Preface

We have written a fundamentally different text for principles of economics, based on two premises:

1. Students are motivated to study economics if they see that it relates to their own lives.
2. Students learn best from an inductive approach, in which they are first confronted with a question and then led through the process of how to answer that question.

The intended audience of the textbook is first-year undergraduates taking courses on the principles of macroeconomics and microeconomics. Many may never take another economics course. We aim to increase their economic literacy both by developing their aptitude for economic thinking and by presenting key insights about economics that every educated individual should know.

Applications ahead of Theory

We present all the theory that is standard in books on the principles of economics. But by beginning with applications, we also show students *why* this theory is needed.

We take the kind of material that other authors put in "applications boxes" and place it at the heart of our book. Each chapter is built around a particular business or policy application, such as (for microeconomics) minimum wages, stock exchanges, and auctions, and (for macroeconomics), social security, globalization, and the wealth and poverty of nations.

Why take this approach? Traditional courses focus too much on abstract theory relative to the interests and capabilities of the average undergraduate. Students are rarely engaged, and the formal theory is never integrated into the way students think about economic issues. We provide students with a vehicle to understand the structure of economics, *and* we train them how to use this structure.

A New Organization

Traditional books are organized around theoretical constructs that mean nothing to students. Our book is organized around the *use* of economics.

Our applications-first approach leads to a fundamental reorganization of the textbook. Students will not see chapters with titles like “Cost Functions” or “Short-Run Fluctuations.” We introduce tools and ideas as, and when, they are needed. Each chapter is designed with two goals. First, the application upon which the chapter is built provides a “hook” that gets students’ attention. Second, the application is a suitable vehicle for teaching the principles of economics.

Learning through Repetition

Important tools appear over and over again, allowing students to learn from repetition and to see how one framework can be useful in many different contexts.

Each piece of economic theory is first introduced and explained in the context of a specific application. Most are reused in other chapters, so students see them in action on multiple occasions. As students progress through the book, they accumulate a set of techniques and ideas. These are collected separately in a “toolkit” that provides students with an easy reference and also gives them a condensed summary of economic principles for exam preparation.

A Truly International Book

International economics is not an afterthought in our book; it is integrated throughout.

Many other texts pay lip service to international content. We have taught in numerous countries in Europe, North America, and Asia, and we use that expertise to write a book that deals with economics in a globalized world.

Rigor without Fear

We hold ourselves to high standards of rigor yet use mathematical argument only when it is truly necessary.

We believe students are capable of grasping rigorous argument, and indeed are often confused by loose argumentation. But rigor need not mean high mathematical difficulty. Many students—even very bright ones—switch off when they see a lot of mathematics. Our book is more rigorous yet less overtly mathematical than most others in the market. We also include a math/stat toolkit to help students understand the key mathematical tools they do need.

A Textbook for the 21st Century

We introduce students to accessible versions of dynamic decision-making, choice under uncertainty, and market power from the beginning.

Students are aware that they live in an uncertain world, and their choices are made in a forward-looking manner. Yet traditional texts emphasize static choices in a world of certainty. Students are also aware that firms typically set prices and that most firms sell products that are differentiated from those of their competitors. Traditional texts base most of their analysis on competitive markets. Students end up thinking that economic theory is unrealistic and unrelated to the real world.

We do not shy away from dynamics and uncertainty, but instead introduce students to the tools of discounted present value and decision-making under uncertainty. We also place relatively more emphasis on imperfect competition and price-setting behavior, and then explain why the competitive model is relevant even when markets are not truly competitive. We give more prominence than other texts to topics such as basic game theory, statistics, auctions, and asset prices. Far from being too difficult for principles students, such ideas are in fact more intuitive, relevant, and easier to understand than many traditional topics.

At the same time, we downplay some material that is traditionally included in principles textbooks but that can seem confusing or irrelevant to students. We discuss imperfect competition in terms of market power and strategic behavior, and say little about the confusing taxonomy of market structure. We present a simplified treatment of costs that—instead of giving excruciating detail about different cost definitions—explains which costs matter for which decisions, and why.

A Non-Ideological Book

We emphasize the economics that most economists agree upon, minimizing debates and schools of thought.

There is probably less ideological debate today among economists than there has been for almost four decades. Textbooks have not caught up. We do not avoid all controversy, but we avoid taking sides. We choose and present our material so that instructors will have all the tools and resources they need to discuss controversial issues in the manner they choose. Where appropriate, we explain why economists sometimes disagree on questions of policy.

Most key economic ideas—both microeconomic and macroeconomic—can be understood using basic tools of markets, accounting identities, and budget sets. These are simpler for students to understand, are less controversial within the profession, and do not require allegiance to a particular school of thought.

A Single Voice

The book is a truly collaborative venture.

Very often, coauthored textbooks have one author for microeconomics and another for macroeconomics. Both of us have researched and taught both microeconomic and macroeconomic topics, and we have worked together on all aspects of the book. This means that students who study both microeconomics and macroeconomics from our book will benefit from a completely integrated and consistent approach to economics.

Chapter 1
What Is Economics?

Fast-Food Economics

You are just beginning your study of economics, but let us fast-forward to the end of your first economics course. How will your study of economics affect the way you see the world?

The final exam is over. You are sitting at a restaurant table, waiting for your friends to arrive. The place is busy and loud as usual. Looking around, you see small groups of people sitting and talking animatedly. Most of the customers are young; this is not somewhere your parents visit very often. At the counter, people line up to buy food. You watch a woman choose some items from the menu and hand some notes and coins to the young man behind the counter. He is about the same age as you, and you think that he is probably from China. After a few moments, he hands her some items, and she takes them to a table next to yours.

Where are you? Based on this description, you could be almost anywhere in the world. This particular fast-food restaurant is a Kentucky Fried Chicken, or KFC, but it could easily have been a McDonald's, a Burger King, or any number of other fast-food chains. Restaurants like this can be found in Auckland, Buenos Aires, Cairo, Denver, Edinburgh, Frankfurt, Guangzhou, and nearly every other city in the world. Here, however, the menu is written in French, and the customer paid in euros (€). Welcome to Paris.

While you are waiting, you look around you and realize that you are not looking at the world in the same way that you previously did. The final exam you just completed was for an economics course, and—for good or for ill—it has changed the way you understand the world. Economics, you now understand, is all around you, all the time.

1.1 Microeconomics in a Fast-Food Restaurant

LEARNING OBJECTIVE

1. What kinds of problems do we study in microeconomics?

You watch another customer go to the counter and place an order. She purchases some fried chicken, an order of fries, and a Coca-Cola. The cost is €10. She hands over a bill and gets the food in exchange. It's a simple transaction; you have witnessed exchanges like it thousands of times before. Now, though, you think about the fact that this exchange has made both the customer and the store better off than they were previously. The customer has voluntarily given up money to get food. Presumably, she would do this only if having the food makes her happier than having the €10. KFC, meanwhile, voluntarily gave up the food to get the €10. Presumably, the managers of the store would sell the food only if they benefit from the deal as well. They are willing to give up something of value (their food) in exchange for something else of value (the customer's money).

Think for a moment about all the transactions that *could* have taken place but did not. For the same €10, the customer could have bought two orders of fried chicken. But she didn't. So even though you have never met the person, you know something about her. You know that—at this moment at least—she prefers having a Coca-Cola, fries, and one order of fried chicken to having two orders of fried chicken. You also know that she prefers having that food to any number of other things she could have bought with those euros, such as a movie theater ticket, some chocolate bars, or a book.

From your study of economics, you know that her decision reflects two different factors. The first is her tastes. Each customer likes different items on the menu. Some love the spicy fried chicken; others dislike it. There is no accounting for differences in tastes. The second is what she can afford. She has a budget in mind that limits how much she is willing to spend on fast food on a given day. Her decision about what to buy comes from the interaction between her tastes and her budget. Economists have built a rich and complicated theory of decision making from this basic idea.

You look back at the counter and to the kitchen area behind it. The kitchen, you now know, is an example of a *production process* that takes inputs and produces output. Some of the inputs are perhaps obvious, such as basic ingredients like raw chicken and cooking oil. Before you took the economics course, you might have thought only about those ingredients. Now you know that there are many more inputs to the production process, including the following:

- The building housing the restaurant
- The tables and chairs inside the room
- The people working behind the cash register and in the kitchen
- The people working at KFC headquarters managing the outlets in Paris
- The stoves, ovens, and other equipment in the kitchen used to cook the food
- The energy used to run the stoves, the ovens, the lighting, and the heat
- The recipes used to convert the ingredients into a finished product

The outputs of KFC are all the items listed on the menu. And, you realize, the restaurant provides not only the food but also an additional service, which is a place where you can eat the food. Transforming these inputs (for example, tables, chickens, people, recipes) into outputs is not easy. Let us examine one output—for example, an order of fried chicken. The production process starts with the purchase of some uncooked chicken. A cook then adds some spices to the chicken and places it in a vat of very hot oil in the huge pots in the kitchen. Once the chicken is cooked, it is placed in a box for you and served to you at the counter. That production process uses, to a greater or lesser degree, almost all the inputs of KFC. The person responsible for overseeing this transformation is the manager. Of course, she doesn't have to analyze how to do this herself; the head office provides a detailed organizational plan to help her.

KFC management decides not only what to produce and how to produce it but also how much to charge for each item. Before you took your economics course, you probably gave very little thought to where those prices on the menu came from. You look at the price again: €5 for an order of fried chicken. Just as you were able to learn some things about the customer from observing her decision, you realize that you can also learn something about KFC. You know that KFC wouldn't sell an order of fried chicken at that price unless it was able to make a profit by doing so. For example, if a piece of raw chicken cost €6, then KFC would obviously make a loss. So the price charged must be greater than the cost of producing the fried chicken.

KFC can't set the price too low, or it would lose money. It also can't set the price too high. What would happen if KFC tried to charge, say, €100 for an order of chicken? Common sense tells you that no one would buy it at that price. Now you understand that the challenge of pricing is to find a balance: KFC needs to set the price high enough to earn a good profit on each order sold but not so high that it drives away too many customers. In general, there is a trade-off: as the price increases, each piece sold brings in more revenue, but fewer pieces are sold. Managers need to understand this trade-off between price and quantity, which economists call *demand*. It depends on many things, most of which are beyond the manager's control. These include the income of potential customers, the prices charged in alternative restaurants nearby, the number of people who think that going to KFC is a cool thing to do, and so on.

The simple transaction between the customer and the restaurant was therefore the outcome of many economic choices. You can see other examples of economics as you look around you—for example, you might know that the workers earn relatively low wages; indeed, they may very well be earning minimum wage. Across the street, however, you see a very different kind of establishment: a fancy restaurant. The chef there is also preparing food for customers, but he undoubtedly earns a much higher wage than KFC cooks.

Before studying economics, you would have found it hard to explain why two cooks should earn such different amounts. Now you notice that most of the workers at KFC are young—possibly students trying to earn a few euros a month to help support them through college. They do not have years of experience, and they have not spent years studying the art of cooking. The chef across the street, however, has chosen to invest years of his life training and acquiring specialized skills and, as a result, earns a much higher wage.

The well-heeled customers leaving that restaurant are likewise much richer than those around you at KFC. You could probably eat for a week at KFC for the price of one meal at that restaurant. Again, you used to be puzzled about why there are such disparities of income and wealth in society—why some people can afford to pay €200 for one meal while others can barely afford the prices at KFC. Your study of economics has revealed that there are many causes: some people are rich because, like the skilled chef, they have abilities, education, and experience that allow them to command high wages. Others are rich because of luck, such as those born of wealthy parents.

Everything we have discussed in this section—the production process, pricing decisions, purchase decisions, and the employment and career choices of firms and workers—are examples of what we study in the part of economics called microeconomics. Microeconomics is about the behavior of individuals and firms. It is also about how these individuals and firms interact with each other through markets, as they do when KFC hires a worker or when a customer buys a piece of fried chicken. When you sit in a fast-food restaurant and look around you, you can see microeconomic decisions everywhere.

KEY TAKEAWAY

- In microeconomics, we study the decisions of individual entities, such as households and firms. We also study how households and firms interact with each other.

CHECKING YOUR UNDERSTANDING

1. List three microeconomic decisions you have made today.

1.2 Macroeconomics in a Fast-Food Restaurant

LEARNING OBJECTIVE

1. What kinds of problems do we study in macroeconomics?

The economic decisions you witness inside Kentucky Fried Chicken (KFC) are only a few examples of the vast number of economic transactions that take place daily across the globe. People buy and sell goods and services. Firms hire and lay off workers. Governments collect taxes and spend the revenues that they receive. Banks accept deposits and make loans. When we think about the overall impact of all these choices, we move into the realm of macroeconomics. Macroeconomics is the study of the economy as a whole.

While sitting in KFC, you can also see macroeconomic forces at work. Inside the restaurant, some young men are sitting around talking and looking at the newspaper. It is early afternoon on a weekday, yet these individuals are not working. Like many other workers in France and around the world, they recently lost their jobs. Across the street, there are other signs that the economy is not healthy: some storefronts are boarded up because many businesses have recently been forced to close down.

You know from your economics class that the unemployed workers and closed-down businesses are the visible signs of the global downturn, or *recession*, that began around the middle of 2008. In a recession, several things typically happen. One is that the total production of goods and services in a country decreases. In many countries, the total value of all the goods and services produced was lower in 2008 than it was in 2007. A second typical feature of a recession is that some people lose their jobs, and those who don't have jobs find it more difficult to find new employment. And a third feature of most recessions is that those who do still have jobs are unlikely to see big increases in their wages or salaries. These recessionary features are interconnected. Because people have lower income and perhaps because they are nervous about the future, they tend to spend less. And because firms are finding it harder to sell their products, they are less likely to invest in building new factories. And when fewer factories are being built, there are fewer jobs available both for those who build factories and for those who work in them.

Down the street from KFC, a large construction project is visible. An old road and a nearby bridge are in the process of being replaced. The French government finances projects such as these as a way to provide more jobs and help the economy recover from the recession. The government has to finance this spending somehow. One way that governments obtain income is by taxing people. KFC customers who have jobs pay taxes on their income. KFC pays taxes on its profits. And customers pay taxes when they buy their food.

Unfortunately for the government, higher taxes mean that people and firms have less income to spend. But to help the economy out of a recession, the government would prefer people to spend more. Indeed, another response to a recession is to *reduce* taxes. In the face of the recession, the Obama administration in the United States passed a stimulus bill that both increased government spending *and* reduced taxes. Before you studied macroeconomics, this would have seemed quite mysterious. If the government is taking in less tax income, how is it able to increase spending at the same time? The answer, you now know, is that the government borrows the money. For example, to pay for the $787 billion stimulus bill, the US government issued new debt. People and institutions (such as banks), both inside and outside the United States, buy this debt—that is, they lend to the government.

There is another institution—called the monetary authority—that purchases government debt. It has specific names in different countries: in the United States, it is called the Federal Reserve Bank; in Europe, it is called the European Central Bank; in Australia, it is called the Reserve Bank of Australia; and so on. When the US government issues more debt, the Federal Reserve Bank purchases some of it. The Federal Reserve Bank has the legal authority to create new money (in effect, to print new currency) and then to use that to buy government debt. When it does so, the currency starts circulating in the economy. Similarly, decisions by the European Central Bank lead to the circulation of the euro notes and coins you saw being used to purchase fried chicken.

The decisions of the monetary authority have a big impact on the economy as well. When the European Central Bank decides to put more euros into circulation, this has the effect of reducing interest rates, which means it becomes cheaper for individuals to get a student loan or a mortgage, and it is cheaper for firms to buy new machinery and build new factories. Typically, another consequence is that the euro will become less valuable relative to other currencies, such as the US dollar. If you are planning a trip to the United States now that your class is finished, you had better hope that the European Central Bank doesn't increase the number of euros in circulation. If it does, it will be more expensive for you to buy US dollars.

Today, the world's economies are highly interconnected. People travel from country to country. Goods are shipped around the world. If you were to look at the labels on the clothing worn by the customers in KFC, you would probably find that some of the clothes were manufactured in China, perhaps some in Malaysia, some in France, some in the United States, some in Guatemala, and so on. Information also moves around the world. The customer sitting in the corner using a laptop might be in the process of transferring money from a Canadian bank account to a Hong Kong account; the person at a neighboring table using a mobile phone might be downloading an app from a web server in Illinois. This globalization brings many benefits, but it means that recessions can be global as well.

Your study of economics has taught you one more thing: the idea that you can take a trip to the United States would have seemed remarkable half a century ago. Despite the recent recession, the world is a much richer place than it was 25, or 50, or 100 years ago. Almost everyone in KFC has a mobile phone, and some people are using laptops. Had you visited a similar fast-food restaurant 25 years ago, you would not have seen people carrying computers and phones. A century ago, there was, of course, no such thing as KFC; automobiles were still a novelty; and if you cut your finger on the sharp metal edge of a table, you ran a real risk of dying from blood poisoning. Understanding why world economies have grown so spectacularly—and why not all countries have shared equally in this growth—is one of the big challenges of macroeconomics.

KEY TAKEAWAY

- In macroeconomics, we study the economy as a whole to understand why economies grow and why they sometimes experience recessions. We also study the effects of different kinds of government policy on the overall economy.

CHECKING YOUR UNDERSTANDING

1. If the government and the monetary authority think that the economy is growing too fast, what could they do to slow down the economy?

1.3 What Is Economics, Really?

LEARNING OBJECTIVE

1. What methods do economists use to study the world?

Economists take their inspiration from exactly the kinds of observations that we have discussed. Economists look at the world around them—from the transactions in fast-food restaurants to the policies of central banks—and try to understand how the economic world works. This means that economics is driven in large part by data. In microeconomics, we look at data on the choices made by firms and households. In macroeconomics, we have access to a lot of data gathered by governments and international agencies. Economists seek to describe and understand these data.

But economics is more than just description. Economists also build models to explain these data and make predictions about the future. The idea of a model is to capture the most important aspects of the behavior of firms (like KFC) and individuals (like you). Models are abstractions; they are not rich enough to capture all dimensions of what people do. Yet a good model, for all its simplicity, is still capable of explaining economic data.

And what do we do with this understanding? Much of economics is about policy evaluation. Suppose your national government has a proposal to undertake a certain policy—for example, to cut taxes, build a road, or increase the minimum wage. Economics gives us the tools to assess the likely effects of such actions and thus to help policymakers design good public policies.

This is not really what you thought economics was going to be about when you walked into your first class. Back then, you didn't know much about what economics was. You had a vague thought that maybe your economics class would teach you how to make money. Now you know that this is not really the point of economics. You don't have any more ideas about how to get rich than you did when you started the class. But your class has taught you something about how to make better decisions and has given you a better understanding of the world that you live in. You have started to think like an economist.

