. Thus, a firm's long run supply curve is flatter (is more price sensitive) than its short run supply curve, as illustrated in the righthand graph of figure 3.4. Note that the market LR supply is flatter than the market SR supply.

That same diagram shows how prices adjust through time in response to an increase in demand for the product or service of interest. In the short run, relatively little can be done to increase supply because of fixed capital (and possibly limits on other factors). In the long run, those factors are adjusted to maximize profits (e.g. to reduce the cost of higher levels of output). At the long run equilibrium price (where the new demand curve D2 crosses the LR supply curve), firms have optimized their production levels and prices fall relative to the short run price.

That lower price tends to increase consumer surplus over what it would have been at the higher price.

Moreover, in this case, the last price movement is not caused by shifts in LR supply or demand, but rather by adjustments to the higher prices and firm outputs generated by adjustments by firms to the new level of demand.

A good recent example of this effect is the supply of N95 masks. The spread of covid 19, a potentially deadly disease, was said to be reduced if one wore an N95 mask. This greatly increased the demand for those masks and supplies were very limited. Prices did not rise fast enough to clear the market, and N95 masks disappeared from store shelves at both drug stores and building supply stores. Gradually, production was ramped up, and they became available again—albeit at prices that were triple or more than their previous prices. During the course of a year or so, production increased again as more equipment was installed, and prices gradually fell back to about their pre-covid levels. And supplies were sufficient that many people had a dozen such masks that they kept in coat pockets and in their glove compartments.

Fortunately for “us” the spike in mask demand was temporary and lasted only two or three years in the U.S. in part (1) because vaccines were developed which reduced infection and death rates somewhat, and in part (2) because Covid 19 fortunately evolved in a manner that made it less deadly.

A Digression on Ricardian vs Marshallian LR Supply

There are two models of LR supply used by economists. The one used in this course (and text) is the Ricardian model. In that model all firms have somewhat different cost function (LR marginal cost curves) and LR output adjustments are mainly by the firms currently in the industry. There may be some marginal producers that enter when prices increase, but they make up a relatively small part of the overall increase in market output. In the Ricardian model firms realize different profits because of their different cost structures.

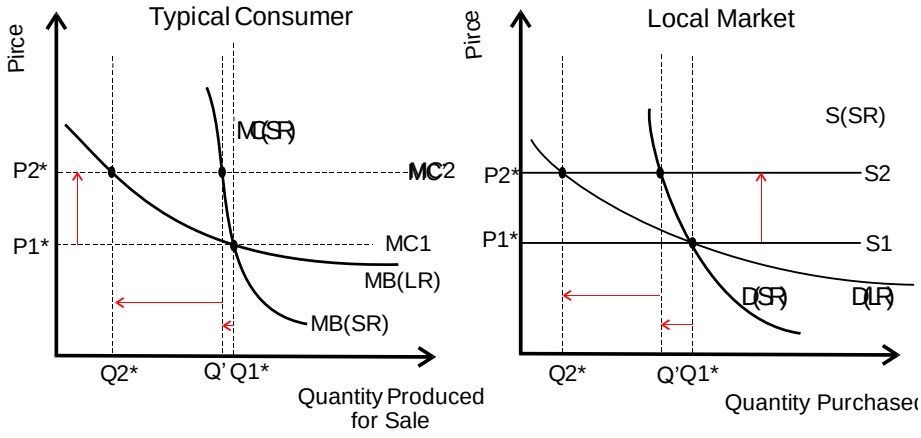
In the Marshallian model all firms in an industry are identical and operate at efficient scale in LR equilibrium. In such a market, LR output adjustments are all ultimately the result of entry and exit into the industry. Here one can imagine, changes in the supply of coffee shops or bars in response to changes in the demand for coffee or booze and nightlife. Coffee shops and bars are (mostly) all about the same size and all use very similar technologies.

In cases, in which firms use different technologies, are of different sizes, and may have better or worse management and access to inputs or good weather, the Marshallian model is inappropriate (even if it does simplify mathematical models) and the Ricardian one is most apt. It is for that reason that the Ricardian models are used in this course. They also simplify the geometry and logic of long run adjustments. Entry and exit do occur, but mostly by the least profitable firms.

LR and SR Consumer Demand

Although most microeconomic textbooks discuss differences between LR and SR supply, relatively few texts discuss differences between LR and SR demand for final goods and services (e.g. demand by consumers). However, there are many consumer demands that are also affected by their holdings of capital goods such as an automobile, house, cell phone, or computer. Such holdings are fixed in the short run although adjustable in the long run and so long run demand tends to be more price sensitive (price elastic) than short run demand for goods that are complements to or inputs for their fixed capital goods.

Figure 3.5 Long and Short Run MB and Demand



For example, suppose that the price of gasoline tripled to European levels. In the short run, gasoline consumption would not change very much because every consumer of gasoline has a particular car, SUV or truck and particular places that "have to" drive them. So, consumer demand for gasoline would not initially change much. There would be some adjustments as leisure drives were reduced and carpooling increased, but basically the result would be very much higher expenditures on gasoline, nearly 3 times as much as before the price rise induced by a supply shock of some kind or new public policy.

In the long run, however, those capital goods could be changed—and smaller cars or hybrid or electric ones purchased to replace their "gas hogs." And, one might move closer to work to reduce commuting expenses or take a somewhat lower paying job closer to home. So, the long run quantity of gasoline demanded at the new higher price would be much less than in the short run.

Figure 3.5 illustrates the effect that would be associated with a "local" public policy shift for a product sold on world markets. In this case, the local supply curve is essentially flat, because it is a small part of a much larger market for gasoline and the effect on gasoline prices is generated by a local regulation or tax. In the short run, there is a small reduction in purchases (from Q1* to Q', but a larger one in the long run, from Q' to Q2*.

In a diagram with upward sloping supply curves, prices would also fall a bit as the adjustment to long run demand occurred. This does not happen in the diagram because of the assumption that the supply effects were local rather than global (induced by local regulations or taxes) rather than global shocks on the world supply of gasoline.

V. Some Conclusions about Price Movements

Microeconomics focuses on the fundamental aspects of choice settings that explain most prices and most changes in prices. It rarely works perfectly, but nearly always works pretty well—and nearly always better than any other theory of prices or prices changes.

The previous chapter demonstrated that ideas associated with the "marginal revolution" that took place during the 1870s can account for the fact individual markets tend not to have excess inventories and yet are able to—in most cases—serve all the customers wanting to purchase a particular product or service.

This chapter has shown that changes in the marginal benefits of most consumers will cause demand curves to shift and the result tends to be new equilibrium prices. If demand increases (shifts to the right), prices tend to rise. If demand decreases (shifts to the left) prices tend to fall. Similarly, changes in the marginal cost associated with producing goods or services for sale in markets tend to cause supply curves to shift. An increase in MC causes supply curves to shift back to the left and tends to cause prices to rise (and sales to fall). A decrease in MC tends to cause supply to increase (supply curves to shift to the right) with the result that—other things being equal—prices tend to fall.

The analysis of such shifts in supply and demand is normally referred to as "comparative statics," because it focuses on equilibrium outcomes rather than the process through which such outcomes emerge. It is comparative in that it compares outputs in prices at a new market equilibrium with those of an older one. It focuses on "statics" because once a new equilibrium emerges, there is no further reason for prices or outputs to change—unless another "shock" changes the average marginal benefits of consumers or their numbers or unless some "shock" affects the marginal cost of SR or LR production, or the number of firms.

Nonetheless, although the models focus on equilibria, the models of this chapter can account for environments in which market prices and outputs are changing all the time. Such changes would occur if there were lots of "shocks" being adjusted to—either simultaneously or one after another.

The models insist that one should focus on the underlying source of gains from trade. Potential gains to trade exist whenever the cost of production is less than the highest price that a subset of consumers is

willing to pay for the good or service of interest. When marginal benefits fall that price (the reservation price) also falls. The latter reduces potential gains to trade and thus trade tends to diminish. When the marginal cost of production falls, new gains to trade tend to emerge and so trade tends to increase.

Fortunately for markets, the latter has been very common for the past 200 years in the West.

The next chapter takes up the effects of markets on personal income and wealth. These are determined by prices in other markets—not markets for final goods, at last not directly, but rather markets for inputs of various kinds.

Chapter 4: Input Markets and Market Equilibria

I. Review: Consumer Tastes and Income

Final goods markets exist because individuals have demands for various goods and services, which means (1) that individuals have the income or credit necessary to pay for them and (2) believe that purchasing some products or services will improve their lives in some way. The term "tastes" is often used by economists to describe the underlying desire for some goods and services. The term "tastes" doesn't really capture the full range of desires that consumers believe goods and services may satisfy or advance, but it is a neutral enough term to convey the idea that individuals may have different "tastes" or "preferences" and that no person's tastes or preferences are privileged—e.g. necessarily better or more important than another's.

Tastes and Income Determine Marginal Benefits

In addition, it provides a basis for understanding that marginal benefits (the highest price that one is willing to pay for successive units of a good) are partly determined by tastes and partly by income. The stronger the preference that one has for a particular good or service the higher are its marginal benefits—other things being equal. And the higher one's wealth or income is, the higher the price that one is willing to pay for successive units of a good tends to be. Together tastes and income largely determine one's marginal benefit curve for all the goods that might be purchased in a market and for many others that cannot be purchased.

One's income is thus an important factor in determining one's marginal benefit curves and thereby the demands for goods that has at a point in time and through time. An increase in one's personal income affects their demand—partly by increase marginal benefit curves and thereby broadening the range of goods that one is likely to purchase.

You will all—hopefully—have more income after you complete your college degree and take a job. That higher income will allow you to pay a higher price for successive units of goods that you desire—including such things as larger apartments, newer cars, more travel, more elaborate clothing, and so on. You will still have to economize—but will do so over a broader range of possibilities. (Those of you who go on to intermediate microeconomics will learn utility-maximizing models of consumer choice that show more clearly why the range of possibilities increases with income.)

What Is Income?

For the most part, income in a commercial society is what you receive for services that you sell to other individuals directly or to economic organizations that value your services. Labor services include services that you may directly provide with body, senses, mind, and education. The latter, acquired skills, are

often referred to as human capital or skilled labor by economists. Economic organizations "demand" various labor services because they help it produce other goods and services that generate profits for its owners.

Most people sell their labor services to economic organizations and realize most of their income from their wages. Thus, labor markets are one of the most important markets for inputs.

If one inherits some wealth or saves and invest over the course of one's life, the service that you sell might also include rental appartements, interests on loans that you make with your financial assets (if you have inherited some or accumulated some by saving). If one inherits, founds, or creates a firm that sell intermediate goods to firms producing final goods, your profits or share of the profits are also part of your income. And, if one owns or owns share in a final goods producing firm, your share of the profits is also part of your income.

In the last case, one is not necessarily selling productive services or inputs that contribute to the output of final goods and services, but in all the other cases, one's income can be regarded as based on sales of inputs. It is such cases that account for most of the income earned in a commercial society, and it is the one focused on in this chapter.

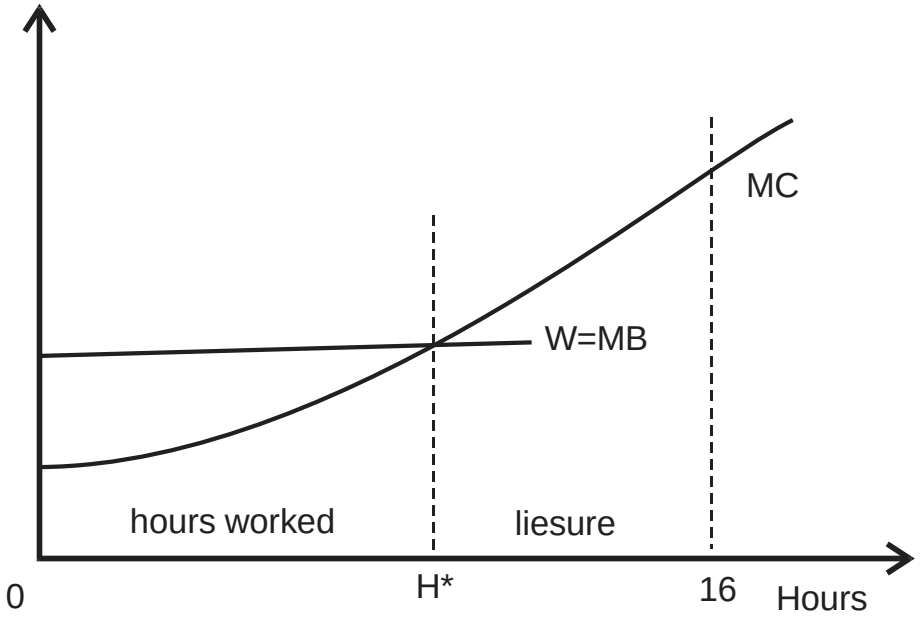
That is to say, individuals in commercial societies acquire most of their income by (1) "hiring themselves out for wages" and (2) selling other inputs into specific processes for producing the "final goods" sold to consumers in the markets focused on in the previous chapters.

In this chapter, we'll analyze both the supply and demand for labor and/or other inputs. We'll also analyze how input markets affect output markets and vice versa. It is such linkages that create much of the network of exchange and production in today's global markets.

II. Opportunity Cost and the Supply of Labor

Scarcity implies that most activities have "opportunity costs." That is to say, when one uses something (as with your time and attention) for one thing (studying economics) you have sacrifices other possible uses for your time (studying calculus or playing video games, hanging out with friends, or taking a nap). A net-benefit maximizing individual will use his or her time for the highest valued use. The second most valuable use is said to be your "opportunity cost." You sacrifice the marginal benefits associated with your second most valuable use of your time in order to realize the marginal benefits of your highest valued use of your time (at least at the "margin"—you might do a bit of both). Figure 4.1 illustrates a net benefit maximizing decision between labor and leisure.

Figurer 4.1 MC and Work Decisions



The marginal cost of labor (hours worked) is generated by referring to an individual's marginal benefit from leisure curve (which is not shown). The idea is that diminishing marginal benefits implies that the last hour of leisure is less valuable than the second to last, which is less valuable than the third to last and so on. Thus, the marginal opportunity cost of hours spent working for money (at wage rate w) is an upward sloping curve reflecting the diminishing marginal benefits (or utility) of leisure.

One works the number of hours where the marginal benefit from additional income, which is represented as a horizontal line equal to the hourly wage—although it could be argued that it should be a downward sloping line (explain why?). As a result, this individual works H* hours and uses the remainder of his or her day for leisure (depicted as 16-H* in figure 4.1).

Deriving and Individual's Supply of Labor Curve

To derive an individual's supply of labor curve, one determines how much labor he or she will supply at different wage rates. To do so, one chooses a wage rate (w) and finds the associated net-benefit maximizing number of hours worked (H*). This process is repeated until the supply curve is traced out. (Normally just a few wage rates are tried, the resulting w, H points plotted, and the rest is interpolated, as in chapter 2's derivation of individual supply and demand curves.)

Notice that if the marginal opportunity cost is upward sloping, then the individual's supply of labor curve will go through exactly the same points as that curve—although the supply curve will be a mapping from wage rates (the vertical axis) into labor hours (the horizontal axis) rather than from the horizontal axis (hours worked) into the vertical axis (marginal opportunity cost of hours worked in $/hour).

Thus, supply functions of labor tend to be upward sloping in wages just as supply curves for final goods tend to be upward sloping in prices.

The relevant wage rate is the highest in "real" terms that the individual may obtain for his or her skills and talents, where the real wage includes both the money wage and the job satisfaction associated with alternative jobs for which he or she is qualified (e.g. has the skills and talents that employers seek for specific types of jobs within their organization. Notice, that this implies that the typical person works just one type of job. It also implies—in most cases—that he or she has specialized by acquiring a specific combination of skills, talent, and other traits that employers find useful such as diligence and honesty.

Labor is far more specialized today than it was 200 years ago, so there are many more types of labor (combinations of skills, talents, and traits sought) today than there were back then. However, this specialization is not complete. If the wage rate for one's combination of skills talents and traits declines, one may look for other types of jobs that require somewhat different combination of skill, talent and traits that one also possesses, or "retool" by acquiring additional or different skills that make one eligible for other types of jobs and careers.

Deriving a Market Supply Curve for Labor

To derive the market supply for a particular type of labor (e.g. for persons with particular combinations of talents and skills), first identify all the persons with the required skills. Then repeat the process undertaken to derive supply curves in chapter 2. (1) Choose a wage rate and find out how many hours all the persons of interest are willing to work at that wage rage. (2) Add up those numbers of hours and plot the wage rate and total number of hours on a separate diagram. (3) Repeat these steps with different wage rates to trace out the market supply of the type of labor that you are interested in.

Since each individual's supply curve tends to be upward sloping, the market supply does as well.

The logic of suppling other inputs to the market's network(s) of production is fundamentally similar to that for labor, although the MC curve may be less induced by an opportunity cost and more by production, harvesting, and transport costs than tends to be the case for labor markets. (Of course, skills are often produced in ways that require money expenditures and other opportunity costs. They are rarely free. Even apprentices "pay" to acquire their skill sets by taking lower wages during their apprenticeships than they could be earning elsewhere.)

III. The Demand for Inputs

Firms demand inputs because they need them into order to produce goods in services demanded by consumers (and other firms in the case of producers of "intermediate goods"). Firm owners may have various reasons for engaging in production—they may enjoy bringing products to market, they may enjoy running their companies, they may take pride in providing consumers with useful goods that they either need or enjoy. However, all firm owners share an interest in realizing profits from their production. This provides their own livelihood (income) and also provides the firm with the resources that it needs to grow (both directly through "savings" and indirectly by generating greater lines of credit).

Thus, the simplest model of firm behavior is that firms are profit maximizers, as assumed in chapter 2. Even though many owners have other interests as well, maximizing profits is a goal shared to a large degree by all firm owners.

This implies that firms will employ the profit-maximizing combination of inputs—given their expectations about prices of their outputs, their prices for inputs, and the productivity of their inputs. It also implies that firms will not hire workers or use other inputs unless they "earn their keep," unless they produce more revenue than they cost.

Marginal Product and Marginal Revenue Product

Every input of interest adds to the output of final goods created by a firm. The increase in the firm's output of final goods (or other goods sold in markets for intermediate goods) is called the input's "**total product**." The increase in final output generated by a "one unit" increase in an input is called its "**marginal product**," which is denoted in this course as MP. Most production processes are subject to "diminishing marginal returns," which means that most inputs have marginal product curves (or marginal product function, MP(Q)) that decline with their usage.

The logic behind this assumption is similar to that used for declining marginal benefits for consumers. Each firm uses the first unit of an input for its most productive purpose, the second unit for the second most productive purpose, and the third for the third most productive purpose, and so on. This process generally implies that marginal product curves are downward sloping just like marginal benefit curves for consumers.

(There are exceptions to this rule, but we'll stick with the "normal" cases for the most part in this course, because as commonplace cases, they are the most likely to be encountered by all but the largest firms.)

Illustration: Production of Doughnuts

Table 4.1: A Production Function and MRP

Quantity of Labor	Quantity of Output	Marginal Product	Marginal Revenue Product
0	0	0	(assume P = 10)
1	20	20-0= 20	20P=20(10)=200
2	38	38-20=18	18P=18(P)=180
3	54	54-38=16	16P=16(10)=160
4	66	66-54=12	12P=12(10)=120
5	76	76-66=10	10P=10(10)=100
6	84	84-76=8	8P=8(10)=80

The first two columns of Table 4.1 characterize a simple 1-input "production function." A production function simply maps various combinations of inputs in the outputs that can be produced via a given technology. Here there is just one input, so increasing output always requires additional units of that

input. If the input is labor, the first and second columns show how much output (column 2) is produced by efficiently using 1, 2, 3, … 6 units of labor (from column 1). Although 1-input production methods are pretty rare—they allow us to show how two other ideas—marginal product and marginal revenue product---affect a firm's demand for inputs.

(One does not have to assume that there is just one input to construct such a table like Table 4.1. One can also do so by holding all other inputs constant and "adding" the input of interest to the mix of available inputs which jointly determine total output in the second column. The one input production process just makes its total, marginal product, and marginal revenue product more obvious.)

Marginal product is the change in total produce (total output) generated by a one unit change in the input of interest (here labor). Column 2 shows how this is calculated. You just take the total output associated with a particular quantity of the input (say 3, which would be 54 units of output (as with 54 dozen doughnuts) and subtract from that output the output realized without the last unit of that input (e.g. the output of 2 units of the input, 38). So, the marginal product of the third unit of labor is 54-38=16. The same process is used to calculate all of the marginal products of labor in column 3.

Firms, of course, want to produce output to sell, but what is most important is the revenue generated by that output. A particular unit of an input is worth employing only if it generates more revenue than it costs to employ (and therefore is profitable to employ). To determine how much revenue is generated by each successive unit of an input, one just multiplies its marginal product by price—this is called an input's **marginal revenue product**. The last column of Table 4.1 computes the marginal revenue product of successive units of labor. It does so by assuming that the market price of the output produced is $10/unit.

One can create a graph of a marginal revenue product curve by using inputs as the horizontal axis and marginal revenue product as the vertical axis. Simply plot each of the points characterized by columns 1 and 4 to do so.[3]

[3] Technically, these **point**s are the actual marginal revenue product "curve" if only whole units of the input can be used. To get a "curve" that you can take areas under requires more work and a bit more math. (Don't worry about this footnote too much if it seems unclear to you.)
A marginal product curve, unfortunately, **cannot** be constructed by simply connecting the points up for cases in which fractions of a unit of inputs can be used. Instead, one should construct another table with different fractions of the input being employed (which in a sense just changes the units of measure, say from hour to minutes of labor). This would give you a series of numbers for the MPR of labor in terms of minutes with a lot more points over the 6 hours of the first figure (and generally smaller numbers).
As one reduces the unit size (say to seconds) one gets more and more points, and they are closer together and begin to resemble a curve. That curve, in the limit, would be an exact marginal revenue product curve—as opposed to an exact series of points that exactly characterize marginal revenue product.
To get a perfectly accurate MRP curve, the best way is to use Calculus. If one can write down a production function mathematically as a function (for example, $Q = aL^b$), then **marginal product** is just the first derivatives of the production function with respect to the input of interest $dQ/dL = abL^{b-1}$. And, **marginal revenue product** is just P times the marginal product function ($MRP = PabL^{b-1}$).

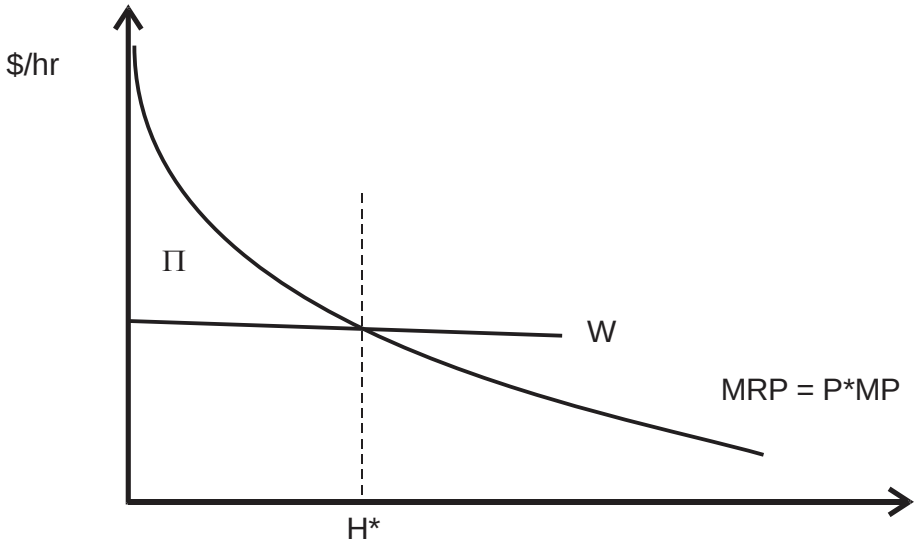
Marginal Revenue Product Curves and the Demand for Inputs

An input's productivity is determined by its marginal product curve. The higher that curve is, the more output is produced by the input of interest. As above, the marginal benefit of an input is the revenue that it directly or indirectly generates for the firm. If one multiplies marginal product by the prevailing market price of the firm's output, the result is an input's **marginal revenue product (MRP(Q) = MP(Q)*P)**.

A marginal revenue product curve describes how much a firm's revenue increases as it uses more and more of the input of interest, as for example more and more of a particular type of labor or a particular type of machine. Every useful input has both a marginal product curve and a marginal revenue product curve associated with it. (Areas under the curve correspond to the revenue generated by a particular range of inputs or all of that input. If one calculates the area under a MRP curve from 0 to L, one gets the total revenue that is produced by L units of the input of interest.)

A firm's marginal revenue product curve for an input is the basis of its demand for inputs. Figure 4.2 illustrates how MRP curve can be used to make decisions about hiring a particular type of labor. The marginal revenue product curve is the downward sloping line. The marginal cost of labor is its wage rate, which is determined by the market equilibrium for this particular type of labor. **Every unit of labor that has a marginal revenue product larger than its marginal cost is profitable for the firm and so will be employed.** In this case, the quantity of labor employed (in man hours) sets its marginal cost (here the equal to the market wage rate) equal to its marginal revenue from the input (here equal to the marginal revenue product of the input).

Figurer 4.2 MRP and Hiring Decisions



Deriving a Firm's Demand for Labor

To derive a firm's demand curve for an input such as a particular type of labor. First, determine the firm's marginal revenue product curve for that input. Second pick a price or wage rate for that input.

Third, find the profit maximizing quantity of labor to hire at the wage rate or price. Third, plot the wage rate or price and the quantity of labor or another input that is employed at that wage or price. Fourth, repeat with another wage rate or price until the demand curve for the input of interest is traced out.

Notice that i**f the MRP curve is downward sloping, the firms demand curve for the input of interest (here labor) will include all the points on the firm's marginal revenue product curve**, although as before the direction of the function is opposite that of the marginal revenue product curve. The demand curve goes from input prices into quantities employed whereas the marginal revenue product curve goes from employment levels (quantities) into revenue per unit of the input of interest.

Notice also, that some inputs may be too expensive to employ, ether because market prices for the input are too high or because its marginal product is too low or because prices of the final good to be sold are too low. **Only units of labor and other inputs that are profitable for the firm are employed.**

Employees and other inputs have to "earn their keep," which is to say they must be profitable to employ otherwise the firm will go out of business as often happens for new firms.

Deriving a Market Demand for an Input from Firm Demand Curves

Like other market demand curves, the market demand for an input is the sum of many different demand curves—here the demand curves of all firms that make use of the input of interest.

The market demand for a particular type of labor or particular type of input is determined more or less in the same way as other market demand curves. Geometrically one does this by (1) choosing a price for the input of interest, (2) finding the quantities of that input that would be demanded by each firm from its input demand curve, (3) adding those quantities up, (4) plotting the price and the total quantity demanded (employed), then (5) repeating with other prices until the market demand curve is traced out.

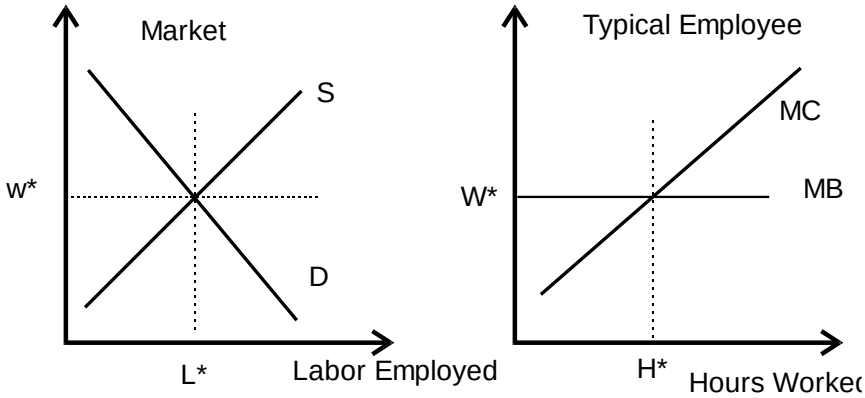
Since each firm's demand for inputs tends to be downward sloping, market demand curves for inputs also are downward sloping—just as demand for consumer goods are—but for some different reasons. It is because, at the margin, there are diminishing returns from using every input at the level of the firm.

Demands for input change as market prices for final goods change and whenever technology changes, and in some cases because of regulations that firms have to satisfy. All such changes affect MRP, either through effects on P (sales price of the firm's output) or through effects on MP (marginal product).

IV. Equilibrium Prices in Input Markets

The equilibrium employment rates of inputs sold in reasonably competitive markets can be characterized with the supply and demand curves developed above. Figure 4.3 illustrates such an equilibrium for a labor market.

Figure 4.3 Equilibrium Wage Rates



From our derivations, we know that the market demand for each type of labor is downward sloping reflecting the downward marginal revenue product curves associated with inputs subject to diminishing marginal returns. We also know that the supply curve for each type of labor is upward sloping reflecting the upward sloping marginal opportunity cost of hours spent hiring oneself out for wages (and the opportunity cost of acquiring various types of skills valued by markets).

Thus, the demand and supply curves have the usual shapes. The areas under those curves can be used to characterize the net benefits of labor suppliers and purchasers of labor in the usual way. Both curves are again based on the net benefits that individual suppliers of inputs (such as labor) and individual purchasers of inputs (such as employees) realize at the market equilibrium. For a labor market, the net benefit realized by workers is the "profit" area of the diagram. Workers supply labor. The profits realized by employers (the buyers) is the usual consumer surplus areas—although that area now represents profits. Firms are purchasers of labor.