KEY TAKEAWAY

- Economists gather data about the world and then build models to explain those data and make predictions.

CHECKING YOUR UNDERSTANDING

1. Suppose you were building a model of pricing at KFC. Which of the following factors would you want to make sure to include in your model? Which factors do you think would be irrelevant?
    a. the age of the manager making the pricing decisions
    b. the price of chicken
    c. the number of customers who come to the store on a typical day
    d. the price of apples
    e. the kinds of restaurants nearby

1.4 End-of-Chapter Material

In Conclusion

Economics is all around us. We all make dozens of economic decisions every day—some big, some small. Your decisions—and those of others—shape the world we live in. In this book, we will help you develop an understanding of economics by looking at examples of economics in the everyday world. Our belief is that the best way to study economics is to understand how economists think about such examples.

With this in mind, we have organized our book rather differently from most economics textbooks. It is built not around the theoretical concepts of economics but around different applications—economic illustrations as you encounter them in your own life or see them in the world around you. As you read this book, we will show you how economists analyze these illustrations, introducing you to the tools of economics as we proceed. After you have read the whole book, you will have been introduced to all the fundamental tools of economics, and you will also have seen them in action. Most of the tools are used in several different applications, thus allowing you to practice using them and gain a deeper understanding of how they work.

You can see this organization at work in our table of contents. In fact, there are two versions of the table of contents so that both students and instructors can easily see how the book is organized. The student table of contents focuses on the applications and the questions that we address in each chapter. The instructor table of contents lists the theoretical concepts introduced in each chapter so that instructors can easily see how economic theory is developed and used in the book.

We have also gathered all the tools of economics into a toolkit. You will see many links to this toolkit as you read the book. You can refer to the toolkit as needed when you want to be reminded of how a tool works, and you can also use it as a study aid when preparing for exams and quizzes.

EXERCISES

1. A map is a model constructed by geographers and cartographers. Like an economic model, it is a simplified representation of reality. Suppose you have a map of your hometown in front of you. Think of

one question about your town that you could answer using the map. Think of another question about your town for which the map would be useless.

2. Which of the following questions do you think would be studied by a macroeconomist and which by a microeconomist? (Note: we don't expect you to be able to answer all these questions yet.)
    a. What should the European Central Bank do about increasing prices in Europe?
    b. What happens to the price of ice cream in the summer?
    c. Should you take out a student loan to pay for college?
    d. What happens when the US government cuts taxes and pays for these tax cuts by borrowing money?
    e. What would happen to the prices of computers if Apple and Microsoft merged into a single firm?

Economics Detective

1. Look at a newspaper on the Internet. Find a news story about macroeconomics. How do you know that it is about macroeconomics? Find a news story about microeconomics. How do you know that it is about microeconomics?

Chapter 2
Microeconomics in Action
2.1 Four Examples of Microeconomics

LEARNING OBJECTIVES

1. What are two ways that you make economic choices all the time?
2. How do economists think about the way people react to a change in a rule?
3. What is the role of markets in an economy?

Here are four short and diverse illustrations of microeconomics you might encounter: deciding what to do with your time and money, buying or selling on eBay, visiting a large city, and reading about a soccer game. After you have finished your study of microeconomics, you will see these concepts very differently from the way you see them now. You may not know it, but your everyday life is filled with microeconomics in action.

Your Time and Money

Wouldn't you rather be doing something else with your time right now, instead of reading an economics textbook? You could be surfing on the Internet, reading blogs, or updating your Facebook profile. You could be reading a novel or watching television. You could be out with friends. But you aren't. You have made a choice—a decision—to spend time reading this chapter.

Your choice is an economic one. Economics studies how we cope with competing demands for our time, money, and other resources. You have only 24 hours each day, so your time is limited. Each day you have to divide up this time among the things you like or need to do: sleeping, eating, working, studying, reading, playing video games, hanging out in your local coffee shop, and so on. Every time you decide to do one thing instead of another, you have made an economic decision. As you study economics, you will learn about how you and other people make such choices, and you will also learn how to do a better job when making these decisions.

Money is also a limited resource. You undoubtedly have many things you would like to buy if money were no object. Instead you must choose among all the different things you like because your money—or, more precisely, your income—is a limited resource. Every time you buy something, be it a T-shirt, a breakfast bagel, or a new computer, you are choosing to forgo something else you could have bought instead. Again, these are economic decisions. Economics is about how you make choices. Whenever there is a limited resource—be it your time, the amount of oil reserves in the world, or tickets to the Super Bowl—and decisions to be made about how to use that resource, then economics is there to help. Indeed, the fundamental definition of economics is that it is the study of how we, as individuals and as a society, allocate our limited resources among possible alternative uses.

eBay and craigslist

Suppose you want to buy an MP3 player. There are many ways you can do this. You can go to a local store. You can look for stores on the Internet. You can also visit sites such as eBay (http://www.ebay.com) or craigslist (http://www.craigslist.org). eBay is an online auction site, meaning that you can look for an MP3 player and then bid against other potential buyers. The site craigslist is like an online version of the classified advertisements in a newspaper, so you can look to see if someone in your town or city is selling the player you want to buy. You can also use these sites if you want to sell something. Maybe you have some old baseball cards you want to sell. Perhaps you have a particular skill (for example, web design), and you want to sell your services. Then you can use sites such as eBay or craigslist as a seller instead of as a buyer.

We have said that economics is about deciding how to use your limited resources. It is also about how we interact with one another, and, more precisely, how we trade with one another. Adam Smith, the founder of modern economics, observed that humans are the only animal that makes bargains: "Nobody ever saw a dog make a fair and deliberate exchange of one bone for another with another dog." [1] Barter or trade—the exchange of goods and services and money—is central to the world we live in today.

Economists often talk about trade taking place in markets. Some exchanges do literally take place in markets—such as a farmers' market where local growers bring produce to sell. Economists use the term more generally, though: a market is any institution that allows us to exchange one thing for another. Sites such as eBay and craigslist create markets in which we can transact. Normally, we exchange goods or services for money. Sometimes we exchange one good or service for another. Sometimes we exchange one type of money for another.

Most of the time, nobody forces you to buy anything, so when you give up some money in return for an MP3 player, you are presumably happier after the transaction than before. (There are some exceptions, of course. Can you think of any cases where you are forced to engage in an economic transaction?) Most of the time, nobody forces you to sell anything, so when you give up your time in return for some money, you are presumably happier after the transaction than before. Leaving aside the occasional mistake or the occasional regret, nearly every voluntary transaction makes both participants better off. Markets matter because they are a means for people to become happier.

Breathing the Air

Welcome to Mexico City! It is a wonderful place in many respects. But not in every way: from the picture you can see that Mexico City has some of the most polluted skies in the world. [2]

Mexico City was not always so polluted. Sadly, economic growth and population growth, together with the peculiarities of geography and climate, have combined to make its air quality among the worst you will encounter anywhere. Other cities around the world, from Beijing to Los Angeles, also experience significant air pollution, reducing the quality of life and bringing with it health risks and other costs.

It is hard to understand economists talking about the beauty and power of markets when you cannot breathe the air. So what is going wrong in Mexico City? Is it not full of people carrying out trades that make them better off? The problem is that transactions sometimes affect other people besides the buyer and the seller. Mexico City is full of gas stations. The owners of the gas stations are happy to sell gasoline because every transaction makes them better off. The owners of cars are happy to buy gasoline because every transaction makes them better off. But a side effect of all these transactions is that the air becomes more and more polluted.

Economics studies these kinds of problems as well. Economists seek to understand where and when markets work and where and when they don't work. In those situations where markets let us down, economists search for ways in which economic policies can help.

Changing the Rules

We have explained that microeconomics studies choices and the benefits and problems that arise from trade. Perhaps most fundamentally, microeconomics studies how people respond to incentives. To illustrate the importance of incentives, here is an example of what can happen when they go wrong.

In February 1994, an extraordinary scene took place during a soccer match in the Caribbean. Grenada was playing Barbados, and with five minutes remaining in the match, Barbados was leading by two goals to one. As the seconds ticked away, it seemed clear that Barbados was going to win the match. Then, three minutes from the end of the game, the Barbados team did a remarkable thing. It intentionally scored an own goal, tying the game at two goals apiece.

After Grenada kicked off again, pandemonium ensued. The Grenada team tried not only to score against Barbados but also to score an own goal. Barbados desperately defended both its own goal and its opponents' goal. The spectacle on the field had very little to do with soccer as it is usually played.

To explain this remarkable sight, we must describe the tournament in which the two teams were playing. There were two groups of teams, with the winner of each group progressing to the final. The match between Barbados and Grenada was the last group game and would determine which two teams would be in the final. The results of the previous matches were such that Barbados needed to win by two goals to go to the final. If Barbados won by only one goal, then Grenada would qualify instead. But the tournament organizers had introduced an unusual rule. The organizers decided that if a game were tied, the game would go to "golden goal" overtime, meaning that the first team to score would win the game, and they had also decided that the winning team would then be awarded a two-goal victory.

As the game was drawing to a close, Barbados realized it was unlikely to get the two-goal win that it needed. The team reasoned that a tie was a better result than a one-goal victory because it gave them roughly a fifty-fifty chance of winning in extra time. So Barbados scored the deliberate own goal. Grenada, once it realized what had happened, would have been happy either winning or losing by one, so it tried to score in either goal. Barbados' strategy paid off. The game finished in a tie; Barbados scored in overtime and went on win the final.

The organizers should have consulted an economist before instituting the rules of the tournament. Economics has many lessons to teach, and among the most important is this: people respond to incentives. The change in the rules changed the incentives that the two teams faced. Because the tournament organizers had not realized that their rules could lead to a situation in which a team preferred a tie to a win, they failed to foresee the bizarre scene on the field. [3]

KEY TAKEAWAYS

- You make economic decisions on the allocation of time by deciding how to spend each minute of the day. You make economic decisions on the allocation of your income by deciding how much to buy of various goods and services and how much to save.
- Economists study how changes in rules lead individual and firms to change their behavior. This is part of the theme in economics that incentives matter.
- Markets are one of the central ways in which individuals interact with each other. Market interactions provide a basis for the trade that occurs in an economy.

CHECKING YOUR UNDERSTANDING

1. When you are choosing how much time to allocate to studying, what incentives affect your decision? Does the decision depend on how much money you have? Does the decision depend on whether you have a quiz or an exam coming up in the course? If your instructor changed the rules of the course—for example, by canceling the final exam—would your choice change?
2. Instead of writing about air pollution in Mexico City, we could have written about water pollution from the 2010 oil spill in the Gulf of Mexico. Would that also be a good example of markets failing?

[1] Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (New York: Modern Library, 1994 [1776]), 14.

[2] "Researchers to Scrutinize Megacity Pollution during Mexico City Field Campaign," *University Corporation for Atmospheric Research*, last modified March 2, 2006, accessed January 22, 2011, http://www.ucar.edu/news/releases/2006/mirage.shtml.

[3] "Football Follies," *Snopes.com*, last modified July 6, 2008, accessed January 22, 2011,http://www.snopes.com/sports/soccer/barbados.asp.

2.2 The Microeconomic Approach

LEARNING OBJECTIVES

1. What is the approach of microeconomics?
2. What are the big questions of economics?

There are several distinguishing features of the microeconomic approach to the world. We discuss them briefly and then conclude with a look at the big questions of economics.

Individual Choice

One element of the microeconomic approach is individual choice. Throughout this book, we explore how individuals make decisions. Economists typically suppose that individuals make choices to pursue their (broadly defined) self-interest given the incentives that they face.

We look at individuals in their roles both as members of households and as members of firms. Individuals in households buy goods and services from other households and—for the most part—firms. They also sell their labor time, mostly to firms. Managers of firms, meanwhile, make decisions in the effort to make their firms profitable. By the end of the book, we will have several frameworks for understanding the behavior of both households and firms.

Individuals look at the prices of different goods and services in the economy when deciding what to buy. They act in their own self-interest when they purchase goods and services: it would be foolish for them to buy things that they don’t want. As prices change, individuals respond by changing their decisions about which products to buy. If your local sandwich store has a special on a breakfast bagel today, you are more likely to buy that sandwich. If you are contemplating buying an Android tablet computer but think it is about to be reduced in price, you will wait until the price comes down.

Just as consumers look at the prices they face, so do the managers of firms. Managers look at the wages they must pay, the costs of the raw materials they must purchase, and so on. They also look at the willingness of consumers to buy the products that they are selling. Based on all this information, they decide how much to produce and what to buy. Your breakfast bagel may be on special because the owner of your local sandwich shop got a good deal on bagels from the supplier. So the owner thinks that breakfast bagels can be particularly profitable, and to sell a lot of them, she sets a lower price than normal.

The buying and selling of a bagel may seem trivial, but similar factors apply to much bigger decisions. Potential students think about the costs and benefits of attending college relative to getting a full-time job. For some people, the best thing to do is to work full time. For others, it is better to go to school full time. Yet others choose to go to school part time and work part time as well. Presumably your own decision—whichever of these it may be—is one you made in your own best interests given your own specific situation.

From this discussion, you may think that economics is all about money, but economists recognize that much more than money matters. We care about how we spend our time. We care about the quality of the air we breathe. We care about our friends and family. We care about what others think of us. We care about our own self-image: what sort of a person am I? Such factors are harder to measure and quantify, but they all play a role in the decisions we make.

Markets

A second element of microeconomics has to do with how individual choices are interconnected. Economics is partly about how we make decisions as individuals and partly about how we interact with one another. Most importantly—but not exclusively—economics looks at how people interact by purchasing and selling goods and services.

In a typical transaction, one person (the buyer) hands over money to another (the seller). In return, the seller delivers something (a good or a service) to the buyer. For example, if you buy a chocolate bar for a dollar, then a dollar bill goes from your hands to those of the seller, and a chocolate bar goes from the seller to you. At the level of an individual transaction, this sounds simple enough. But the devil is in the details. In any given (potential) transaction, we can ask the following questions:

- How many? Will you buy 1, 2, or 10 chocolate bars? Or will you buy 0—that is, will the transaction take place at all?
- How much? How much money does the buyer give to the seller? In other words, what is the price?

You will see in different chapters of this book that the answers to these questions depend on exactly how buyers and sellers interact. We get a different answer depending on whether there are many sellers or only a few. We get a different answer if the good is sold at a retail store or at an auction. We get a different answer if buyers and sellers can or cannot negotiate. The exact way in which people exchange goods and services matters a great deal for the *how many?* and*how much?* questions and thus for the gains from trade in the economy.

The Role of Government

We have pointed out that individuals acting in their own self-interest benefit from voluntary trade. If you are not forced to buy or sell, then there is a presumption that every transaction makes the participants happier. What is more, markets are often a very effective institution for allowing people to meet and trade with one another. In fact, there is a remarkable result in economics that—under some circumstances—individuals acting in their own self-interest and trading in markets can manage to obtain *all* the possible benefits that can come from trading. Every transaction carried out is for the good, and every good transaction is carried out. From this comes a powerful recommendation: do whatever is possible to encourage trade. The phrase *under some circumstances* is not a minor footnote. In the real world, transactions often affect people other than the buyer and the seller, as we saw in our example of gas stations in Mexico City. In other cases, there can be problems with the way that markets operate. If there is only a small number of firms in a market, then managers may be able to set high prices, even if it means that people miss out on some of the benefits of trade. Later in this book, we study exactly how managers make these decisions. The microeconomic arguments for government intervention in the economy stem from these kinds of problems with markets. In many chapters, we discuss how governments intervene in an attempt to improve the outcome that markets give us. Yet it is often unclear whether and how governments should be involved. Pollution in Mexico City illustrates how complex these problems can be. First, who is responsible for the pollution? Some of it comes from people and firms outside the city and perhaps even outside the country. If pollution in Mexico City is in part caused by factories in Texas, who should deal with the problem: the Mexico City government, the Mexican government, the US government, or the Texas state legislature? Second, how much pollution should we tolerate? We could shut down all factories and ban all cars, but few people would think this is a sensible policy. Third, what measures can we use to combat air pollution? Should we simply place limits on production by firms and the amount of driving? Should we use some kind of tax? Is there a way in which we can take advantage of our belief that people, including the managers of firms, respond to incentives?

There are two traps that we must avoid. The first is to believe that markets are the solution to everything. There is no imaginable market in which the residents of Mexico City can trade with the buyers and sellers of gasoline to purchase the right amount of clean air. The second trap is to believe that the government can fix every market failure. Governments are collections of individuals who respond to their own incentives. They can sometimes make things better, but they can sometimes make things worse as well.

There is room for lots of disagreement in the middle. Some economists think that problems with markets are pervasive and that government can do a great deal to fix these problems. Others think that such problems are rare and that governmental intervention often does more harm than good. These disagreements result partly from different interpretations of the evidence and partly from differences in politics. Economists are as prone as everyone else to view the world through their own ideological lens. As we proceed, we do our best to present the arguments on controversial issues and help you understand why even economists sometimes come to differing conclusions about economic policy.

Incentives

Perhaps our story of the Barbados-Grenada soccer game did not seem related to economics. Economists believe, though, that the decisions we make reflect the incentives we face. Behavior that seems strange—such as deliberately scoring an own goal in a soccer game—can make perfect sense once you understand the underlying incentives. In the economic world, it is often governments that make the rules of the game; like the organizers of soccer tournaments, governments need to be careful about how the rules they set can change people's behavior.

Here is an example. In some European countries, laws are in place that give a lot of protection to workers and keep them from being unfairly fired by their employers. The intentions of these laws are good; some of their consequences are not so beneficial. The laws also make firms more reluctant to *hire* workers because they are worried about being stuck with an unsuitable employee. Thus these laws probably contribute to higher unemployment.

Incentives affect all transactions. When you buy a breakfast bagel on sale, both you and the owner of the sandwich shop are responding to the incentives that you face. The owner responds to the lower price of bagels. You respond to the lower price of the sandwich. Economists think that we can understand a great deal about people's behavior if we have a good understanding of the incentives that they face.

Notice that not everyone makes the same choices. There are two main reasons for this:

- People have different desires or tastes. Some people like bagels; others hate them. Some people like being students; others would prefer to work rather than study.
- People have different incentives. Some people face very different job prospects and thus make different decisions about schooling. If you have this great idea for a new web product (for example, the next Google or Facebook), then you might be wise to spend your time on this project instead of studying.

The Big Questions of Economics

To conclude our introduction to microeconomics, let us look at the big picture of what happens in an economy. An economy possesses some *resources*. These include the time and abilities of the people who live in the economy, as well as natural resources such as land, mineral deposits, and so on. An economy also possesses some *technologies*. A technology is a means of changing, or transforming, one set of things into other things. For example, we have a technology for making tea. This technology takes cold water, energy, and dried leaves and transforms them into a hot beverage. Finally, an economy, of course, contains its people, and these people like to consume things. Economics studies all aspects of this process. It considers the following:

- **What** goods and services are produced in an economy?
- **How** are these goods and services produced?
- **Who** gets to consume these goods and services?