As in final goods markets, **prices affect the behavior of both sides of every input market**. In labor markets, the wage that emerges affects both each worker's supply decisions and also each firm's employment decisions.

V. Simultaneous Equilibria in Output and Input Markets

We are now in a position to illustrate how equilibrium market prices coordinate the decisions of firms, input sellers, and consumers in final goods markets simultaneously. We are also in position to analyze how changes in market conditions in either the input or final goods markets affect the others.

To fully characterize even a relatively simple equilibrium between the input and output markets requires six diagrams. One market diagram for each level, plus two diagrams representing supplier and demander interests at each level. Figures 4.4a and 4.4b below illustrate such an equilibrium. **Notice that every supplier and demander is in equilibrium at the same time that the overall markets are. They have all maximized their own net benefits given the prices that the face.**

The output price affects the demand for inputs by affecting the marginal revenue product curve for each firm in the relevant input market, and therefore affects the location of the input demand curve in the

relevant input market (here a labor market). The price of the input, here wage rates, affects the marginal production cost of all firms in the relevant output market(s) and thereby the location of the market supply curve for the output of interest (and every other output market in which such inputs are used).

Thus, any change in conditions that affects input prices or final goods prices will tend to cause adjustments by consumers, firms, and input providers in both markets. In this way, input and output prices do a lot of the work to assure that decisions of consumers, firms, and input providers are consistent with one another—they are all the ones that produce the equilibrium outcomes (supply equal demand) in all the markets connected.

Comparative Statics in Market Systems

Comparative statics in these systems are often complicated to draw, because so many effects have to be taken into account. So, we'll begin by analyzing the effects of a relatively simple "shock" to the equilibrium depicted in Figures 4.4a and 4.4b.

Suppose that the demand for the final good increases because consumer income increases or because the price of a substitute increases. Such changes cause the price of the final good to rise, which in turn causes the marginal revenue product curves for each of the firms in that market to increase, which in turn increases the demand for all inputs, here labor, out which tends to increase input prices—at least a bit—which in turn increases the quantity of labor and other inputs supplied to firms.

These cost effects in turn affect the supply curve in the final goods market by raising marginal costs a bit, which is an effect that we've not considered before, because we've assumed that the final goods producers are a tiny part of the demand for inputs. In such cases, supply is unaffected by shifts and demand because input prices are hardly affected if at all. However, if the producers of the final goods are important users of some of their inputs, then price affects may be significant and supply curves will also be affected by shifts in demand because of indirect effects on input prices.

If the output market is "small" such effects may be so small that they can be ignored, but if the industry of interest is a relatively large employer, then the effect of an increase in demand for particular inputs on input prices should be taken into account.

(Graphically depicting these effects is left as an exercise for the students and other readers.)

Figure 4.4a Equilibrium in an Input Market

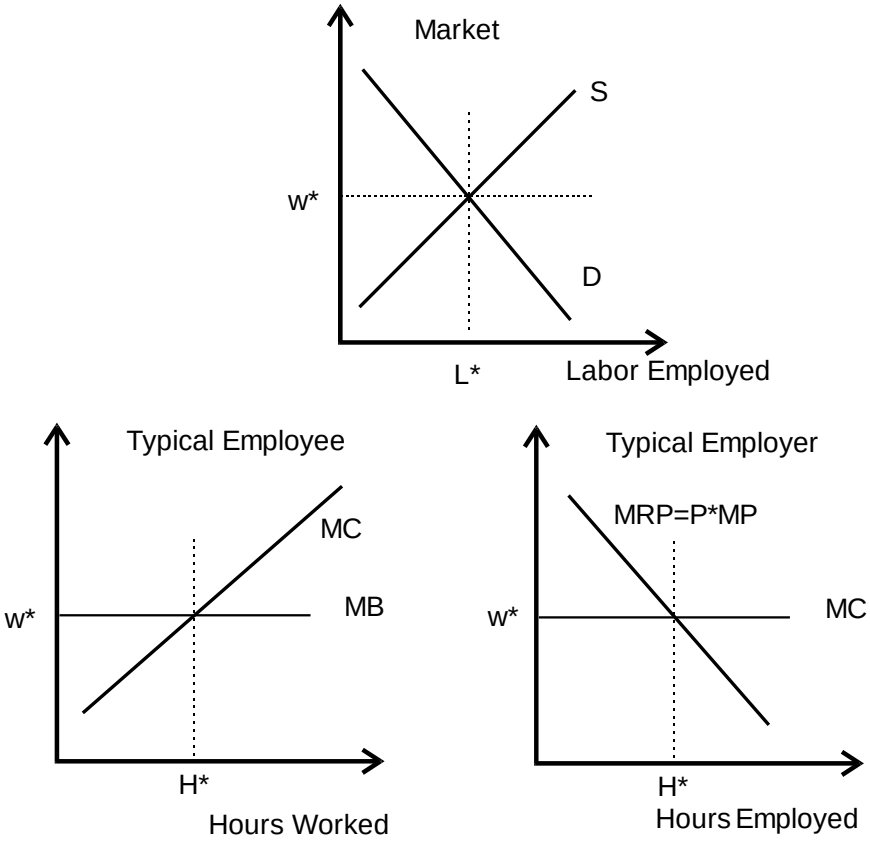
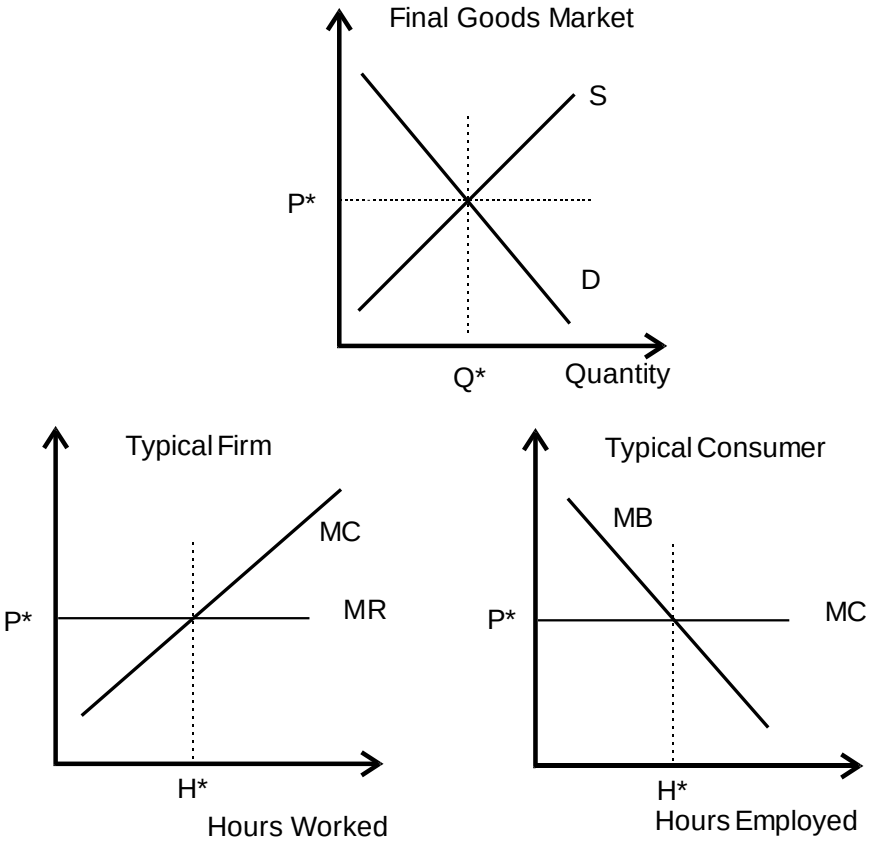


Figure 4.4b Equilibrium in a Final Goods Market



Chapter 5: Economic Effects of Entry Barriers and Taxes

I. Market Equilibrium Revisited

A Short Overview of the first 4 Chapters

In Chapter 2 , we demonstrated that if people behave as "rational" net benefit maximizers, then demand curves tend to slope downwards and supply curves tend to slope upwards, and prices are a means through which all the millions of decisions between consumers and producers can be coordinated.

A firm's marginal cost curve is affected by input costs, technology and the time available to make adjustments (short and run). A consumer's marginal benefit curves change when income, the prices of substitutes, or complements change—or when tastes or circumstances change such as weather. Change in any of these variables (parameters of a firm or consumer's choice setting) shifts the supply or demand curve of typical firms and consumers.

These effects in turn tend to effect **market supply and demand curves therefore market outcomes** whenever most consumers or most firms are affected by the same changes (prices, outputs, employment, income, etc.).

Note that these properties of markets were all **deduced from just a few assumptions about firms and consumers. They were derived** from the assumptions that consumers and firms are rational in the sense that they have goals (net benefits, utility, profits ...) that try to advance, and that they adopt means that are more or less "cost effective" to advance their goals. This allows both firms and consumers to be modelled as if they routinely maximize their net benefits from market activities.

A series of intuitively plausible illustrations were used to show that the conclusions reached help explain a good deal about the markets that we observed. These case studies show that the implications of the models are themselves reasonable. Other statistical evidence is also largely consistent with these models, although such tests are beyond the scope of this course. (For those that choose to go on in economics, such tests are covered in a course called econometrics, which teaches the statistical methods used by economists to test these and other models.)

For the most part, this course, and microeconomic theory in general, is a deductive exercise based on rational choice models. And thus "deduction" is the "main dish" in this course.

Not much of what we have covered was "intuitively obvious," **except after the fact**. Once the tools are understood, you can see how "natural" many of the conclusions are and how neatly all the pieces (decisions of suppliers, demanders, and input providers) all fit together.

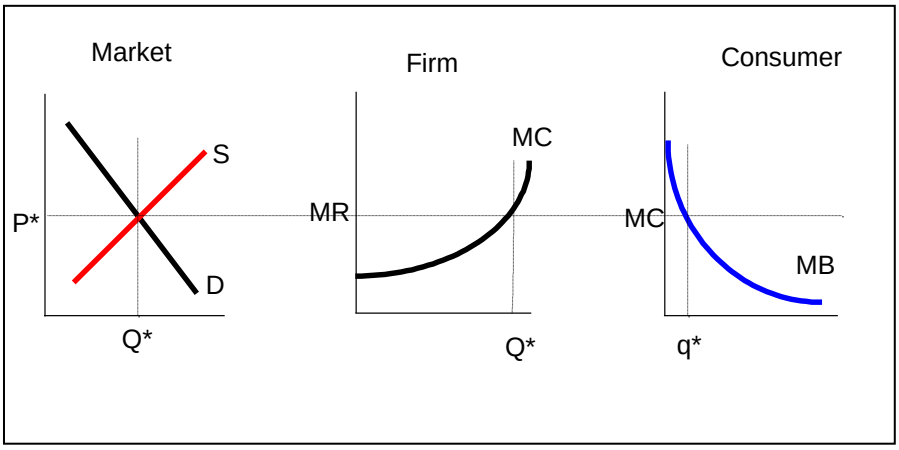
However, these models and conclusions took nearly a century to be worked out. And, they were not really worked out until after commerce became central to life in industrialized societies.

What we are doing in this course is developing the logical foundations for a very complete model of life in a commercial society. It will help understand much about specific markets, the way market networks operate, and the reason why contemporary markets are efficient systems for satisfying human desires for material comforts, entertainment, and leisure.

Markets and Efficiency

After we modeled market equilibrium, we showed that in most "ordinary" circumstances markets tend to produce outputs that maximize social net benefits. These results followed from our ability to use demand curve to represent consumer marginal benefits and to use supply curves as industry marginal costs.

In cases where that is possible, the price at which demand equal supply produces the quantity where marginal social benefits equal marginal social costs. That quantity is the social net benefit maximizing quantity.



There are a variety of assumptions that we used to derive demand and supply curves, but the most important of these for our "efficiency" or "welfare" conclusions was the assumption that all firms and all consumers are "price takers."

Each side of the market has been assumed to simply adjust to the prevailing prices and maximizes their own net benefits (consumer surplus or profits). It is this "price taking" behavior that we used to derive supply and demand. It is this "price taking" behavior that generates the particular supply and demand equilibrium that was efficient.

Of course, if no one can control prices, it's not clear where prices come from--but this has been left out of the theory so far. It's only been explained insofar as inventories (surpluses and surfeits) tend to cause prices to change (fall or rise) until an equilibrium is reached.

The chapter 6 other less competitive market environments will be analyzed. That subfield of economics is normally called "Industrial Organization." It analyzes in somewhat more detail the characteristics do firms and consumers need to have if they are to behave more or less as price takers? And it explores what happens if those characteristics are lacking?

II. Public Policies and "Violations" of the "Competition Assumptions"

In some cases, less the perfectly competitive markets emerge because of market forces such as economies of scale that are large enough that only a handful of efficiently sized firms can survive. In other cases, markets are less competitive than they could be as a result of public policies and regulations adopted by the governments.

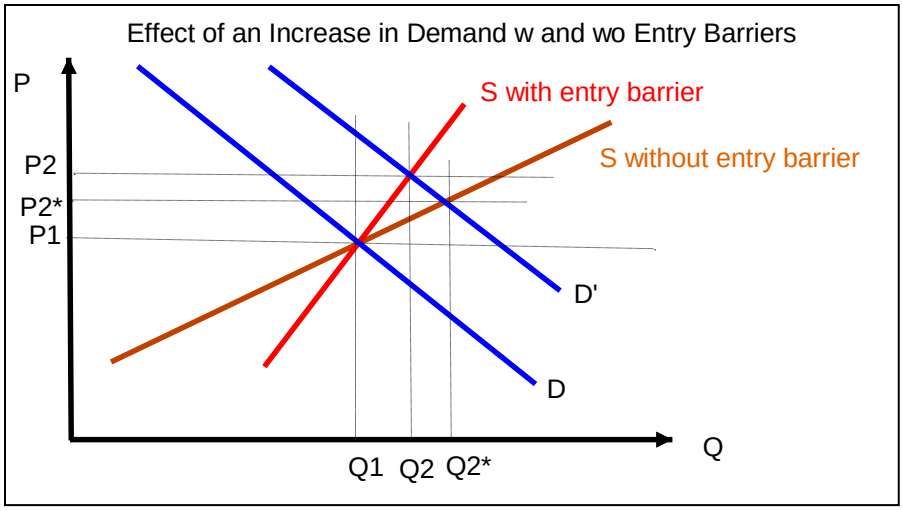
How Entry Barriers Affect Market Equilibria

Entry barriers reduce long run supplies in Ricardian markets by blocking the entry of high-cost producers and new low-cost producers that might have been formed through technological innovation by those initially outside the equilibrium. Entry barriers also prevent the Marshallian type of exit and entry from taking place, which tends to reduce supplies when demand is growing to levels below competitive levels, or to increase supplies above competitive levels if exit is blocked.

There are many types and instances of barriers to entry. Regulations, for example, may reduce the number of firms (number of suppliers) to levels below the Marshallian equilibrium. Doctor salaries (and other income) remain very high, well above other that of other graduate degrees) because there are limits on the number of doctors that can be "produced" (e.g. Graduate from medical schools) every year. [The US has about 20% fewer doctors per capita than other Western countries.]

Other regulations such as zoning tend to reduce supplies of low cost (small) housing units in many cities, making prices higher than it would otherwise have been. Barriers to entry also exist in various foreign trade markets, which often limit imports of goods and services (as in Japan with respect to rice and beef, and in the US with respect to sugar).

- [Who benefits from such regulations?]
- [Draw (i) a Marshallian and (ii) a Ricardian market with a barrier to entry and analyze the effects of the barrier on profits and consumer surplus.]
- [Is the AACSB an organization that creates artificial barriers to entry? Why or why not?]



Note that as demand increases price rise more in the market with entry barriers than would have been the case without the barrier to trade. This causes profits to be higher, but consumer surplus to be lower. It also tends to reduce social net benefits to levels below those associated with the "open" market.

- [As an exercise, redraw the above diagram and label all the net benefit areas.]

Effects of transactions costs (imperfect information and search costs)

Transactions and information costs imply that firms will not all be forced to sell at the same price, because (a) consumers may not know where the lowest priced source of a good or service is, and (b) because consumers will take account of the cost of waiting in line at the lowest cost sources of the goods. [In effect their true price is the "posted price" plus their search and transactions costs.]

Positive search costs imply that prices may vary somewhat (within a fairly narrow band) according to the degree of information and waiting costs.

Reconsider the example of two firms selling at two different prices:

- Puzzle: Suppose two firms sell identical products and firm A sets a price that is 10% higher than firm B. What happens?
- In this case, as customers go to firm B, lines form, which increases the effective price of shopping at B. Some will return to A and pay a higher money price in order to save time.
- (It could be argued that A and B now sell different products.)
- Firm A no longer has to its price or go out of business, but if its price were really high, it is still the case the consumers might go to B even given its waiting costs.

- **Waiting costs thus imply that some price variation will exist**, but that prices will stay within relatively narrow bands--based on those waiting costs. If the price is too far above or below the average price, it will pay to change suppliers, even with waiting costs.
- [Puzzles: what effect on the distribution of prices do you think that internet shopping has had? Does the same effect occur in Marshallian and Ricardian Markets?]

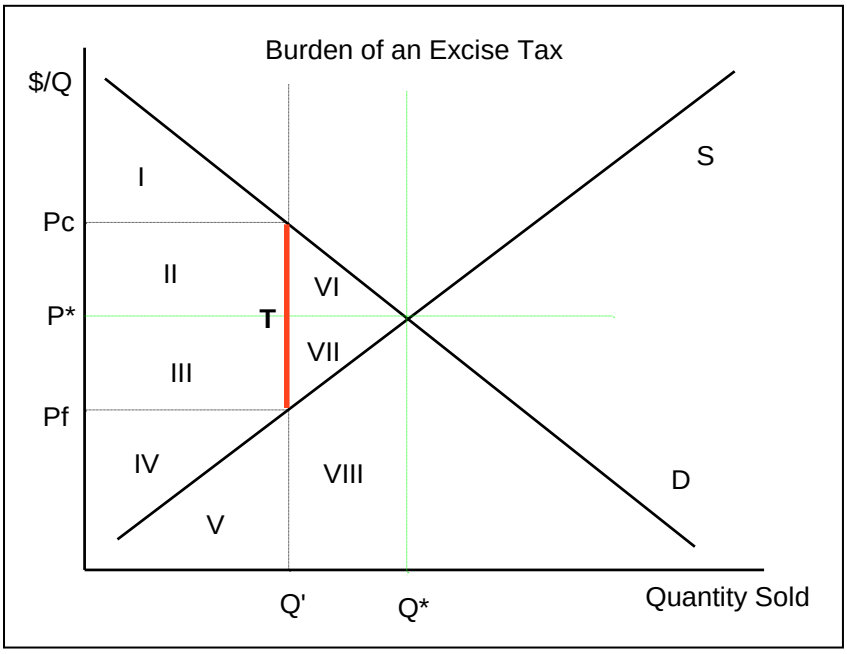
Effects of small numbers of firms or consumers

As the number of firms or consumers falls from dozens or hundreds to just a handful, it become more likely that a single firm's output decision or a single consumer's purchase decision will have a clear, observable, effect on market supply or demand. In those cases, it is not likely that firms or consumers will behave as price takers. They will understand that their supply or purchase decisions will affect market prices.

When the number of firms or consumers becomes so small that pricing taking behavior becomes implausible, then we shift from "competitive" market models to other models of market behavior such as the monopoly (single firm), duopoly (two firms), and monopolistic competition (lots of firms selling similar but not identical products) models.

In those cases, markets still tend to clear, but by conscious decisions by firms, rather than as an unintended consequence of inventory adjustments. ***We take up other "market structures" in chapter 6.***

Taxation: The Economic Effects of an Excise Tax e



Suppose that a market is initially in an equilibrium without taxes, so that demand equal supply at P*. In this case, there is no "tax wedge" between the price paid by consumers, Pc, is the same as that received by firms, Pf; so, Pf=Pc=P*.

Now, suppose that an excise tax of **T** dollars per unit is imposed on each unit of the good sold in this market, as for example is done with tire sales in the US. This changes the transaction from ones involving two parties (firms and consumers) into one's involving three parties (firms, consumers, and the government).

After the tax is imposed, P* is no longer the market clearing price. To see this, note that if **T** is simply added to P* by firms, consumers will purchase too little at their new price (Pc = P* + T) to match supply, which would remain at Q*.

On the other hand, if firms simply "ate" the tax, they would provide too little of the good to meet demand (at their after-tax price of Pf = P* - T). Supply would fall and demand would remain at Q* if Pc = P* and Ps = P* - T.

To clear the market, thus, firms have to receive less than P* per item sold, and consumers have to pay more than P*.

- At the new equilibrium output, the demand curve will be exactly T dollars above the supply curve, and Qd(Pf + T) = Qs(Pf).

This equilibrium output is shown in the diagram. At Q', supply equals demand, if the price paid by consumers is exactly **T** dollars higher than the amount firms receive (Pf = Pc - T).

- Q' units of the good are sold, with Q'<Q*.
- At this equilibrium, there is a sense in which the tax has simply been passed onto consumers, because Pc = Pf + T.

However, there is another sense in which the **burden of taxation is shared** by firms and consumers, because both consumer surplus and profits have been diminished by the tax!

- **Consumer Surplus falls** from area I + II + VI (before the tax at Q*) to just area I after the tax is imposed and output falls to Q'.
- Similarly, **Profit falls** from III + IV+ VII (before the tax at Q*) to area IV (after the tax at Q').
- The burden on consumers is II + VI, and that on firms is III + VII.

Note that this distribution of the loss of consumer and firm net benefits occurs regardless of who actually writes the check to the state or federal treasury.

- Price movements ultimately determine the actual division of burden between firms and consumers.
- If firms send in the check, their effective "payment" is reduced by the increase in price paid by consumers.
- If consumers write out the checks, their effective "payment" is reduced by the price decrease absorbed by firms.

The amount of revenue raised by the tax is **T***Q'.

- Q' units are sold and each pays a tax of T dollars.
- The total tax revenue, TQ', can be represented in the diagram area II + III in the diagram.
- (Note that II + III is the area of a rectangle T tall and Q' wide.)

Notice that the tax revenue is smaller than the "surplus" lost by taxpayers (firms and consumers in the affected market).

- The reduced profit plus the reduced consumer surplus equals {II + VI} + {III + VII}.
- The total burden of this tax is VI + VII larger than the tax revenue.
- This area of "excess burden" is sometimes referred to as the deadweight loss of an excise tax.

Both the extent of the deadweight loss and the distribution of the tax burden vary with the slopes of the supply and demand curves.

J. Generally, more of the burden falls on the side of the market with the least price sensitive curves.

- If the demand curve is less elastic than the supply curve, more of the burden falls on consumers than on firms.

- In the extreme case in which market demand is completely inelastic or the industry supply curve is completely elastic, all of the burden falls on consumers!
- On the other hand, if the demand curve is very elastic, because good substitutes exist, or the supply curve is relatively inelastic then more of the burden tends to fall on the firm.
- In the extreme case in which the market supply of the product of interest is completely inelastic or consumer demand is perfectly elastic, all of the burden falls on suppliers.

The excess burden of a tax tends to increase with the price sensitivity (slope or elasticity) of the demand and supply curves.

- Recall that both supply and demand tend to be **more elastic in the long run than in the short run**, consequently, the excess burden of taxation tends to be larger in the long run than in the short run.

III. Applications, Puzzles, and Policy Analysis:

Contrast the effects of a regulation that creates a barrier to entry (or exit) with one that imposes a cost increasing production technology (as often are associated with environmental regulations).

Many professions have licensing requirements of various kinds. All of these tend to create entry costs. Do they all have the same effects on supplier net income (profits) in the long run? Why or why not?

Countries often have rules and regulation that make it more difficult for foreign providers of goods and services to enter a nation's markets. Show how such regulations affect long run supply in Ricardian and Marshallian markets. It can be argued that this "supply" effect is often greater in Ricardian than in Marshallian markets--explain why.

To what extent are college education requirements simply an entry barrier?

Are college degrees simply a form of information that reduces search costs?

In addition to barriers to entry, regulations often reduce the range of prices and price adjustments that can take place in a particular market.

How do price controls (ceilings and floors) affect market equilibria in the short and long run?

[See your lecture notes for other illustrating diagrams and realistic examples than covered in the web notes.]

Chapter 6: Industrial Organization: Market Structure and Market Outcomes

I. Industrial Organization

There is a sub area of microeconomics called "industrial organization" that studies why some markets behave differently than others. In competitive markets, there are lots of buyers and sellers, and prices tend to equal marginal production costs. In other markets, prices may be consistently above marginal cost, and their associated producers may innovate at different rates, they may have less entry and exit, and they may employ very specialized types of labor and capital goods.

Here one might consider the "old fashioned" (pre internet) industries of steel, automobiles, electronics, retail sales, medical services, military equipment, etc. Or some of the newer computer-based firms such as Microsoft, Google, and Facebook.

One of the main differences among industries (and regions) is the number of firms and/or consumers. Some industries have far fewer firms (are more "concentrated") than others. Sometimes, purchasers (customers) are less numerous (more concentrated) than others or differ in important respects. For example, military products are sold principally to governments.

In this chapter, we'll take a look at three or four types of market structures that are less competitive than the market types that we've previously examined. These are market structures in which the "price taking" characteristic of firms and/or consumers is unlikely to hold.

At this point in the course, we have discussed both Marshallian and Ricardian theories of competitive markets where the primary difference was in assumptions about firms. Marshall assumes that they are all identical, Ricardo assumes that they are all a bit different although they all produce the same products. Another difference is Marshall's emphasis on entry and exit to clear markets in the long run, whereas this process is less important in the Ricardian version of very competitive markets. In Marshall's long run equilibrium all firms earn just the "ordinary" rate of return on their investments. In Ricardo's version, profits will vary among firms because of differences in their production costs associated with differences in the talents of their owners, location, and workforce.

The main focus of this chapter is on less competitive markets in which firms can no longer be modeled as price takers.

Perhaps surprising, it is the least competitive market types reviewed in this chapter that are easiest to model with our neoclassical tools are monopoly (one seller), monopsony (one buyer), and monopolistic competition.

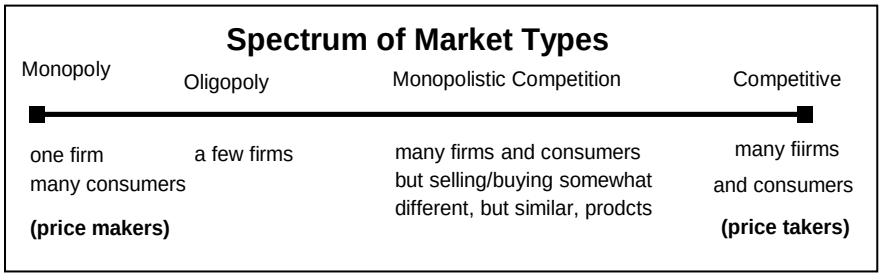
And this is where we'll focus most of our examination of market structures.

Other intermediate cases require other tools to analyze and have less "sharp" predictions about prices, outputs, and efficiency. Those require game theory, and we'll give those less attention—partly because the results are less clear cut.

The number of firms and consumers in an industry is partly endogenous. That is to say, it is often a consequence of demand, the technologies of production, network economies, opportunities for risk sharing, and informational limits. It is also partly a matter of firm strategies.

The degree of market concentration is also affected by economic regulation (patents, licensing, and antitrust laws), and efforts by firms to coordinate their pricing and output decisions (cartels).

A useful **rule of thumb** is that as the **number of rivals** shrinks, the **degree of price and quality competition falls**. As a consequence, prices no longer equal marginal production costs as they do in competitive markets. As numbers of rivals fall, it also becomes easier for firms to form **cartels**.



Of course, in practice, more than the number of firms is important, but as a first approximation, the number of firms can be used to represent the degree of monopoly power held by suppliers.

The rule of thumb is "The more firms, the less price setting ability an individual firm tends to have."

However, it should be acknowledged that there are exceptions to this rule. Sometimes even markets with just a few firms generate very competitive markets. The types of entrepreneurs matter—if each entrepreneur attempts to maximize market share by providing lower prices, higher quality, or better service, then the result may resemble that of competitive markets (with prices approximately equal to marginal cost—e.g. no monopoly markup).

II. Monopoly

The easiest type of market to model is one in which there is a single firm that produces a unique product or service that sells to many consumers of its goods or services. In this case, a monopoly firm

completely controls the supply of its product and so can price and produce and price its output anyway that it wants.

However, a monopolist still wants to maximize its profits. Thus, a monopolistic firm will not sell its output at the highest possible price or produce the maximum or minimum quantities that it can, but it will produce the quantity and sell it at the price that maximizes profits.

The geometry of maximizing profits looks a bit different for a monopolist than for a competitive firm, but a monopolist will still produce the output that sets marginal cost equal to marginal revenue. However, its marginal revenue is no longer a horizontal line determined by market prices. Instead, it reflects the demand for its product or services. Generally, a monopolist's marginal revenue curve is a downward sloping curve or line that reflects market demand for its products—but lies below the demand curve over most of its range.

If the monopolist's demand curve is a straight line (is linear) then its marginal revenue curve is also a straight line and is exactly halfway between the demand curve and the vertical axis.

This result is not intuitive—but rather emerges from a bit of calculus, which is undertaken below for interested students.