These questions concern the *allocation of resources*.

The *what* in the first question reflects the choice among the multitude of goods and services an economy could produce. Think for a moment about the clothes you are wearing right now, the food you have eaten today, and the activities you undertake during a typical day. Someone made those clothes; someone prepared that food. Somehow, society must decide how much of each type of good and service to produce.

The *how* in the second question reflects competing ways to produce goods and services. Take a basic commodity such as rice. A large amount of rice is produced in the United States on large-scale, mechanized farms. A large amount of rice is also produced in Vietnam, but the production methods are very different. In Vietnam, people do much more work manually rather than by machine. A big part of the *how* question is deciding what mix of resources and what technologies should be used to produce goods and services. The answer in a rich country such as the United States is frequently different from the answer in a poor country such as Vietnam. Indeed, the answer may be different in different states in the United States or in the same place at different times.

The *who* in the third question concerns the distribution of goods and services in the economy. Suppose you were responsible for the distribution of all goods and services to your family. If there are 4 people in your family and each consumed 50 products in a typical day, you would have to make about 200 allocation decisions each day. It would be a very hard task. Yet somehow the economies of the world allocate billions of products to billions of people.

These three questions are answered in the world partly through individual decisions. The way in which you allocate your time each day is part of the allocation of resources in the economy. If each of us lived alone, engaging in subsistence farming and not interacting with others, then we would each determine our own allocation of resources. Because we interact with others, however, these questions are also answered in part by the way in which society is organized. Most of us produce only a few goods but consume many. We *specialize in production* and *generalize in consumption*. To do so, we must exchange what we produce with others. Most of these exchanges take place as a result of individual decisions in different kinds of markets. It is the operation of these countless markets that determines the allocation of goods and services in the economy. Remarkably, these markets somehow coordinate the decisions of the billions of people in the world economy.

Some of these exchanges are controlled by the government. In some economies, the government plays a very active role; in others, it intervenes less. When a government makes decisions about the allocation of resources, this is another mechanism in the production of goods and the distribution to individuals.

KEY TAKEAWAYS

- The approach of microeconomics starts with the decisions of an individual about the allocation of time and income. The impact of incentives on individual choices is a key part of economics. The approach of microeconomics then looks at the interactions of individuals directly and in markets.
- Economics answers the questions of what goods and services are produced, how they are produced, and who consumes them.

CHECKING YOUR UNDERSTANDING

1. We said that most people *specialize in production* and *generalize in consumption*. What goods or services (if any) do you produce? What are the most important goods and services that you consume?
2. Police protection is a service provided by most governments. What are the*what*, *how*, and *who* aspects of the provision of this service?

2.3 End-of-Chapter Material

In Conclusion

Our book is built around economic topics. Examples of these topics include the decisions you make in your everyday life, auctions such as those you see on eBay, whether you can make money on Wall Street, where jobs come from, and health care. As we introduce and discuss these applications, we remain keenly aware of the key themes in microeconomics: individuals responding to incentives, markets as the basis for interactions among firms and households, and the role of government intervention.

Throughout this book, we emphasize the measurement and interpretation of economic data. Understanding how to read charts and tables of economic data is a critical skill for anyone who wants to be a sophisticated consumer of economic and political news.

Mastering microeconomics involves both understanding the tools that microeconomists use and knowing how and when those tools should be applied. In this book, you will learn about these tools by example; you will see them in use as we study different questions in economics. At the same time, you will learn about many topics that should interest you as engaged and aware citizens of the world. We hope that, after reading this book, you will both better understand what it is that economists do and be better informed about the world in which we all live.

There is a considerable amount of core material in microeconomics that we use repeatedly as we tackle different problems. We highlight these core elements in the chapters and also gather them together in the toolkit. You can read any and every chapter in the book without necessarily having to refer to the toolkit, but you may often find it to be a helpful reference.

EXERCISES

1. Think about the last item of clothing you bought for yourself. How much did you spend on it? List three other things that you like and could have bought with (approximately) the same amount of money. Why did you decide to buy the clothing rather than one of the things you just listed?
2. How have you spent the previous 24 hours? How much time did you spend sleeping? How much time did you spend working? What else could you have done with your time? Why are you reading this chapter instead of doing something else with your time?
3. Think about a game or sport that you enjoy. What rule of that game could be changed? How would this change in the rules affect the way in which the players behave?

4. When we discussed individual choice, we talked mainly about the choices of an individual person. However, in economics we often talk about the choice of a household consisting of two or more people. In what ways are the choices of a household different from the choices of an individual? In what ways are they similar?
5. Can you think of examples of economic choices that are made by the government?

Economics Detective

1. We explained the social problem of air pollution in Mexico as a situation where markets have failed to bring about good outcomes. Instead of writing about pollution, we could have written about other social problems, such as crime, illiteracy, or obesity. Browse the Internet to find another example of a social problem—either from this list or something else that interests you. Write one paragraph that explains the problem and another that discusses if and how the government might solve the problem.

Chapter 3
Everyday Decisions
You and Your Choices

Economics is about you. It is about how you make choices. It is about how you interact with other people. It is about the work you do and how you spend your leisure time. It is about the money you have in your pocket and how you choose to spend it. Because economics is about your choices plus everyone else's, this is where we begin. As far as your own life is concerned, you are the most important economic decision maker of all. So we begin with questions you answer every day:

- What will I do with my money?
- What will I do with my time?

Economists don't presume to tell you what you *should* do with your time and money. Rather, studying economics can help you better understand your own choices and make better decisions as a consequence. Economics provides guidelines about how to make smart choices. Our goal is that after you understand the material in this chapter, you will think differently about your everyday decisions.

Decisions about spending money and time have a key feature in common:*scarcity*. You have more or less unlimited desires for things you might buy and ways that you might spend your time. But the time and the money available to you are limited. You don't have enough money to buy everything you would like to own, and you don't have enough time to do everything you would like to do.

Because both time and money are scarce, whenever you want more of one thing, you must accept that you will have less of something else. If you buy another game for your Xbox, then you can't spend that money on chocolate bars or movies. If you spend an hour playing that game, then that hour cannot be spent studying or sleeping. Scarcity tells us that everything has a cost. The study of decision making in this chapter is built around this tension. Resources such as time and money are limited even though desires are essentially limitless.

Road Map

We tackle the two questions of this chapter in turn, but you will see that there are close parallels between them. We begin by looking at spending decisions. Although we have said that money is scarce, a more precise statement is that you have limited *income*. (Economists usually use the term "money" more specifically to mean the assets, such as currency in your wallet or funds in your checking account, that you use to buy things.) Because your income is limited, your spending opportunities are also limited. We show how to use the prices of goods and services, together with your income, to analyze what spending decisions are *possible* for you. Then we think about what people's wants and desires look like. Finally, we put these ideas together and uncover some principles about how to make choices that will best satisfy these desires.

Your decisions about what to buy therefore depend on how much income you have and the prices of goods and services. Economics summarizes these decisions in a simple way by using the concept of *demand*. We show how demand arises from the choices you make. Demand is one of the most useful ideas in economics and lies at the heart of almost everything we study in this book.

Finally, we turn to the decision about how to spend your time. Again, we begin with the idea that your resources are limited: there are only so many hours in a day. As with the spending decision, you have preferences about how to spend your time. We explain the principles of good decision making in this setting. Based on this analysis, we introduce another central economic idea—that of *supply*.

Economics is both *prescriptive* and *descriptive*. Economics is prescriptive because it tells you some rules for making good decisions. Economics is descriptive because it helps us explain the world in which we live. As well as uncovering some principles of good decision making, we discuss whether these are also useful descriptions of how people actually behave in the real world.

3.1 Individual Decision Making: How You Spend Your Income

LEARNING OBJECTIVES

1. What are an individual's budget set and budget line?
2. What is an opportunity cost?
3. How do people make choices about how much to consume?
4. What features do we expect most people's preferences to have?
5. What does it mean to make rational choices?

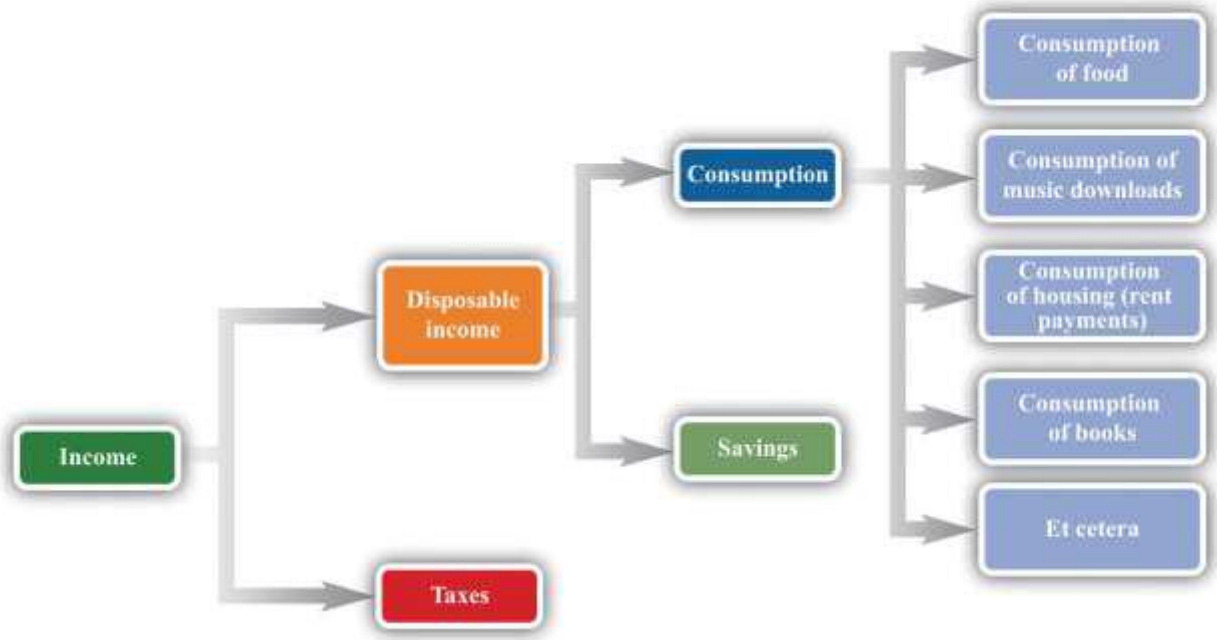
We start with the decision about how to spend your income. We want to know what possibilities are available to you, given that your income is limited but your desires are not.

The Budget Set

We describe your personal decision making on a month-by-month basis (although we could equally well look at daily, weekly, or even annual decisions because the same basic ideas would apply). Suppose you receive a certain amount of income each month—perhaps from a job or a student grant. The government takes away some of this income in the form of taxes, and the remainder is available for you to spend. We call the income that remains after taxes your disposable income.

You may want to put aside some of this income for the future; this is your savings. The remainder is your consumption, which is your spending on all the goods and services you buy this month: rent, food, meals out, movies, cups of coffee, CDs, music downloads, DVD rentals, chocolate bars, books, bus rides, haircuts, and so on. Figure 3.1 "What You Do with Your Income" shows this process.

Figure 3.1 What You Do with Your Income



Here is a schematic view of what happens to your income.

This view of your paycheck involves several economic decisions. Some of these are decisions made by the government. Through its tax policies, the government decides how much of your income it takes from you and how much is left as disposable income. You make other decisions when you allocate your disposable income among goods and services today and in the future. You choose how to divide your disposable income between consumption this month and saving for the future. You also decide exactly how much of each good and service you purchase this month. We summarize your ability to purchase goods and services by your budget set.

Toolkit: Section 17.1 "Individual Demand"

The budget set is a list of all the possible combinations of goods and services that are affordable, given both income and the prices of all goods and services. It is defined by

total spending ≤ disposable income.

Begin by supposing you neither save nor borrow. We can construct your budget set in three steps.

1. Look at spending on each good and service in turn. For example, your monthly spending on cups of coffee is as follows:

   spending on coffee = number of cups purchased × price per cup.

   A similar equation applies to every other good and service that you buy. Your spending on music downloads equals the number of downloads times the price per download, your spending on potato chips equals the number of bags you buy times the price per bag, and so on.
2. Now add together all your spending to obtain your total spending:

   total spending = spending on coffee + spending on downloads + … ,

   where … means including the spending on every different good and service that you buy.
3. Observe that your total spending cannot exceed your income after taxes:

   total spending ≤ disposable income.

You are consuming within your budget set when this condition is satisfied.

In principle, your list of expenditures includes every good and service you could ever imagine purchasing, even though there are many goods and services you never actually buy. After all, your spending on Ferraris every month equals the number of Ferraris that you purchase times the price per Ferrari. If you buy 0 Ferraris, then your spending on Ferraris is also $0, so your total spending does include all the money you spend on Ferraris.

Imagine now that we take some bundle of products. *Bundle* here refers to any collection of goods and services—think of it as being like a grocery cart full of goods. The bundle might contain 20 cups of coffee, 5 music downloads, 3 bags of potato chips, 6 hours of parking, and so on. If you can afford to buy this bundle, given your income, then it is in the budget set. Otherwise, it is not.

The budget set, in other words, is a list of all the possible collections of goods and services that you can afford, taking as given both your income and the prices of the goods and services you might want to purchase. It would be very tedious to write out the complete list of such bundles, but fortunately this is unnecessary. We merely need to check whether any given bundle is *affordable*or not. We are using affordable not in the casual everyday sense of "cheap" but in a precise sense: a bundle is affordable if you have enough income to buy it.

It is easiest to understand the budget set by working though an example. To keep things really simple, suppose there are only two products: chocolate bars and music downloads. An example with two goods is easy to understand and draw, but everything we learn from this example can be extended to any number of goods and services.

Suppose your disposable income is $100. Imagine that the price of a music download is $1, while the price of a chocolate bar is $5. Table 3.1 "Spending on Music Downloads and Chocolate Bars" shows some different bundles that you might purchase. Bundle number 1, in the first row, consists of one download and one chocolate bar. This costs you $6—certainly affordable with your $100 income. Bundle number 2 contains 30 downloads and 10 chocolate bars. For this bundle, your total spending on downloads is $30 (= 30 × $1), and your total spending on chocolate bars is $50 (= 10 × $5), so your overall spending is $80. Again, this bundle is affordable. You can imagine many other combinations that would cost less than $100 in total.

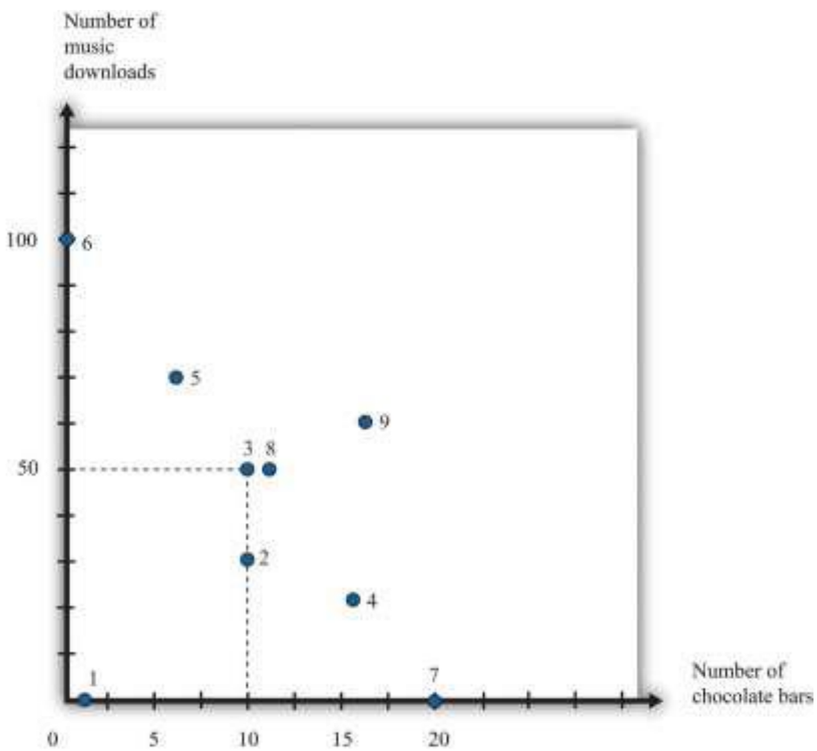
Table 3.1 Spending on Music Downloads and Chocolate Bars

Bundle	**Number of Downloads**	**Price per Download ($)**	**Spending on Downloads ($)**	**Number of Chocolate Bars**	**Price per Chocolate Bar ($)**	**Spending on Chocolate Bar ($)**	**Total Spending ($)**
1	1	1	1	1	5	5	6
2	30	1	30	10	5	50	80
3	50	1	50	10	5	50	100
4	20	1	20	16	5	80	100
5	65	1	65	7	5	35	100
6	100	1	100	0	5	0	100
7	0	1	0	20	5	100	100
8	50	1	50	11	5	55	105
9	70	1	70	16	5	80	150
10	5,000	1	5,000	2,000	5	10,000	15,000

Bundles 3, 4, 5, 6, and 7 are special because they are affordable if you spend *all* your income. For example, you could buy 50 downloads and 10 chocolate bars (bundle 3). You would spend $50 on music downloads and $50 on chocolate bars, so your total spending would be exactly $100. Bundle 4 consists of 20 downloads and 16 chocolate bars; bundle 5 is 65 downloads and 7 chocolate bars. Again, each bundle costs exactly $100. Bundle 6 shows that, if you chose to buy nothing but downloads, you could purchase 100 of them without exceeding your income, while bundle 7 shows that you could buy 20 chocolate bars if you chose to purchase no downloads. We could find many other combinations that—like those in bundles 3–7—cost exactly $100.

Bundles 8, 9, and 10 are *not* in the budget set. Bundle 8 is like bundle 3, except with an additional chocolate bar. Because bundle 3 cost $100, bundle 8 costs $105, but it is not affordable with your $100 income. Bundle 9 costs $150. Bundle 10 shows that you cannot afford to buy 5,000 downloads and 2,000 chocolate bars because this would cost $15,000. There is quite literally an infinite number of bundles that you cannot afford to buy.