For students that have not had calculus or are not interested in the math behind the shape and location of a monopolist's MR curve, skip the next section.

The Calculus of Marginal Revenue Curves (optional)

For those who know a little calculus, **here is how one derives a marginal revenue curve** for the simple (non-discriminating) monopolist that faces a linear (straight line) demand curve.

- Suppose that demand is **Q = a – bP. This function can be rewritten to describe how a monopolist's selling price changes with output:**

  P = a/b - Q/b.

  This tells the monopolist how its price changes as it produces more output.

- The firm's total revenue, R, is PQ, which in this case can be written as

  $$R = PQ = \mathbf{(a/b - Q/b)Q} = aQ/b - Q^2/b$$

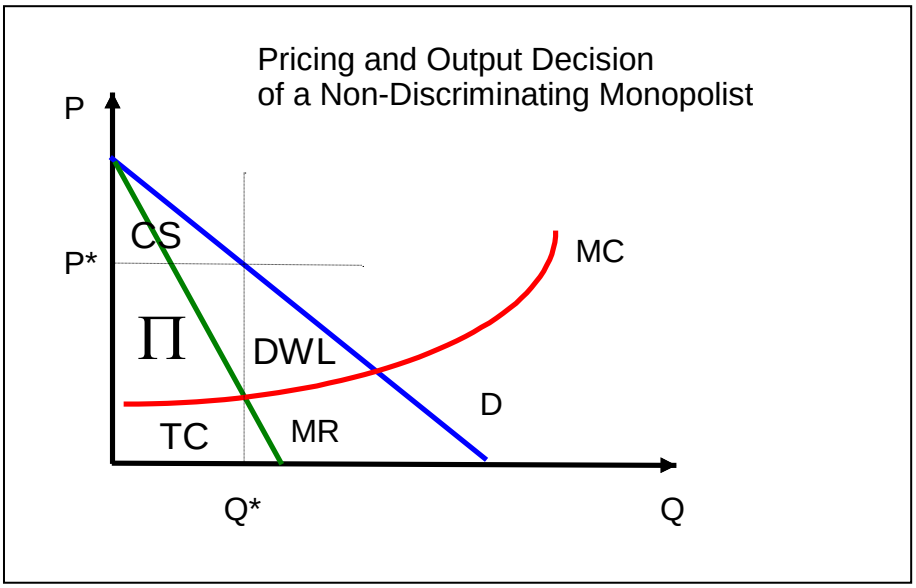
  using the pricing equation that we derived from the monopolist's demand curve.

- Differentiating R with respect to Q gives us the firm's marginal revenue function or curve, which is **dR/dQ = a/b - 2Q/b**.
- Note that this is a straight line. It starts at the same point on the vertical axis (a/b) but falls twice as fast as the demand curve we started with, -2Q/b instead of -Q/b.

- So, the marginal revenue curve is exactly halfway between the demand curve and the vertical axis.

The figure below illustrates the geometry of the profit-maximizing output and price for a monopolist.

- Note that the monopolist still produces the output where MC=MR, because it is this output that maximizes profits.
- However, its price is not equal to marginal revenue at that quantity. Instead, the price at which Q* units can be sold is found by looking at the demand curve at Q*.
- That price is greater than its marginal revenue.



- The monopolist can sell all Q* units of its product at price P*, which it the price implied by the demand curve.
- At a price of P*, consumers will want to purchase Q* units.
- The areas for consume surplus and profit can be calculated in the usual way.
- Consumer surplus is smaller than in competitive markets and profits are larger, because the selling price is above marginal cost at Q* rather than equal to it.
- Deadweight loss occurs because there are unrealized gains to trade at the monopolistic equilibrium. Thus, social net benefits are smaller than they would have been in an equivalent competitive market. (This plays an important role in economic support for anti-trust policy, which is taken up later in this chapter.)
- Nonetheless, the quantity supplied equals the quantity demanded in equilibrium, as in competitive markets.

Overview of differences between competitive and monopolistic markets.

- For example, prices no longer equal a firm's marginal cost of production at Q*. Nor is price equal to the firm's marginal revenue.
- Monopolists charge a price higher than marginal cost, because such prices increase their profits. The difference between marginal cost (at Q*) and the selling price is called the **monopoly mark up**.
- The output level chosen no longer maximizes social net benefits.
- There is now a dead weight loss. Social net benefits are maximized where the MC curve (SMC) crosses the demand curve (SMB), but a monopolist's output is a somewhat less than that.
- Consumer surplus is smaller, and profits are larger than they would have been in a competitive market with a price equal to the marginal cost at the point where the MC curve crosses the demand curve (at Q**).
- As an exercise, draw several monopoly diagrams with different slopes for the demand curve. Find the implied MR curve, output level, and price. Then label the consumer surplus and profits associated with the Monopolist's output and pricing decision.
- Next, analyze the social net benefits that could have been realized at the point where SMB = SMC. (Hint: assume that Demand is SMB and the firm's MC curve is SMC.)
- Discuss: is the area labelled DWL Would it be better labelled as "unrealized gains to trade?"

III. (Optional) Monopoly Markets with Price Discrimination

The above model and all of our previous models have assumed that all consumers pay the same price for their goods and services: the same price for every unit of the service sold.

However, firms can sometimes charge different prices to different consumers and/or different prices for different units of output sold to the same customer, profits can often be increased by doing so.

- Selling the same product at different prices is called **price discrimination**.

A **perfectly discriminating monopolist** would sell every single unit of its output at the highest price possible—which is to say at the highest price that a consumer is willing to pay.

- If it can actually do this, this implies that the demand curve is its marginal revenue curve.
- Recall that the market demand curve can be thought of as the marginal benefit curve for all consumers in the market; marginal benefit curves plot the highest prices that consumers are willing to pay for a particular unit of a good; so, the highest price that can be gotten for each successive unit of the good to be sold is captured by the demand curve.

- (In the non-discriminating case, the demand curve is the average revenue curve rather than the marginal revenue curve. [Explain why.].)
- If a monopolist can perfectly discriminate, the firm will sell the output level where its MR curve (now the demand curve) crosses its MC curve.
- Note that this is the social net benefit maximizing quantity.
- However, notice also, **that consumers gain no consumer surplus in this extreme case**.
- In this limiting case, all the net benefits are captured by the perfectly discriminating monopolist.
- Fortunately—at least to this point—such an equilibrium is informationally impossible. The monopolist would have to know the reservations prices of each of its potential customers.

IV. Natural Monopolies

Monopoly markets can occur for many reasons, including innovation, distance, patent rights, regulation, and luck.

What economists call "natural monopolies" occur when the market is too small to support more than one efficiently sized firm. This may occur, for example, when MC slopes downward over the range of interest (out to beyond the point where MC and the demand curve cross).

In Ricardian markets, entry stops when no firm exists that can profitably enter this market. This condition is pretty similar to that of entry in the Marshallian model, but the Ricardian case allows for variation in the types of firms in a market or industry. Generally, in Ricardian markets, entry requires a new firm to have a technological (cost) advantage over the least profitable firm already in the market.

If a firm's mc is sufficiently downward sloping, the market would tend toward natural monopoly from the Ricardian perspective, and also from the Marshallian perspective. Such markets will support just one efficient firm.

These sorts of monopolies are fairly common in many rural areas where only a single gas station, grocery store, church, or pub can be supported by local demand.

At the state, national, or world level, natural monopolies are less common, but a few exist. For example, it can be argued that Microsoft's operating system and Google's search engine come close to those conditions for the U.S. and for much of the world.

- As an exercise draw a monopoly price and output diagram for a natural monopolist. (Remember that MC is downward sloping over the entire range of the market demand for natural monopolists.)
- Note prices, outputs, profits, and deadweight loss.
- Are such monopolists always profitable?

- In a separate diagram, show that a natural monopolist would run at a loss if it priced its output at marginal cost.
- Think about how a monopolist might price its services if it was worried about entry. It is quite likely that a monopolist would lower its price if it was concerned about entry, which would reduce the DWL associate with such monopolies.

Natural monopolies are often said to exist in electricity and telephone (land line) services. To increase consumer surplus, such local monopolists are often **regulated** by state utility commissions, who set prices that can be charged for electricity, which has the effect of making monopolists price takers.

- As an exercise, show how price regulations may increase outputs and consumer surplus in monopoly markets, although they tend to reduce profits. Note that this case is different from price regulation in competitive markets.
- (George Stigler, the winner the 1982 Nobel prize in economics, argued that **regulators tend to be "captured"** by those who they regulate. In that case, regulation would not change the natural monopoly outcome. Explain why this could be the case.)

V. Artificial Monopolies: Cartels and Regulatory Monopolies

Of course, not all monopolists are "natural." A group of firms may organize a "pricing club," cartel, or trust and try to coordinate pricing and output decisions, rather than compete with each other for customers. OPEC is an example of an international cartel of producers of petroleum (oil).

- **A perfectly organized cartel** would function as if were a monopolist and its members would coordinate their production decisions to obtain the monopoly price for the output sold.
- However, such cartels are difficult to organize because firms have incentives to "cheat" on their cartel agreements by producing more than their allowed output and/or by trying to sell their output at a price below that set by the cartel.
- Explain why that might be profitable for a cartel member.

Another artificial source of monopoly is **patent laws**. Patent law gives inventors a temporary monopoly (14 years) as a reward for inventing something of value. This, of course, tends to produce a deadweight loss in the short run, but the monopoly profits spur inventions and other innovations, which generate dynamic benefits that are greater than the static costs. (The effect of innovation on economic growth is taken up in chapter 7.)

- Whether this is always true is sometimes debated. A sufficiently broad patent may block a good deal of independent innovation that otherwise would have occurred. It is possible that patents have different effects in different kinds of markets.

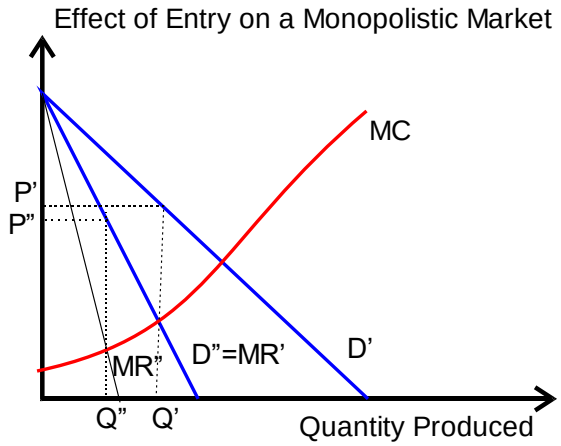
- In some markets, patents spur innovation, but in others it simply tends to produce lucky monopolies.
- Discuss why this could be true. (This may be easier to do after the next chapter is covered.)

VI. Can there be monopolists in the long run?

Both the Ricardian and Marshallian models of competitive markets implies that firms will try to enter and compete with monopolists--unless the market is a natural monopoly (because of economies of scale) or is protected by a governmental regulation (patent). Except in those cases, monopoly markets tend to be temporary phenomena.

The effect of entry is to gradually alter the market structure towards the competitive ones studied in the first half of the course.

As firms selling similar (but not necessarily identical products) enter the market the demand for this general type of product is shared by more and more firms. Demand curves may become flatter, or the original demand curve may simply be divided up among more firms. In either case, profits and prices tend to fall as entry takes place.



The figure above illustrates the case where the demand curve is simply divided up. If the original market structure was monopolistic, the new demand curve facing the original monopolist was its former MR curve (e.g. the new demand curve is halfway to the vertical axis from the original demand curve, because its new rival has taken half of the market (by assumption).

This implies a new MR curve halfway between the new demand curve and the vertical axis, MR". Which in turn (given an upward sloping MC curve) implies a lower price.

The lower price implies greater consumer surplus for the entire market and somewhat lower profits for the industry as a whole. As entry continues, a temporary monopoly may be replaced with a competitive market, if the market is large enough to support a large number of efficient firms.

- Labeling the areas for consumer surplus and profit is left as an exercise. It is essentially the same as in the original monopoly diagram for each of the two demand curves, D' and D", and marginal revenue curves, MR' and MR".

VII. (Optional) Monopoly Buyers: Monopsonists

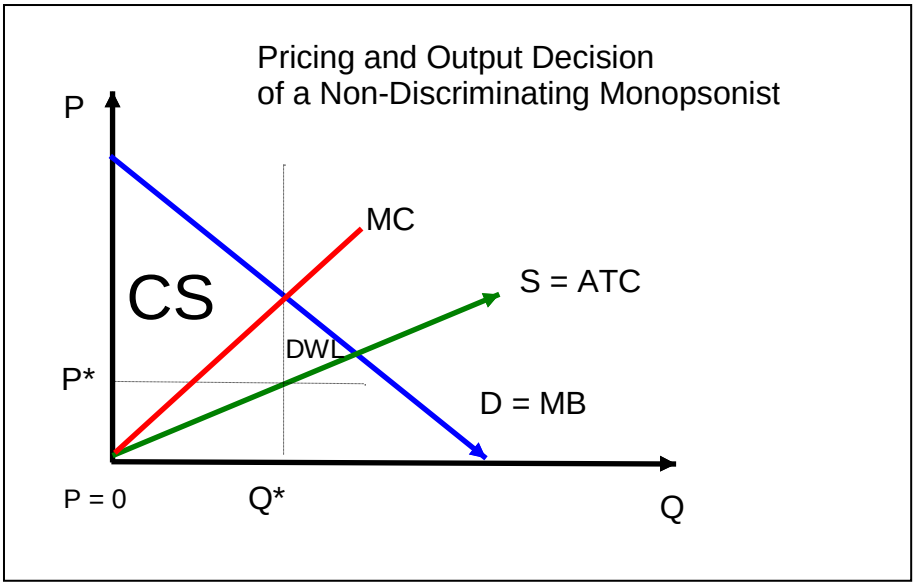
The effects of monopoly power on the buyer side of the market resembles that of the classic monopoly price and output decision.

A **monopsony buyer** will want to purchase the output level that sets his, her or its marginal benefits equal to its marginal costs. But it will realize that its purchase decisions will affect market prices and take that effect into account.

A monopsonist's MC curve can be derived as follows. Suppose that supply is simply Q = aP. This can be rewritten as P = Q/a.

Total cost for a buyer is $C = PQ = Q^2/a$. Marginal cost is the derivative of the total cost function with respect to Q, which in this case is MC = 2Q/a.

Note that the monopsonist's MC curve rises twice as fast as the supply curve that it faces.



Note that the monopsonist also first identifies the net benefit maximizing quantity to purchase given its upward sloping MC curve and downward sloping MB curve.

Given the quantity, it offers to pay just the amount that will induce this quantity to be supplied by firms in the industry. This maximizes buyer net benefits, although it does not maximize social surplus (why?).

- A non-discriminating monopsonist will produce a deadweight loss similar to that of a monopolist.
- Some potential gains to trade will go unrealized.

As with monopolists, there can be "natural" or "unnatural" monopsonists, and, as with sellers, again "price clubs" or cartels of purchasers may be useful ways to increase buyer net benefits. And again, price discrimination may be useful for buyers and may reduce social net losses.

VIII. Monopolistic Competition

The idea that monopolists might compete with other firms who produce similar but not identical products was proposed and analyzed by Edward Chamberlin in 1933. It is a far more common market structure today than it was back in the 1930s.

Edward Chamberlin's idea (monopolistic competition) refers to markets where several firms, perhaps dozens, ell similar but not identical products, as true of many contemporary producers and retailers.

The products in monopolistically competitive markets are good, but not perfect, substitutes for one another.

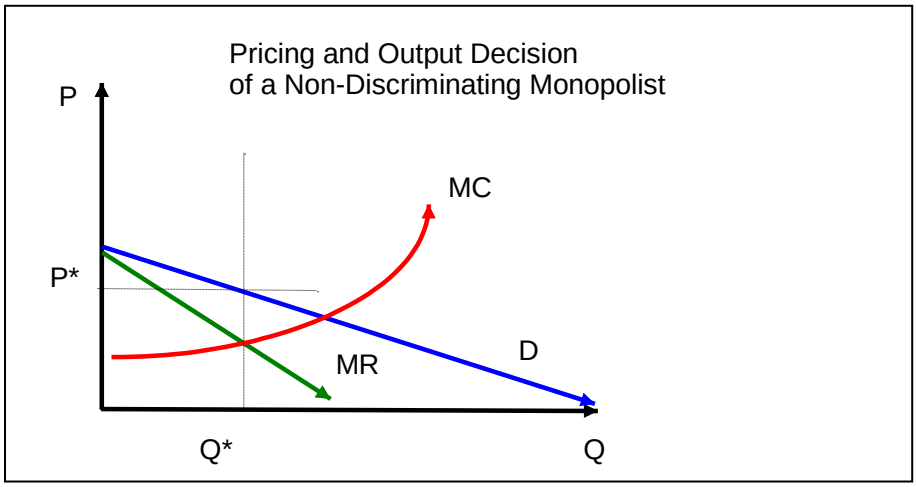
The availability of substitutes changes both the slope and the extent of the demand curves faced by firms in this industry. Demand curves are flatter (more price elastic) than they were in the original monopoly case, which tends to make the market resemble a competitive market, because "mark ups" (the extent to which P>MC) tend to be smaller and profits decline toward "normal" rates of return.

In old town Morgantown, for example, there is one Japanese, one Chinese, and one Tai restaurant. They are good substitutes for one another, but not perfect substitutes, so each has some monopoly power, but it is limited by the prices and products of its competitors.

- Each firm's monopoly power depends on the extent to which other firms produce goods that consumers regard to be similar--that is to say, good substitutes.
- Entry into a monopolistically competitive market occurs by firms producing different, but similar products.
- As the available substitutes become better (that as more close substitutes become available, demand curves shift down and become "flatter" and monopoly power falls (although not to zero).
- (Note that the flatter is a firm's own demand curve, the more his situation resembles that of a firm in a perfectly competitive market.) As better and better substitutes for one's own

product are produced and sold, the market becomes more and more like perfectly competitive markets.

- Monopolistic competition is very common in today's commercial societies, because firms often produce somewhat different products that are substitutes for others sold by other firms, but no perfect substitutes, because consumers care about the differences.



- In long run equilibrium, there are enough firms producing similar, but different products, that no new entrant can join the market and obtain enough sales to be profitable.
- This condition sometimes relies upon fixed costs to produce an equilibrium, but this is not really necessary. U shaped long run MC curves can achieve this as well.

Draw a zeros profit monopoly outcome as an exercise. Hint: this will require a J-shaped MC curve.

The main point, however, is that with monopolistic competition each firm's demand curves is "lower" and "flatter" (more price sensitive), so firms in monopolistically competitive markets have less monopoly power (less ability [or interest] in charging consumers more than the marginal cost of production). In the linear case, the markup may still be 2 to 1, but the selling price is closer to MC in dollars than a similar demand curve with a steeper slope would have been.

Both monopolistically competitive markets and monopoly markets have persons or organizations in them who can directly determine prices. This differs from perfectly competitive markets where all persons are price takers. (And there are no price makers.) Nonetheless, **the comparative static results are similar to those found for competitive markets.** Prices tend to rise if MC increases or if Demand increases. Prices tend to fall if MC decreases or Demand decreases.

Thus, many of the qualitative results from "demand and supply" based analysis will apply to monopoly markets as well, although monopolists tend to realize (on average) somewhat greater average profits

than competitive firms do, because they are no longer selling (the last unit of) their output at marginal cost.

MR>MC at Q* for monopolists where Q* is the profit maximizing output.

An exception to this rule of thumb occurs when a monopolist is able to engage in perfect price discrimination, in which case the last unit sold will have a price equal to MC, although all other units will be sold at above MC.

How much greater price is than marginal cost depends on the slope of the demand curve. The less price sensitive demand is (the steeper it is), the more prices will be above marginal cost (in dollars per unit).

(Other dynamic considerations, and price discrimination will also affect pricing decisions, but these are neglected in this summary.)

IX. Antitrust Law: Using the Law to Block Shared and other Unnatural Monopolies and Monopolistic Practices

Concern with monopoly firms emerged in the 19th century as industrialization occurred and as national transportation networks allowed single firms or small groups of firms (often organized as cartels or "trusts") to dominate production of important goods and services such as steel, aluminum, and railroads, etc.

Prior to the nineteenth century, most monopolies were created by governments and sold off as a revenue sources. These state monopolies were also of concern, but fewer were created after mass protests in the seventeenth century.

In the nineteenth century, as production became more capital intensive (e.g. used larger machines and factories), fewer efficient sized forms could be supported by even national markets, which made it easier for firms to meet and attempt to coordinate their pricing and output decisions.

It did not, however, make such agreements binding (they were not enforced by common law), and so economists debate the extent to which cartels actually were able to realize monopoly profits, because members of a cartel face strong temptations to produce more than their "quota" established by their cartel.

The Sherman Antitrust Law

The first national antitrust law adopted in the US was the Sherman Antitrust Act of 1890. (Named after senator John Sherman, an Ohio Republican, who was the main author of the bill.)

The Sherman Act essentially makes cartels and other methods of monopolization illegal.

- "Section 1 prohibited contracts, combinations, and conspiracies in restraint of trade.

    - Section 2 prohibited monopolization and conspiracies and attempts to monopolize.
    - Other provisions of the act imposed criminal sanctions for its violation but also authorized injunctive suits by the Justice Department and treble-damage suits by private parties."
- The current (amended) wording of the Sherman Act can be found at: http://www.antitrustupdate.com/statutes/shermanact/st-sherman1-4.html

The Sherman Antitrust Act was initially applied to all manner of market cooperation aimed at controlling prices or that created "unfair" marketing practices—including efforts by unions and farm cooperatives.

Like most new areas of public law, how the law should be applied was not initially entirely clear. So, it was left up to the courts to work out how to apply the law to specific cases.

Thus, antitrust law is partly a product of the legislature and partly a product of court decisions, a few of which we'll look at in this last major section of Chapter 6.

The Clayton Act and FTC Act

Concerns over whether the Sherman Act was being applied in a reasonable (politically reasonable?) manner led to two additional anti-trust acts: The **FTC** (Federal Trade Commission) **Act** of 1914, and the **Clayton Act** also in 1914.

> "Under this Act, the Commission is empowered, among other things, to (a) prevent unfair methods of competition, and unfair or deceptive acts or practices in or affecting commerce; (b) seek monetary redress and other relief for conduct injurious to consumers; (c) prescribe trade regulation rules defining with specificity acts or practices that are unfair or deceptive, and establishing requirements designed to prevent such acts or practices; (d) conduct investigations relating to the organization, business, practices, and management of entities engaged in commerce; and (e) make reports and legislative recommendations to Congress"

The FTC act of 1914 created a new federal agency and gave it responsibility for enforcing antitrust law. It forbade "unfair methods of competition" including "tie in" sales and "exclusive" dealing.

- The FTC acts **exempts** banks, airlines, common carriers, from its rules. (why?)
- FTC decisions were to be final unless appealed to the Supreme Court.
- The amended text of the FTC act can be found at: http://www.stolaf.edu/people/becker/antitrust/statutes/ftc.html

The Clayton Act of 1914 forbade price discrimination, stock acquisitions, and interlocking directorships, which could be used to coordinate pricing and output decisions but were neither monopolies nor trusts.

- The Clayton Act, like the Sherman Act, also allows those harmed to sue for damages, and actually recover triple damages.
- Sec 17 of the Clayton Act exempts labor unions and (non-profit) farm cooperatives from antitrust suits.
- (Baseball was exempted after a 1922 Supreme Court decision.) See http://www.nytimes.com/1994/12/24/sports/baseball-antitrust-exemption-history.html for a nice overview of that decision.
- The amended text of the Clayton Act can be found at: http://www.stolaf.edu/people/becker/antitrust/statutes/clayton.html

These three antitrust laws remain the main legislative basis for antitrust lawsuits and criminal actions in the United States. Others followed but were relatively less important. For example, in 1950, these three acts were augmented by the **Celler-Kefauver Act**, which addresses mergers that may reduce competition. The Antitrust acts after 1914 were generally amendments of the Sherman, FTC, or Clayton Acts and so normally appear in the text of the contemporary (amended) texts of these acts.

The main civil remedy in the first anti-trust laws was the provision for **triple damages** for a firm that successfully wins an anti-trust case. Triple damages create strong incentives for damaged (and other) firms to launch civil suits charging monopoly practices. (This provision could be "efficient" in the sense of the punitive damages of tort cases, if only a fraction of monopoly damages is every brought to court, otherwise it simply encourages more lawsuits.)

During the 1955, 1974, and 1990, the various criminal penalties (fines) for anti-trust law violations were increased, although the triple (treble) damages provisions were kept. Such criminal proceedings could be initiated by the FTC or the Department of Justice.

Richard Posner's book on Anti-trust law includes a table that list the number of anti-trust cases brought by the Department of Justice. See his Table 1. That table shows that after the Sherman Act was adopted, relatively few cases were brought by the Department of Justice (aka DOJ), just 15 cases were brought in the first ten years, 42 in the next ten years, and 126 between 1910 and 1919.

- More and more cases were brought each decade except during the Great Depression, peaking in the 1980-1989 period (Reagan Presidency) with 741 cases, followed by 609 cases in 1990-1999 (Clinton Administration).
- The cases are roughly 2-fifths civil cases and 3-fifths criminal cases.
- Average fines have been increasing through time in nominal terms, rising from about 20K during the 1910-1929 period to about 325K in the 1970-1989 period, and then rising dramatic during the 1990s to nearly 5 million dollars. (See Posner's table 2).

X. APPENDIX I: Some Famous Antitrust Cases

Antitrust law evolved through a long series of court decisions, especially ones made by the Supreme Court. There are essentially two main lines of argument.

(1) First, that some practices and levels of concentration are "**per se**" in violation of the antitrust acts and so illegal.
(2) Second, that only practices that "**unreasonably** constrain competition or restrain markets" are illegal. These vary case by case according to what is "reasonable" for firms in the market of interest. Generally, "reasonable" monopoly power tends to increase, rather than decrease, consumer surplus.

Both interpretations came to be more and more influenced by economic arguments, with the result that the central issue often became (i) the extent of market concentration and (ii) whether a particular practice increased or diminished competition (and/or consumer net benefits.)

- **Richard Posner (2010-10-22). Antitrust Law, Second Edition Provides a lengthy defense of the "reasonable practices" view of proper applications of anti-trust law**. [Posner is a law professor at the University of Chicago and a judge on the 7th US Court of Appeals in Chicago.]
- [Economists who specialize in "industrial organization" often earn large fees to appear in monopoly cases as "expert witnesses."]

This is not to say that the court always gets it right (economically), but it is to say that the trend in the decades after the Reagan administration has been toward a "reasonability" standard (anti-competitive standard), rather than a per se standard.

Other economists and lawyers support "per se" laws because they are clearer and less subject to manipulation in court. "Per se" rules have recently made a comeback within the Biden administration.

See, for example, Dennis Mueller (1996) for a somewhat less optimistic take on US antitrust law that favors using "per se" rules in most cases.

The Standard Oil Case

The Sherman Act changed antitrust law from traditional common law practices. A good illustration of the new national approach to monopoly generated by the Sherman Act is developed in the **Appeal to the Supreme Court in the Standard Oil Case**. [Source http://www.law.cornell.edu/supct/html/historics/USSC_CR_0221_0001_ZS.html]

The following are excerpts from the majority opinion.

The debates in Congress on the Anti-Trust Act of 1890 show that one of the influences leading to the enactment of the statute was doubt as to whether there is a common law of the United

States governing the making of contracts in restraint of trade and the creation and maintenance of monopolies in the absence of legislation.

While debates of the Congress enacting it may not be used as means for interpreting a statute, they may be resorted to as a means of ascertaining the conditions under which it was enacted.

The terms "restraint of trade," and "attempts to monopolize," as used in the Anti-Trust Act, originated in the common law, and were familiar in the law of this country prior to and at the time of the adoption of the act, and their meaning should be sought from the conceptions of both English and American law prior to the passage of the act.

The original doctrine that all contracts in restraint of trade were illegal was long since so modified in the interest of freedom of individuals to contract that the contract was valid if the **resulting restraint was only partial in its operation and was otherwise reasonable.**

At common law, monopolies were unlawful because of their restriction upon individual freedom of contract and their injury to the public and at common law, and contracts creating the same evils were brought within the prohibition as impeding the due course of, or being in restraint of, trade.

The early struggle in England against the power *to create* monopolies resulted in establishing that those institutions were incompatible with the English Constitution.

At the time of the passage of the [Sherman] Anti-Trust Act, the English rule was that the individual was free to contract and to abstain from contracting and to exercise every reasonable right in regard thereto, except only as he was restricted from voluntarily and unreasonably or for wrongful purposes restraining his right to carry on his trade. Mogul Steamship Co. v. McGregor, 1892, A.C. 25.

Standard Oil (the appeal brief is available at:
http://www.law.cornell.edu/supct/html/historics/USSC_CR_0221_0001_ZS.html

J. D. Rockefeller created Standard Oil in 1870 and largely through that firm became the world's richest man and America's first billionaire by corning the US market for refined oil products and also through large oil and pipeline holdings. He also managed to obtain preferential rates for railroad shipping. In 1890, it controlled 88% of the refined product market, and continued to increase its share of production and sales.

Rockefeller and his major partners invested a good deal of their dividends in Railroad stocks, which may account for his ability to gain preferential rates for shipping relative to other refined oil producers.

The company began in Ohio, where the first American Oil boom occurred.

In 1885, SO moved from Ohio to NY and then on to NJ, because of its more lenient corporate law.