Figure 3.2 Various Bundles of Chocolate Bars and Downloads

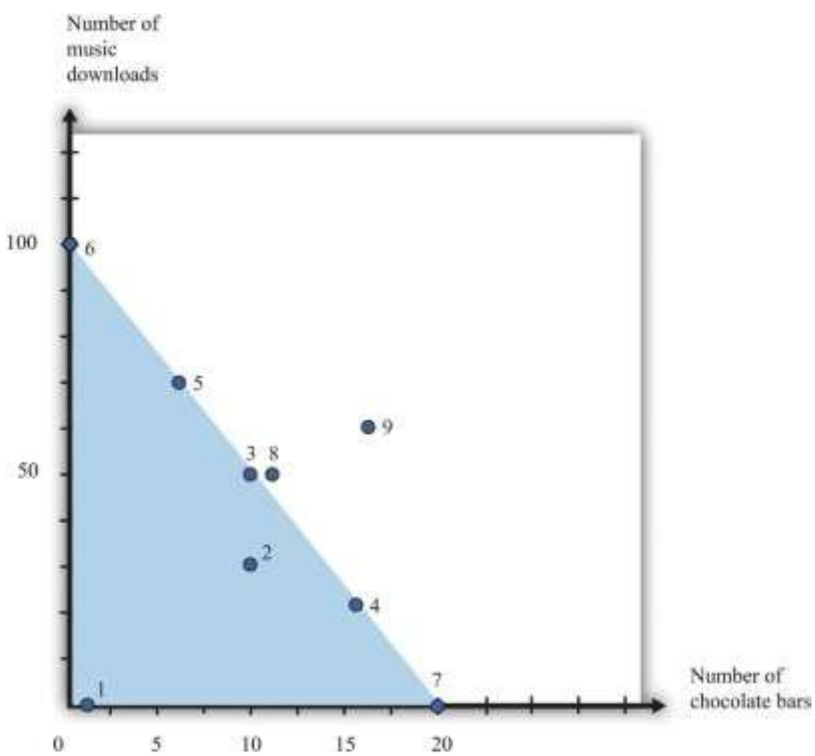


This figure shows the combinations of chocolate bars and music downloads from Table 3.1 "Spending on Music Downloads and Chocolate Bars".

Figure 3.2 "Various Bundles of Chocolate Bars and Downloads" illustrates the bundles from Table 3.1 "Spending on Music Downloads and Chocolate Bars". The vertical axis measures the number of music downloads, and the horizontal axis measures the number of chocolate bars. *Any point on the graph therefore represents a consumption bundle—a combination of music downloads and chocolate bars*. We show the first nine bundles from Table 3.1 "Spending on Music Downloads and Chocolate Bars" in this diagram. (Bundle 10 is several feet off the page.) If you inspect this figure carefully, you may be able to guess for yourself what the budget set looks like. Look in particular at bundles 3, 4, 5, 6, and 7. These are the bundles that are just affordable—that cost exactly $100. It appears as if these bundles all lie on a straight line, which is in fact the case. All the combinations of downloads and chocolate bars that are just affordable represent a straight line.

Meanwhile, the bundles that are affordable with income to spare—like bundles 1 and 2—are below the line, and the bundles you cannot afford—like bundles 8, 9, and 10—are above the line. Building on these discoveries, we find that the budget set is a triangle (Figure 3.3 "The Budget Set").

Figure 3.3 The Budget Set



*The bundles that are affordable are in the **budget set**, shown here as a triangle.*

Every point—that is, every combination of downloads and chocolate bars—that lies on or inside this triangle is affordable. Points outside the triangle are not affordable, so they are not in the budget set.

What Have We Assumed?

We now have a picture of the budget set. However, you might be curious about whether we have sneaked in any assumptions to do this. This is a *Principles of Economics* book, so we must start by focusing on the basics. We do our best throughout the book to be clear about the different assumptions we make, including their importance.

- We have assumed that there are only two products. Once we have more than two products, we cannot draw simple diagrams. Beyond this, though, there is nothing special about our downloads-and-chocolate-bar example. We are using an example with two products simply because it makes our key points more transparent. We can easily imagine a version of Table 3.1 "Spending on Music Downloads and Chocolate Bars" with many more goods and services, even if we cannot draw the corresponding diagram.
- We assume that you cannot consume negative quantities of downloads or chocolate bars. In our diagram, this means that the horizontal and vertical axes give us two sides of the triangle. This seems reasonable: it is not easy to imagine consuming a negative quantity of chocolate bars. (If you started out with some chocolate bars and then sold them, this is similar to negative consumption.)
- An easier way to look at this is to add any money you get from selling goods or services to your income. Then we can focus on buying decisions only.
- By shading in the entire triangle, we suppose that you can buy fractional quantities of these products. For example, the bundle consisting of 17.5 downloads and 12.7 chocolate bars is inside the triangle, even though iTunes, for example, would not allow you to purchase half a song, and you are unlikely to find a store that will sell you 0.7 chocolate bars. For the most part, this is a technical detail that makes very little difference, except that it makes our lives much easier.

- We have supposed that the price per unit of downloads and chocolate bars is the same no matter how few or how many you choose to buy. In the real world, you may sometimes be able to get quantity discounts. For example, a store might have a “buy two get one free” offer. In more advanced courses in microeconomics, you will learn that we can draw versions of Figure 3.3 "The Budget Set" that take into account such pricing schemes.
- We assume no saving or borrowing. It is easy to include saving or borrowing in this story, though. We think of borrowing as being an addition to your income, and we think of saving as one more kind of spending. Thus if you borrow, the budget set is described by

  total spending ≤ disposable income + borrowing.

  If you save, the budget set is described by

  total spending + spending ≤ disposable income.

The Budget Line

Continuing with our two-goods example, we know that

spending on chocolate = number of chocolate bars × price of a chocolate bar

and

spending on downloads = number of downloads × price of download.

When total spending is exactly equal to total disposable income, then

(number of chocolate bars × price of a chocolate bar) + (number of downloads × price of download) = disposable income.

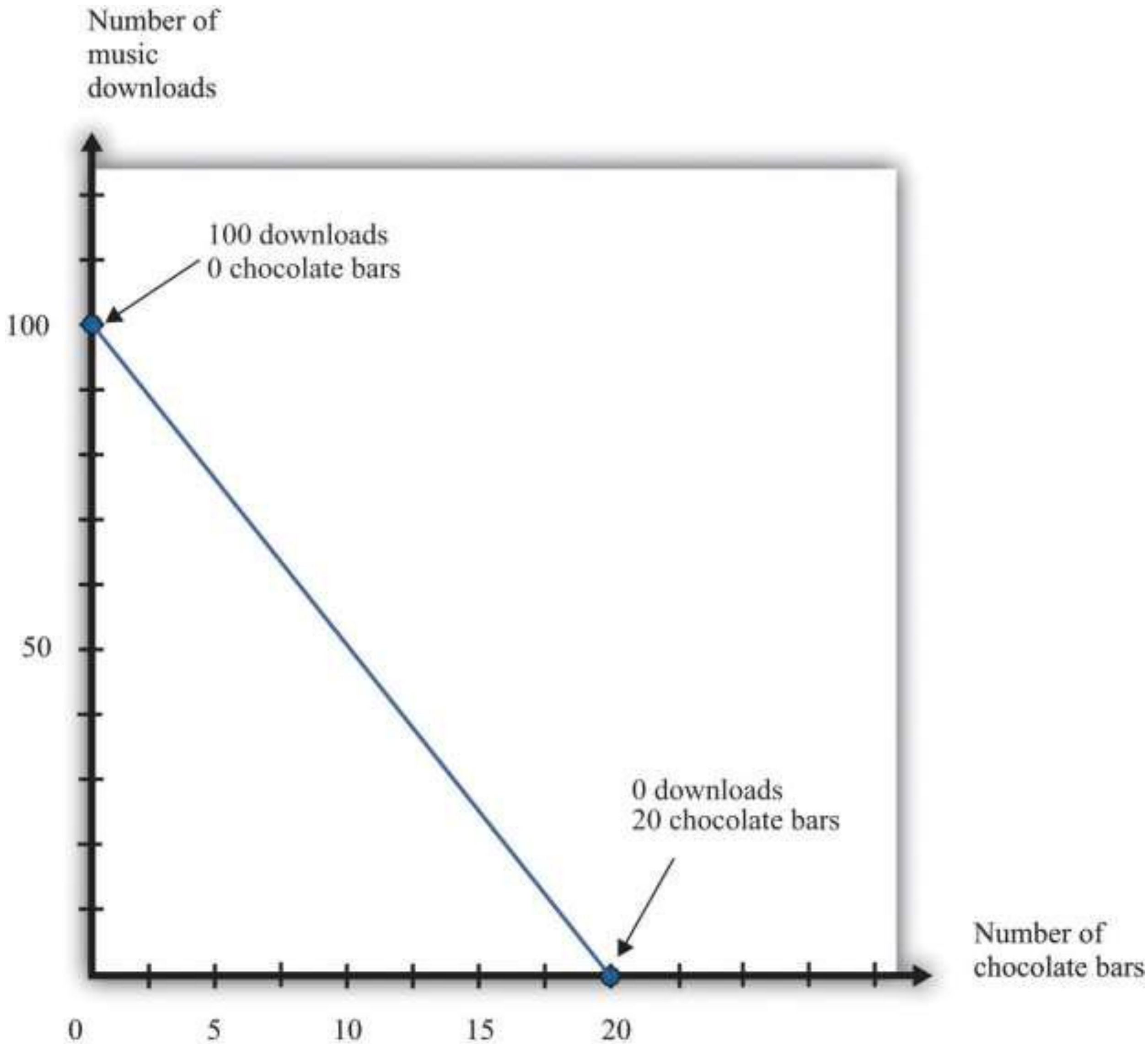
Toolkit: Section 17.1 "Individual Demand"

The budget line lists all the goods and services that are affordable, given prices and income, *assuming you spend all your income*.

The difference between the definitions of the budget set and the budget line is that there is an inequality in the budget set and an equality in the budget line:

total spending = disposable income.

Figure 3.4 The Budget Line



*The bundles that are exactly affordable are on the **budget line**.*

In the two-goods example, the budget line is the outside edge of the budget set triangle, as shown in Figure 3.4 "The Budget Line". What information do we need to draw the budget line? If we know both prices and the total amount of income, then this is certainly enough. In fact, we need only two pieces of information (not three) because basic mathematics tells us that it is enough to know two points on a line: once we have two points, we can draw a line. In practice, the easiest way to draw a budget line is to find the *intercepts*—the points on each axis. These correspond to how much you can obtain of each product if you consume 0 of the other. If you don't buy any chocolate bars, you have enough income to buy 100 downloads. If the number of chocolate bars is 0, then the budget line becomes

number of downloads × price of download = disposable income,

so

$$\text{number of downloads} = \frac{\text{disposable income}}{\text{price of download}} = \frac{100}{1} = 100.$$

Similarly, if the number of downloads is 0,

$$\text{number of chocolate bars} = \frac{\text{disposable income}}{\text{price of a chocolate bar}} = \frac{100}{5} = 20.$$

So we have two points on the budget line: (1) 0 chocolate bars and 100 downloads and (2) 0 downloads and 20 chocolate bars.

Another way to describe the budget line is to write the equation of the line in terms of its intercept (on the vertical axis) and its slope: [1]

$$\text{number of downloads} = \frac{\text{disposable income}}{\text{price of download}} - \frac{\text{price of a chocolate bar}}{\text{price of downloads}} \times \text{number of chocolate bars}.$$

The intercept is $\frac{\text{disposable income}}{\text{price of downloads}}$, which answers the following question: "How many downloads can you obtain if you buy no chocolate?" As we have already seen, this is 100 in our example.

The slope is

$$-\frac{\text{price of a chocolate bar}}{\text{price of downloads}},$$

which answers the following question: "What is the rate at which you can trade off downloads for chocolate bars?" In our example, this is −5. If you give up 1 chocolate bar, you will have an extra $5 (the price of a chocolate bar), which allows you to buy 5 more downloads.

The negative slope of the budget line says that to get more downloads, you must give up some chocolate bars. The *cost* of getting more downloads is that you no longer have the *opportunity* to buy as many chocolate bars. More generally, economists say that the opportunity cost of an action is what you must give up to carry out that action. Likewise, to get more chocolate bars, you must give up some downloads.

The opportunity cost of buying a chocolate bar is that you do not have that the money available to purchase downloads. The idea of opportunity cost pervades economics.

You may well have heard the following quotation that originated in economics: "There is no such thing as a free lunch." This statement captures the insight that everything has an opportunity cost, even if it is not always obvious who pays. Economists' habit of pointing out this unpleasant truth is one reason that economics is labeled "the dismal science." [2]

We said that a goal of this chapter is to help you make good decisions. One ingredient of good decision making is to understand the trade-offs that you face. Are you thinking of buying a new $200 mobile phone? The cost of that phone is best thought of, not as a sum of money, but as the other goods or services that you could have bought with that $200. Would you rather have 200 new songs for your existing phone instead? Or would you prefer 20 trips to the movies, 40 ice cream cones, or $200 worth of gas for your car? Framing decisions in this way can help you make better choices.

Your Preferences

Your choices reflect two factors. One is what you can afford. The budget set and the budget line are a way of describing the combinations of goods and services you can afford. The second factor is what you like, or—to use the usual economic term—your *preferences*.

Economists don't pretend to know what makes everyone happy. In our role as economists, we pass no judgment on individual tastes. Your music downloads might be Gustav Mahler, Arctic Monkeys, Eminem, or Barry Manilow. But we think it is reasonable to assume three things about the preferences that underlie your choices: (1) more is better, (2) you can choose, and (3) your choices are consistent.

More Is Better

Economists think that you are never satisfied. No matter how much you consume, you would always like to have more of something. Another way of saying this is that every good is indeed "good"; having more of something will never make you less happy. This assumption says nothing more than that people don't usually throw their income away. Even Bill Gates is not in the habit of burning money.

“More is better” permits us to focus on the budget line rather than the budget set. In Figure 3.3 "The Budget Set", you will not choose to consume at a point inside the triangle of the budget set. Instead, you want to be on the edge of the triangle—that is, on the budget line itself. Otherwise, you would be throwing money away. It also allows us to rank some of the different bundles in Table 3.1 "Spending on Music Downloads and Chocolate Bars". For example, we predict you would prefer to have bundle 3 rather than bundle 2 because it has the same number of chocolate bars and more downloads. Likewise, we predict you would prefer bundle 8 to bundle 3: bundle 8 has the same number of downloads as bundle 3 but more chocolate bars.

By the way, we are not insisting that you must *eat* all these chocolate bars. You are always allowed to give away or throw away anything you don’t want. Equally, the idea that more is better does not mean that you might not be sated with one particular good. It is possible that one more chocolate bar would make you no happier than before. Economists merely believe that there is always*something* that you would like to have more of.

“More is better” does *not* mean that you necessarily prefer a bundle that costs more. Look at bundles 7 and 9. Bundle 7 contains 0 downloads and 20 chocolate bars; it costs $100. Bundle 9, which contains 70 downloads and 16 chocolate bars, costs $150. Yet someone who loves chocolate bars and has no interest in music would prefer bundle 7, even though its market value is less.

You Can Choose

Economists suppose that you can always make the comparison between any two bundles of goods and services. If you are presented with two bundles—call them A and B—then the assumption that “you can choose” says that one of the following is true:

- You prefer A to B.
- You prefer B to A.
- You are equally happy with either A or B.

Look back at Table 3.1 "Spending on Music Downloads and Chocolate Bars". The assumption that “you can choose” says that if you were presented with any pair of bundles, you would be able to indicate which one you liked better (or that you liked them both equally much). This assumption says that you are never paralyzed by indecision.

"More is better" allows us to draw some conclusions about the choices you would make. If we gave you a choice between bundle 3 and bundle 8, for example, we know you will choose bundle 8. But what if, say, we presented you with bundle 4 and bundle 5? Bundle 5 has more downloads, but bundle 4 has more chocolate bars. "You can choose" says that, even though we may not know*which* bundle you would choose, you are capable of making up your mind.

Your Choices Are Consistent

Finally, economists suppose that your preferences lead you to behave consistently. Based on Table 3.1 "Spending on Music Downloads and Chocolate Bars", suppose you reported the following preferences across combinations of downloads and chocolate bars:

- You prefer bundle 3 to 4.
- You prefer bundle 4 to 5.
- You prefer bundle 5 to 3.

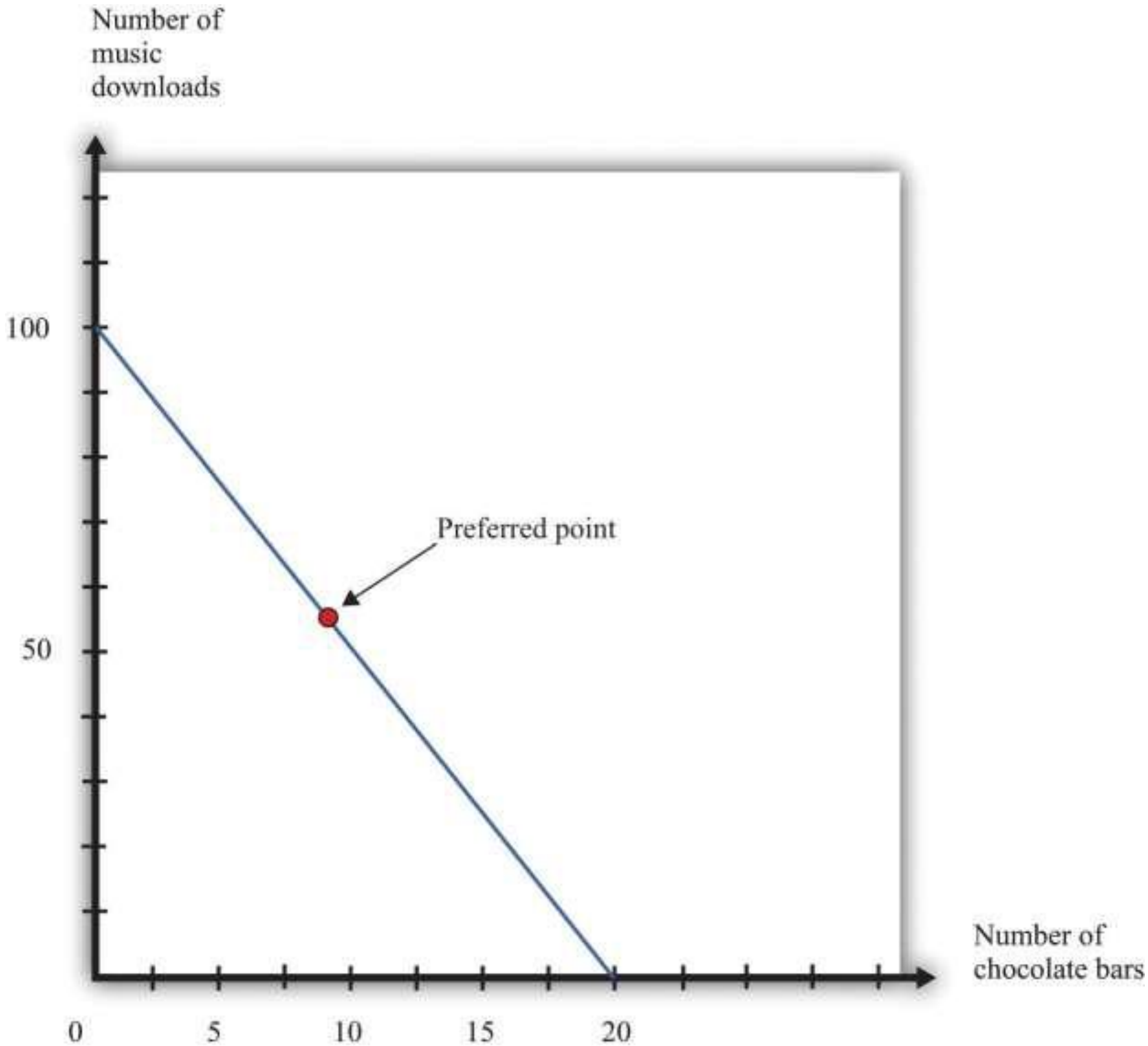
Each choice, taken individually, might make sense, but all three taken together are not consistent. They are contradictory. If you prefer bundle 3 to bundle 4 and you prefer bundle 4 to bundle 5, then a common-sense interpretation of the word "prefer" means that you should prefer bundle 3 to bundle 5.