SO produced so much refined product, that it exceeded US demand and created major export markets and SO outlets in Europe and Asia.

In 1909, the US Justice Department sued SO and ordered it to be broken into 34 companies.

> "Rebates, preferences, and other discriminatory practices in favor of the combination by railroad companies; restraint and monopolization by control of pipe lines, and unfair practices against competing pipe lines; contracts with competitors in restraint of trade; unfair methods of competition, such as local price cutting at the points where necessary to suppress competition; [and] espionage of the business of competitors, the operation of bogus independent companies, and payment of rebates on oil, with the like intent."
>
> "The evidence is, in fact, absolutely conclusive that the Standard Oil Company charges altogether excessive prices where it meets no competition, and particularly where there is little likelihood of competitors entering the field, and that, on the other hand, where competition is active, it frequently cuts prices to a point which leaves even the Standard little or no profit, and which more often leaves no profit to the competitor, whose costs are ordinarily somewhat higher."

In May 1911, the US Supreme Court upheld the lower court judgement and declared SO to be an "unreasonable monopoly," and ordered it to be broken up into 34 firms.

Among the larger firms created are the present-day Exxon, Chevron, Amoco, and Mobil Oil.

SO (ESSO) continues to operate in Europe and many other parts of the world.

(Surprisingly, total share prices rose after the breakup, making Rockefeller even richer!)

By the time of the **breakup SO's share of refined product production had fallen** from around 90% in 1900 to around 65% in 1911.

US Steel

US Steel was founded in 1901 in Pittsburgh by Andrew Carnegie, JP Morgan, Charles Schwab, and E. H. Gary. It was essentially a conglomeration of steel and steel product producing companies.

It grew to be the world's first company worth more than a billion dollars. Mergers and acquisitions continued, and it began to look to many as if US Steel completely dominated the market for steel and steel products.

US Steel built the town/city of Gary Indiana in 1906, the site of one of the world's largest steel mills.

During its formative period the company was dominated by Gary (its CEO), who exercised influence throughout the American steel industry through his famous “Gary dinners,” attended by the heads of major steel producers; out of the meetings came agreements on cooperative pricing and marketing that stabilized a once wildly fluctuating market. Gary opposed “unreasonable” competitive practices as well as labor organizers. (Brittanica.com)

From its inception it was the dominant steel producer in the US, with a market share of well over 50%. It’s market share remained more or less flat or shank somewhat between 1901 and 1911, although industry output increased by 25% during that period. (A. Cotter 1916: 224)

In 1911, the **department of justice began antitrust proceedings** against US Steel.

> In 1920, the US Supreme Court decided that US Steel **was not** a monopoly and so its conduct did not come under the Antitrust laws.
>
> Held that the power attained by the United States Steel Corporation, much greater than that of any one competitor, but not greater than that possessed by them all, did not constitute it a monopoly.
>
> The fact that a corporation, alleged to be an illegal combination, during a long period after its formation, persuaded and joined with its competitors.
>
> Its efforts, **at times successful and at times not, to fix and maintain prices in violation** of the Anti-Trust Act, dos not warrant present relief against it if the illegal practices were transient in purpose and effect, were abandoned before the suit was begun because of their futility and not for fear of prosecution, and have not since been resumed, and if no intention to resume them or dangerous probability of their resumption is shown by the evidence. Pp. 251 U. S. 444 et seq.
>
> Purpose and effect of the Steel Corporation's acquisition of control of the Tennessee Coal & Iron Company considered in the light of President Roosevelt's prior approval of the transaction and his testimony concerning it. P. 251 U. S. 446.
>
> Upon the question whether the power possessed by the Steel Corporation operated per se as an illegal restraint, held that testimony of its officers, its competitors, and hundreds of its customers to the effect that competition was not restrained and that prices varied or remained constant according to natural conditions must be accepted as clearly outweighing a generalization advanced by government experts that constancy of prices during certain periods evinced an artificial interference. P. 251 U. S. 447.
>
> **An industrial combination short of a monopoly is not objectionable under the act** merely because of its size -- its capital and power of production -- or merely because of a power to restrain competition, if not exerted. Pp. 251 U. S. 447, 251 U. S. 450 et seq.

Alcoa

Alcoa was founded as the Pittsburgh Aluminum Co by a group of young entrepreneurs (Hall, Cole, Hunt, and others) in 1888, shortly after Charles Hall discovery of a new method for recovering Aluminum from Bauxite ore in 1886, based on a patent for the process (finally issued in 1889.)

It expanded its operations to include fabrication as well as recovery of aluminum from ore in 1890.

Between 1888 and 1897, the price of aluminum fell from 8$/lb to 36 cents/lb.
(http://www.alcoa.com/usa/en/alcoa_usa/history.asp)

Because of its patent on innovations in both production and fabrication, Alcoa had **a virtual monopoly** on US production and produced 60% of world output.
(http://www.alcoa.com/usa/en/alcoa_usa/history.asp)

In 1907, the company was renamed the Aluminum Company of America (ALCOA).

Raising funds for expansion required selling shares, and the Mellon family gradually became the largest shareholder--controlling about a third of Alcoa's shares.

In 1937, the FTC launched an antitrust suit against Alcoa.

The Justice Department believed that Alcoa had violated the Sherman Act on three counts: making restrictive covenants, engaging in alleged acts of unfair competition and participating in foreign cartels.
(http://www.alcoa.com/usa/en/alcoa_usa/history.asp)

The FTC believed Alcoa tried to monopolize bauxite, attempted to monopolize the waterpower of the world, dominated and controlled the foreign market for aluminum in the US, and engaged in injurious price cutting.

Alcoa won the trial on all 130 counts.

But the Government won the appeal.

Review by the Supreme Court was impossible, since four of the justices had been involved in prior antitrust suits against Alcoa.

A special act of Congress was necessary to give the 2nd Circuit Court of Appeals the weight of a Supreme Court opinion in this matter.

In 1944, the court found Alcoa controlled over 90% of the US market for aluminum ingot. This proportion alone was sufficient to support a violation of the Sherman Act, regardless of intent to monopolize.

The decision was made by **Judge Learned Hand**, included the following:

"It was not inevitable that it [Alcoa] should always anticipate increases in the demand for ingot and be prepared to supply them. Nothing compelled it to keep doubling and redoubling its capacity before others entered the field. It insists that it never excluded competitors; but we can think of no more effective exclusion than progressively to embrace each new opportunity as it opened, and to face every newcomer with new capacity already geared into a great organization, having the advantage of experience, trade connections and the elite of personnel."

"90 percent is enough to constitute a monopoly; it is doubtful whether 60 to 64 percent would be enough; and certainly, 33 percent is not."

[Some lawyers and economists regard this characterization to be the "per se" rule as opposed to the "rule of reason" interpretation of monopoly as "unreasonable restraint of trade or competition." The debate between the "per se" rule and the "rule of reason" approach played an important role in antitrust suits for the rest of the 20th century.]

In 1947, Alcoa made the argument to the court that there were two effective new entrants into the aluminum market – Reynolds and Kaiser – as a result of demobilization after the war and the government's divestiture of defense plants. In other words, the problem had solved itself and no judicial action would be required.

On this basis, the district court judge ruled against divestiture in 1950, but the court retained jurisdiction over the case for five years, so that it could look over Alcoa's shoulder and ensure that there was no re-monopolization.

ATT

Graham Bell invented the telephone in 1875 and received two patents on the telephone in 1876.

These were used to launch the Bell Telephone company in 1877.

Service expanded fairly rapidly with the first calls between Chicago and NY occurring in 1892, and the first transatlantic calls in 1927.

Because of its near monopoly over telephone service in the US, AT&T was the target of many antitrust actions over the decades, although settlements of various kinds were normally worked out, which left the company in one piece.

In 1974, the Department of Justice launch an antitrust suit against AT&T, which was finally decided in 1984, and caused the breakup of the "Bell System" into 7 different regional telephone firms and a long-distance provider (AT&T).

The breakup led to a surge in competition in both long-distance service and in telephone technology.

(However, many of the new firms were allowed to merge 15-20 years later, which reduced the seven independent firms to just two or three by 2012.)

[For additional cases and discussion of antitrust law, see my law and economics website.]

Chapter 7: Entrepreneurship, Innovation, and Economic Development

I. A Short Overview of the Implications of the Analysis to This Point

From the previous chapters, you are familiar with the effects of market structure on equilibrium prices and outputs.

i. In competitive markets, prices equal marginal cost and marginal cost equal revenues.
ii. In noncompetitive markets prices exceed marginal cost, but marginal cost still equals marginal revenue.
iii. These are consequences of rational decision makers attempting to maximize their net benefits (consumer surplus and/or profits).
iv. You should also know that input and output markets are connected and that as demand for outputs rises, demand for inputs also rise, and thus input prices also tend to rise, and the quantity of inputs used in particular industries also tends to increase.
v. In equilibrium each input earns its marginal revenue product (at the margin).
vi. Such connections among markets imply that prices help to coordinate the decisions of firms, consumers, and input providers—and tend to draw resources into markets where they are most demanded.

Market equilibria all have prices that set demand equal to supply, matches inputs to markets, and competition among producers and the quest for profits tend to minimize the cost of production.

In competitive markets, social surplus (social net benefits) is maximized (in the absence of externalities). In less competitive markets, social surplus is increased by the trade that takes place, but not necessarily maximized—the latter being the economic justification for antitrust policies. The division of that surplus varies by market type, with consumers getting the lions share in perfectly competitive markets and somewhat less in monopolistic markets. (How much less depends on the degree of price discrimination.)

Trade is a voluntary activity, so it takes place only if both consumers and firms expect to benefit from their purchases and sales of the goods traded. It is this property that assures that social surplus tends to increase as trade expands.

All of our models and the analysis based on them has assumed that individuals are all the same in the sense that they all try to maximize net benefits—although they may disagree about how to best do so. The latter explains why people buy different things, why firms produce different things, and why they do so in different ways. This is not to suggest that people make a lot of mistakes, but simply that although they are the same in one sense (maximizing net benefits), but they are also different in their

preferences, understanding of a good life, have different skills and access to capital, and so choose different goods, careers, and organize production in different ways.

We have also studied factors that may disrupt a market equilibrium.

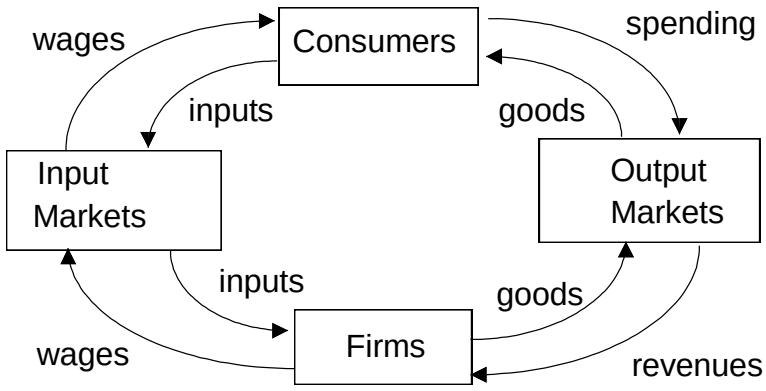
For example, consumer demand tends to increase for most goods (normal and superior goods) when consumer income increases (although it may fall for inferior goods). Thus, as consumer income increases on average, the demand for most goods increases and their prices tend to increase. (When demand shifts out to the right, prices tend to rise if supply curves are upwards sloping and remain where they were.) Demand also tends to increase if the price of a substitute increases or if the price of complements fall.

Market equilibria are also disrupted by changes in the price of inputs or in the opportunity cost of the folks that own the firms that make the things and services sold in markets. If the price of inputs increases, then the supply curve tends to shift to the left, which causes market prices to increase.

Another factor that can shift the supply curve is technology. As technology improves, the cost of producing goods and services tends to fall which shift the supply curve out to the right, reducing prices and increasing total sales of the products that make use of the improved technology.

If technology is stable and no capital accumulation occurs, the logic of rational choice implies a static economy. All markets clear simultaneously (including input markets). And, stable income and production costs yield a stable pattern of consumption and prices. In effect, everyone simply repeats what they did yesterday.

Joseph Schumpeter called such an economy an "**evenly rotating economy**. See the diagram below to see why.



This chapter shifts our attention away from such stable—non growing—markets towards ones where the overall scope of trade expands through time, more or less on market at a time.

There are two main models of economic growth. One focuses on capital accumulation and the other focuses on innovation. The first of these approaches worked out is the one the focuses on capital accumulation, and that is where we'll start.

II. Capital Accumulation and Growth in a Simple Neoclassical Model

Whenever additional capital in an economy has accumulated, the evenly rotating economy becomes a "spiral," because productivity and income tends to increase as capital is accumulated.

A. Capital is accumulated, for example, when a firm uses some of its profits to purchase additional equipment, expand its factories, or hires more educated, skilled, or talented people so that it can expand production—as would tend to happen if the owner(s) expects future demands to increase and prices to rise for their existing products.
B. Capital is also be accumulated when consumers save part of their income, places in banks or other intermediaries, which lend out the money to firms that use the money to expand production. An increase in savings, thus tends to increase capital stocks and outputs.
C. New capital may be either **physical capital** (machines, computers, equipment) or **human capital** (knowledge and skill). What you are doing here at WVU is accumulating more human capital—which hopefully will increase your productivity and lifetime wages and/or quality of life.

In a setting where labor and capital are used together to produce goods and services, **an increase** in either the physical capital used in production, or the human capital used in production tends to **increase the marginal product of labor** and therefore tends to **increase wage rates where the new capital is employed**—other things being equal..

The increase in income associate with a higher marginal revenue product, then increases the demand for all normal and superior goods (although it reduces them for inferior goods). The result is an increase in the economy's total output measured in real value terms (social net benefits) and in terms of revenues from sales (real gross national product, RGNP). **Such an economy is no longer "evenly rotating."** A continual increase in its capital stock creates a spiral of ever-increasing output.

It is this capital-accumulation model of growth that many economists refer to when they discuss "classical growth models." Classical growth models rely upon capital accumulation to propel an economy forward (e.g. to increase average income).

In principle, there are many different kinds of capital, just as there are many inputs and outputs, and these models only capture a few key relationships.

Economies are complex systems of many inputs, products, firms, and consumers, most of which are connected to each other. Macro-economists normally simplify this system by assuming that there are just two inputs (capital and labor) and one final output. (We'll leave further discussion of such models to your macroeconomics classes, but it is important to understand that classical growth models are essentially micro economic models that are based on rational choices and capital accumulation.)

Thought Problems and Exercises. As an exercise, draw a series of diagrams to illustrate market prices, a typical firm's output decision, a typical consumer's purchase decisions, and the market for labor.

K. Note that average wages determine income for the average consumer, and that w* = MRP = P* x MP

L. Now assume that the marginal product of labor increases because human or physical capital increases.

M. Show the effects on consumer income, market demand, and market equilibrium.

N. How does this change affect the distribution of profits and consumer surplus in your model?

O. How does this change increase total social surplus and total sales in your model?

P. How would your results change if you shift from Marshallian to Ricardian models of long run supply?

Q. Repeat for another set of diagrams with different slopes.

III. Innovation and Growth (Schumpeter: Creative Destruction)

In the classical model, growth implies more and more of the same final goods and services. There are larger supplies of most existing outputs and so higher real incomes for the persons living and working in an economy. **However, there is more to economic growth and development than simply more of the same. The range of alternative products also increases.**

Joseph Schumpeter was an Austrian economist who took a position at Harvard in 1932. For the most part, he wrote in the period between WWI and WWII and argued **that innovation is an important engine of economic growth,** and that **innovation is generated by entrepreneurs.** Schumpeterian entrepreneurs are thus innovators.

Schumpeter's entrepreneurs create new products and new production processes. By doing so, they "disrupt" or "creatively destroy" previous market equilibria. New products, for example, usually affect demands for other related products (both substitutes and complements).

New production methods affect demands for inputs through effects on marginal product. In some cases, they disrupt very long-standing patterns of life, as when the Automobile (at about this time) displaced horses as the main mode of transport.

Schumpeter called this process of innovation and market response the process of **creative destruction**. Major innovations often have major impacts on the pattern of life in a community. Automobiles, for example, greatly altered patterns of life in communities, among communities, and the landscapes of communities themselves.

Road networks in old town centers had to be widened in many cases (and paved).

Buggy whip manufacturers lost markets and had to find bug whip makers had to find new jobs.

Old stable and feed networks, which extended right into downtown areas, had to find new uses for their buildings and supply networks.

New facilities were built to fuel, repair, and service automobiles (sometimes in the old stables, but often in completely new buildings with completely new designs).

Schumpeter's theory of economic development implies that economic growth is not simply "more of the same," rather it is new products and new methods of production. Contemporary economics agrees with Schumpeter that innovation is a significant source of growth, perhaps more than half of it.

Schumpeter further argued that large profitable firms often undertake much of the innovations. Thus, there are often dynamic advantages of monopoly power that ordinary static models of monopolies ignore. These dynamic benefits may more than offset their static deadweight losses.

This point is still debated.

R. After all, Apple computer began as a very small firm in a garage. Similarly, Facebook started off in college dorms, etc. etc. Many innovative firms began as very small enterprises.

S. However, once up and running and profitable, both firms engaged in a good deal of innovation, bringing new products and services to markets. Thus, mature innovative firms often continue to innovate.

Practice Exercises and Thought Questions. Draw a series of graphs to represent the market for Automobiles and Horses in 1900 (when automobiles were still luxury goods).

Now imagine that Henry Ford invents the assembly line and reduces the price of cars by 60% through improved production methods and introduces the model T in 1908.

T. Show the effects of this innovation on the markets for automobiles and horses.

U. Repeat for the markets for land line telephones and cell phones (Blackberries, Nokias, etc.) after the introduction of the I-phone by Apple in 2007.

IV. Risk and Uncertainty in Economic Activities (Knight: Entrepreneurs as Risk Takers)

Frank Knight had a somewhat different take on the role of entrepreneurs. He was a professor at the University of Chicago in the early twentieth century and during the same period as Schumpeter. His work was better known in the United States because he wrote in English rather than German.

Knight's analysis of competitive economies, profits, and growth focuses on risk and uncertainties, rather than creativity. He argues that some risks can be understood, and others cannot. Those that can be understood can be insured against, but those that cannot, cannot be insured against. The latter—which he referred to as "uncertain" rather than risky investments—can be a source of unusually large profits (and losses) even in a perfectly competitive market.

Note that Schumpeter's innovative entrepreneurs take risks in markets where they cannot be sure of success.

V. Note also that innovation can generate risks for market participants, especially major ones of the creative destruction variety.

Most neoclassical models and classes ignore risk and uncertainties and focus on price theory, income, and scarcity in a world where there are few risks and both firms and consumers are very well informed. (This was true of our models covered until now in this course.) However, as Knight reminds us, much about life in the real world is uncertain, including the weather, illness, accidents, and close elections. All these events affect markets by affecting the productivity or wealth (marginal product of land, labor, capital, or their value) and thereby marginal costs.

All kinds of random and surprise factors can influence patterns of demand or production costs vary year-by-year, season by season, and from time to time. Such random "shocks" imply that market demand and supply are also somewhat random and thus so are market prices.

Price, outputs, and incomes may still be fully determined by marginal product and marginal benefits as in our models to this point, but because these vary somewhat because of random shocks, prices, outputs, and income will also be **partly random**. (Some demands are predictably seasonal, as with the demand for ice scrapers and snowshoes, but we'll ignore these for now.)

Coping with Risks

Knight points out that a variety of steps can be taken by firms and consumers to deal with risks including the accumulation of reserves (rainy day funds), purchases of loss reducing capital goods (umbrellas), and diversifying portfolios (growing more than one crop, holding more than one kind of stock, producing more than one kind of product, etc.)

Another way to reduce the risk of losses of various kinds is to purchase of insurance. (A topic taken up in more detail in the next chapter.) Any well understood and bounded risk can be insured. Thus, Knight's analysis predicts the emergence of insurance markets. Risk averse persons will purchase such insurance because they are willing to pay a positive price to limit their losses.

Most insurance limits losses associated with risks by paying insured persons some amount of money when an unpleasant surprise occurs: a car accident, a house fire, a nasty illness, etc.

Large insurers can provide insurance more cheaply than smaller firms, because of the law of large numbers in statistics (e.g. sample variance falls with sample size, as developed in the next chapter). The latter implies that insurance markets are unlikely to be perfectly competitive, although diminishing returns in the effects of sample size together with administrative costs imply that insurance markets are probably not natural monopolies.

Frank Knight argued that markets for insurance tend to be competitive, which implies that the average profits realized by selling insurance will be approximately the average rate of return in such markets (e.g. for supplier of other goods in competitive markets). That is to say, **in the long run**, insurers in equilibrium earn only their opportunity cost rates of return, and thus no pure economic profits (on average). Knight, thus, assumes a Marshallian model of long run supply of insurance.

Given this, he argues that people who purchase insurance receive it at least cost--at an actuarially fair price. The actuarially fair price is $C = Pr*L + H$, where Pr is the probability of a particular loss, L is the amount paid by the insurance if that loss occurs, and H is the firm's overhead for selling insurance and settling claims.

Knightian Uncertainty

Knight also pointed out that there are also "risks" that cannot be fully understood. **He called these risks "uncertainties."** Uncertainties, he argued, cannot be insured.

Thus, uncertainties can produce profits and losses even in Marshallian markets with perfect competition, because uncertainty cannot be insured. Thus, entrepreneurs bear the entire risk of both success (high profits) and loss (bankruptcy). Taking on such risks personally sometimes allows entrepreneurs to do a lot better than average, to earn more than their opportunity cost rate of return, e.g. true economic profits.

- These unknowable risks became known as "**Knightian Uncertainty**," in recognition of Knight's insight.

Knight argued, for example, that many adopting new production methods and bringing new products to markets are often **uncertain investments, rather than merely risky ones where the probabilities of success and loss were well understood**.

In Knight's theory of markets and entrepreneurship, **entrepreneurs are risk takers**, because they all undertake uncertain investments and so make large profits if they are right, but they may earn large losses (or become bankrupt) if they are wrong. (Note that Schumpeter's entrepreneurs are innovators as well as risk takers. They are a subset of Knightian entrepreneurs.)

In a Marshallian world, true profits can only occur because some risks cannot be known, and entrepreneurs are brave (or foolish) enough to **take the risks** that produce them. Otherwise, average returns will be the normal competitive one of the Marshallian models. Only entrepreneurs can earn true economic profits or losses, on average, because they take risks that other investors and firm owners will not, often without much competition. (Part of the profits realized by successful Knightian entrepreneurs may be from temporary monopoly profits.)

V. Markets as Coordinating Systems (Kirzner: Finding and Eliminating Market Opportunities and Inconsistencies)

Israel Kirzner, who wrote in the post WWII period, suggests that markets do not always clear by themselves. Sometimes they need a bit of help from entrepreneur speculators.

Kirzner suggests that the inventory adjustments of suppliers and shopping efforts of buyers are not always enough to clear markets within a region, nation, or world. Alert speculators--those who buy low and sell high--are often necessary for markets to clear at a single price.

Kirznerian entrepreneurs are especially insightful businessmen. They may not be risk takers in the Knightian sense, nor particularly innovative in the Schumpeterian sense, but they also undertake activities not included in neoclassical models. They simply take advantage of market opportunities that others have missed. Their efforts tend reduce the extent of *unrealized gains from trade* and bring both prices and the usage of economic resources toward their theoretical values. Kirzner's entrepreneurs are simply persons who are **more alert** than average to market opportunities (inconsistencies).

Speculators move markets towards a single price equilibrium by purchasing goods (or assets) in markets where it is cheap (underpriced) and selling it in markets where the goods are more expensive (overpriced). Of course, this process works only for portable goods and services--of which there are many. Some immovable goods can be made portable by selling partial rights to them, as in stock markets where shares of immovable assets and their associated profits (net revenues) are bought and sold.

This process of exploiting market inconsistencies (**speculation**) Kirzner calls Entrepreneurship.

By exploiting disequilibrium conditions, Kirzner argues that entrepreneurs make profits and **move markets toward equilibrium**, rather than away from them, as in Schumpeter's analysis. Entrepreneurs in Kirzner's sense are "grease" that makes the gears of supply and demand operate more smoothly.

This contrasts with the roles of entrepreneurs in Schumpeter's view are "shocks" to market equilibria that move markets to new equilibria. Kirzner also argues by pushing prices toward their true equilibrium levels, the resource allocation of markets as a whole also become more efficient (more likely to minimize the cost of production).

Within the context of the social net benefit maximizing normative theory, we would conclude that entrepreneurs help to maximize social surplus (social net benefits)--although Kirzner would resist using the concept of SNB.

Thought Questions. All three concepts of the entrepreneur provide important explanations of economic growth and development that are not usually included in the (neoclassical) competitive model.

Explain why price takers cannot be entrepreneurs in any of these three senses.

W. The entrepreneurial models are, however, consistent with monopolistic competition models. Explain why.

X. Are all small business owners entrepreneurs? If not, why not. If so, in what sense? Discuss.

VI. Entrepreneurs and Evolutionary Economic Development

Kirzner and Schumpeter's entrepreneurs are in a sense "super men and women" or super-informed, very creative, men and women who recognize possibilities that no one else does.

There is another way to think of these activities (innovation, speculation, and risk taking) allows us to dispense with such all-knowing or brilliant entrepreneurs. We can replace such imaginary persons with entrepreneurs that make **hypotheses about neglected market opportunities and by following up on their hypotheses, they test their theories. If they are correct they profit, if not they go out of business.**

With such ideas in mind, Viktor Vanberg and James Buchanan (1991) suggest that markets should be thought of as **experimental laboratories**.

In effect, Vanberg and Buchanan combine the three theories of entrepreneurship. Their entrepreneurs are all, in a sense, like scientists testing hypotheses—not about properties of nature or society, but about market opportunities. Such entrepreneurs are experimenters, who test their hypotheses about markets by launching products, firms, and purchasing assets of various kinds. If they are right, they earn profits, as in Knight's theory. If they are wrong they lose money because their idea fails to attract sufficient support from consumers.

Prior to their investments, Buchanan and Vanberg suggest that it is impossible to really know whether their experiment will succeed.

Will the new Panera's be profitable or not?

Does Morgantown need another Duncan Doughnuts or not?

Is there a market for Dick Tracy watches ala Samsung and Apple smart watches or not?

Are stock markets too high or too low?

Fredrich Hayek--writing after Knight and Schumpeter, but before Vanberg and Buchanan—developed a similar argument and reaches a somewhat broader conclusion.

Market Prices as Information Aggregators

Hayek suggests that markets are social mechanisms for inducing the efficient use of resources including knowledge that varies among individuals. Market prices, products, and production methods reflect a broad subset of the information of all market participants. No single persons or small group could have all the information that firms, consumers, and entrepreneurs take account of (jointly).

Thus, open competitive markets nearly always outperform centrally planned economies in markets for private goods. This is not to suggest that markets are perfect, but simply to say that they (usually) make more effective use of information that is dispersed in the minds of millions of entrepreneurs, workers, and consumers than other forms of organization. This enables markets (often through entrepreneurial decisions to take risks, innovate, and speculate) to both more fruitfully allocate the resources available to a society at a given level of technological development and to develop technologically at higher rates than possible with central planning.

> Of course, markets would work less well without a well-functioning legal system that clearly defines and enforces property, contract obligations, and liabilities...what some economists refer to as “property rights.” We’ll take some of those issues up towards the end of this course (in chapter 10)

Chapter 8: Random Events, Expected Values, and Market Outcomes

I. Information Problems and Rationality

With the exception of the Knightian analysis of entrepreneurship, all of the analysis and models to this point in the course have assumed that information is "perfect" in the sense that consumers knew enough to maximize their net benefits from market transactions and that firms knew enough to maximize their profits from market transactions. As a place to start the analysis of markets, this is a completely reasonable place to begin.

If trades are largely repeated, consumers will know or have accurate estimates of the average quality of the goods purchased. Firms will have accurate estimates of the costs required to produce their products and bring them to market. And, input providers will have accurate estimates of the prices that their goods and services will command in their various labor, raw, material, and intermediate goods markets.

In an evenly rotating economy, experience provides one with essentially all of the data that one needs to be "locally" informed—which is to say, to know or have good estimates of the qualities and prices of the goods and services on offer that one is most likely to purchase or produce. In such settings, market participants can maximize their net benefits from trade—even if they occasionally make mistakes.

In such cases, models that assume that consumers and firms have very good or perfect information are reasonable, and the conclusions drawn from them will be correct on average. Although those same models would not be able to explain the existence of some types of markets or mistakes, the models would be useful, reliable, ways to think about most market transactions.

However, in settings where such information is not available, is too costly, or prices move more or less randomly because of variations in weather, input prices, innovation, or public policies, neither firms nor consumers will know exactly what prices or the quality of goods and services are, and mistakes—in a sense worked out in this chapter—will be more common. Such fluctuations need to be taken into account by both consumers and firms.

What this chapter does is to (1) analyze cases in which some aspects of a good's quality, prices, or costs are randomly distributed, and so exact knowledge is not possible. The cases that are easiest to handle are ones where the fluctuations can be characterized with a probability distribution (as argued by Frank Knight). Economists generally assume that in such cases, consumers and firms maximize average or "expected" net benefits rather than net benefits per se. (2) It also discusses what some term "rational ignorance," which differs from risk, but also may affect consumer choices, planning, and their demand for services.

Chapter 9 (the next chapter) takes up choice settings in which choices that take significant time to execute—which is to say, it examines choices among plans or long-term commitments rather than immediate actions.

In the cases explored in most of this chapter, market participants do not know exactly the quality of the goods on offer, nor the lowest price at which they can be purchased, but they do know something about the probability function that generates market prices. So, some of their expectations about prices are mistaken, but if the probability function is known reasonably well, they will, on average, maximize their net benefits in such choice settings. This is really as well as they can do, if prices or other choice relevant variables are generated via a random process.

Randomness generates unavoidable risk and, in some cases, also uncertainty. The use of "expected values" for making decisions in those settings is the most systematic way to model rational decision making by consumers and firms. It is the effect of risky settings on decision makers and markets that is the focus of this chapter. The same method can be used to analyze about choices under Knightian uncertainty, although in such cases the probability function can only be guessed or intuited rather than actually known.

Economic and other social systems are complex, and always a bit chaotic at the margin, because a variety of random factors including weather, disease, innovation, and geopolitical risks constantly jostle market decisionmakers and thus induce equilibrium prices to move about.

Thus, the choice settings covered in this and the next chapter are more realistic in that they explicitly take account of risks (and to some extent uncertainties) that occur in the ordinary course of life in commercia societies. On the other hand, the assumption that individuals maximize expected net benefits is a bit less general than simply assuming that individuals maximize net benefits as they themselves understand them. It takes a bit more sophistication to use expected values or median values than to maximize net benefits in familiar predictable circumstances.

Before moving on to these more realistic settings, it should be kept in mind that if people are on average correct in their expectations, then the models studied previously in the course also will be correct on average—although they cannot explain the existence of errors, risk, or uncertainty, nor the existence markets for goods and services that help to ameliorate risk and uncertainty or to increase knowledge. For example, insurance markets make sense only in risky environments. Libraries and universities exist only because people are less than perfectly informed. Self-help books make sense only if people are at least a little uncertain about their true preferences. Accountants are unnecessary every firm owner knew exactly what their marginal costs and marginal revenues were for every output and market condition. Planning advice is useful only if someone does not know all of the circumstances that will have to be addressed in the future.

In a contemporary market system, Schumpeterian innovation is one of the main causes of both risks and uncertainties. Innovations often affect many prices. They may reduce the value of one's human capital

(skills) or by change other relative prices in ways that were not anticipated (or anticipatable) .Innovations, such as libraries, universities, and the internet, can reduce ignorance and makes one's expectations more reliable predictors of the future, but also reduce income from "expert" advisors who may help others make better decisions. Even the beneficial effects of innovations may undermine long-standing patterns of life. Thus, dealing with random events, risk, and uncertainty are all natural parts of life in a contemporary commercial society. They are part of the reason that so many people change jobs and careers in their adult lives.

Both risks and uncertainties tend to be greater when one plans over relatively long periods of time, because the future can never be estimated exactly. Chapter 9 explores some of the consequences of or rational decisions that take account of both risk and time when making such long-term commitments. Examples include efforts to obtain a college degree, building marriage, a new factory, and undertaking research to create new products such as new medicines, self-driving cars, AI, or new apps.

II. Statistical Methods for Calculating Averages: Expected Values and Expected Net Benefits

The choice settings of interest in this chapter are ones in which perfect predictions are impossible because of random factors that have to be taken into account or because the scope of one's knowledge is insufficient to fully understand the processes that generate choice-relevant factors.

Some ideas from statistics help us (and individual decision makers) to think sensibly about possible outcomes and their average values when the probabilities of particular outcomes are known or knowable. The four ideas from statistics that we'll be using to model choices in "risky" settings are **probability distributions, expected (or average) values, variance, and sample averages**.[4]

A **probability function** lists all the possible outcomes of a random process and assigns probabilities to those outcomes that characterize the frequency that the random process of interest generates those values. The sum of the probabilities always adds up to 1. This is partly by assumption (as a defining feature of a probability distribution) and partly as a matter of logic—the probability that something will happen, where something is one of the known possibilities is exactly 1. So, the sum of the probabilities of each possible event or value must also equal 1.

A probability function maps possible events into probabilities. For example, the rolling of a fair dice of the conventional cube-shaped form with numbers 1 through 6 on its sides has a relatively simple probability function—namely P(1)=1/6, P(2)=1/6, P(3)=1/6, P(4) = 1/6, P(5)=1/6, and P(6)= 1/6. Such distributions are said to be "uniform" because the probability of each possibility is the same. "Normal" distributions, in contrast, are bell-shaped and individual possibilities (or range of possibilities) often have different probabilities. In this chapter, we'll mainly use probability functions where there are just two

[4] Probability theory began about five hundred years ago, although some ideas about probabilities were worked out by games of chance that are much older than that..

possibilities, because such probability functions (sometime called Bernoulli probability functions) are sufficient to illustrate some of the main implications of risky settings.

The two statistical ideas that we'll use most are the ideas of a "**true average"** or expected value, and a **sample average.** If you roll a dice 10 times, write down the individual numbers, add them up and divide by 10, you have calculated a particular **sample average**. In most cases, you'll have calculated a number that in most cases is around 3.5. 3.5 is the true average value generated by rolling a dice in very large samples or calculated using the **expected value** formula. The larger the number or rolls that you tabulate the closer the typical sample average tends to be to 3.5.

That is to say, that the larger the sample is the smaller is the variation in the sample averages that you calculate around the true average value. The "average" deviation of randomly generated series of numbers around the process's true average is called its "standard deviation". The standard deviation squared is the **variance of a distribution**, which is often easier to calculate than its standard deviation. The variance of a sample average tends to fall as the sample size increases.

Expected Values

In cases in which the probability function is known, one can also use mathematics or arithmetic to figure the average value that you would tend to observe from very large samples. For example, in cases where the probability of a every possible random event (P = P_1, P_2, P_3, …P_N) and the values are known (V= V_1, V_2, V_3, …. V_N), or can be accurately estimated, the average value or **"expected value"** can be calculated with the following formula.

$$V^e = \sum_{i=1}^{N} p_i v_i$$

Economists for the most part use the term "expected values" for that calculation, although the term average value would more accurately describe the value calculated, V^e.

The expected value of a single role of a dice is thus:

$$V^e = \sum_{i=1}^{N} p_i v_i =$$

(1/6)(1) + (1/6)(2)+(1/6)(3)+(1/6)(4) + (1/6)(5) +(1/6)6 = **3.5**

Notice that the "expected value" is actually impossible in this case, but 3.5 does predict the average value observed for very large samples of dice rolls. Thus, expected values are not always the "typical" value.

In continuous distributions like the normal distribution that are symmetric, the expected value tends to be in the middle of the distribution and is usually among the most frequently observed values. For the normal distribution, the average value is a typical value, rather than an impossible one. Indeed, it is the value that is most likely to occur.

III. *Using Expected Values to Make Decisions in Settings where Outcome Are at Least Partly Random (e.g. Risky Circumstances)*

This section illustrates how one can use expected values to help make decisions in settings where the results are random and generated by a reasonably well-understood probability function.

For example, consider a career choice after graduation. Suppose that you are offered a job at a new firm (a startup). If the company survives for more than 10 years you'll make 750,000 in wages and have stock options worth another 500,000 in stock options. If the startup fails, you'll have to switch careers which may not be easy and realize no stock options. BLS statistics suggest that only about 35% of firms survive 10 years. You also have a good job offer with another offer from an established firm that will pay out 1000000 in wages over the ten-year period (counting average raises).

To simplify, we'll ignore taxes and time discounting (which is taken up later in the next chapter). The expected value of the startup—which is the risky option—can be written as $V^E = (P^f)(V^f) + (P^s)(V^s)$, *where* P^f *is the probability of failure and* P^s *is the probability of success. There are only two possibilities so* $P^f + P^s = 1$. *(The superscripts are just "notation" [a way to distinguish the probability of success (*P^s*) from the probability of failure (*P^f*). They are not exponents in this case.)*

The BLS (Bureau of Labor and Statistics) statistics on startups suggest that $P^s = .35$, *so* P^f *must equal .65. If we substitute these into the expected value formula, we get:*

$$V^E, = (P^f)(V^f) + (P^s)(V^s) = (.65)(750{,}000) + (.35)(750{,}000 + 500{,}000) = 925{,}000.$$

In this case, the expected value of the startup job is less than the certain value of the conventional firm. So, if you want a job that maximizes your expected income, in this case, you should take the offer from the conventional firm (1,000,000>925,000).

Notice that if the odds of success were quite a bit higher, perhaps because of the talent of the entrepreneur or because the idea behind the new firm is especially good, you might reach the opposite conclusion. Suppose, for example, that the probability of success is $P^s = 0.60$ *and of failure is* $P^f = 0.40$. *In that case, the expected value of joining the startup is:*

$$V^E = (P^f)(V^f) + (P^s)(V^s) = (.40)(750{,}000) + (.60)(750{,}000 + 500{,}000) = 1{,}050{,}000.$$

Since there is a greater probability of success, the stock options become more valuable and are now enough to compensate for the lower salary paid by the start up.

Risk Aversion and Risk Premiums

When individual use expected money values to make decisions, they are said to be ***risk neutral****. They use expected income—for example—to choose among occupations that are otherwise very good substitutes for one another. There are no adjustments to take account of risk. Such a person would be indifferent*

between the job at the start up and the job with the job with the regular firm above, if the expected salaries including stock options were the same.

However, not everyone is risk neutral. Many folks would want to would prefer a "safe bet" to a "risky bet" if the expected values of the startup job and the conventional job are the same. Such a person would choose the conventional job. To take a job at the start up would require a "risk premium." They would take the startup job only if it had a sufficiently greater expected value. That extra amount is the risk premium.

Whenever individuals "demand" a premium (a risk premium) to accept a risk, they are said to be ***risk averse****. The more risk averse a person is, the larger the risk premium that he or she demands to accept a risky job, plan, or type of asset.*

Folks that prefer a bit of risk if the high side reward is great enough are said to be ***risk preferring.*** *Risk preferers are willing take on a lot of risk even if there is only a small chance of a large prize. (A lot of heroes on TV are risk preferers—but far more of them "win" on TV than actually would given the odds of success claimed by the script writers. Although to be fair to the script writers, uncertainty builds tension and the release of tension when the hero wins is enjoyed by their viewers—partly because it is unrealistic and so in a sense special—like winning a lottery. Knight's entrepreneurs are also risk preferers.)*

IV. *Buying Products of Random Quality*

Consider a consumer's decision to purchase products when the probability of a defect is known (or can be estimated). A consumer's willingness to pay for a product can be characterized with his or her total benefit curve for that product. The highest price that a consumer is willing to knowingly pay for Q units of a good is a just bit less than the total benefit generated by that good (his or her reservation price), measured in the currency used in his or her country—dollar, euro, yen, etc.—or in units of satisfaction or pleasure—utility. Successive units of goods and services normally produce additional benefits and so total benefits tend to rise with the number of units obtained.[5]

Recall that the additional benefit generated by one additional unit of a good or service is called its marginal benefit and can be interpreted as the highest price that a person will knowingly pay for the Q-th unit of a good. Economists normally assume that marginal benefits fall with quantity, so marginal benefit curves tend to slope downward, and total benefit curves tend to rise more slowly as the quantity acquired increases. The highest valued use receives the first unit, the second highest, the second, and so on.

[5] This subsection is largely taken from Solving Social Dilemmas (2022), with some revision to make the discussion developed there more relevant for this course.

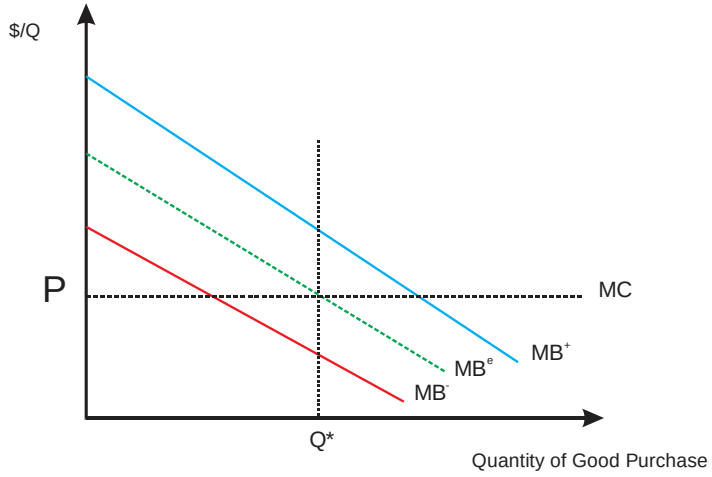
Next we introduce some quality uncertainty. When the quality of a product varies, the marginal benefit received from a given unit cannot be known with certainty unless defects are obvious at the point of sale.

To simplify a bit, we'll assume that only two degrees of quality exist: perfect and defective. Successive units of goods or services have one marginal benefit associated with them when they are perfect and another when they are defective. The marginal benefit curve associated with defective units lies below that of perfect units.

To further simplify, assume that consumers are risk neutral—which means that consumers use statistical averages when determining the "expected value" of a particular unit of a particular good. If probability of a "perfect" unit of a good is P and a "defective" unit of the good is 1–P, the expected or average marginal benefit (MB) from a particular unit of the good is $P*MB(Q)^{+} + (1–P)MB(Q)^{-}$ where $MB(Q)^{+}$ is the marginal benefit of the Q-th unit of the good when it is perfect and $MB(Q)^{-}$ is its marginal benefit when defective.

Figure 8.1 illustrates a consumer's decision about the number of units of a good to purchase when its per unit price (marginal cost) is P. The expected marginal benefit curve (MB^{e}) lies between the marginal benefit curves of the perfect and defective units of the product. If P=0.5, then the expected MB is exactly midway between the MB^{+} and MB^{-} curves. As the probability of perfect units increases, the MB^{e} curve moves closer and closer to the MB^{+} curve.

Figure 8.1 Maximizing Expected Consumer Surplus



The risk-neutral consumer purchases all of the units for which the expected marginal benefits are greater than their marginal costs, which implies that Q* units of the good are purchased by the consumer illustrated. Economists refer to the difference between the total benefit received from Q units of a good and the total cost of Q units of a good as its "net benefit" or "consumer surplus" [CS(Q) = TB(Q)–TC(Q)].

The total expected benefits of Q units of the good is the area under the MB^e curve from 0 to Q, and the total cost of Q units is the area under the marginal cost (MC) curve from 0 to Q, so the expected consumer surplus realized by purchasing Q* units of the good is the triangular area between the MB^e and MC curves from 0 and Q*. The area of this triangle is a measure of a consumer's expected net gains from purchasing Q* units of the good. However, the amount realized is randomly distributed around the expected value, because the exact proportion of good and bad units varies with the sample taken (the actual units bought). In this case, this is just a matter of good or bad luck, not ignorance on the part of the consumer if the probabilities of good and bad units have been accurately estimated or is known with certainty.

The quantity characterized by the intersection of the MB^e and MC curves, Q*, is the quantity of the good that maximizes this buyer's expected consumer surplus—which is to say the average CS realized by a long series of such purchases.[6]

Notice that as the average quality increases, the expected marginal benefit curve shifts toward the higher curve (in blue)—the one characterizing perfect units, and the quantity purchased increases. If the average quality falls, then the expected marginal benefit curve shifts back toward the lowest MB curve (in red) and the consumer's purchases fall.

Notice, we are assuming that different degrees of quality occur randomly. They are not efforts by the seller to defraud their consumers—although such cases do exist. They are simply consequences of variations in quality associated with their production methods. An agricultural product may receive more or less sun, be growing on a more or less fertile bit of land, be genetically a bit better or worse tasting instance of the fruit or vegetable being purchased. The equipment used to produce the good of interest may be more or less precise and so the items produced may vary in quality.

However, if the typical firm in an industry increases the average quality of the goods produced, the demand for their products would increase because the expected marginal benefit increase for their customers. Contrariwise, if the average quality declines, their market demand curve would shift back to the left (fall) because consumers realize smaller expected marginal benefits from their purchases of the good or service of interest.

---

[6] In the case usually assumed by economists, consumer surplus is maximized at the quantity where marginal benefit equals marginal cost. This follows from elementary calculus. If CS(Q)=TB(Q)–TC(Q) then CS is maximized when the first derivative of CS with respect to Q is zero or when CS'=TB'–TC'=0. TB' is marginal benefit and MC' is marginal cost. The mathematics are not important for this chapter but are the simplest way to explain why finding the quantity that sets MB=MC maximizes consumer surplus—the net gains from trade for a consumer.

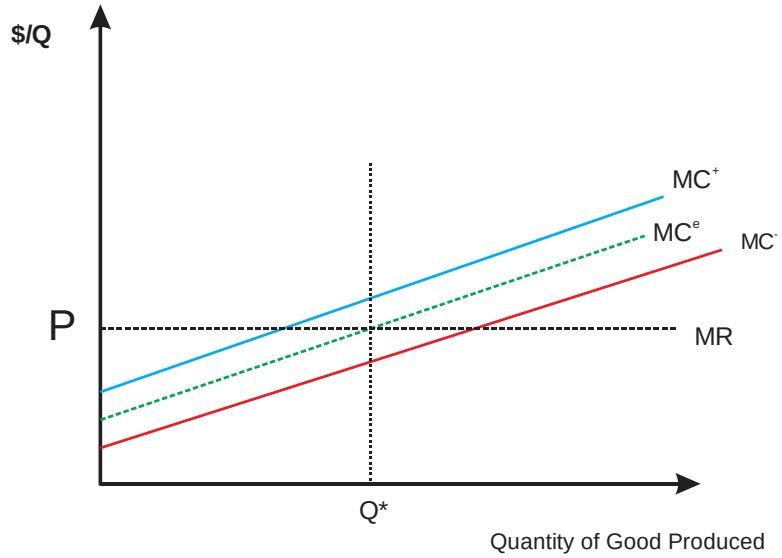
This provides firms with an incentive to invest in quality control whenever the additional revenue produced by increased quality and sales is more than sufficient to pay for the cost of the quality control. Perfection, however, is normally too expensive to be worth pursuing.

V. *Producing Products with Random Costs*

The choice faced by firms with random costs is very similar to that of consumers facing random quality. If there are just two states of the world that influence a firm's marginal cost function, and the probability of being in the high-cost state is known (or can be estimated with reasonable accuracy), a risk neutral firm will produce at the output level that maximizes ***expected profits*** *rather than certain profits as in the models in prior chapters.*

Figure 8.2 two illustrates such a choice for a "price taking" firm, which is a firm that has many rival producers for essentially the same good or service. It has a horizontal marginal revenue equal to the prevailing market price for the good or service that it produces and sells. Uncertainty of this variety can be the result of weather, which may disrupt the firm's production in various ways—but in a more or less predictable manner—as for example very cold weather may make it more difficult to harvest trees from a tree farm, or particularly stormy weather may interfere with fishing, or early frosts or droughts may reduce a farm's output level for given expenditures on fertilizer and tilling for particular crops.

Figure 8.2 Maximizing Expected Profit



Geometrically, expected marginal costs are a curve (or line) between those associated with the high marginal cost and low marginal cost states of the world. Producers maximize expected profits by producing outputs where ***expected marginal costs*** *equal (expected) marginal revenues.*

The higher the probability of a "high cost" setting, the closer the MC^e curve is to the higher of the two marginal cost curves (MC^+) and the smaller is the firm's expected profit maximizing output. Changes in technology that makes crops more robust may have the opposite effect, moving expected costs toward

the lower MC curve (by increasing the average output associated with expenditures on seed, tilling, and fertilizer etc.). Such innovations reduce rather than increase downside risks.

VI. *Producing Products with Random Prices*

Market demand or market supply may also be affected by random factors that together jointly determine the price at which a firm's output is sold. Prices in such cases vary in an unpredictable random manner. This, for example, is true of all agricultural products. The world-wide market supply of agricultural products is affected by weather patterns throughout the world. For many other products, variation in demand cause prices for various raw materials to vary widely, because their supply is relative inelastic (their supply curves are nearly vertical) in the short run.

In cases where products take time to be produced, firms cannot simply adjust their day-to-day output for the "spot" price of the service that they are selling. Their time-intensive production processes require firms to produce quantities and at costs that they expect to be most profitable, but they cannot really know for sure what the prevailing market price will be at the time that their products are ready to be sold.

The first case examined in this section is straight forward. If there are just two prices that might prevail in the market of interest, a high one and low one, then their expected marginal revenue is simply the expected or average price that prevails, as with $P^e = f\,P^- + (1\text{-}f)P^+$*, where* P^- *is the low-price choice setting and* P^+ *is the high price choice setting*[7]*. The expected marginal revenue curve is just a horizontal line at the expected price,* P^e*. Firms that are aware of the price variation will produce output levels that maximize profits at the expected price. (As a practice exercise, draw a diagram of this choice setting, illustrate the choice using expected marginal revenues and its effect on a firm's expected profits.)*

The case in which a firm faces its own downward sloping demand curve is also fairly easy to characterize when the revenue effects are caused by random shifts in the demand curve rather than changes in its slope. If the demand curve shifts in a random way between high demand and low demand states, the expected (average) demand curve and their associated expected marginal revenue curves will be in between those in the high and low demand states.

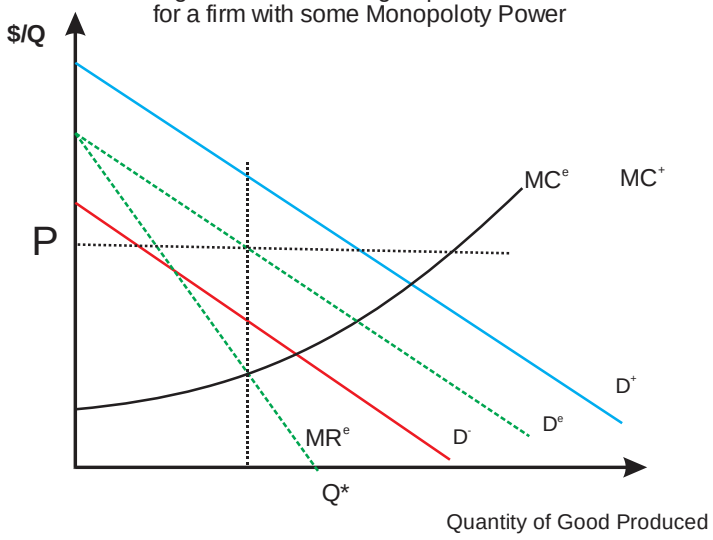
For example, if each demand curve occurs with probability 0.5, then the expected demand curve will be halfway between the high and low demand curves and the expected marginal revenue curve will be halfway between the marginal revenue curves associated with their respective demand curves.

Figure 8.3 illustrates such a choice setting, but it includes only the expected marginal revenue curve to reduce the number of lines that readers need to keep track of. (Recall that with a straight-line demand curve, its associated marginal revenue curve is downward-sloping line halfway between the vertical axis

[7] Keep in mind that these models can easily be extended to settings where lots or outcomes are possible. The formula for calculating expected values is sufficient for any countable number of outcomes. Two outcome illustration are used to simplify the discussion and drawings—and without much loss of generality.

and the demand curve of interest. In this case depicted, the expected marginal revenue curve is halfway between the vertical axis and the expected demand curve.

Figure 8.3 Maximizing Expected Profit for a firm with some Monopoloty Power



The firm chooses the output that sets expected marginal revenue equal to expected marginal cost. It then sells its output at the price associated with the actual conditions, which will vary according to which demand curve it actually faces after its production has taken place. On average this will be the price associated with the expected (average) demand curve. The average price is the one that is depicted in the diagram. The actual selling prices will tend to be higher or lower than the average price at which the product is sold.

Effects of Risks on the Extent of Markets

Notice that in each of the cases analyzed so far in this chapter, the effect of "downside" risk is to reduce sales and output, as consumers and firms "hedge" their "bets," and buy less or produce less than they would have in circumstances where only the "good" outcome are possible. ***Thus, these models and their associated diagrams can be used to understand why downside risks tend to reduce the size of markets and extent of market networks.***

The effects of risk on the size and extent of markets can be reduced in various ways. *For example, decision makers may adopt more flexible consumption or production plans that allow one to rapidly adjust to new circumstances. Travelers may, for example, always pack an umbrella for a long trip. Firms may engage in short-term rather than long-term contracting. Both may accumulate rainy day funds to help cope with unexpected unpleasant surprises.*

VII. *Risk Premiums and the Demand for Insurance*

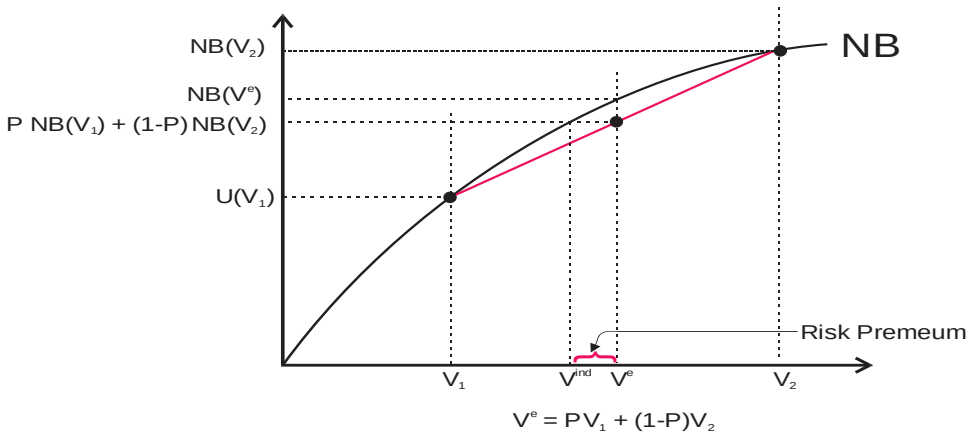
In cases in which a risk can be insured (and there are a variety of insurance and insurance like products that can do so other than the ones sold by insurance companies) it will often be worthwhile to purchase

insurance and so reduce the risks associated with long term commitments, such as purchasing an automobile or house.

Risk-averse individuals are willing to pay a "safety premium" to realize a certain return rather than a risky return. That safety premium provides the basis for their ***demands for insurance****.*

Figure 8 provides the geometric logic and geometry behind the existence of risk premiums and thereby on the demand for insurance. The curve depicted is the ***total net-benefit curve****. Diminishing marginal benefits and more or less constant marginal costs, imply that the total Net Benefit Curve has a shape similar to this one up to the point where maximum net benefits are realized with certainty. (When the payoffs are risky, a risk averse person will purchase fewer units than that best case NB-maximizing amount, so it is this portion of the NB curve that is relevant here.)*

Figure 8.4 Expected Net Benefits,
Risk Aversion, and Risk Premiums



The net benefit curve of a risk-averse individual (or profit curve for a risk averse firm) is upward sloping and curved so that it becomes flatter as the choice variable (value, quality, or quantity) increases (e.g. as the values along the horizontal axis increase). Such curves are said to be ***strictly concave****. It turns out the more rapidly NB curves flatten out as one moves to the right, the larger their risk premium is, and so the more risk averse an individual is.*

The diagram illustrates the case where a risky purchase or investment may have two values in the future (as, for example, the value of a house that might or might not catch fire or be affected by flood or landslide). It will either have value V_1 or value V_2 –which one is determined by some random process, with P being the probability of V_1 and (1-P) being the probability of V_2.

As P varies from one to zero, the expected net benefit traces out the straight line that lies below the NB curve (the net benefit curve). The expected net benefit is the highest price that a risk-averse consumer would pay for this risky asset. Notice that this amount is always below the NB line because of the

curvature of the NB curve. If P is 0.5, the expected or average net benefit will be at exactly the midpoint of that line connecting the NB(V_1) and the NB(V_2).

Note also that the average value of the asset (V^e) lies between V_1 and V_2 on the horizontal axis. The net benefit of the various possible outcomes differs from NB(V^e) because of the way the consumer surplus (or reservation prices) varies with different values of V for risk averse individuals. (If V^e always equaled NB^e, the individual would be ***risk neutral****, and the NB line would be a straight line in the range depicted.)*

A risk-averse person is only willing to pay an amount that is less than the average value of the risky asset. That is to say, he or her risk aversion implies that the expected net benefit of a risky investment is less than a certain investment that returned the average value, V^e, with certainty. This is shown on the graph by the fact that at V^e, the expected net benefit is lower than the value on the NB curve at V^e.

The equivalent certain value to the risky investment can be found by going from NB^e over to the NB curve and then straight down to the V axis. That value is labeled V^{ind}, because the individual would be indifferent between that amount (with certainty) and the risky investment.

Notice that that value is less than V^e. ***The difference between Ve and V^{ind} is the risk premium that an investor would require to accept the risky investment*** *(V_1 with probability P and V_2 with probability 1-P) rather than a certain one that returns V^{ind}.*

It is this risk premium that creates the demand for insurance. *It is the highest price that an investor, consumer, or firm would be willing to pay to totally avoid the risk associated with the investment (asset) characterized in the diagram. Thus, it might also be called the individual's "safety premium," the highest amount that he, she, or it is willing to pay to avoid a risk rather than bear it. If full insurance can be purchased for less than that amount the investor would be better off buying it than bearing the risk personally.*

Thus, the risk aversion and its associated risk premiums is the ultimate source of the demand for insurance.

Individuals are not willing to pay more than their "safety premium" for complete insurance but would be happy to pay less than that (because then he or she would gain expected net benefits from the insurance.