Consistency means that your preferences must not be contradictory in this way. Put another way, if your preferences are consistent and yet you made these three choices, then at least one of these choices must have been a mistake—a bad decision. You would have been happier had you made a different choice.

Your Choice

We have now looked at your opportunities, as summarized by the budget set, and also your preferences. By combining opportunities and preferences, we obtain the economic approach to individual decision making. Economists make a straightforward assumption: they suppose you look at the bundles of goods and services you can afford and choose the one that makes you happiest. If the claims we made about your preferences are true, then you will be able to find a "best" bundle of goods and services, and this bundle will lie on the budget line. We know this because (1) you can compare any two points and (2) your preferences will not lead you to go around in circles.

Figure 3.5 Choosing a Preferred Point on the Budget Line



An individual's preferred point reflects opportunities, as given by the budget line, and preferences. The preferred point will lie on the budget line, not inside, because of the assumption that more is better.

In Figure 3.5 "Choosing a Preferred Point on the Budget Line", we indicate an example of an individual's preferred point. The preferred point is on the budget line and—by definition—is the best combination for the individual that can be found in the budget set. At the preferred point, the individual cannot be better off by consuming any other *affordable* bundle of goods and services.

There is one technical detail that we should add. It is possible that an individual might have more than one preferred point. There could be two or more combinations on the budget line that make an individual equally happy. To keep life simple, economists usually suppose that there is only a single preferred point, but nothing important hinges on this.

Rationality

Economists typically assume rationality of decision makers, which means that people can do the following:

- evaluate the opportunities that they face
- choose among those opportunities in a way that serves their own best interests

Is this a good assumption? Are people really as rational as economists like to think they are? We would like to know if people's preferences do satisfy the assumptions that we have made and if people behave in a consistent way. If we could hook someone up to a machine and measure his or her preferences, then we could evaluate our assumptions directly. Despite advances in neurobiology, our scientific understanding has not reached that point. We see what people do, not the preferences that lie behind these choices. Therefore, one way to evaluate the economic approach is to look at the choices people make and see if they are consistent with our assumptions.

Imagine you have an individual's data on download and chocolate bar consumption over many months. Also, suppose you know the prices of downloads and chocolate bars each month and the individual's monthly income. This would give you enough information to construct the individual's budget sets each month and look for behavior that is inconsistent with our assumptions. Such inconsistency could take different forms.

- She might buy a bundle of goods inside the budget line and throw away the remaining income.

- In one month, she might have chosen a bundle of goods—call it bundle A—in preference to another affordable bundle—call it bundle B. Yet, in another month, that same individual might have chosen bundle B when she could also have afforded bundle A.

The first option is inconsistent with our idea that "more is better." As for the second option, it is generally inconsistent to prefer bundle A over bundle B at one time yet prefer bundle B over bundle A at a different time. (It is not *necessarily* inconsistent, however. The individual might be indifferent between bundles A and B, so she doesn't care which bundle she consumes. Or her preferences might change from one month to the next.)

Inconsistent Choices

Economists are not the only social scientists who study how we make choices. Psychologists also study decision making, although their focus is different because they pay more attention to the processes that lie behind our choices. The decision-making process that we have described, in which you evaluate each possible option available to you, can be cognitively taxing. Psychologists and economists have argued that we therefore often use simpler rules of thumb when we make decisions. These rules of thumb work well most of the time, but sometimes they lead to biases and inconsistent choices. This book is about economics, not psychology, so we will not discuss these ideas in too much detail. Nevertheless, it is worth knowing something about how our decision making might go awry.

On occasion, we make choices that are apparently inconsistent. Here are some examples.

The endowment effect. Imagine you win a prize in a contest and have two scenarios to consider:

1. The prize is a ticket to a major sporting event taking place in your town. After looking on eBay, you discover that equivalent tickets are being bought and sold for $500.
2. The prize is $500 cash.

Rational decision makers would treat these two situations as essentially identical: if you get the ticket, you can sell it on eBay for $500; if you get $500 cash, you can buy a ticket on eBay. Yet many people behave differently in the two situations. If they get the ticket, they do not sell it, but if they get the cash, they do not buy the ticket. Apparently, we often feel differently about goods that we actually have in our possession compared to goods that we could choose to purchase.

Mental states. We may be in a different mental state when we buy a good from when we consume it. If you are hungry when you go grocery shopping, then you may buy too much food. When we buy something, we have to predict how we will be feeling when we consume it, and we are not always very good at making these predictions. Thus our purchases may be different, depending on our state of mind, even if prices and incomes are the same.

Anchoring. Very often, when you go to a store, you will see that goods are advertised as "on sale" or "reduced from" some price. Our theory suggests that people simply look at current prices and their current income when deciding what to buy, in which case they shouldn't care if the good used to sell at a higher price. In reality, the "regular price" serves as an anchor for our judgments. A higher price tends to increase our assessment of how much the good is worth to us. Thus we may make inconsistent choices because we sometimes use different anchors.

How should we interpret the evidence that people are—sometimes at least—not quite as rational as economics usually supposes? Should we give up and go home? Not at all. Such findings deepen our understanding of economic behavior, but there are many reasons why it is vital to understand the behavior of rational individuals.

1. Economics helps us make better decisions. The movie *Heist* has dialogue that sums up this idea:

D. A. Freccia:	You're a pretty smart fella.
Joe Moore:	Ah, not that smart.
D. A. Freccia:	[If] you're not that smart, how'd you figure it out?
Joe Moore:	I tried to imagine a fella smarter than myself. Then I tried to think, "what would he do?"

2. Most of us are "not that smart"; that is, we are not smart enough to determine what the rational thing to do is in all circumstances. Knowing what someone smarter would do can be very useful indeed. [3] Further, if we understand the biases and mistakes to which we are all prone, then we can do a better job of recognizing them in ourselves and adjusting our behavior accordingly.

3. Rationality imposes a great deal of discipline on our thinking as economists. If we suppose that people are irrational, then anything is possible. A better approach is to start with rational behavior and then see if the biases that psychologists and economists have identified are likely to alter our conclusions in a major way.
4. Economics has a good track record of prediction in many settings. A lot of the time, even if not all the time, the idea that people behave rationally seems more right than wrong.

More Complicated Preferences

People may be rational yet have more complicated preferences than we have considered.

Fairness. People sometimes care about fairness and so may refuse to buy something because the price seems unfair to them. In one famous example, people were asked to imagine that they are on the beach and that a friend offers to buy a cold drink on their behalf. [4] They are asked how much they are willing to pay for this drink. The answer to this question should not depend on where the drink is purchased. After all, they are handing over some money and getting a cold drink in return. Yet people are prepared to pay more if they know that the friend is going to buy the drink from a hotel bar rather than a local corner store. They think it is reasonable for hotels to have high prices, but if the corner store charged the same price as the hotel, people think that this is unfair and are unwilling to pay.

Altruism. People sometimes care not only about what they themselves consume but also about the well-being of others. Such altruism leads people to give gifts, to give to charity, to buy products such as "fair-trade" coffee, and so on.

Relative incomes. Caring about the consumption of others can take more negative forms as well. People sometimes care about whether they are richer or poorer than other people. They may want to own a car or a barbecue grill that is bigger and better than that of their neighbors.

More complicated preferences such as these are not irrational, but they require a more complex framework for decision making than we can tackle in a*Principles of Economics* book. [5]

KEY TAKEAWAYS

- The budget set consists of all combinations of goods and services that are affordable, and the budget line consists of all combinations of goods and services that are affordable *if you spend all your income*.
- The opportunity cost of an action (such as consuming more of one good) is what must be given up to carry out that action (consuming less of some other good).
- Your choices reflect the interaction between what you can afford (your budget set) and what you like (your preferences).
- Economists think that most people prefer having more to having less, are able to choose among the combinations in their budget set, and make consistent choices.
- Rational agents are able to evaluate their options and make choices that maximize their happiness.

CHECKING YOUR UNDERSTANDING

1. Suppose that all prices and income were converted into a different currency. For example, imagine that prices were originally in dollars but were then converted to Mexican pesos. Would the budget set change? If so, explain how. If not, explain why not.
2. Assume your disposable income is $100, the price of a music download is $2, and the price of a chocolate bar is $5. Redo Table 3.1 "Spending on Music Downloads and Chocolate Bars". Find (or create) three combinations of chocolate bars and downloads that are on the budget line. Find a combination that is not affordable and another combination that is in the budget set but not on the budget line.
3. What is the difference between your budget set and your budget line?

[1] To derive this equation, go back to the budget line and divide both sides by the price of a download:

$$\text{number of chocolate bars} \times \text{price of a chocolate bar}/\text{price of a download} + \text{number of downloads} = \text{disposable income}/\text{price of a download}.$$

Rearranging, we get the equation in the text.

[2] Although economists may dislike this characterization of their profession, they can take pride in its origin. The term was coined by Thomas Carlyle about 150 years ago, in the context of a debate about race and slavery. Carlyle criticized famous economists of the time, such as John Stuart Mill and Adam Smith, who argued that some nations were richer than others not because of innate differences across races but because of economic and historical factors. These economists argued for the equality of people and supported the freedom of slaves.

[3] The quote comes from the Internet Movie Database (http://www.imdb.com). We first learned of the scene from B. Nalebuff and I. Ayres, *Why Not?* (Boston: Harvard Business School Press, 2003), 46.

[4] See Richard Thaler, “Mental Accounting and Consumer Choice, *Marketing Science* 4 (1985): 199–214.

[5] We say more about some of these ideas in Chapter 12 "Superstars".

3.2 Individual Demand

LEARNING OBJECTIVES

1. What is a demand curve, and what is the law of demand?
2. What is the decision rule for choosing how much to buy of two different goods?
3. What is the decision rule for choosing to buy a single unit of a good?

Now that we have a framework for thinking about your choices, we can now explain one of the most fundamental economic ideas: demand. Here we focus on the demand of a single individual. [1] We use two different ways of thinking about your demand for a good or a service. One approach builds on the idea of the budget set. The other focuses on how much you would be willing to pay for a good or service. In combination, they give us a detailed understanding of how economic decisions are made.

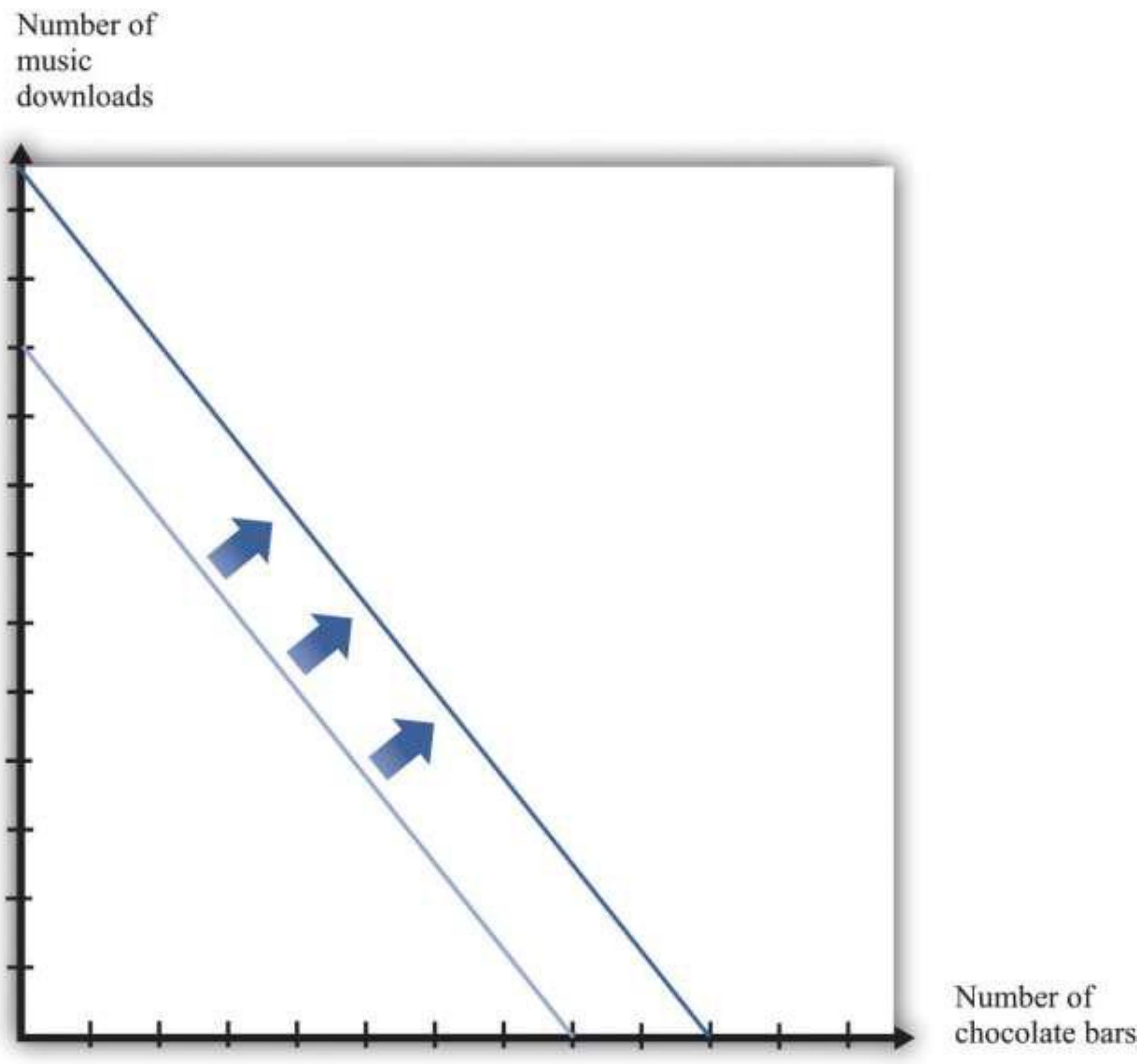
Individual Demand for a Good

As you visit stores at different times, you undoubtedly notice that the prices of goods and services change. At the same time, your income may also change from one month to the next. So if we were to look at your budget set monthly, we would typically find it changing from one month to the next. We would then expect that you would choose different combinations of goods and services from one month to the next.

To keep things simple, suppose we are still in a world of two goods—downloads and chocolate bars—and that you do no saving. We will describe your demand for chocolate bars. (If you like, you can think of downloads as representing all the other goods and services you consume.) Given prices and your income, you pick the best point on the budget line. Look again at the "preferred" point inFigure 3.5 "Choosing a Preferred Point on the Budget Line". One way to interpret this point is that it tells us how many chocolate bars you will buy, given your income and given the price of a chocolate bar and other goods. Using this as a reference point, we now ask how your choice will change as income changes and then as the price of a chocolate bar changes.

Changes in Income

Figure 3.6 An Increase in Income

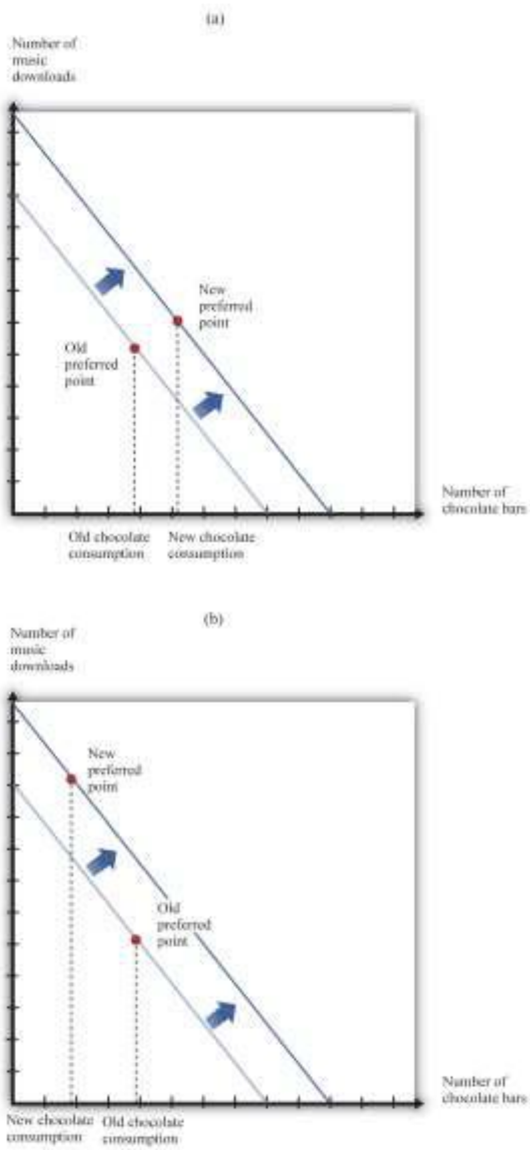


An increase in income shifts the budget line outward.

Imagine that your income increases. Figure 3.6 "An Increase in Income" shows what happens to the budget line. Higher income means that you can afford to buy more chocolate bars and more downloads, so the budget line shifts outward. The slope of the budget line is unchanged because there is no change in the price of a chocolate bar relative to downloads.

What happens to your consumption of chocolate bars? There are two possibilities (Figure 3.7 "The Consequences of an Increase in Income"): the increase in income leads you to consume either more chocolate bars or fewer chocolate bars. Both are plausible, and either is possible. (Of course, you might also choose exactly the same amount as before.) We might think that the more normal case is that higher income would lead to higher chocolate bar consumption. If a good has a property that you will consume more of it when you have higher income, we call it a normal good.

Figure 3.7 *The Consequences of an Increase in Income*



In response to an increase in income, two things are possible: the consumption of chocolate bars may increase (a) or decrease (b).