VIII. *The Supply of Insurance*

In principle, there are gains from trade between more risk averse people and less risk averse people, since less risk averse people are willing to pay a higher price for a risky asset (with a well-understood risk) than a more risk averse person. So, one market response to risk, is that risky assets tend to "move" from more risk-averse persons and firms to less risk-averse firms. For example, risk-averse persons would be out bid for ocean-front properties along the South Atlantic Coast of the United States because of the risk of hurricanes. The risk averse lovers of oceans would prefer property that is inland a bit, but perhaps

in walking distance of the beach, or a really solidly built beach-side house over one that is a bit ramshackle.

Another effect of risk is the emergence of insurance markets. *We'll focus on the standard ones that consumers deal with rather than various futures contracts and similar investments that some investors use to reduce their risk. It turns out that insurance markets can be quite profitable because of the statistical properties of large samples. So, to understand the basic logic of an insurance company, first we must understand a bit about those statistical properties.*

On the Properties of Sample Averages

Every random variable has a probability function associated with it that assigns probabilities to the possible values of the random variable of interest.[8] *Every probability function, thus, has a range of values. A measure of the breadth of that range is called the* ***variance*** *of a probability function. Given the probability function over variable X,* ***the variance of that probability function can be calculated as****:*

$$Var(X) = \sum_{i=1}^{N} p_i(X_i - X^e)^2$$

A related measure of the breadth of the outcomes associated with a probability function is its standard deviation, which is simply the square root of a probability function's variance. $S.D.(X) = (\mathrm{Var(X)})^{0.5}$ *. (Greek letters are sometimes used for the mean or average value of a probability function as with mu, μ, and the standard deviation of a probability distribution is often written as sigma,* ***σ****, and its variance as sigma squared,* ***σ²****. But we do not need that notation for the purposes of understanding the supply of insurance.)*

What is important about variance (and stand deviations) for insurance markets is that the smaller those values are, the narrow is the range in which most values of the probability distribution fall and the more an asset resembles one without risk.

It turns out that the larger a sample size is, the more narrowly the values of a sample average are distributed. The variance of a sample mean, $\left(\frac{\sum X}{N}\right)$, falls as the sample size, N, increases.

$$Var\left(\frac{\sum X}{N}\right) = Var(X)/N = \left(\frac{1}{N}\right)\left(\sum_{i=1}^{N} p_i(X_i - X^e)^2\right)$$

[8] Technically this is only true of probability functions that have only discrete values. For random variables that have continuous values, a probability density function is used to characterize the probabilities that values fall within a range. The "normal distribution" is an example of such a probability density function. However, the intuition that is associated with probability functions extends pretty well to probability density functions, so

Selling insurance is like sampling, because each insurance policy for a similar house or asset, simply another instance of the random process that is generating the risk (fire, flood, loan default, etc.). *So, a company the sells lots of insurance for risks that are well understood, pays out, on average, the sample average of its collection of insurance policies. The insurance company's risk is basically the variance about its sample average, which becomes smaller and smaller as more and more insurance policies are sold.*

So, for large insurance companies, the payout for insured losses is, on average very predictable. It is approximately the true average value of the losses generated by the random phenomena being insured (as with fire or flood insurance). In other words, "pooling risks" essentially eliminates the risk (random nature) of the loss being insured for by the insurance company. (The initial risk sometimes said to be eliminated or greatly reduced by pooling.)

The Profitability of Insurance Companies

This sampling effect allows even risk averse firms to profit by selling insurance if they can sell insurance and settle claims at a reasonable cost. Unlike the owner of a house, whose may burn down or not in a given year, the insurance company essentially pays out the average or expected loss every year.

Since those demanding insurance are willing to pay up to the average loss plus the safety premium as long as its selling and settling costs are lower than the risk premiums that its customers are willing to pay, selling insurance can be profitable. In such cases, there are potential gains to trade between homeowners and insurance companies. Competition between insurance companies keeps their prices down to more or less ordinary rates of return for the products that they sell. Once up and running, insurance markets often resemble ordinary competitive markets.

The emergence of insurance markets is one of the implications of commonplace risks such as those associated with owning a house or car—or even offering loans (banks often insure against non-performance). Such markets would be unnecessary in a world in which everything was certain and well understood by consumers and firms.

IX. *Rational Ignorance, Markets, and Specialization*

Prior to World War II, there was very little consideration of the role of information in choices by consumers and firms and thereby on the market equilibria. In 1961, George Stigler wrote an important paper on the economics of information in which he suggests that gathering information is analogous to sampling a random distribution. Sampling allows an individual to know with greater precision the random process generating market prices and many other things. The more one samples, the more precise is that understanding—the more precise is one's estimates of the mean, variance, and trends in the phenomenon of interest.

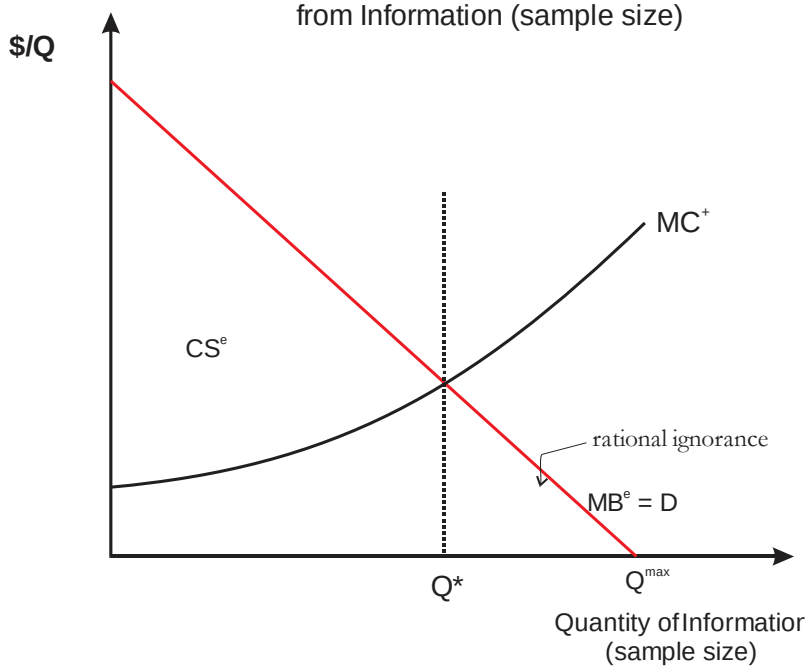
The extent of that information can affect market equilibria. For example, if a sufficient number of consumers search for the best price, they tend to induce market prices to converge on the lowest ones

that they have been able to find. This is one of the mechanisms that tends to support the "one" price equilibria of the demand and supply models that we studied in the first half of the course. On the other hand, if consumers do not search for good prices, there will be less demand pressure on firms to sell at low prices and the range of prices among shops will be broader. Although this seems obvious now, Stigler won a Noble price, partly for that paper.

The basic idea is that seeking information has expected benefits and expected costs, so individual's rationally economize on information. (A very similar idea was worked out by Anthony Downs at about the same time with respect to voter information about candidates and public policies.) As a consequence, neither consumers nor producers nor voters nor students, etc. know as much as possible about most things. They remain rationally ignorant of many potentially useful things because they only have so much time to devote to collecting and analyzing data.

Figure 8.5 illustrates the logic of an individual's choice of the amount of information to gather from a randomly generated distribution that he or she is aware of. The marginal cost line represents the opportunity cost of using time to gather and process the information of interest (for example, about the price of cell phones or their capabilities). The marginal benefits are "expected" because one never really knows what one will learn until one does so.

Figure 8.5 Maximizing Expected Net Benefits from Information (sample size)



Rational ignorance (and natural ignorance—ignorance that one has not chosen) is an important phenomenon in any setting in which information is not naturally very complete and reliable. In some cases, ignorance can induce systematic errors. In others, ignorance simply implies that one's estimates are less accurate (lower variance) than they could have been, although they are still unbiased (the expected values are correct rather than different from the actual value).

In some cases, the marginal opportunity cost is also informational—that is, the best alternative use of one's time may be learning about something else. If you have T hours to divide up across informational investments, a rational student will allocate time so that the marginal benefit from time is the same for each use. This may imply very different amounts of time devoted to different types of study because the expected marginal benefits may differ substantially.

For example, the marginal benefits from learning more about one's expected major has relatively larger marginal benefits in terms of adding to skills that you are likely to find useful in your future career than in studies that have no direct entertainment value or effect on one's future earnings in one career.

Markets make use of the information that individuals have and what they lack. Knowledge and skills that are scarce relative to demand tend to have higher prices associated with them. This in turn encourages other students to specialize in acquiring such knowledge and skills. In the very long run, prices may thus cause the market value of skills that require equal investments to acquire to converge on roughly the same wage rates unless the capacity to acquire skills is itself scare, because of variations in talent, diligence, or regulations that make it difficult for some persons to acquire the skills with the highest return.

Specialization, as an economic activity, often is associated with focusing most one's attention on specific types of information that are useful for one's career (or expected career). Thus, to a considerable extent, it tends to reflect choices about which kinds of knowledge to acquire and which types one chooses to remain ignorant of.

Chapter 9: Intertemporal Choice

I. Introduction

Most neoclassical models are timeless in the sense that "time" is left out of the model. That is not because time is never important, but because for some purposes leaving time out of a model or analysis does not undermine our understanding of the puzzle or phenomena being addressed. Economic choices and consequences often in the present or near future. If a consumer decides that he or she will this month's wages in a particular way, the fact that the actions associated with that decision take place during the month (or perhaps in the next month) does not materially influence the optimization process that generates the choices made.

However, there are cases in which time matters. Time cannot be ignored when actions are taken today that affects one's possibilities in the future—if one is rational and forward looking. Indeed, the phrase "forward looking" implies that rational beings take some account of the effects of one's present actions on one's future possibilities. For example, a consumer's decision to spend a certain amount of money in the future may affect the extent to which he or she works (and how they work) today. One may save up for a vacation, an automobile, a house, a college education, or retirement. Such long-erm plans simultaneously affect consumption decisions today and in the future. If one decides to save money today, then less money is spent today than otherwise would have been spent, and more will be spent in the future.

Or, a consumer may engage in the opposite type of behavior. He or she may borrow against future income to pay for capital goods (computer, automobile, house, etc.) or for ordinary consumption today. Such choices increase current expenditures and reduce future ones (apart from loan payments). In the average course of a life, wage rates rise during the first 20-30 years of one's career, so an individual's future wages are very likely to be higher in the future than in the present. Borrowing allows one to shift a bit of future income to the present.

On the other hand, a theory that models such choices has to account for the fact that there is generally a somewhat greater reduction in future consumption than gained in the present because interest payments will have to be paid until the loan is paid off.

The same logic applies to decisions made by economic organizations. Firms and other organizations may also divert current income into future enhancing uses. For example, they may invest in new equipment or attempt to build a cash reserve today that can be used to address nasty surprises that may arise in the future, or simply to smooth out predictable fluctuations in their cash flow during the course of their planning horizon. Firms may also borrow against their assets or future profits to pay for capital goods that will be used in production. Those capital goods must, in a sense, pay for themselves, which is to say that they must increase profit flows through time sufficiently to be used to pay off the loan—unless the decision was unwise, or bad luck (uncertainty) reduces the demand for their products.

Many businesses have sales patterns that are connected with the seasons, and firms try to hold onto their employees for the entire year. The routines required for efficient team production benefit from the stability of the team members. Workers, on the other hand benefit from more predictable cash-flow which means that they will accept a lower wage than they otherwise would have (otherwise a risk premium would have been required). Most firms use "a wages fund to keep their teams on the payroll. They save some net income in the most profitable times in the year to pay their employees in the least profitable times. Examples of markets with cyclic demands include the market for toys, holiday foods and beverages, and markets for housing (because of school year effects).

This chapter develops models of choice and markets that characterize such intertemporal decision making.

To do so requires somewhat "stronger" assumptions about how individuals and firms make intertemporal decisions than needed in part 1 of the course. The focus remains on net-benefit maximization. However, as in the previous chapter, a specific assumption is added to analyze how individuals think about the future. Namely, it is assumed that consumers and firms use the **present discounted value** method to think about future benefits, costs, and net benefits—the algebra of which emerges naturally when it is recognized that both borrowing and saving have opportunity costs.

However, as with the assumption that individuals use "expected values" to make decision in risky settings, the assumption that individuals use "present discounted values" for intertemporal choices is reasonable and a useful "first approximation," but it is not as general as the basic "maximize net benefit" assumption used in the first half of this course. Some individuals will take account of time and risks in other ways, but most will behave in a manner that is more or less consistent with these models of forward-looking choices in more or less risky settings.

II. Markets for Savings and Borrowing

In the chapter on capital accumulation and growth, we briefly modelled markets for savings and loans. For this chapter, we'll need to dig a bit deeper into how such markets operate.

Borrowing is impossible unless there are savers, because borrowing requires the current purchasing power (income) of some consumers and firms to be shifted to other consumers and firms. If no one wants to save, no one will be able to borrow. On the other hand, the interest and other returns that one can realize by saving is possible only because there are borrowers that are willing to pay a premium to have money now, rather than the future.

Together savers and borrowers create a market for both savings and loans. The price paid for savings is normally called "interest" or "interest rate." The higher the return on savings (the higher is the interest rate, adjusted for risk and expected inflation) the more inclined savers are to put aside money. On the other hand, for borrowers, the interest rate is the cost of borrowing, the higher interest rates are, the less money they are inclined to borrow.

Because there are many places where one can save and many places where one can borrow, ordinary supply and demand curves can be used to characterize this market. Figure 9.1 characterizes a simple "direct" market for savings and loans.

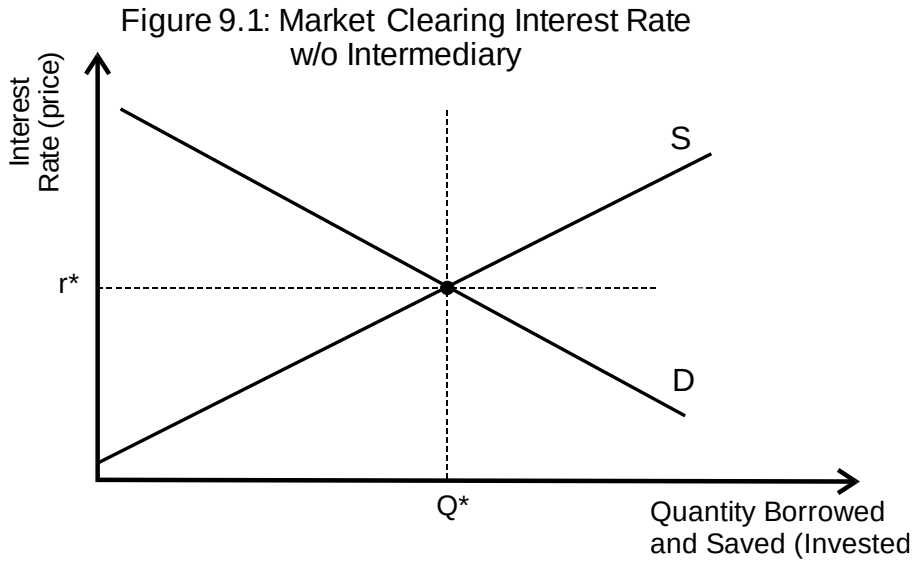


Figure 9.1 represents the equilibrium that emerges for a given class of loans. If this is the imaginary "risk free" market for savings and loans than the interest rate that emerges is called the risk-free interest rate. It simply represents the value of having "a dollar" today rather than a dollar in the future. That difference is sometimes called the subjective rate of time discount or discount rate. In the "direct" case, the interest rate earned on savings is the same as that paid by borrowers and the interest rate is the discount rate of the marginal saver and borrower. (Other savers have lower discount rates and other borrowers have higher discount rates.) If the loans are risky because a significant fraction of borrowers will not pay back their loans, then a risk premium is added to the risk-free interest rate to take account of the risk of default.

Direct saving and loans take place in bond markets where savers buy bonds from seller of bonds. However, not all saving and loans are done directly. Often there are third parties involved that serve as intermediaries between savers (lenders) and borrowers. For most consumers, banks are (or were) them most common intermediary between savers and borrowers.

Banks provide a variety of services to both lenders (savers) and borrowers. For example, banks often "guarantee" a fixed interest rate for savers, which reduces their risks. They can do so, because they make lots of loans and can accurately judge to risk of default on loans that they make. In this market, banks are like insurance companies, and their expectations about losses from defaults have a much lower variance than a direct lender would tend to have. Banks also provide "liquidity" for their depositors (saver=lenders), which is to say, rather than having all of their money tied up in loans to

borrowers, banks keep a cash reserve that can be used to pay off savers that want to withdrawal their savings as cash or use it to pay bills (as with checking accounts).

Similarly, those borrowing can borrow at lower rates (especially if they have assets that can be used to guarantee the loans [collateral]) for the same reason. If banks can judge risks accurately, their overall return from their entire loan portfolio is quite stable and on average, the expected value of the loans made. Banks, thus, require a smaller risk premium than would be required in direct lending from savers.

(Here we are ignoring other risks associated with government policies and business cycles, topics taken up in macroeconomics.)

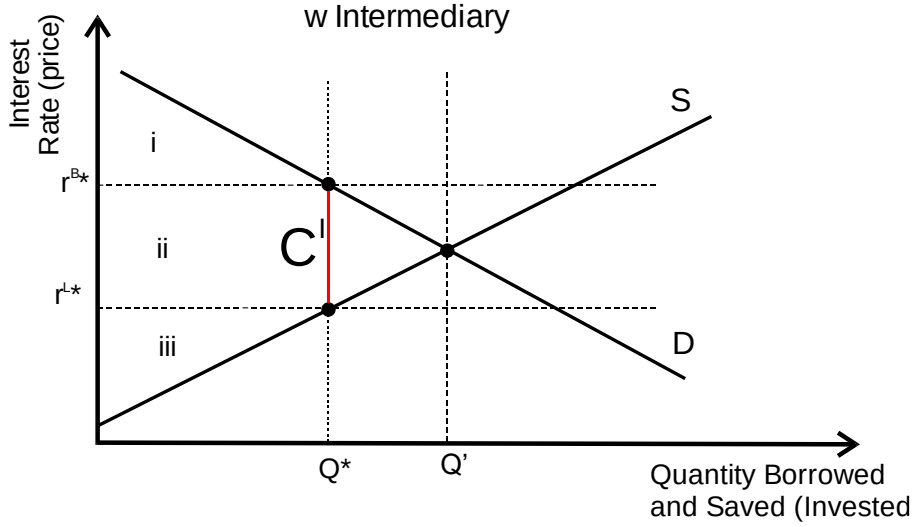
Intermediaries have costs and expect at least an average return on their investments. These markets are competitive, and market forces tend to limit their returns as in the markets studied in the first part of the course.

Nonetheless, the costs of their services come between the interest rate paid to depositors (saver-lenders) and borrowers. A diagram of the savings and loan markets when intermediaries are used (as they usually are) resembles that of the tax diagrams developed in chapter 5—there are three parties to the transactions in this market—buyers (borrowers), suppliers (savers), and intermediaries. Figure 9.2 illustrates such a saving and loan market in equilibrium.

The marginal cost of intermediary services is denoted as C^{I}. Area "i" is the net benefit or profits of those taking out loans and "iii" is the net benefit or profit of those saving/investing through the intermediary (here banks). C^{I} reduces both savings and loans over what they would be in a world without transactions costs, but given real-world transactions costs and risks, they actually tend to increase the size of saving and loan markets over what they otherwise would have been if each saver-investor had to find borrowers, assess the risks, and write up contracts.

"Middlemen" in financial markets normally take on at least some of the risk from such intertemporal transactions and reduce transactions costs for both borrowers and lenders. It is their risk reduction and liquidity services that induce both savers and borrowers to use their services.

Figure 9.2: Market Clearing Interest Rate w Intermediary



If all loans were risk free, few borrowers or lenders would use the services of an intermediary. However, when risks are non-trivial and liquidity is useful, their services can be very useful to both.

If the risks of default were higher, the supply curve would be lower (be to the left of the supply curve depicted) which would cause interest rates to be higher for both lenders and borrowers. The rates of return and cost associated with loans would reflect the risk premium demanded by savers (and intermediaries) associated with different kinds and numbers of loans and the costs of attracting depositors and providing them with liquidity. (For example, house and automobile loans are pretty similar, but loans to startups firms are all quite different and generally far riskier.)

Shifts in the demand and supply of loans and savings affect market interest rates in much the usual way. But the shifts in those two curves are driven by somewhat different considerations than in the previous cases. Borrowing by firms reflects expected net benefits from capital investments. Savings by investors and savers reflect expectations about alternative rates of return and risks associated with different investments, and also future consumption plans. For example, if average longevity increases, the supply of savings tends to increase because forward looking individual will put aside more funds for retirement. Changes in the risk of default will also shift supply curves, as would increases in personal income.

It turns out that interest rates play a role in determining both the cost and benefits of decisions with respect to savings and loans—partly through effects on the value placed on future consumer surplus and profits.

Most of the rest of this chapter focuses on the use of the **present discounted value** formulae for intertemporal calculations of benefits and costs.

III. Intertemporal Choice: Time Discounting and Present Discounted Values

The most general way to think about "present discounted value" is to think about the amount in the present (PV) that you would be indifferent to having rather than some other value (F) in in T years. One way to estimate this, is to think in money terms. That is, one can calculate the amount of money (PV) that one would have to invest today to have F dollars T years in the future.

We'll initially use two letters to indicate present discounted value (PV) and normally write "present value" rather than present discounted value from now on.

If the interest rate or rate of return is r, one can just apply the compound interest formula to determine how much money an investment of PV at interest rate r will generate in T years—its future value denoted as F_T. **$PV (1+r)^T = F_T$**

Solving for PV yields **$PV = F_T/(1+r)^T$** which is the basic formula for calculating the present discounted value of some value F that is realized T years in the future. Notices that the present value of F_T is always smaller than the actual value in future dollars because interest rates are greater than zero, r>0.

To make the calculation more concrete, suppose that F is \$20,000 that T=2 and R=3% or 0.03. In that case,

$$PV = (20{,}000)/(1.03)^2 = \$18{,}851.92$$

Notice that the PV of future value F goes down when the interest increases and/or when the time period increases.

The PV of \$20,000 in two years at an interest rate of 5% is

$PV = (20{,}000)/(1.05)^2 = \$18{,}140.59$

The PV of \$20,000 in ten years at an interest rate of 5% is

$PV = (20{,}000)/(1.05)^{10} = \$12{,}278.27$

If one thinks in purely financial or money terms, one would be indifferent between \$12,278.27 today and \$20,000 in 20 years if the 20-year interest rate is 5%.. This assumes that no inflation occurs (or that F_T is stated in inflation adjusted terms in current dollars) and that there is no risk involved about whether the future amount will be paid or not.

When one takes account of inflation either everything should be in inflation adjusted (real) terms (including the interest rate and future values). The real interest rate is the nominal or state rate of interest less the average annual inflation rate over the period of interest. Or everything should be in nominal (ordinary dollar) terms. When there is the risk that amount F will not be paid, then one needs to also take account of the risk using the expected value methods that were developed in chapter 8.

Many decisions involve long term flows of costs and benefits that need to be evaluated by a decision maker or group of decisionmakers. These flows are easiest to compare if one can construct a common measure for the purposes of comparison. The present value of a series of benefits and/or costs through time is one such measure. It is the amount, **P**, that you could deposit in a bank at interest rate **r** and used to replicate the entire stream of benefits or costs, F_1, F_2, F_3, ... F_T. That is to say, you could go to the

bank in year 1, and withdraw the amount (B_1) for that year, return in year 2, pull out the relevant amount for that year (B_2) and so on ... . The present discounted value of a series of future amounts is simply the sum of the present values of each element of the series—which is calculated as above.

DEF: Let **Ft** be the value of some asset or income flow "**t**" time periods from the present date. Let **r** be the interest rate per time period over this interval.

The present discounted value of F_t, now written as $P(F_t)$ is

$$P(F_t) = F_t/(1+r)^t$$

The present value (now written as P to reduce notation) of a series of future income flows (which may be positive or negative) over **T** years when the interest rate is **r** (as a fraction) per period is simply the sum of the present values of each element of the series of future amounts:

$$\boldsymbol{P = \sum_{t=1}^{T} F_t/(1+r)^t}$$

The present discounted value of any series of values is the sum of the individual present values of each element of the series.

This formula always "works" but it is somewhat cumbersome to use as the planning period, T, becomes relatively large. Another useful formula is one that characterizes the present discounted value of a **steady flow** of values on off into the future for the next T years. In cases where a **constant value** is received through time, e.g. $V_1 = V_2 \ldots = Vt \ldots = V_T = v$, a bit of algebra allows the above present value formula to be reduced to:

$$\mathbf{P = v\,[\,((1+r)^T - 1)/r\,(1+r)^T]}$$

Derivation of the Above Formula (optional)

This formula can be derived as follows:

Multiply $\boldsymbol{P = \sum_{t=1}^{T} V_t/(1+r)^t}$ by (1+r) which yields: $(1+r)\boldsymbol{P = \sum_{t=0}^{T-1} V_t/(1+r)^t}$

Subtract P from (1+r) P which yields: $rP = v\,[\,1/(1+r)^0 - 1/(1+r)^T)]$

(Note that all the terms in the two sums are the same except for the first and last one, so the middle terms all cancel out.)

Recall that $1/(1+r)^0 = 1$ so $rP = v\,[\,1 - 1/(1+r)^T)]$

Putting the lefthand term over a common denominator yields

$$rP = v\,[\,(1+r)^T - 1]\,/\,[(1+r)^T]$$

Dividing both sides by r yields **$P = v\,[(1+r)^T - 1]\,/\,[r\,(1+r)^T]$** QED.

A Useful Extension of the Formula for Calculating the PV of a Constant Flow of Benefits or Costs or Net Benefits to Infinite TIme Periods

Note that this constant flow of benefits (or costs) formula **has a very simple limit** as T approaches infinity, namely:

$$\mathbf{P = v/r} \ .$$

This is another very convenient formula. There are many long-term investments and regulatory policies that have very long lives that can be thought of as infinitely lived investments as a "first approximation". The P=v/r formula allows the present values of such flows of cost or benefits to calculated very easily.

IV. Illustrative Applications

(1) These formulae can, for example be used for net revenue analysis. Suppose that a windmill can be built that cost $1,000,000 and will produce $50,000/year in electricity for 40 years. Is the windmill profitable to construct if the interest rate is 5%/year?

Use the PV formula: $P = v\,[\,((1+r)^T - 1)/r\,(1+r)^T]$

The PV of the future profits are $P = 50{,}000[((1.05)^{40} - 1)/(.05)(1.05)^{40}\,] =$ $857,954.31

So, the answer in this case is NO

What if the interest rate is 2%/year? In this case PV = $1,367,773.96

In this case the answer is YES

(Real) interest rates matter. Note that the net benefits in the distant future are worth far less when r = 0.05 than when r = 0.02

Note that if the dam would provide electricity forever, then $P = v/r = \$50{,}000/0.05 = \$1{,}000{,}000$

In that case the dam project exactly breaks even (ignoring any maintenance expenses)

But, also note that the all the years after year 40 add relatively little to the present discounted value of the future benefits.

(2) Suppose that Al can afford to pay $5000/year in car payments for 5 years toward a new automobile. If the bank's opportunity cost rate of return is 7%, what is the largest amount that the bank will loan Al given his budget?

Use the PV formula: $P = v\,[\,((1+r)^T - 1)/r\,(1+r)^T]$

$P = 5000\,[\,((1.07)^5 - 1)/r(1.07)^5] = \$20{,}500.99$

That is the bank's opportunity cost of tying up P dollars during the 5 years the loan will be repaid.

V. Risky Intertemporal Choices: Combining Present Value and Expected Value Calculations

The present value and expected value formula can be combined to deal with uncertain flows of future benefits and costs or uncertain future income levels.

For example, consider the purchase of a lottery ticket in a "million dollar" lottery game. Suppose that the winner receives $50,000/year for twenty years, the interest rate is 5%, the probability of winning is 1/1,000,000 and the lottery ticket costs 1 dollar. Suppose also that there are just two outcomes: winning and losing.

The present value of winning the lottery is the present value of $50,000/year for twenty years. Substituting into the present value formula for a constant flow of future benefits yields:

$$(50{,}000)\ [\ (1.05)^{20} - 1)\ /\ (\ .05\ (1.05)^{20}\ )] =$$

$$(50{,}000)(12.4622) = \$623{,}110.52$$

when the current interest rate is 5%/year. This is, of course, much less than the $1,000,000 value that lottery sponsors usually claim for the prize of such contests.

The expected present value of such a lottery ticket requires calculating the expected value of the ticket The probability of winning is 1/1000000 and the probability of losing is 999999/1000000, so the expected value of the ticket is:

[1/1000000][623,109.52-1.00] + [999999/1000000][-1.00] = -$0.37

Notice that this ticket is a "bad bet." It has a negative expected discounted value. (By the way, this hypothetical lottery is a better deal than most state lotteries, which have expected present values of less than -$0.50)

VI. Applications to Normative Policy Analysis: Benefit-Cost Analysis

One of the most widely used tools of policy analysis is benefit-cost analysis. In principle, benefit-cost analysis attempts to determine whether a given policy or project will yield benefits sufficient to more than offset its costs.

Cost-benefit analysis, ideally, attempts to find policies that maximize social net benefits measured in dollars. (Every diagram that includes a dead weight loss triangle is implicitly using cost benefit analysis.) Economists use this approach to characterize externality and monopoly problems. It is also used to criticize ideal and less than ideal public policies and taxes. Unfortunately, the data do not always exist

for these calculations to be made. The most widely used methods for dealing with uncertainty and time in Benefit-Cost analysis is to use various combinations of “Expected Value” and “Present Value” calculations.

Cost-benefit analysts carefully estimate the benefits, costs, and risks (probabilities) associated with of alternative policies through time. If several policies are possible, cost-benefit analysis allows one to pick the policy that adds most to social net benefits (in expected value and present value terms) or that has the highest social rate of return. If only a limited number of projects can be built or policies adopted, then one should invest government resources in the projects or regulations that generate the most net benefits (the highest rates of return in terms of social net benefits). One can also use cost-benefit analysis to evaluate alternative environmental policies.

When many projects can be adopted, the policy question is essentially a yes or no question is: Does the policy of interest generate sufficient benefits (improved air quality, health benefits, habitat improvements etc.) to more than offset the cost of the policy (the additional production costs borne by those regulated plus any dead weight losses and the administrative cost of implementing the policy)?

The *net-benefit maximizing* norm implies that both good projects, and good regulations, should have **benefit-cost ratios** that exceed one, B/C > 1. That is to say, the benefits of a project should exceed its costs to be worth undertaking. However, many of the goods and services generated by environmental regulations *are not sold in markets* and so *do not have prices* that can be used to approximate benefits or costs at the margin.

These "implicit prices" can be estimated, but the estimates may not be very accurate. Thus, a good deal of the policy controversy that exists among environmental economists is over the proper method of estimating non-market benefits and costs. For example, the recreational benefits of a national forest may be estimated using data on travel time. However, this estimate is biased downward. We know that the benefit must be somewhat greater than the opportunity cost of driving to the forest! Survey data can also be used, but people have no particular reason to answer truthfully (or carefully) to such questions as how much would you be willing to pay to access "this national forest," "to protect this wetland," or to "preserve this species?"

In cases where the benefits and costs are not entirely predictable, the probability of benefits and costs also have to be estimated. The probabilities assigned to the various outcomes also are often difficult to estimate.

Thus, although arguably better than nothing, benefit-cost analysis tends to be quite inaccurate (e.g., estimates of net benefits typically have high variance). So instead of attempting to find the best (social net benefit maximizing) policies, cost benefit analysis often simply attempts to determine whether the benefits of a policy exceed its costs. A policy is said to improves a situation if it generates Benefits greater than its Costs. This is, of course, a normative statement—one based loosely on the utilitarian school of philosophy.

In spite of all these difficulties, benefit-cost analysis has several advantages as method of policy analysis. It forces the consequences of policies to be systematically examined. It provides "ballpark" estimates of the relevant costs and benefits of regulations for everyone who is affected by a new regulation or program.

A Relatively Simple Illustration of an Environmental Cost-Benefit Analysis Suppose that Acme produces a waste product that is water soluble and that its current disposal methods endanger the local ground water. Acme saves $5,000,000/year by using this disposal method, rather than one which does not endanger the ground water. What is the present discounted value of Acme's savings (much of which is passed on to consumers) if the interest rate is 10% and Acme expects to use this method for 30 years?

The easiest method is to use the formula

$$\mathbf{P = v\,[(1+r)^T - 1] \,/\, [r\,(1+r)^T]}$$

Here: $P = (5{,}000{,}000)\,[\,((1+.10)^{30} - 1)]/[(.10)\,(1+.10)^{30}] =$ $49,574,072.44

One could also approximate the present value of Acme's cost savings using the present value of an infinite series formula (P=F/r) which yields (5,000,000/0.1 = $50,000,000.00. Note that this simpler calculation produces nearly the same answer, and so is often a good way to check one's math.

Suppose that an environmental law is passed which requires firms like Amex to adopt the more costly but safer technology. If the fine assessed is $10,000,000, what probability of detection and conviction will Amex adopt the safer technology if its discount rate (interest rate) is 10% ? The expected fine in a given year has to be greater than the expected cost savings, Thus, P*10,000,000 > 5,000,000 in order for the fine to affect Acme's choice. (In this case the interest rate is not necessary for finding the solution because it is assumed that violations would be detected, and fines paid annually. Although, we could also use present values for both the penalties and cost savings.) The smallest probability of punishment that "works" is 0.5, because this makes the expected fine equal to the expected cost savings.

Suppose that administering the enforcement regime costs $1.000,000/year that produces a 0.75 probability of punishment. What is the smallest annual external damage that can justify the program? Given the fine and probability of being caught and punished, we know that this program will induce Acme to clean up, so the only important question is when the present value of the damages (net of administration costs) avoided are greater than the present value of the extra costs borne by Acme (and its consumers).

Intuitively, we can see that if the damage per year (D) less the administrative costs ($1,000,000/year) are greater than the cost imposed then the program is worthwhile in cost-benefit terms. (D - $1,000,000 > $5,000,000). This implies that the damages must be greater than $6,000,000 per year. If

the damages vary a bit through time, then we would need to use present and expected values to figure this out.

In that case the present value of the damages avoided minus the present value of the administrative costs would have to be greater than the present value of the cost increase imposed on Acme (and its consumers). If the damages were random, perhaps because rainfall is random, then we would have to compare the expected damage reductions (net of administrative costs) with the cost of "cleaning up."

For example, suppose that on rainy days the "dirty" waste disposal system causes $20,000,000of damages and that on dry days, the "dirty" waste disposal causes no damages to the local ground water supply. Suppose that it rains one third of the time. In this case the expected damages from the "dirty" waste disposal system are: D^e = (.33) ($20,000,000) + (.67) (0) = $6,666,666 per year.

In this case the cost of eliminating the damage is the cost of the clean-up (more expensive waste disposal system) plus the administrative costs ($5,000,000 +$1,000,000) while the benefits are the expected reduction in damages: ($6,666,666 per year). The **expected present value** of the social net benefits from the program over thirty years can be calculated with formula $P^e = v\ [\ ((1+r)^T - 1)/r\ (1+r)^T]$ given a planning horizon (T) and discount rate (r). Let T= 30 and r = 10% again.

P^e = ($666,666) $[((1+0.1)^{30} - 1)/(0.10)\ (1+0.1)^{30}]$ = ($666,666) (9.4269)

Thus, P^e = $6,284,603.40

Given all these details, this program will produce a bit more than 6.28 million dollars of expected net benefits over a thirty-year period (in present value terms).

VII. Some Practice Exercises

1. Suppose that Al wins the lottery and will receive $100,000/year for the next twenty-five years. What is the present value of his winnings if the interest rate is 6%/year?, 5%/year, 3%/year? How much more would a prize that promised $100,000/year forever be worth?
2. Suppose that Al can purchase lottery tickets for $5.00 each and that the probability of winning the lottery is P. If Al wins, he will receive $100,000 dollars per year for 20 years. The twenty-year interest rate is 3%/year.

   What is the highest price that Al will pay for a ticket if he is risk neutral? Determine how Al's willingness to pay for the ticket increases as P, the probability of winning, increases and as the interest rate diminishes.
3. Suppose that Amex produces a waste product that is water soluble and that its current disposal methods endanger the local ground water. Amex saves

$1,000,000/year by using this disposal method rather than one which does not endanger the ground water. What is the present discounted value of this waste disposal technology to Amex if the interest rate is 6%? if it is 4%?

4. Suppose that an environmental law is passed which requires firms like Amex to adopt the more costly but safer technology. If the fine assessed is $2,000,000, what probability of detection and conviction will Amex adopt the safer technology if its discount rate is 5%? if it is 10% ?

5. Suppose that global warming is caused (at the margin) by CO_2 emissions and that to reduce CO_2 emissions enough to affect future temperatures requires policies that will reduce economic output by 5% per year. U. S. GNP is currently about 15 trillion dollars and is expected to grow by about 2.5% per year in the future. How large do expected damages have to be to justify such an aggressive environmental policy?

   Hint 1: in this case, the future value of GNP is $Yt = 15*(1+.025)^t$, because of economic growth, which works like compound interest. The reduction in non-environmental income in year t is thus $Vt = (.05)15*(1+.025)^t$

   Hint 2: This implies that present values can be calculated using the summation formula $P = \Sigma (Vt/(1+r)^t$ by substituting for $Vt = (.05)\ 15*(1+.025)^t$

   That is to say, $P = \Sigma ((.05)\ (15 \text{ trillion})\ (1+0.025)^t/(1+0.05)^t$

   Hint 3: more generally one can write this expression as $P = \Sigma\ (Vo\ (1+g)^t/(1+r)^t$ where g is the economic growth rate, r is the discount rate (interest rate), and Vo is the initial value of the "thing" that is growing at rate g.

   Hint 4: It turns out that in a present value problem with an infinite planning horizon, one can use a relatively simple formula to calculate the present values of a series of values that grow by a constant percentage each year:

   P = Vo / (r-g) where Vo is the initial value, r is the discount rate (or interest rate), and g is the long-term growth rate.)

   [**Now you can calculate** the present discounted value of the cost of reducing CO_2 emissions, which is approximately 30 trillion dollars, given all the assumptions made. The hard part in Cost Benefit analysis is estimating the costs and benefits, once that is done, the rest is straight forward.]

Chapter 10: Extensions and Foundations: Law, Public Policy, Norms, and Economic Welfare

I. Introduction

To this point we have assumed that markets have significant social support provided by a society's laws, norms, and public policies. Many of these types of support can be regarded as prerequisites to a modern commercial society. Others can be regarded as complements to commercial activities which tend to increase the size of markets and market networks.

For example, if there is no private property with an "bundle of ownership rights" that can be shifted from one person to another or from one organization to another, then trade will not be possible—or at least will be difficult and risky. If the prevailing norms regard all commercial activities to be evil or corrupting, only the least moral persons in society will engage in commerce, and markets will tend to be much smaller than they could have been, and market transactions riskier than they could have been with more normative support or less normative resistance. Similarly, public policies may restrict markets in a wide variety of ways, with various entry barriers, licensing requirements, taxes, and barriers to entry. Such policies tend to make markets smaller than they could have been, and average incomes lower than they could have been for reasons developed in the last few lectures of the first part of this course.

This chapter uses the ideas from the last two chapter and chapter 5 to demonstrate how the legal, political and social systems in the broader society influence the size of markets and market networks. They do so through effects on various risks faced by consumers and firms that influence their decisions to produce or purchase goods and services.

We'll do so by demonstrating that rational choice models can shed light on a variety of non-market activities that indirectly affect the extent of markets in a given society. For example, we can use the expected net benefit maximizing model to analyze the extent of crime and some of the market-relevant risks associate with it We'll also show how rational choice models can be used to model policy choices in a democracy, and the effects of norms on both those and other choices. Lastly, we'll show how these ideas (as previously applied in chapters 5 and 6) can be used to evaluate the relative merits of public policies.

Rather than regarding such factors to be beyond economics, a "multi-disciplinary" approach includes such factors in economic analysis. And, it shows why such "non-economic" factors effect both market equilibria and economic growth rates.

For the most part, these extended rational choice models are associated with economic innovations of the last third of the twentieth century. Several innovative economists were often rewarded with Nobel prizes in economics during that period. Examples include Gary Becker, James Buchanan, Douglas North, and Elinor Ostrom.

So, although, these sorts of behavior are not among the core models of markets that we have developed in first parts of the course, they are nonetheless relevant for economics, and help to explain the variation in market size and efficiency among countries today. They also provide a partial explanation for why markets change through time. Today's markets are far more extensive than they have ever been before, partly because of improvements in public policies and greater normative support.

Economic development is not just about capital accumulation and innovation. The extent of commerce is also affected by a society's legal system (and law enforcement), its public policies, and the types of norms that are most prevalent in it. Changes in any of these factors can accelerate or retard economic development.

II. Crime and the Scope of Markets (Rational Criminals)

Gary Becker of the University of Chicago pioneered both the economics of crime and the use of rational choice models (net benefit maximizing or utility maximizing models) to analyze other socio-economic topics such as marriage and drug addiction.

We'll develop a somewhat simplified version of his analysis of crime in this section of chapter 10, and we'll link it to markets through the effects that property crimes impose on persons not engaged in criminal activities. Economic crimes affect the riskiness buying, selling, and owning particular goods. Examples include burglary, theft, and fraudulent practices. By affecting the riskiness of market transactions, such crimes affect producer and/or buyer choices, and thereby market equilibria.

The economic analysis of crime assumes that (most) potential criminals are rational in the sense that they maximize their expected net benefits as they perceive them. They may be among the risk preferring part of society with high discount rates, but that does not imply that they are necessarily irrational in the economic sense. They may still be forward looking and make decision that they believe will systematically improve their welfare—as they understand it.

We'll focus on economic crimes rather than violent ones, because decisions to engage in thievery and fraud closely resemble "ordinary" economic decisions.

The economic advantage (marginal revenues) associated with stealing, fraud, or selling illegal goods and services vary with the price of the goods of interest. The greater the resale value of a good and easier it is to steal and sell stolen goods, the larger are the expected net revenues associated with using one's time for theft rather than—for example, working at a minimum wage job, or in the case of white-collar crimes, at a relatively high-paying job.

Reductions in one's honest income occur as one diverts one's time and attention from honest work to criminal activities. Thus, part of the cost of engaging in crime is one's **opportunity cost**.

Another part is the **expected penalty** associated with the crime or rate of crime undertaken. Every time a crime is committed, there is some probability that one will be caught, convicted, and punished (P). Punishment clearly diminishes one's criminal net revenues. We'll call the penalty F as with a fine. Although relatively few criminal penalties are fines, one can think of F as the money loss equivalent to being in jail for a period of months or years and/or subject to other punishments.

The expected penalty cost of the crime(s) is just P*F, the probability of being caught, convicted, and punished times the (average) punishment meted out for the crime(s) of interest. The probability that one is not caught, convicted, and punished is (1-P).

In most cases both P and F increase with the number of crimes committed, so the expected punishment tends to increase with the extent of one's criminal activities. As a consequence, the marginal expected cost of engaging in crime rises with the number of crimes undertaken.

This can be written out mathematically If penalty F is a function of n and the probability of being caught, convicted, and punished also increases with *n*, then F = f(n) and P=p(n) where n is the number of crimes committed. Thus, F^e = f(n)p(n) is the expected fine, which increases with n, because both parts of this calculation increase with n.

The marginal cost from reduction in honest income is either flat or also increasing in *n* as one diverts more of one's "working" hours from honest work to the criminal activity. Fewer marketable skills are likely to be accumulated as one works less.

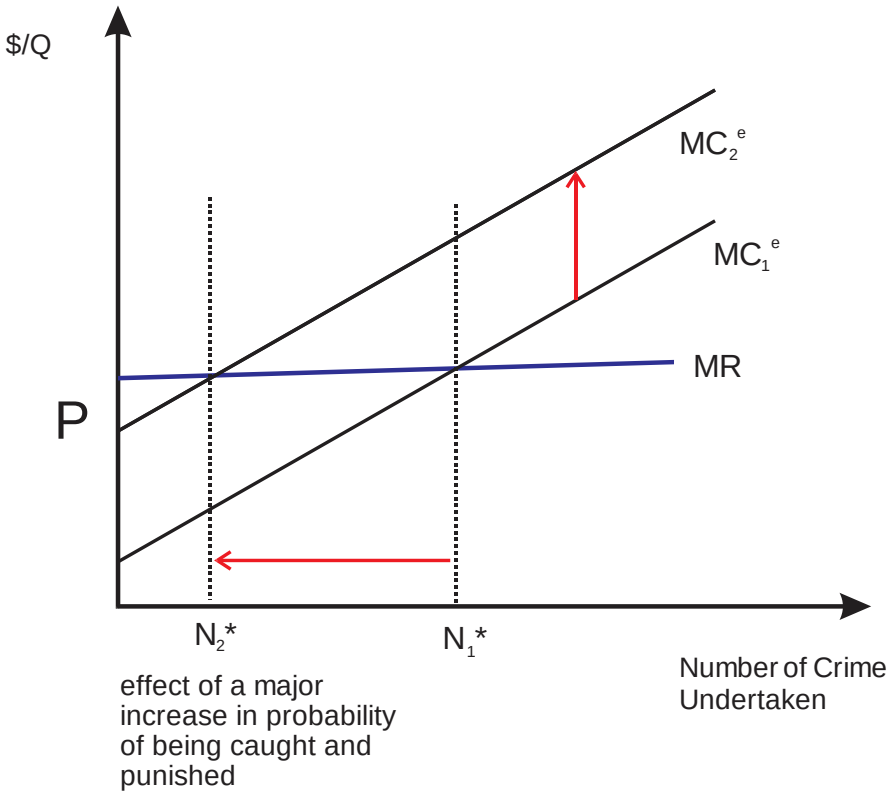
The overall cost of crime is the sum the opportunity cost of crime, the cost of equipment used in the crimes undertaken, and the expected penalty schedule associated with such crimes.

The revenues generated by theft and similar crimes vary with the resale value of the items stolen, which we'll assume is constant. In such cases, a thief's marginal revenue curve tends to be relatively flat. (The same logic applies to sales of illegal goods and services.)

Given all the above, the rational criminal's choice resembles that of a firm that has uncertain production costs—a case that we analyzed in chapter 8. He or she attempts to maximize net revenues by varying his or her "output" of crimes in a setting where production costs are randomly generated (although in this case, the probability of high costs and extent of those costs are affected by his or her output decisions, that is to say the potential criminal's crime rate (n)).

Figure 10.1 illustrates the criminal's rational engagement in criminal activities and also the effect of a significant increase in the probability of being caught, convicted, and punished (which increases the expected marginal cost of criminal activity).

Figure 10.1 Maximizing Expected Profit



The effect of property crimes on market equilibria and growth rates occurs through effects on risks for consumers and firms. Theft, for example, reduce the probability that after a consumer buys something, he or she actually gets to use it for enjoyment or other purposes in the future. After a TV, cell phone, or computer is stolen, "it" is no longer available to the consumer for the uses he or she had intended when he or she purchased it.

Similarly, theft at a retail outlet implies that the quantity that one has produced (or otherwise acquired) may not all be available for sale, which reduces the expected marginal revenue associated with the production of such goods. It also reduces the expected profit maximizing quantity that a firm will produce if it attempts to maximize expected profits.

For reasons developed in chapter 8, an increase in the risk of losses reduces demand for such goods and an increase in the downside risk from lower revenues reduces the production of such goods. Both of these effects tend to reduce the size of markets through downward (leftward) shifts of either the market demand curve or the market supply curve or both for the goods most often stolen.

By increasing expected marginal producer costs and/or reducing expected consumer marginal benefits, property crimes tend to reduce the size of markets and market networks. In general, the higher the risk of an economic crime, the smaller market networks tend to be.

Contrariwise, policies that tend to reduce crime property crime rates tend to increase the size of markets.[9]

(As an exercise, draw the consumer and producer diagrams that illustrate why lower expected marginal benefits or lower expected marginal revenues tend to have these effects. These will resemble those developed in Chapter 8 for firms with uncertain production costs and consumers purchasing goods with uncertain quality).

III. Choosing Public Policies in Democracies with Rational Voters

A variety of public policies can affect markets. We showed this in chapter 5 and also, indirectly in the previous section of this chapter. Creating policies that reduce competition in markets, tends to increase firm profits, reduce consumer surplus and sales in the long run. An increase in police enforcement, prosecution, and penalties can decrease crime rates, which reduces risks for consumes and firm owners, and thereby tends to increase the size and growth rates of markets.

The next section of chapter 10 models how policies are selected in democratic governments. As in the previous section, more can be said about this than possible in part of a chapter. The aim is again to introduce students to some of key ideas. (More is developed in upper-level classes on Law and Economics and in Public Economics courses.)

The most widely used model of majority-rule politics is the **median voter model**. In a variety of electoral settings, self-interested behavior implies that the "median voter" will get his or her way. We first show why this tends to be true in direct democracy and then show why it tends to be true in representative democracy.

Suppose that three individuals: Al, Bob and Cathy are to make a decision about where to eat lunch using majority rule. Al prefers a restaurant where lunch can be had for $5.00, Bob wants one where lunch costs around $10.00 and Cathy, a gourmet, prefers one costing around $20.00. For convenience assume

[9] Resale of stolen item stolen may offset some of the income effects and lost gains from trade, but not fully, because the sale and purchase of such goods is also illegal, which makes them risky to acquire and sell, which lowers prices and the net benefits realized by both “fences” and consumers of such stolen good resellers. Also diverting resources from productive activities to the illegal ones tends by itself, to reduce the overall quantity of goods and services for sale in the combined legal and illegal markets for profitably stealable goods.

that, given any two options, each will prefer the restaurants whose price for lunch that is closest to their preferred one.

- "Spatial voting" occurs when an individual's marginal benefit and marginal cost curves are straight lines, as may be shown in class if time permits.

In our illustrating example, price is used as proxy for the quality of the restaurant and/or its service and ambiance. Given this, consider some votes on various alternative restaurants.

Table 10.1: The Median Voter Model of Democracy				
(options)	**Al (5)**	**Bob (10)**	**Cathy (20)**	**Result**
8 vs 13	8	8	13	8
7 vs 12	7	12	12	12
5 vs 20	5	5	20	5
9 vs 14	9	9	14	9
9 vs 10	9	10	10	10
10 vs 11	10	10	11	10

Note that Bob always votes in favor of the outcome that wins the election. (The B column of Votes Cast by Bob and the Outcome column are EXACTLY the same.)

Note also that exactly the same number of individuals prefer a more expensive dinner as prefer a less expensive dinner than Bob.

- *Bob is the median voter. (*He is the voter with the *median* ideal point in the distribution of voter ideal points.)
- The median "ideal point" or "preference." is the one in the middle in the sense that the same number of persons want more of this good or service as want less of it all things considered.

Note also that the last two votes demonstrate that the median voter's ideal point can beat every other possible alternative in pair-wise voting. The median voter's ideal point is "majority preferred" to alternatives just a bit larger or just a bit smaller than his or her ideal point.

There are sometimes said to be two versions of the median voter theorem.

- The **Weak Form** of the *median voter theorem* says that the median voter always casts his vote for the policy that is adopted. [In the example above, Bob always votes for the alternative that wins.]

- The **Strong Form** of the *median voter theorem* says that the median voter always gets his or her most preferred policy.
- In the example above, Bob's preferred expenditure or quality level, $10, will defeat any other restaurant. As long as it is on the ballot, it wins. Thus, the median voter's ideal tends to emerge as an equilibrium in a series of votes.

Representative Democracy and The Median Voter

The previous illustration shows that the median voter determines the electoral outcome in direct democracies. If there are just two options, he or she always votes for the policy that is ultimately adopted. We next show that the median voter is also very important--pivotal--in representative democracy.

To make our analysis of elections for high office(s) more straightforward, we will assume that Voters all vote for the candidate (or policy) that is "closest" to them in the policy dimension.

The assumption of spatial voting allows competition between candidates for government office to be analyzed with a diagram that characterizes the distribution of voter ideal points (the distribution of persons that actually vote).

Voters can have a wide variety of ideal points. When the number of voters is relatively small, one can either construct a bar chart with different ideal points along the bottom and numbers of voters represented as the height of the bars. If there are a large number of voters and possible positions, one can also represent the distribution with a diagram analogous to a continuous functional form of a frequency distribution.

In a "frequency distribution" representation, the areas under the curve between two values, say G_1 and G_2, represent the number of voters whose ideal points are between G_1 and G_2 (including G_1 and G_2).

- The frequency distribution of voter ideal points has policy alternatives along the bottom (X) axis and number of voters per ideal point on the horizontal axis.
- The assumption of spatial voting allows us to determine how all the voters cast their votes when there are two candidates, or two policy options being voted on.

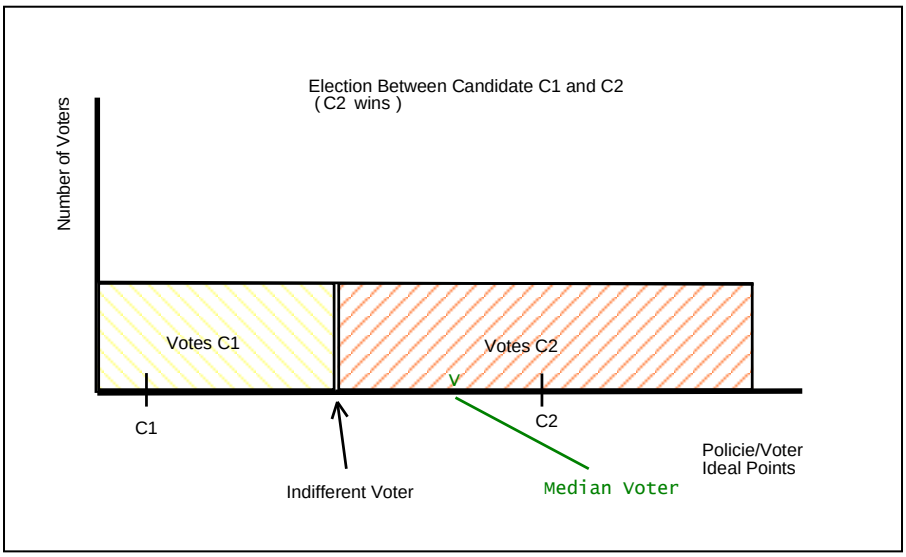
- Every "spatial voter" will vote in favor of the candidate whose position is closed to their own ideal point.

Voters who are exactly halfway between the two "alternatives" being voted on will be **indifferent** between them and are assumed either not vote or to vote by flipping a coin.

Voters to the left of the "indifferent voters" will vote for the policy or candidate on the left, and those to the right of the "indifferent voters" will vote for the policy or candidate on the right.

The illustration below assumes that candidates 1 and 2 have taken policy positions C_1 and C_2, and that voters vote for the candidate closest to their ideal point.

The distribution of voter ideal points in the figure below is assumed to be a "uniform" distribution--although essentially the same results would arise for other distributions, as shown in class. The uniform distribution is adopted here to make drawing the diagram a bit simpler. As drawn, it turns out that Candidate C_1 loses this election. **He or she gets fewer than half of the votes cas**t. This is implied by the fact that the area labeled C1 is much smaller than the area labeled C2.



How could Candidate C1 have done better? Clearly he or she should have chosen a policy position further to the right. That is to say, **he or she should have chosen a position closer to the median voter**.

It turns out that the candidate who is closest to the median voter's ideal point will always win the election, because that voter will always receive AT LEAST HALF OF THE VOTES.

Thus, if candidates are free to adjust their policy position to attract votes, they will each try to be closer to the Median Voter's ideal point than the other candidate. As both candidates try to do this, this induced the candidates to converge toward the median voter's ideal point. This tends to make the candidates take more similar positions on the issue(s) of greatest interest to voters, which is to say less extreme policy positions.

In equilibrium, this kind of competition for votes implies that both candidates will take essentially the same position, namely that of the median voter.

- At this equilibrium, the candidates take essentially the same position, so they receive approximately the same number of voters.
- If both candidates take exactly same position, the expected outcome is a tie. One or the other candidate will win, because voters vote randomly (by flipping a coin or the equivalent), but just as a matter of luck rather than be having adopted a better policy position.
- At this equilibrium, the median voter gets exactly what he or she wants.
- In such cases, *the strong from of the median voter theorem holds!*.
- This is not because the median voter is "special" or "admirable" but simply because of competition among candidates for votes.

The median voter outcome is simply the **Nash equilibrium** of pragmatic competition by candidates or parties for votes along a single ideological or policy dimension.[10]

The Median Voter and Public Policy

One important insight that follows from the median voter model is that to understand the scale of government programs, one has to look at both the benefit and cost sides of programs from the point of view of the median voter.

In most cases, the median voter is approximately the voter with median characteristics. That is to say he or she is a voter of median age, median income, median education, median family size, median political ideas and so forth.

This allows this model to be used to predict public policies using estimates of median voter demands for public services. **Geometrically, the median voter's choice looks a good deal like the choice of consumers in a grocery store. Their policy ideal is where his or her marginal benefits from the policy equals its marginal cost.** This implies that any change that alters the median voter's marginal benefits or marginal costs associated with the government policy of interest will also change his or her ideal—and insofar as the strong from of the median voter theorem holds, it will also change public policies.

The median voter will not ordinarily be the same as the median member of the community because not all persons are equally likely to vote! In the US it turns out that the median voter is a bit older, richer,

[10] Nash equilibrium is a concept from game theory, which is a bit beyond the scope of this course. Game theory is a field of applied mathematics that examines how rational players participate in various "games." The "games" are not usually of the parlor game or app variety. Instead, games in this sense are social settings in which the outcomes that one observes are joint products of the decision made by all of the "game's" participants.

and better educated than the median member of the group of persons eligible to vote, because poor, young, and less educated person vote less frequently than older, richer, and more educated persons.

To the extent that the Median Voter gets what he or she wants, anything that changes the median voter's preferred policy will affect government policies. This includes factors such as income, age, ideology, marginal tax costs, and information about the costs, benefits and risks associated with policies.

The median voter model is not a complete model of policy formation—it simply explains how policies that are driven by elections tend to emerge from systems of government based on majority rule. Other factors that may also affect public policies include:

- Agency Problems (representatives may not do what they promise during the election)
- Rent Seeking: special interest groups may be able to influence policy choices by lobbying, providing campaign contributions, or bribing officials.
- The bureaucracy itself may also affect policies both by lobbying for particular policies and through their discretion over how to implement policies.