Under some circumstances, higher income leads to *lower* consumption. For example, suppose you are surviving in college on a diet consisting of mostly instant noodles. When you graduate college and have higher income, you can afford better things to eat, so you will probably consume a smaller quantity of instant noodles. Economists call such products inferior goods. If a particular product exists in several qualities (cheap versus expensive cuts of meat, for example), we often find that the low-quality version is an inferior good. A good might be normal for one consumer and inferior for another. More precisely, therefore, we say that a *good* is inferior if, on average, higher income leads to lower consumption.

Economists make a further distinction among different kinds of normal goods. If you spend a *larger fraction of your income* on a particular good as your income increases, then we say that the good is a luxury good. Another way of saying this is that for a luxury good, the percentage increase in consumption is bigger than the percentage increase in income.

We can also define these ideas in terms of the income elasticity of demand, which is a measure of how sensitive demand is to changes in income. For an inferior good, the income elasticity of demand is negative: higher income leads to lower consumption. For a normal good, the income elasticity of demand is positive. And for a luxury good, the income elasticity of demand is greater than one.

Toolkit: Section 17.2 "Elasticity"

For more discussion of normal, inferior, and luxury goods, see the toolkit.

The distinctions among these different kinds of goods are crucial for managers of firms. To predict sales, managers need to know whether the products they are selling are normal, inferior, or luxury goods. Firms that sell inferior goods tend to do well when the economy as a whole is doing poorly and vice versa. By contrast, firms selling luxury goods will do particularly poorly when the economy as a whole is performing poorly.

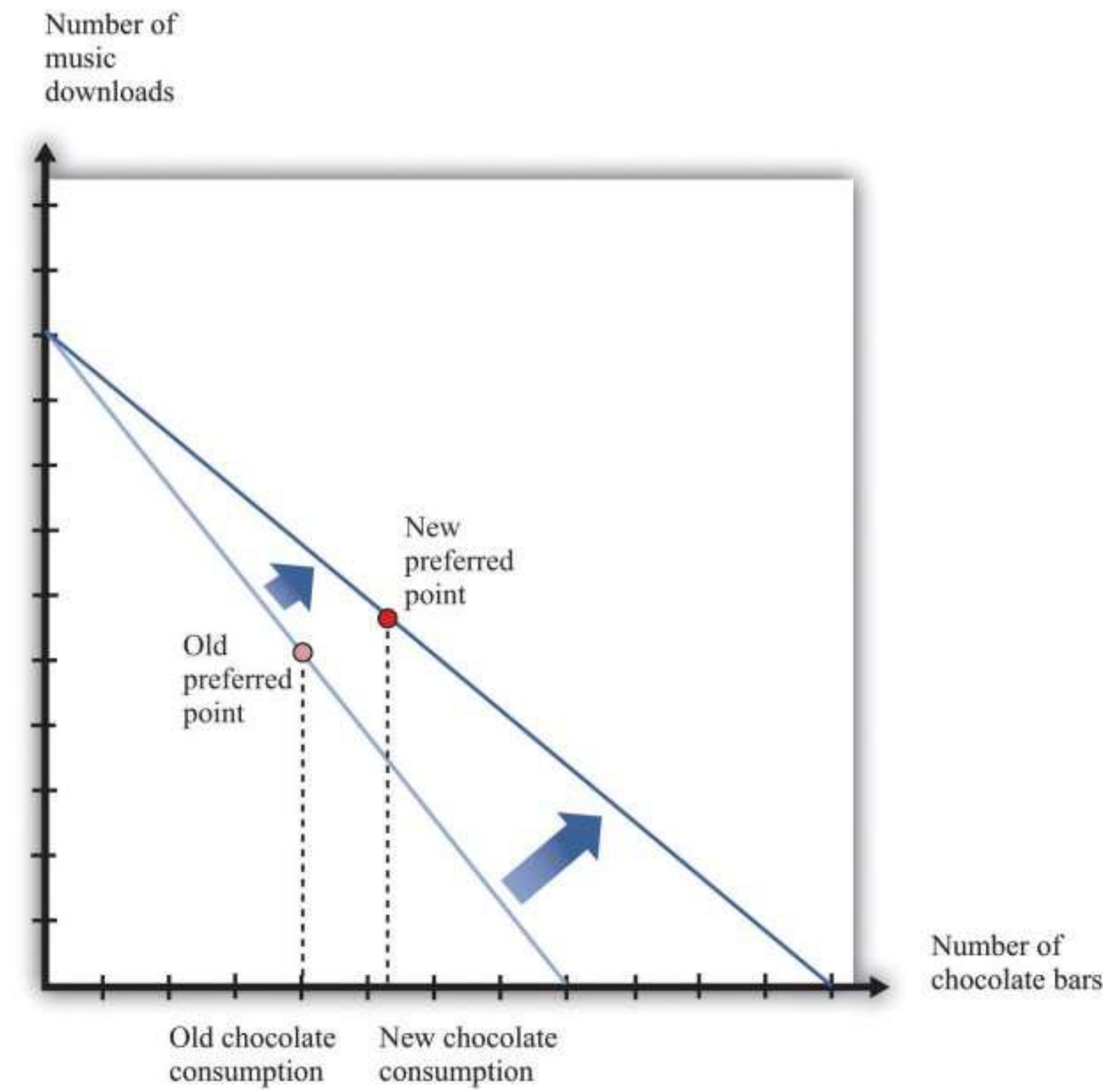
Changes in Price

Now we look at what happens if there is a change in the price of a chocolate bar. Suppose the price decreases. First, let us analyze what this means in terms of our picture. Remember that the intercepts of the budget line tell you how much you can have of one good if you consume none of the other. If you consume no chocolate bars, then a decrease in the price of a chocolate bar has no effect on your consumption. You can consume exactly the same number of downloads as before. The intercept on the vertical axis therefore does not move. However, a decrease in the price of a chocolate bar means that, if you consume only chocolate bars, then you can have more than before. The intercept on the horizontal axis moves outward. One way to see this is to remember that this intercept is given by

total incomeprice of a chocolate bar.

Figure 3.8 "A Decrease in the Price of a Chocolate Bar" illustrates a decrease in the price of a chocolate bar. The new budget line lies outside the old budget line. Any bundle you could have bought at the old prices is still affordable now, and you can also get more. A decrease in the price of a chocolate bar, in other words, makes you better off. In addition, the slope of the budget line changes: it is flatter than before. The slope of the budget line reflects the way in which the market allows you to trade chocolate bars for other products. If you choose to consume fewer downloads, the reduction in the price of a chocolate bar means that you get relatively more chocolate bars in exchange.

Figure 3.8 A Decrease in the Price of a Chocolate Bar



A decrease in the price of a chocolate bar causes the budget line to rotate. The result is an increase in the quantity of chocolate bars consumed.

Figure 3.8 "A Decrease in the Price of a Chocolate Bar" also shows a new consumption point. In response to the decrease in the price of a chocolate bar, we see an increase in the consumption of chocolate bars. The idea that people almost always consume more of a good when its price decreases is one of the fundamental ideas of economics. Indeed, it is sometimes called the law of demand. It is certainly intuitive that lower prices lead people to consume more. There are two reasons why we expect to see this result.

First, if a good (for example, chocolate bars) decreases in price, it becomes cheaper *relative to other goods*. Its opportunity cost—that is, the amount of other goods you must give up to get a chocolate bar—has decreased. The lower opportunity cost means that there is a substitution effect away from other goods and toward chocolate bars. Second, a decrease in the price of a chocolate bar also means that you can afford more of *everything*, including chocolate bars. Provided chocolate bars are a normal good, this income effect will also lead you to want to consume more chocolate bars. If chocolate bars are inferior goods, the income effect leads you to want to consume fewer chocolate bars.

In Figure 3.8 "A Decrease in the Price of a Chocolate Bar", the substitution effect is reflected by the fact that the budget line changes slope. The flatter budget line tells us that the opportunity cost of chocolate bars in terms of downloads has decreased. The income effect shows up in the fact that the new budget set includes the old budget set. You can consume your previous bundle of goods and still have some income left over to buy more.

Based on the idea of the law of demand, we can construct your individual demand curve for chocolate bars. We do so by drawing the budget set for each different price of a chocolate bar, seeing how much you buy, and then plotting this data. For example, we might find that your purchases of chocolate bars look like those in Table 3.2 "Demand for Chocolate Bars". If we plot these points on a graph and then “fill in the gaps,” we get a diagram like Figure 3.9 "The Demand Curve". This is your demand curve for chocolate bars. It tells how many chocolate bars you would purchase at any given price. The law of demand means that we expect this curve to slope downward. If the price increases, you consume less. If the price decreases, you consume more.

Table 3.2 Demand for Chocolate Bars

Price per Bar ($)	Quantity of Chocolate Bars Bought
1	12
2	6
3	4
4	3
5	2.4
6	2

Figure 3.9 The Demand Curve

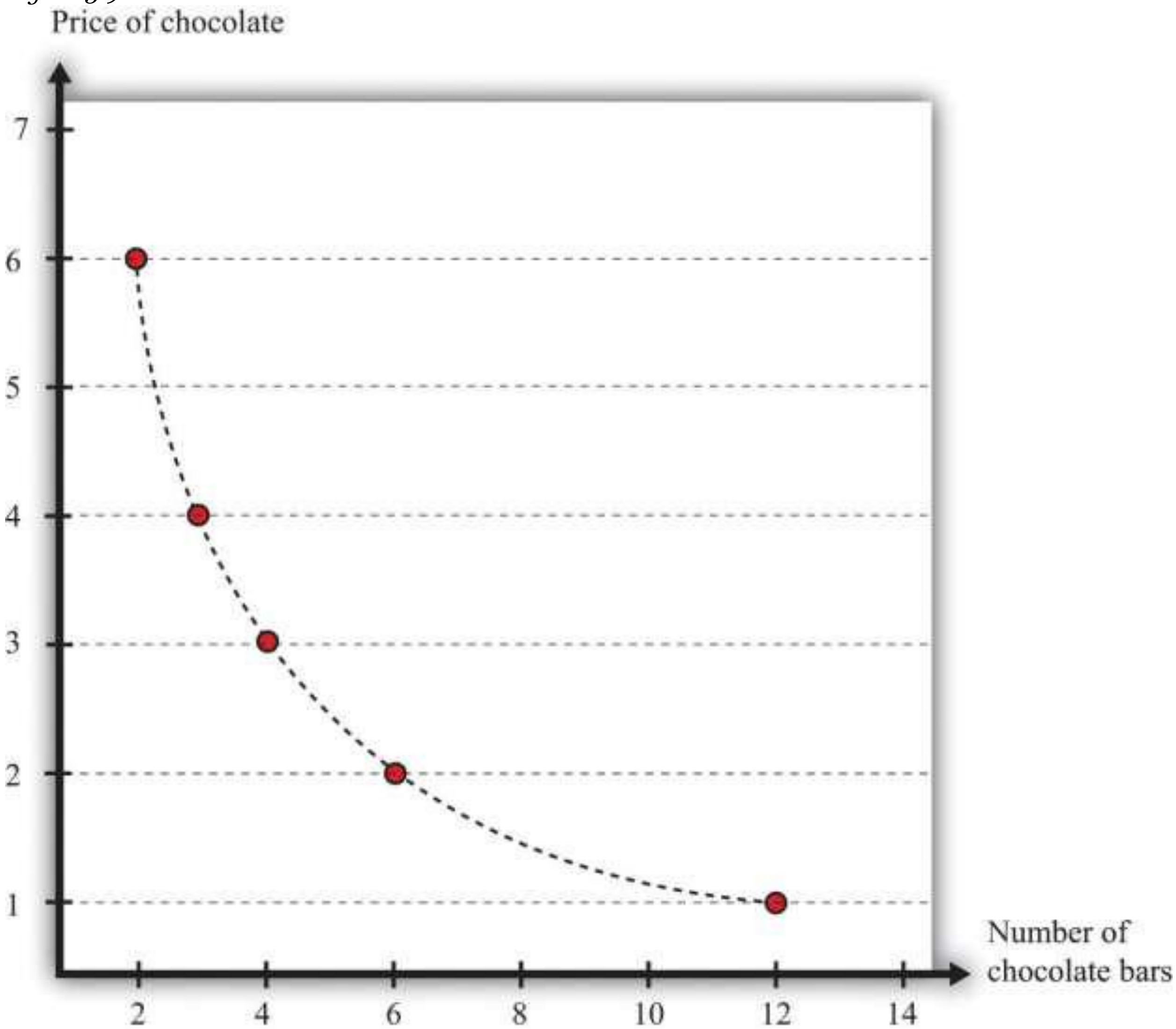


Table 3.2 "Demand for Chocolate Bars" contains an example of some observations on demand. At different prices from $1 to $6, we see the number of chocolate bars purchased. If we fill in the gaps, we obtain a demand curve.

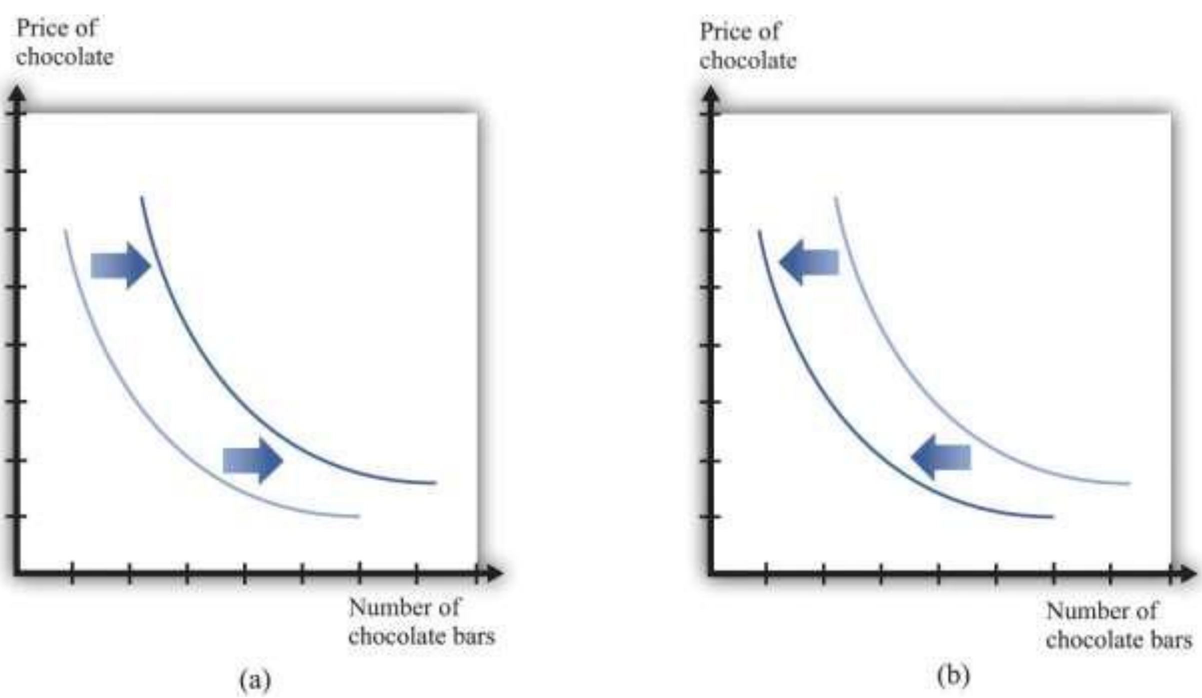
Toolkit: Section 17.1 "Individual Demand"

The *individual demand curve* is drawn on a diagram with the price of a good on the vertical axis and the quantity demanded on the horizontal axis. It is drawn for a given level of income.

We must be careful to distinguish between movements along the demand curve and shifts in the demand curve. Suppose there is a change in the price of a chocolate bar. Then, as we explained earlier, the budget line shifts, and the quantity demanded of both chocolate bars and downloads will change. This appears in Figure 3.9 "The Demand Curve" as a *movement along the demand curve*. For example, if the price of a chocolate bar decreases from $4 to $3, then, as in Table 3.2 "Demand for Chocolate Bars", the quantity demanded increases from three bars to four bars. The demand curve does not change; we simply move from one point on the line to another.

If a change in anything other than the price of a chocolate bar causes you to change your consumption of chocolate bars, then there is a *shift in the demand curve*. For example, suppose you get a pay raise at your job, so you have more income. As long as chocolate bars are a normal good, this increase in income will cause your demand curve for chocolate bars to shift outward. This means that, at any given price, you will buy more of the good when income increases. In the case of an inferior good, an increase in income will cause the demand curve to shift inward. You will buy less of the good when income increases. We illustrate these two cases in Figure 3.10 "Shifts in the Demand Curve: Normal and Inferior Goods".

Figure 3.10 Shifts in the Demand Curve: Normal and Inferior Goods



(a) If income increases and chocolate bars are a normal good, then the individual demand curve will shift to the right. At every price, a greater quantity of chocolate bars is demanded. (b) If

income increases and chocolate bars are inferior goods, then the individual demand curve will shift to the left. In the event of a decrease in income, the two cases are reversed.

Exceptions to the Law of Demand

The law of demand is highly intuitive and is supported by lots of research for all sorts of different goods and services. We can take it as a reliable fact that in almost all circumstances, the demand curve will indeed slope downward. Yet we might still wonder if there are any exceptions, any cases in which the demand curve slopes upward. There are indeed a few such exceptions.

- **Giffen goods.** We explained that the law of demand comes from both a substitution effect and—for normal goods—an income effect. But for inferior goods, the income effect acts in the opposite direction to the substitution effect: a decrease in price makes people better off, which is an incentive to consume less of the good. Theoretically, it is possible for this income effect to be stronger than the substitution effect. Although conceivable, Giffen goods are extremely rare; indeed, economists are unsure if they actually exist outside of textbooks. For the income effect to overwhelm the substitution effect, the good in question must form a very large part of the overall consumption bundle. This might arise in extremely poor economies where people spend a large part of their income on a staple food, as was found in a recent experiment conducted in rural China. [2] The researchers gave families subsidies to buy rice, making it cheaper, and found that consumption of rice did indeed decrease.
- **Status goods.** Some luxury products are purchased mainly for their status appeal—for example, Rolls-Royce automobiles, Louis Vuitton handbags, and Gucci shoes. The high prices of these goods contribute to their exclusivity. Price is an *attribute* of a good, so a higher price can make a good seem more, not less, attractive. Again, it is theoretically possible that this could lead the demand curve to slope upward, at least for some range of prices. Although high prices increase the appeal of status goods, it is rare for this effect to be strong enough to outweigh the more basic income and substitution effects.
- **Judging quality by price.** Implicitly, we have supposed that people are well informed about the products they buy. In many cases, though, we must purchase goods and services with only imperfect knowledge of their quality. In this situation, we may use price as an indicator of the

quality of a good. Of course, marketers are well aware that we do this and will often try to use price as a signal of the quality of their brand or products. The upshot is that people may be more willing to buy at a higher price.