Rational ignorance, a topic that we covered briefly in chapter 8, implies that the median voter (and other voters) is not usually as well informed as would be ideal—so voters can make errors—even systematic errors—when their information is limited, even if they make the best use of the information that they have.

However, candidates that are known to have cheated and done poorly at overseeing the bureaucracy will be more likely to lose the next election than those that have not, because the median voter will not have gotten what he or she wanted from that candidate. So, the assumption that elections matter is not unrealistic—and generally works as a first approximation.

IV. Internalized Norms and Market Equilibria

Internalized norms are ideas about “proper behavior” that individuals have come to believe. Such ideas may be “internalized” as a consequence of an individual’s own efforts to understand the nature of “good” behavior, or it may have been, in a sense, absorbed from others or the mass media without much conscious thought.

In either case, internalized norms tend to affect behavior. In our models of rational choice, they do so by influence either an individual’s perception of the marginal benefits of an activity, product, or policy. Or, they do so by influence their perceived marginal costs of engaging in a particular activity or voting for a particular policy or candidate.

In the model of criminal behavior developed earlier in this chapter, a person that aspires to be a “good person” and believes that such a person does not engage in criminal activities (except perhaps during

emergencies) will have a higher marginal cost for engaging in criminal activities than one that has not internalized such a norm. He or she will feel guilty when engaging in criminal activity in addition to the other costs. This makes such persons less likely to be criminals than one's that have not internalized an "honesty" or "law abiding" norm.

Similarly, a person that has internalized a "diligence" norm or a "work ethic" will be less inclined to do a poor job when at work producing goods and services for sale than one that has not internalized such norms. The latter would be inclined to goof off whenever not monitored, whereas the person that has internalized a work ethic will not. The latter makes employees with a work ethic or diligence norm more valuable to firms than those that have not internalized such norms—other things being equal.

To illustrate how such norms may affect market equilibrium, markets for products with uncertain quality are revisited in this subsection—where the variation in quality is now affected by the norms (or lack of them) of the persons working at the firms producing the goods of interest.

Norms and the Average Quality of a Firm's Output

Virtually all products, from automobiles to zucchini, have properties that cannot be perfectly assessed by consumers at the point of sale. Much of this uncertainty is irreducible. The production processes that create and distribute products and services are partly stochastic. Accidents happen. Even the best machines become less reliable through time. Workers may be more or less attentive. As a consequence, defective units are always produced along with normal units.

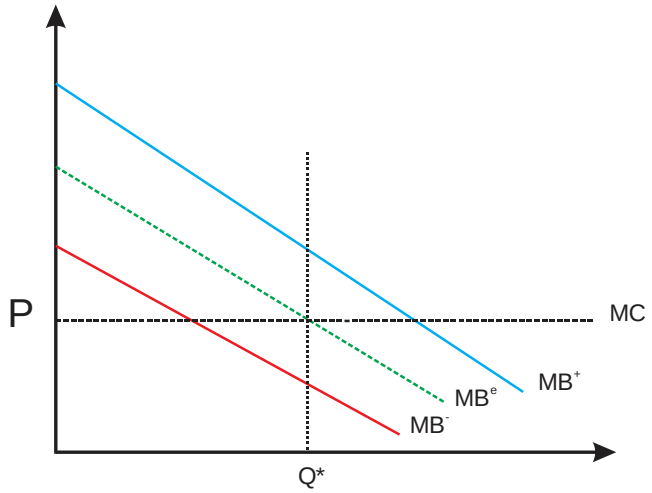
Consider a consumer's decision to purchase products when the probability of a defect is known (or at least reasonably well estimated). The product has relatively high marginal benefits associated with it when all of its features are non-defective. The marginal benefit curve is lower when defective units are received. Uncertainty about the quality implies that the average quality and there for marginal benefits from the product lies between the extremes of entirely flawless and entirely flawed units of the product or service.

To simplify, assume that consumers are risk-neutral, and that the probability of defective units of the good produced and sold is a decreasing function of the average work ethic of the personnel in firm(s) manufacturing the product.

Initially, we'll assume that consumers regard all firms to be the same, and we'll also assume that the work ethic at each firm is the same (Ei=Ej). In addition, we'll also treat each successive unit purchased by every firm as an independent draw from the distribution of product quality. To simplify, we'll also assume that there are just two quality levels, perfect (+) and defective (-). The average marginal benefit of the Nth unit is simply $(1\text{-}P)MB^{+}(N) + (P)MB^{-}(N)$, where the "+" superscript denotes units without defect and the "-" superscript denotes defective units.

The expected marginal benefit curve (MB^e) lies between the marginal benefit curves of the perfect and defective units of the product. If P=0.5, then the expected MB is exactly midway between the two actual MB curves.

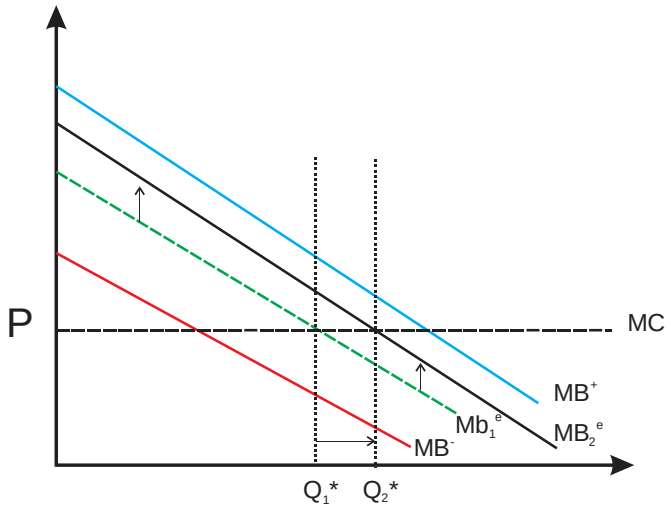
Figure 10.1 Maximizing Expected Consumer Surplus



At the market price of P*, an expected consumer surplus maximizing consumer purchases the quantity that equates the expected marginal benefit to its marginal cost. Figure 10.1 illustrates geometry of this choice setting and the quantity purchased, Q*.

An increase in the average extent of internalized ethics that increase the average quality of the products sold increases the expected marginal benefit from all purchases. As the probability of a defect falls, the MB^e shifts toward the MB^+ curve, and purchases of this consumer increases. Such effects may be the result of more diligent work effort by production workers (E”>E’) or improved quality control by managers.

Figure 10.4 Effect of an Increase in Productive
Ethical Dispositions on a Typical Individual's Demand



This result follows regardless of whether firms or consumers realize that internalized ethical dispositions affect the average quality of industry output or not. An increase in the average diligence and care of employees (E">E') reduces the probability and/or extent of product defects. This reduces the risks of exchange for consumers and so tends to increase the extent of commerce (Q">Q').

The opposite follows in cases in which the internalized norms of employees change in a manner that weakens interests in careful production and handling of goods. A more haphazard production and assembly of the products and services sold tends to increase the probability of defects. Such changes may arise because of changes in the pool of employees themselves or by shifts in the locations of factories from one moral community to another.

Ethics and Prices in Competitive Markets without Quality Differentiation

In the usual textbook characterization of competitive markets, no single firm has a reputation that is different from any other, because they are produced using the same methods with more or less similar personnel and other inputs. In such cases, a profit maximizing firm would take all the steps that it is aware of to reduce the cost of producing the goods or services of interest, but not necessarily pay much attention to quality control.

If cost saving steps tend to reduce average for the industry as well as the firm, a race to the be the least costly producer tends to take place. This race reduces average quality and consumer demand for this

product.[11] In extreme cases, markets for such products may disappear for reasons similar to those associated with the problem of fraud in the previous chapter. If consumer expects the average seller to sell uniformly defective products with little no marginal benefits, then there is no point in purchasing the goods produced by such markets.[12]

The extreme lemons problem outcome can be avoided in several ways, but most of these arguably have ethical foundations. For example, there may be cost saving measures that firms refuse to adopt because they reduce average quality below levels acceptable to their own internalized norms of conduct. Minimum quality standards may be adopted through industrial councils or laws. Firms may be required to exchange defective units for others or to offer money back guarantees. Such laws, of course, improve average quality only if government employees tend to be more diligent (ethical) than those of the average firm in the problematic industry. Corrupt enforcement would not have the desired effect.

If the Protestant reformation increased the average diligence of employees, employers and government officials, it would have reduced the extent of the lemons problem in perfectly competitive markets. If so, Weber's explanation for the greater economic development of northern Europe after the Protestant reformation would have occurred even without its other effects on capital formation or direct market rewards for ethical dispositions.

Ethics and the Emergence of Quality Differentiation in Output and Labor Markets

We now shift to a setting in which consumers can recognize differences in the average quality of the products sold by individual firms. Suppose that employees differ with respect to an ethical disposition that tend to improve average quality at whatever firm they work at. Suppose also that it is impossible for firms to distinguish among their employees and so differences in diligence go unrewarded. In this case, neither the wage nor employment rates of persons with productive ethical predispositions would any be higher than those lacking such dispositions.

Distinguishing among Firms

Nonetheless, some firms will get a bit lucky and employ well above average numbers of diligent employees. Others will get unlucky and employ well below average numbers of diligent employees.

---

[11]Figure 8.2 can be used to represent markets as well as consumer choices, if one assumes constant returns to scale in production with respect to both defective and perfect units and identical consumers. Note that the market shrinks but does not necessarily disappear unless the marginal benefits associated with defective units of the good also diminish, that is to say the "rejects" come to be more frequent and have even lower quality than initially.

If sufficient differences in output quality emerge so that allow consumers can distinguish between the highest and lowest defect firms, the market may separate into two markets even if the products themselves remain indistinguishable from one another. Instead, particular producers or brands may be used as a proxy or estimator for the average quality of the good. This may occur even when differences in average quality emerge entirely through chance.

Such brand or reputation-based estimators for quality are, of course, imperfect. A firm's reputation can be used as to estimate the quality of its outputs only as long as it continues to produce higher quality products. In this case, firm identity (name brands) would be a relatively good estimator for product quality as long as employee turnover is relatively low.

If the good firms cannot satisfy the demand for their product at the preexisting market prices, they may raise prices without losing their customers, because of the lower risk of defects. In this way, prices at "good" and "bad" stores may come to differ, even though the products themselves are indistinguishable from one another. Although, the product remains homogeneous as far as the consumers are concerned, difference in defect rates have generated two markets.

When some firms become known for producing relatively more of the "good" version of the product and fewer of the "bad" version, such firms may come to be referred to as the "good" firms, where "good" reflects the accidental higher average virtue of the firm's employees. Good firms produce good products and look after their consumers. Bad firms produce bad products and are indifferent to the effects of their products on their consumers. Indeed, the words "goods" and "services" also have ethical connotations.

It is possible that the ethical differences in the personnel of the firms that survive gradually led to these words being used to describe a typical firm's outputs; other words could have been used.

Identifying High Quality Suppliers: Ethics and Third-Party Assessments

A single consumer will not usually purchase enough of the products of interest from all the firms in the market to be able to distinguish between the high quality and low-quality firms. Some method of aggregating the experiences of a large number of buyers across firms is usually necessary.

Markets themselves provide various signals of quality. For example, if individual consumers follow a rule like "only return to stores at which I have received high quality goods and service in my previous purchase, otherwise try a new store," relatively high-quality suppliers would have largest numbers of return customers. As more consumers leave the low-quality firms and try the high-quality ones, the market shares of high-quality firms would increase.

As the greater average quality of the large suppliers becomes noticed, size of firm may be used as another quality estimator, although again an imprecise one. Large suppliers with somewhat higher prices would tend to be (and be believed to be) more reliable sources of products than smaller ones--

again without necessarily any conscious strategy on the part of the larger store, but simply their initially better than average personnel.

Alternatively, social networks may be relied upon. Consumers may consult with one another and use "word of mouth" to distinguish among firms.[13] To the extent that the information that informs such informal recommendations is reasonably large and honest, it also creates support for high quality firms. Contemporary web vendors often have consumer comment and ranking information on their websites. If, however, the information is dishonest or not grounded in experience, little of value would be learned in this way.

Still another method of identifying high quality firms is the use of expert opinion. There are economies of scale in sampling and testing. An honest "recommendation firms" can produce more useful information than can be gathered from one's friends and neighbors or that can be deduced from the size of a firm's clientele. Both of these are partly random phenomena in the environment of interest and so "noisy" signals.

Unfortunately, third party information is only as good as the honesty and diligence of the persons providing it. Unethical firms might hire pragmatists to write testimonials about their products. They might create contests in which their products always win prizes. They may create organizations that assess product quality in which their products usually come out on top. Indeed, most firms routinely use such practices in their advertising campaigns.[14]

Consumers might, thus, look for independent expert assessments of quality. However, independent organizations may also sell their quality assessments to firms. For example, an independent automobile magazine might improve assessments of the cars of manufacturers that purchase the most advertising in them. The testers themselves could be rewarded (bribed) to tout a manufacturer's cars or to report relatively negative assessments of their rival's products. The assessments of nonprofit organizations may similarly reflect manufacturer "donations" to the organization's fundraising campaigns.

As far as consumers are concerned, honest competent assessments tend to look essentially the same as dishonest sloppily conducted assessments. "Expert assessments" are simply another product available with higher and lower defect rates that cannot easily be appraised by consumers at the point of sale. As

[13] See Paula Fitzgerald (1995) for evidence that word of mouth accounts of quality are highly influential.

[14] Nelson (1974) suggest that the size of an advertising campaign can itself be used as a proxy for quality, insofar as it makes the most economic sense to spend one's advertising dollars on products most likely to sell in the long run.

1

true of other products, the reliability of third-party assessments of quality are more likely to be higher when the person's undertaking them are honest and diligent, other things being equal.

When consumers believe that particular organizations tend to be staffed by such person and/or have adopted internal rules that tend to encourage honest accurate assessments, the assessment organizations with relatively stronger reputations for diligence and accuracy will tend to supplant those with lesser reputations, other things being equal.[15]

Ethics, thus, plays several roles in the processes through which market forces can increase the average quality of products (or reduce the cost of otherwise equivalent products). And, this is the true even in settings in which ethical dispositions, per se, are not directly rewarded or supported by markets.

The Demand for and Distribution of Ethical Employees

Of course, differences in quality of output are not likely to remain an accident in the long run. When quality pays, pragmatic firm owners will investigate the source of their higher profits and take steps to assure that their relatively high profits continue. In the present context this will induce employers to attempt to distinguish among the ethical dispositions of their existing employees and their new hires.

In the United States, *Consumer Reports*, has a reputation for high quality reviews of all sorts of products. Their nonprofit nature implies that firms cannot bribe them to overrate their merchandise.

For profit magazines and websites also undertake product assessments, but somewhat less reliably, because they are open to influence by their advertisers.

However, they can only bias their assessments within limits without losing their readership and thereby their advertisers. The signal of quality from such sources is thus somewhat unreliable but can still be useful. Objective information may be honestly produced and provided, as when car magazines provide evidence of noise levels in decibels, standardized acceleration rates, top speeds, gas mileage, etc. Other subjective characteristics may be shaded to favor their advertisers as with style or ease of use assessments. Nonetheless, a reputation for honest, diligent, assessments clearly increase readership and advertising revenues from the most honest firms, albeit at the cost of lower revenues from less honest firms.

When the latter spend more on advertising than the former, such private sources of information tend to be unreliable, and the magazines remain in business for reasons other than their quality assessments such as the quality of their prose and photos. Insofar as magazine subscribers can distinguish between informative and non-informative magazine, a spectrum of more or less informative magazines may be supported by markets.

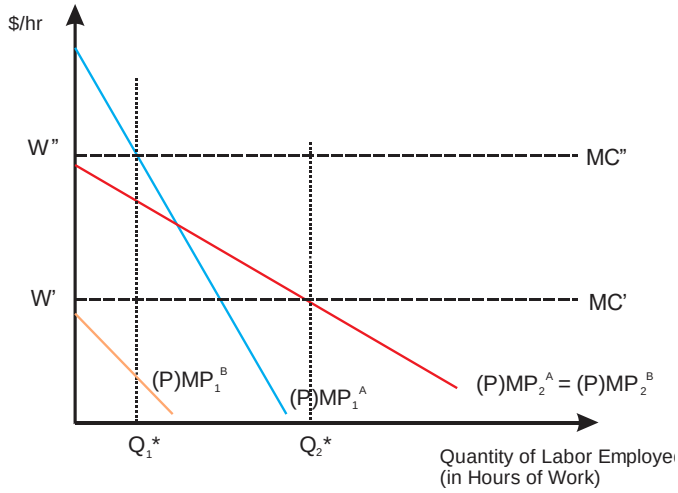
Markets will tend to economize on virtue if it is relatively scarce. Not every virtue on a philosopher's list is likely to increase production, reduce defect rates, or increase the quality of output, but those that do would be sought and, if necessary paid for. Not all industries benefit equally from honest or diligent employees, because effort and diligence are more easily observed in some production processes than others. Similarly, not all positions within a firm benefit from the same virtues or to the same degree.

On the Distribution of Ethical Persons Within an Organization

The figure below illustrates hiring decisions for two different positions within a single firm. The lower marginal revenue product curve represents a position in which monitoring is relatively easy and so the marginal product of honesty in that position is relatively low. The higher marginal revenue product curve represents a position in which monitoring is difficult or honesty especially important (MP">MP').

Output prices are the same (P), because both employees are assumed to be members on the same production team. The upward slope of the marginal cost (MC) curve reflects the scarcity of relatively virtuous persons with the skill set necessary for the jobs of interest. To facilitate comparisons across the two positions, assume that the supply of more or less ethical persons available is similar for both positions.

Figure 10.5 Hiring More (A) or Less (B) Ethical Persons for Two Different Positions (1 or 2)



Given a range of potential employees and their costs, the employer is willing to pay for different degrees of virtue for the two positions. For some tasks, honesty is more important than for others (MP">MP') and honesty pays (W">W').

For more on the effects of internalized norms on markets, see Congleton R. D. (2022). *Solving Social Dilemmas: Ethics, Politics, and Prosperity*. Oxford University Press. Or take his Moral Foundations of Capitalism course.

V. Welfare Economics and Externalities

Welfare economics is the normative strand of microeconomics. It is derived from ideas from utilitarian philosophy which had a significant influence on economics as it developed in the 19th and 20th centuries. As a normative framework, welfare economics uses "social net benefits" as a norm. We have done so a bit in this course, as when we analyzed the deadweight losses of monopoly and taxes. Deadweight losses in both cases were instances in which some potential gains from trade were left unrealized—in other words, they were cases in which social net benefits were not maximized.

Generally, competitive markets tend maximize social net benefits and monopoly (ignoring their Schumpeterian advantages) tend to produce social net benefits, but do not maximize them.

Another case in which markets may not maximize social net benefits is that in which there are "externalities," the topic dealt with in this section of chapter 10.

The Nature and Geometry of Externalities

DEF: An *externality* occurs whenever a decision made by an individual or group has effects on others not involved in the decision. That is to say, an externality exists whenever some activity imposes spillover costs or benefits on other persons not directly involved in the activity being analyzed.

The existence of externality "problems" follows from the normative framework that focuses on maximizing social net benefits. Given these (positive) predictions about firms and consumers, there are often "externality problems," which is to say market outcomes that fail to maximize social net benefits. How large the problem is varies with the size of the externalities (marginal external costs and/or benefits).

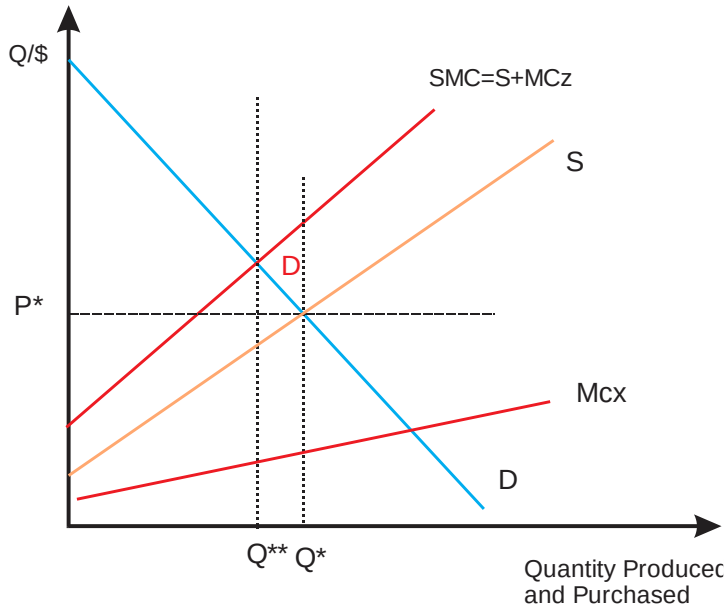
The problem from the point of view of welfare economics is not externalities themselves, but rather that the wrong level (too much or too little) of the externality generating activity gets produced to maximize social net benefits.

- Externality problems occur because individuals are assumed to be self-interested and so tend to ignore spillover costs and benefits associated with their choices.
- As a consequence, some relevant costs or benefits are not taken into account by firms and/or consumers.

Figure 10.6 illustrates the geometry of a "negative" externality—negative externalities occur when production or consumption generate "spillover costs" that fall on individuals outside the market of

interest. Examples include all forms of pollution, and a variety of other costs such as congestion, allergic reactions to some types of flowers, and unpleasant noises.

Figure 10.6 Negative Externalities



From the first part of the course, we know that market demand curves can be used to characterize the marginal benefits realized by consumers and the supply curve can be used to characterize the marginal cost of production for industry.

In the absence of externalities, these curves represent all relevant costs and benefits, so the demand curve can be used to represent social marginal benefits (SMB) and the supply curve can be used to represent social marginal costs (SMC). Social net benefits are maximized at the quantity where the SMB and SMC curves cross intersect.

That is to say, social net benefits are maximized at the quantity where SMB equals SMC, which in the absence of externalities is the output produced by competitive markets.

However, the existence of externalities implies that demand and supply curves no longer fully characterize the social marginal benefits or social marginal costs in the market of interest.

In a supply and demand (market) diagram, the effect of "**spillover costs**" is represented by drawing in a new curve that characterizes the external marginal costs generated by the production of consumption of the good of interest. (Cases where spillover costs are generated by the production of consumption of the good of interest are called **negative externalities**, because they tend to reduce social net benefits.)

The predicted market outcome (Q*) is not affected by the existence of the new marginal external cost curve, because both firms and consumers are assumed to ignore the externality generated.

- Markets still "clear" where $Q^D(P^*) = Q^S(P^*)$
- But the market output no longer maximizes social net benefits.

Finding the SMC curve when there are external costs (negative externalities).

Given an external marginal cost curve (MCx), one simply adds that curve to the supply curve to characterize the true social marginal cost of production.

- To do so, first pick a quantity, Q. Then, find the industry's marginal cost from the supply curve and add the external marginal cost to it.
- Geometrically, this involves adding the **vertical distances** from the Q axis to the supply curve and to the MCx curves.
- SMC(q) = MCind(q) + MCext(q)
- Repeat with another quantity and continue to do so until the social marginal cost curve (SMC) is traced out).

The level of the activity (Q**) that **maximizes social net benefits** is (generally) found where the social marginal benefit of the activity equals its social marginal cost curve. There is said to be "an externality problem" whenever the market equilibrium output (Q*) differs from the output that maximizes social net benefits (Q**).

Externality "problems" are a result of a **normative evaluation**. They normally are said to exist when social net benefits are not being maximized.

- See figure 10.6 above.
- The social net benefits that could be realized by reducing output to Q** is labeled with a red "*D*."

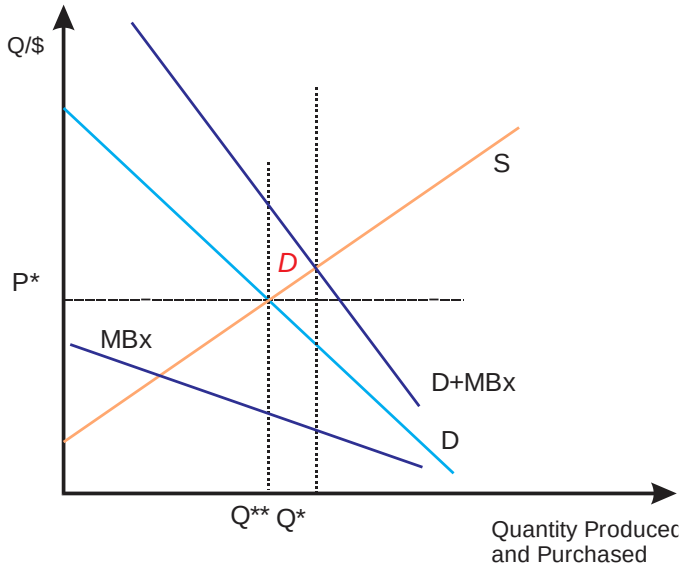
Finding the SMB curve when there are external benefits (positive externalities).

The process of finding a social marginal benefit curve is very similar in cases where there are external benefits rather than external costs that "should be" accounted for. To represent the "spillover benefits" we add an external marginal benefits curve to the market diagram. That curve is labeled, MBx in figure 10.7 below.

In cases where there is a spillover benefit, the demand curve no longer represents all the benefits from production. The social marginal benefit curve includes the benefits realized by consumers and the

spillover marginal benefits. To determine the **social marginal benefit curve** (SMB) one adds the MBx curve to the demand curve.

Figure 10.7 Positive Externalities



The geometry of deriving a SMB curve so is similar to that used to generate a SMC curve when there are spillover costs,

- First pick a quantity, q.
- Next, find the consumer marginal benefit from the demand curve.
- Next find the marginal spillover benefits from the MBx curve.
- Then add the two vertical distances together to fine a point on the SMB curve.
- SMB(q) = MBcon(q) + MBext(q)
- The social net benefits realized by producing at Q** instead of Q* is denoted with a red *D*.
- Repeat this process for other quantities until the SMB curve is traced out.

There is said to be an **externality problem** whenever the market equilibrium (Q*) differs from the output that maximizes social net benefits (Q**). Since Q* is not equal to Q** in figure 10.7, there is an **externality problem**.

Generally, any activity that imposes external losses (marginal costs) on third parties at "the margin" will be carried out at levels **greater than** those which maximizes the social net benefits from the activity. Any activity that imposes external benefits on others (marginal benefits) will be undertaken at levels **less than** that which maximizes social net benefits.

This is partly a positive prediction about behavior--that spill over costs and benefits are ignored by those controlling the activity, here consumers and firms. In other words, it assumes that no norms exist that moderate externality problems—this may not be the case. This would not necessarily be true if internalized norms induced people to take account of externalities, but generally economists assume that such norms either do not exist are too weak to solve most externality problems.

Solutions

There are a variety of possible solutions to externalities. Governments may tax the behavior that generates negative externalities or subsidize the behavior that generates positive externalities (Pigovian taxes or subsidies). Those affected may join together and bargain with those generating the externality to solve them (Coasian bargaining). Or, in some cases, the externality may cost more to address than solutions generate in benefits, so the best course is to ignore the externality.

More on all these issues is taken up in courses on Public Economics.

VI. On the Usefulness of Multi-Disciplinary Analysis of Markets

Most of the economic theories developed in this course (and in other principles of economics courses) are based on models that abstract from many details in order to reduce very complex phenomena down to their basics or their essential properties—holding other variables constant. Most of the time it does not matter too much if those other variables change a bit. The main conclusion will still hold. But problems can arise when the “other things” change in a major way—or when looking at different societies, the other things are very different.

In such cases, the “other things” cannot be ignored if one wants to understand the extent and efficiency of markets in a society through time or among societies at a moment in time.

This chapter has shown how the expected net benefit maximizing model of human decision making can be used to think systematically about some of the “other” things being held constant in most micro-economic analyses.

It has shown how and why such factors can affect the extent of markets and how they can account for differences in the effectiveness of markets as a means to advance broadly shared interests in material comforts, free time, and safety (reduced risks).

Much more can be said about all of the topics covered in this course, which is, after all, just an introduction to the field of economics. However, the overview provided in the webnotes, and your class notes should provide you with a much better understanding of how markets operate than you began the course with—where prices come from and how they help to coordinate the decisions of millions of individuals scattered around the world, but connected through market networks. They also show—at least a bit—why public policies can reduce or expand the scope of market networks and thereby the average quality of life in communities that rely on markets for both necessities and luxuries.

This chapter has added to that analysis by showing why other features of a given society or community also matter—as with crime rates, public policy choices, and the internalized norms of the persons living in a community. One does not always have to take such factors into account to have a coherent view about how a given market operates—but when markets seem to fail, it is often useful to think about how some or all of these other factors may have played a role in those failures.

Such causal elements of market performance would be missed in a more narrowly economic analysis and many conclusions would tend to be mistaken.

VII. Coda

This concludes the web text for the course. That text was written to provide a broad introduction to microeconomics—one that goes beyond the usual principle of economics course by taking account of entrepreneurship, risk, time, and non-economic factors that tend to affect markets equilibria, and thereby the extent of markets, market networks, and growth rates.

I have enjoyed sharing this material with you and hope that you found it of interest.

If you return to the introductory chapter and reread that, I think that you'll see that you have learned a lot during the course. Nonetheless, as usual, there is always more to learn.

The course is an introduction to a wealth of material, and for those who found the material important and/or interesting, many other "upper level" economics courses are provided at WVU that would be worth checking out during your studies here at WVU.

Best regards,

Prof Congleton