In these three situations, it is conceivable that we might observe a higher price being associated with a higher quantity of the good being purchased. You should not overestimate the significance of these cases, however, because (1) most products do not fall into any of these categories, and (2) the substitution effect is still in operation for all goods, as is the income effect for all normal goods.

Changes in Other Prices

So far, we have said that the number of chocolate bars you want to buy is affected by income and the price of a chocolate bar. Changes in the prices of other goods also have an impact. In general, an increase in a price of another good could cause the demand for chocolate bars to increase or decrease.

Goods are substitutes if an increase in the price of one good leads to increased consumption of the other good. CDs and music downloads are one example: if the price of CDs increases, you will obtain more music through downloads.

Goods are complements if an increase in the price of one good leads to decreased consumption of the other good. For example, DVDs and DVD players are complements. If the price of DVD players decreases, more people will buy DVD players. As a result, more people will want to buy DVDs.

Toolkit: Section 17.2 "Elasticity"

To learn more about substitutes and complements, see the toolkit for formal definitions. (These definitions are presented in terms of the *cross-elasticity of demand*, which is a measure of how responsive the quantity demanded is to changes in the price of another good.)

The Valuation Approach to Demand

There is another way of thinking about demand. Instead of focusing attention on the budget set and the budget line, we can think more directly about your preferences. Imagine you are asked the following question in an interview:

What is the maximum amount you would be willing to pay for one chocolate bar?

In answering this question, you should not worry about what would be a reasonable or fair price for a chocolate bar or even about the price at which this chocolate bar might actually be available. You should simply decide how much you want chocolate bars. For example, you might think that you would really like to have at least one chocolate bar, so you would be willing to pay up to $12 for one bar—that is, you would be happier with one more chocolate bar and up to $12 less in income. This is your valuation of one chocolate bar.

Now we ask you some more questions.

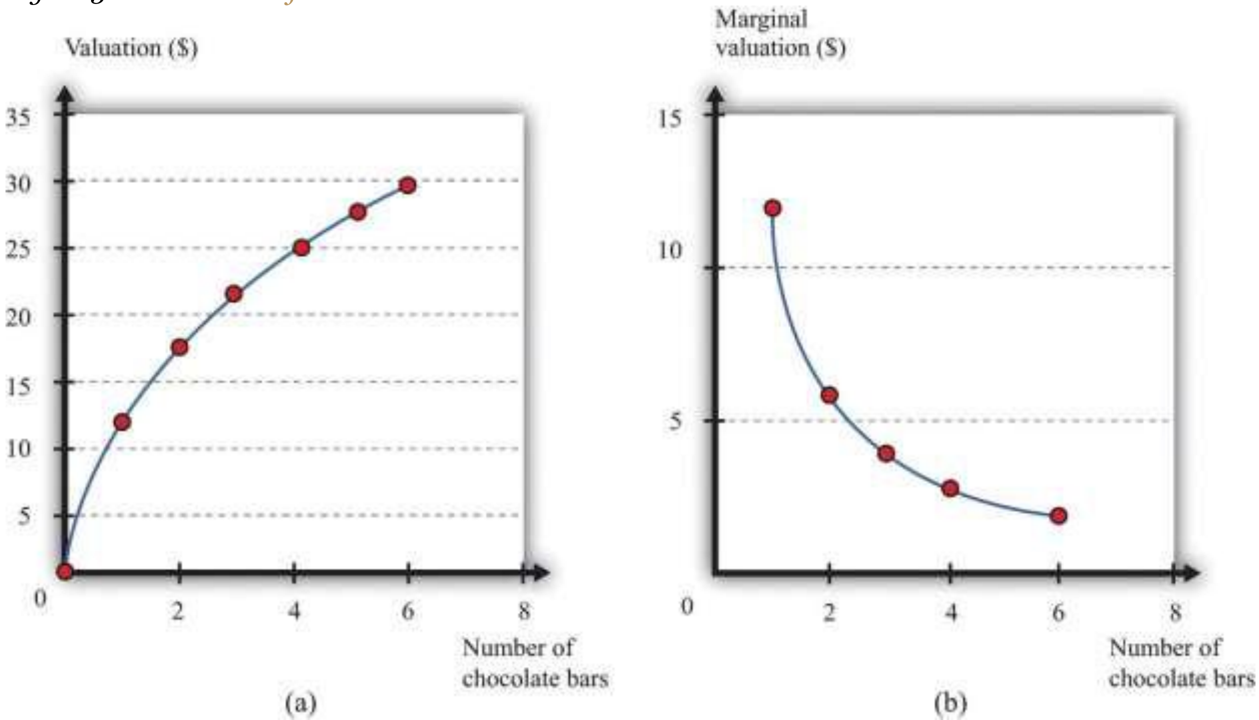
What is the maximum amount that you would be willing to pay for two chocolate bars? Three chocolate bars?…

Perhaps you decide you would be willing to pay $18 for two bars, $22 for three bars, and so on. If we kept asking such questions, we might get the first two columns of Table 3.3 "Valuation and Marginal Valuation". We can also plot this valuation (part [a] of Figure 3.11 "Valuation of a Good"). Your valuation is increasing; you always like having more chocolate bars because "more is better."

Table 3.3 Valuation and Marginal Valuation

Quantity of Chocolate Bars Bought	**Valuation ($)**	**Marginal Valuation ($)**
0	0.00	
1	12.00	12.00
2	18.00	6.00
3	22.00	4.00
4	25.00	3.00
5	27.40	2.40
6	29.40	2.00

Figure 3.11 Valuation of a Good



Your ***valuation*** *of a good is the maximum amount that you would be willing to pay, purely on the basis of your desire for the good.*

Because you are willing to pay $12 for one bar and $18 for two bars, we know you would be willing to pay an *additional* $6 for the second chocolate bar. Similarly, if you have two chocolate bars, you would be willing to pay an additional $4 for a third bar. We call the *change* in your valuation yourmarginal valuation (see the third column of Table 3.3 "Valuation and Marginal Valuation", which is graphed in part [b] of Figure 3.11 "Valuation of a Good"). Notice that marginal valuation decreases as the quantity of chocolate bars increases. The change in your valuation gets smaller as you obtain more chocolate bars. We can see the same thing in part (a) of Figure 3.11 "Valuation of a Good" from the fact that the valuation curve gets flatter as the quantity of chocolate bars increases.

Toolkit: Section 17.1 "Individual Demand"

An individual's valuation of some quantity of a product is the maximum amount the individual would be willing to pay to obtain that quantity. An individual's marginal valuation of some good is the maximum amount the individual would be willing to pay to obtain one *extra* unit of that product.

It usually seems reasonable to think that marginal valuations will indeed decrease in this way. If you don't have any chocolate bars, then the first bar is worth a lot to you—$12 in our example. But if you already have five chocolate bars, then the sixth bar is worth only $2 to you. As you obtain more and more of any given product, each additional unit is less and less valuable. For most people, we expect that most products will exhibit such *diminishing marginal valuation*.

Part (b) of Figure 3.11 "Valuation of a Good" may seem familiar. It is the demand curve for chocolate bars you saw previously in Figure 3.9 "The Demand Curve". This is not an accident or a coincidence. There is a simple decision rule to tell you how much you should buy: when your marginal valuation is greater than the price, you should buy more of the good, stopping only when the marginal valuation of the good has dropped to the level of the price. For example, suppose that chocolate bars are selling for $3.99. You should definitely buy the first chocolate bar, because it is worth $10 to you and will cost you only $3.99. You should buy the second bar as well because it is worth an additional $8 to you; likewise you should buy the third and fourth bars. You don't buy the fifth bar because it is worth only $3 to you, which is less than what it costs. Thus your decision rule is as follows:

- Buy until the marginal valuation of a good equals the price of the good.

Because the demand curve, by definition, tells you how much you buy at a given price, it is the same as the marginal valuation curve.

Combining the Two Approaches to Demand

We have presented two different ways of thinking about consumer decisions, but the underlying choice is the same. To see how the two approaches are linked, rewrite the decision rule for chocolate bars as buy until

$$\frac{\text{marginal valuation of chocolate bars}}{\text{price of a chocolate bar}}=1.$$

You have a similar decision rule for downloads: buy until

$$\frac{\text{marginal valuation of downloads}}{\text{price of a download}}=1.$$

Combining these two equations, we see that

$$\frac{\text{marginal valuation of chocolate bars}}{\text{price of a chocolate bar}}=\frac{\text{marginal valuation of downloads}}{\text{price of a download}},$$

which we can rearrange as

$$\frac{\text{marginal valuation of chocolate bars}}{\text{marginal valuation of downloads}}=\frac{\text{price of a chocolate bar}}{\text{price of a download}}.$$

The ratio on the right-hand side of this expression should look familiar. Earlier, we found that the slope of the budget line is

$$-\frac{\text{price of a chocolate bar}}{\text{price of a download}}.$$

So the marginal valuation of a chocolate bar divided by the marginal valuation of downloads equals minus the slope of the budget line.

What does this mean? The budget line tells us the rate at which the market allows you to trade off chocolate bars for downloads. If you consume one fewer chocolate bar, the number of dollars you will get is equal to the price of a chocolate bar. These dollars will buy you $\frac{\text{price of a chocolate bar}}{\text{price of a download}}$ downloads. The ratio of marginal valuations describes how you view the trade-off between chocolate bars and downloads. If you are making good decisions about how to spend your money, then the rate at which you are happy to trade off chocolate bars for downloads equals the rate at which the market allows you to make such trades. If this were not true, then you could make yourself happier by choosing a different bundle on the budget line.

Making Decisions at the Margin

To make good decisions, you need to understand the trade-offs you are making. To put it another way, you need to recognize that every purchase has an opportunity cost, which is summarized by the budget line. If you want more chocolate bars, you must consume less of something else. You also need to find the right point on the budget line—the point that makes you happiest. Most of the time, economists simply assume that you are able to make this decision correctly on the basis of the three assumptions about your preferences that we introduced earlier.

You can also use this theory to help you think about the decisions you make. Suppose you are facing the budget line we discussed earlier and plan to buy 8 chocolate bars and 60 downloads (as in Figure 3.5 "Choosing a Preferred Point on the Budget Line"). In principle, you need to compare that bundle with every other combination on the budget line. In practice, it is enough—most of the time at least—to compare it with nearby bundles. For example, if you prefer this bundle to 7 chocolate bars and 65 downloads, and you also prefer this bundle to 9 chocolate bars and 55 downloads, then you can be reasonably confident that you have found the best bundle. If a small change won't make you happier, then neither will a large change.

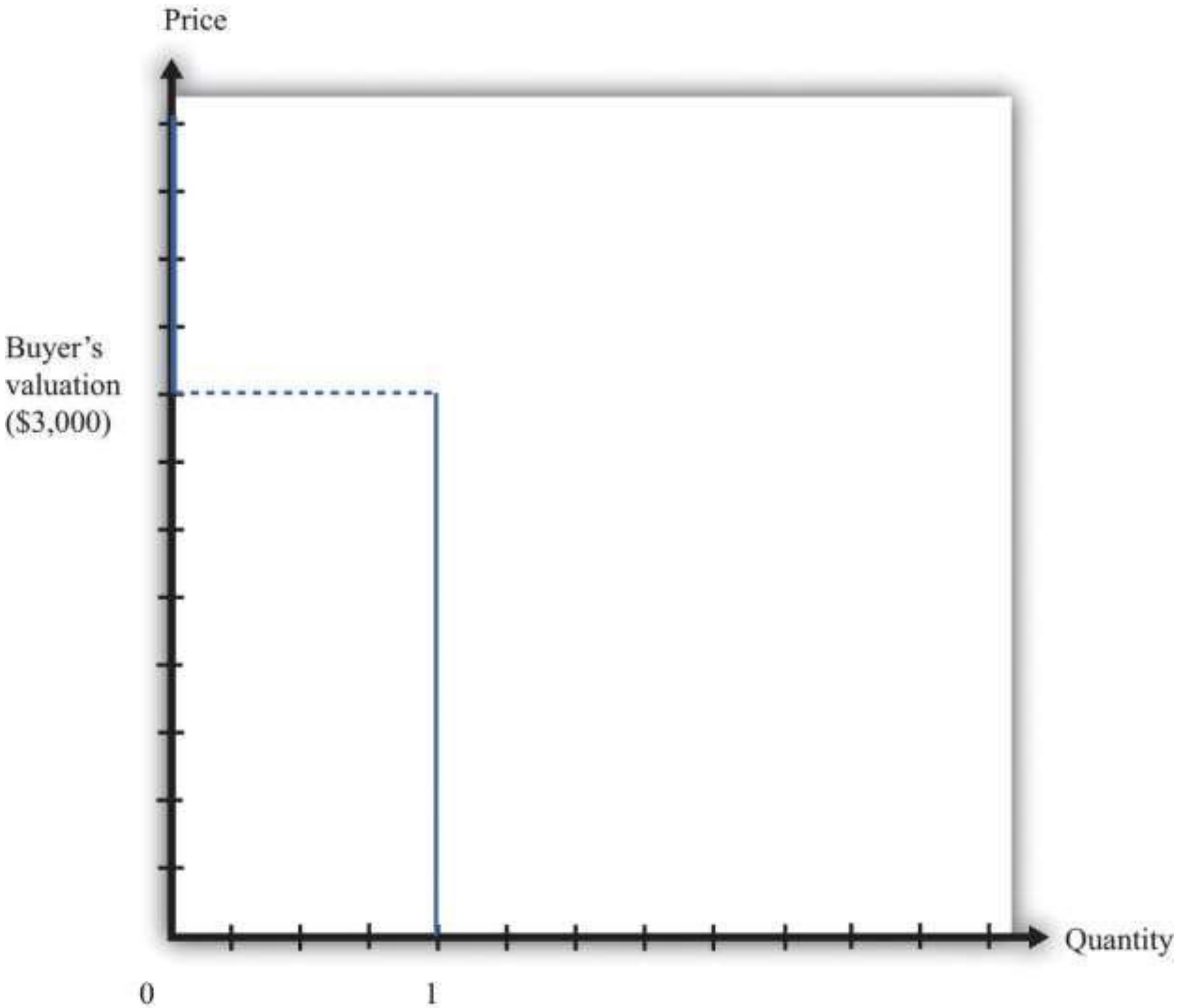
Unit Demand

So far we have considered situations where you might buy multiple units of a good—for example, 20 music downloads or 5 chocolate bars. To keep things simple, we also supposed that you could buy fractional amounts, such as 20.7 downloads, or 3.25 chocolate bars. This assumption gives a decision rule for purchase: buy until marginal valuation equals price.

Some purchase decisions are better thought of as "buy or don't buy." Large, infrequent purchases fall into this category. Think, for example, about the decision to buy a new car, a new microwave, or an expensive vacation. You won't buy five microwaves because they are cheap. The decision rule for purchase is even easier in this case: buy if your valuation of the good exceeds the price of the good. In fact, this is really no different from our earlier decision rule. Because you are only ever thinking about buying one unit, your valuation and your marginal valuation are the same thing (look back at the first two rows of Table 3.3 "Valuation and Marginal Valuation"). And because in this case it does *not* make sense to suppose you can buy fractional amounts of the good, you cannot keep buying until your marginal valuation decreases all the way to the price.

If a buyer is interested in purchasing one and only one unit of a good, theunit demand curve tells us the price at which he is willing to buy. Below his valuation, he buys the good. Above his valuation, he does not buy the good. A unit demand curve is shown in Figure 3.12 "Unit Demand". In this example, the buyer's valuation is $3,000.

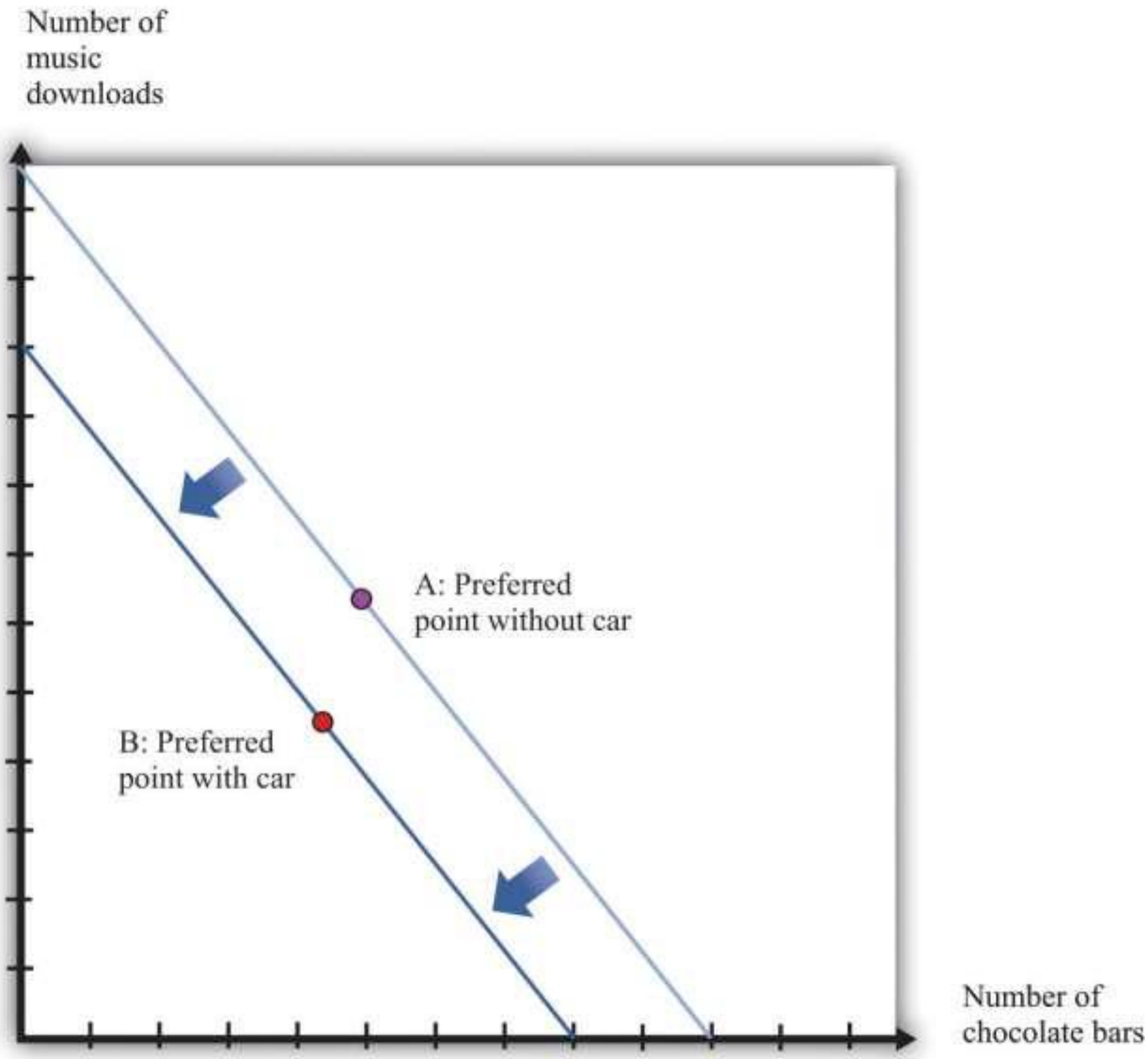
Figure 3.12 Unit Demand



The buyer follows this decision rule: "Buy if the price is less than valuation." If the price is greater than $3,000, the buyer will not purchase. The quantity demanded is zero. If the price is less than $3,000, the buyer will purchase one unit. No matter how low the price decreases, the buyer will not want more than a single unit. This is an example of unit demand.

To see where such a valuation could come from, look at Figure 3.13 "The Valuation of a Car". Suppose you are thinking about buying a car. The figure shows our standard downloads-and-chocolate-bar diagram, except that there are two budget lines. The outer budget line applies in the case where you do not buy the car. You have a preferred point in terms of downloads and chocolate bars (A). If you buy the car, you have less income to spend on everything else. The effect is to shift your budget line inward, in which case you have a new preferred point (B). The thought experiment here is to decrease your income and shift your budget line inward until you are equally happy with the two bundles. The change in your income—the amount by which the budget line must shift—is your valuation of the car. Your valuation, in other words, is the opportunity cost of the car: if you buy the car, you can only consume bundle B rather than bundle A.

Figure 3.13 The Valuation of a Car



If you don't buy a car, your preferred mix of chocolate bars and downloads is at point A. If you buy a car, then you no longer have that income available to spend on chocolate bars and downloads. Your budget line shifts inward, and you consume at your preferred point B. Now imagine that you *are equally happy at point A and point B. Then the difference in income is equal to your valuation of the car. Thus the valuation of a car is its opportunity cost in terms of other goods.*

Budget Studies

Economists' theories are all well and good. But we do not actually get to see people's preferences or marginal valuations. We observe what people actually do. To see our theory in action, we can look at *household budget studies*. These are surveys where government statisticians interview households and ask them how they spend their income. For example, Table 3.4 "Budget Shares in the United States" contains data on US consumer expenditures for the years 2005, 2007, and 2009.

You can see that, on average, households spend a little more than 45 percent of their income on food and housing. Insurance is also a large category, with about 11 percent of income being spent on it. [3] Interestingly, the budget shares do not change much over the three years despite the differences in income and spending. From this we see that, although individual goods may be inferior or luxury goods, such differences are largely offset when we look at broad categories of goods or services.

The table also contains data for households under age 25. [4] We can compare the spending patterns of this group against all households. Not surprisingly, the younger group earns less than the average household. Also, this younger group often spends more than it earns, indicating that younger people are borrowing, on average. The younger group spends more on alcohol, transportation, and education and much less on health care and insurance than the average household. This makes sense given the health status of young individuals as well as their demand for education.

Table 3.5 "United Kingdom Budget Study" is a UK budget study for households headed by young people (under the age of 30) in 2009. It shows how these households allocated their expenditures over a week. [5] We can compare these figures against those for young people in the United States. (We need to be careful in making comparisons because the categories for spending are not exactly the same across surveys. Still, it is useful to explore these differences.) In the United Kingdom, spending on food and housing is much lower for these younger households than in the United States. Health also has a much lower expenditure share.

KEY TAKEAWAYS

- The demand curve of an individual shows the quantity of a good or service demanded at different prices, given income and other prices.
- The law of demand—which holds for almost all goods and services—states that the demand curve slopes downward: as the price of a good decreases, the quantity demanded of that good will increase.
- When you are making an optimal choice between two goods, the rate at which you want to trade off the two goods—at the margin—should equal the rate at which the market allows you to trade off the two goods.
- You should buy one more unit of a good whenever your marginal valuation of the good is greater than the price.
- When you are willing to buy at most one unit of a good (unit demand), your valuation and your marginal valuation are identical, so you should purchase the good as long as your valuation of that good is greater than the price.

CHECKING YOUR UNDERSTANDING

1. Think about your own preferences. Can you think of a good that—for you—is a substitute for a chocolate bar? Can you think of a good that is a complement?
2. Draw a version of Figure 3.6 "An Increase in Income" to show a decrease in income.
3. Create a version of Figure 3.7 "The Consequences of an Increase in Income"that shows music downloads as an inferior good. Why can't you draw a version of the figure where both music downloads and chocolate bars are inferior goods?

[1] In Chapter 5 "eBay and craigslist", we develop the idea of the market demand curve, which combines the demands of many individuals.

[2] See Robert Jensen and Nolan Miller, "Giffen Behavior: Theory and Evidence" (Harvard University, John F. Kennedy School of Government Working Paper RWP 07-030, July 2007).

[3] Chapter 4 "Life Decisions" discusses why we buy insurance.

[4] This is based on the age of the reference person in the household, who is the individual who owns or rents the property.

[5] The British source is as follows: Office for National Statistics, *Family Spending, 2010,* table A.11, accessed January 24, 2011, http://www.statistics.gov.uk/downloads/theme_social/family-spending-2009/familyspending2010.pdf.

3.3 Individual Decision Making: How You Spend Your Time

LEARNING OBJECTIVES

1. What is the time budget constraint of an individual?
2. What is the opportunity cost of spending your time on a particular activity?
3. What is the meaning of real wage?
4. What is the labor supply curve of an individual, and how does it depend on the real wage?

So far we have discussed how you choose to spend your money. There is another decision you make every day: how to spend your time. You have 24 hours each day in which to do all the different things you want to do: work, sleep, eat, study, watch television, surf the Internet, go to the movies, and so on. Time, like money, is scarce. Given that you have only 24 hours to allocate in every day, how do you decide which activities to spend your time on? This problem is very similar to the allocation of your budget, with one key difference: you cannot save or borrow time in the way that you can save or borrow money. There are exactly 24 hours in each day—no more, no less.

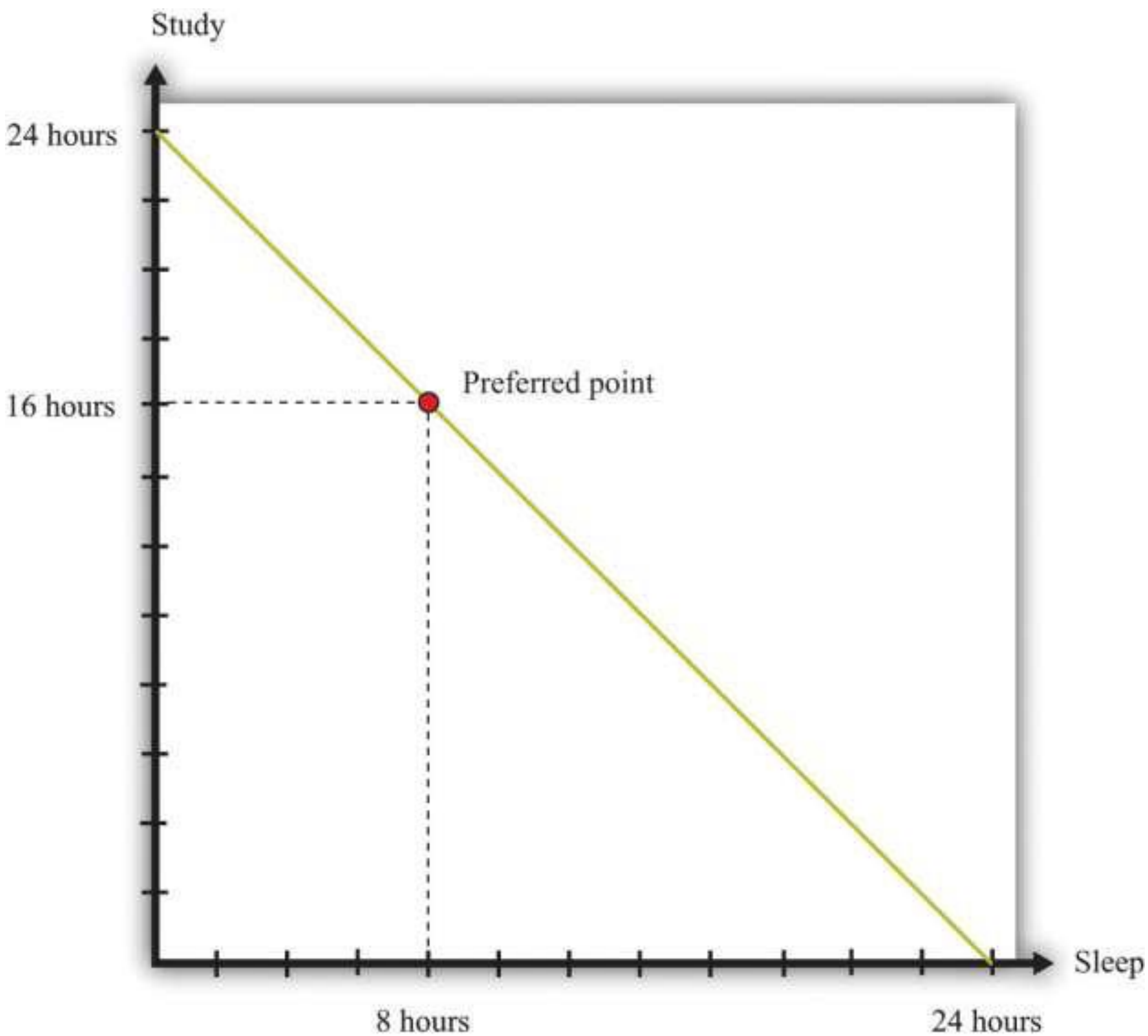
Choosing among Different Uses of Your Time

We begin with the most fundamental time allocation problem for all students: choosing between studying and sleeping. As before, we keep things simple by thinking about only two possible uses of your time. You are given 24 hours in the day to study and sleep. How should you allocate your time?

As with the allocation of your income, there are two aspects of this problem. First, there is a budget constraint—only now it is your time that is scarce, not money. Second, you have your own preferences about sleep and study time. Your ability to meet your desires is constrained by the scarcity of your time: you must trade off one activity for another.

The time budget constraint is the restriction that there are only 24 hours in the day. It is shown in Figure 3.14 "The Time Budget Constraint" and is the counterpart to the budget line in our earlier discussion. Any point in this figure represents a combination of sleep and study time. The sum of sleep and study time must equal 24 hours (remember we are supposing that these are the only ways you spend your time). Thus your allocation of your time must lie somewhere on this line.

Figure 3.14 The Time Budget Constraint



The time allocation line shows your options for dividing your time between study and sleep.

Figure 3.14 "The Time Budget Constraint" also shows one possible choice that you might make: allocating 8 hours to sleep and 16 hours to studying. The choice of this point reflects your desires for sleep and study. As with the spending decision, we pass no judgment, as economists, on the actual decision you make. We suppose you typically make the choice that makes you the happiest.

At your preferred point, your choice to sleep for 8 hours means that your study time *must* equal 16 hours; equivalently, your choice to study for 16 hours means that you must sleep for 8 hours. Any increase in one activity must be met by a reduction in time for the other. The opportunity cost of each hour of sleep is an hour of study time, and the opportunity cost of each hour of study time is an hour of sleep. If you choose this point, you reveal that you are willing to "pay" (that is, give up) 8 hours of study time to obtain 8 hours of sleep, and you are willing to pay 16 hours of sleep to obtain 16 hours of study time.

As with consumption choices, it is often enough to look at small changes to evaluate whether or not you are making a good decision. The opportunity cost of a little more sleep is a little less study time. If you are making a good decision about the allocation of your time, then the extra sleep is not worth the extra study time. Suppose you are contemplating a particular point on the time budget line, and you want to know if it is a good choice. If a very small movement away from your chosen point will not make you happier, then—in most circumstances—neither will a big movement. By a very small movement, we mean sleeping a little less and studying a little more *or* studying a little less and sleeping a little more. If you are making a good decision, then your willingness to substitute sleep for study time is exactly the same as that allowed by the time constraint.

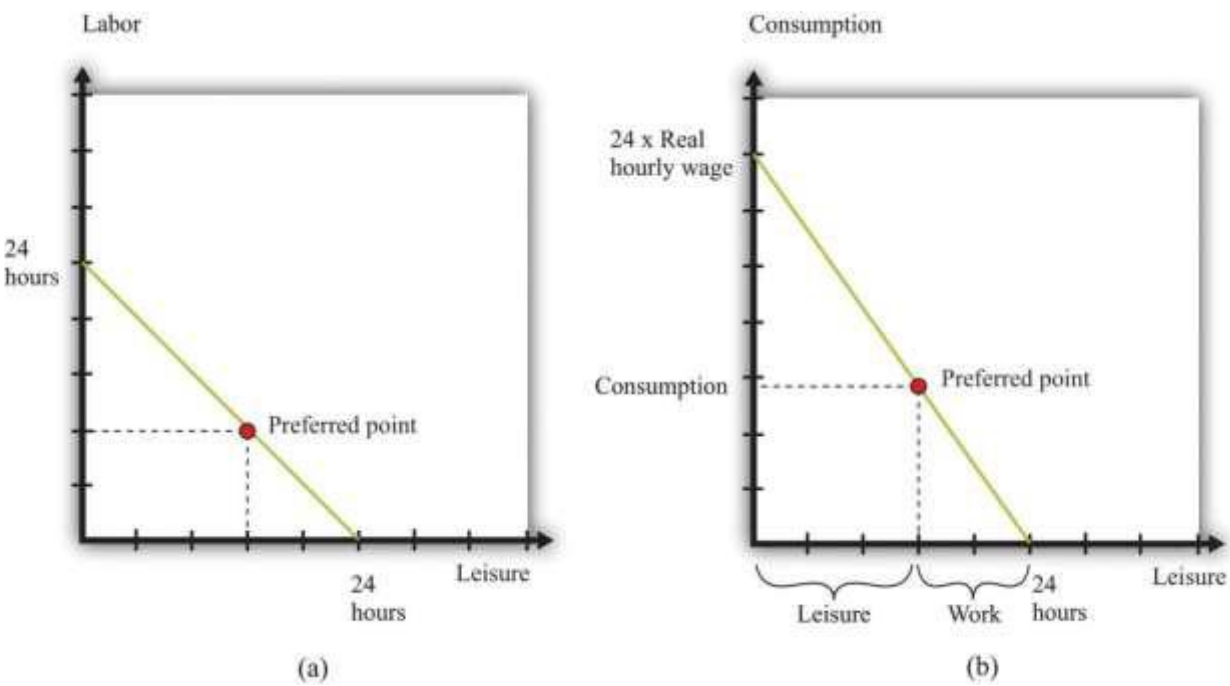
Individual Labor Supply

Sleeping and studying are uses of your time that are directly for your own benefit. Most people—perhaps including yourself—also spend time working for money. So let us now look at the choice between spending time working and enjoying leisure. Our goal is to determine how much labor you will choose to supply to the market—which is equivalently a choice about how much leisure time to enjoy, because choosing the number of working hours is the same as choosing the number of leisure hours. Your choice between the two is based on the trade-off between enjoying leisure and working to earn money that allows the purchase of goods.

Once again—to make it easy to draw diagrams—we suppose that these are the only uses of your time. Part (a) of Figure 3.15 "Choosing between Work and Leisure" presents the allocation of time between work and leisure. As with the sleep-study choice, there is a time budget constraint, and you have preferences between these two ways to allocate time. Your best choice satisfies the same property as before: you allocate time such that no other division of your time makes you happier.

What makes this different from the sleep-study choice is the valuation of your time. We can think of sleep as a good thing in that you generally prefer more to less. Likewise, we can think of study as a good thing in that—even if you don't always enjoy it—you perceive a gain to spending time studying. So Figure 3.14 "The Time Budget Constraint" is like our earlier diagrams with downloads and chocolate bars: it has a good thing on each axis. Now, people presumably prefer more leisure to less: leisure is a "good," like chocolate bars, blue jeans, or cans of soda. But we have drawn part of (a) Figure 3.15 "Choosing between Work and Leisure" as if work is also a good thing. Most people, however, see work time as a "bad" rather than as a "good." Even people who like their work would almost always prefer to work a little less and have a little more leisure time.

Figure 3.15 Choosing between Work and Leisure



(a) The time budget line shows your options for dividing your time between labor and leisure. However, we generally think of labor as a "bad" rather than as a "good." (b) Now the choice is between consumption and leisure. For each hour of your time, you earn the nominal hourly wage. If you divide this by the price level, you get the real wage. The real wage tells you how many goods and services you can enjoy for one hour of work.

The gain from working, of course, is that you earn income, allowing you to purchase goods and services. Each extra hour of your work allows you to buy more goods and services. Conversely, if you want more leisure time, you must give up some goods and services. Thus the choice between labor and leisure is linked to the choice about how many chocolate bars and other goods you buy. The income we take as given in describing your budget set typically comes from your decision to supply labor time. (Of course, you may have other sources of income as well, such as loans or grants.)

Part (b) of Figure 3.15 "Choosing between Work and Leisure" takes the labor-leisure choice and converts it into a choice between leisure and consumption. Here, consumption refers to all the goods and services you consume. We lump together all the products you consume, just as we lump together all your different forms of leisure (sleep, study, watching television, and so on). As before, time is measured on the horizontal axis: there are 24 hours to the day, which must be split between leisure time and labor time. On the vertical axis, we measure consumption.

To get the budget constraint for this picture, we begin with the time budget constraint:

leisure hours + labor hours = 24.

The value of an hour of time in dollars is given by the wages at which you can sell your time. Multiplying the time budget constraint by the wage gives us a budget constraint in dollars:

(leisure hours × wage) + wage income = 24 × wage.

Wage income is equal to the number of hours worked times the hourly wage. Because wage income is used to buy goods, we can replace it by total spending on consumption, which is the price level times the quantity of consumption goods purchased:

(leisure hours × wage) + (price level × consumption) = 24 × wage.

This is the budget constraint faced by an individual choosing between leisure and consumption. Think of it as follows: The individual first sells all her labor at the going wage, yielding the income on the right-hand side. With this income, she then "buys" back leisure and also buys consumption goods. The price of an hour of leisure represents the wage rate, and the price of a unit of consumption goods represents the price level.

Toolkit: Section 17.3 "The Labor Market"

The real wage is the *relative price* of labor in terms of consumption goods:

real wage =nominal wageprice level.

Dividing the time budget constraint by the price level, we get the budget in line in part (b) of Figure 3.15 "Choosing between Work and Leisure".

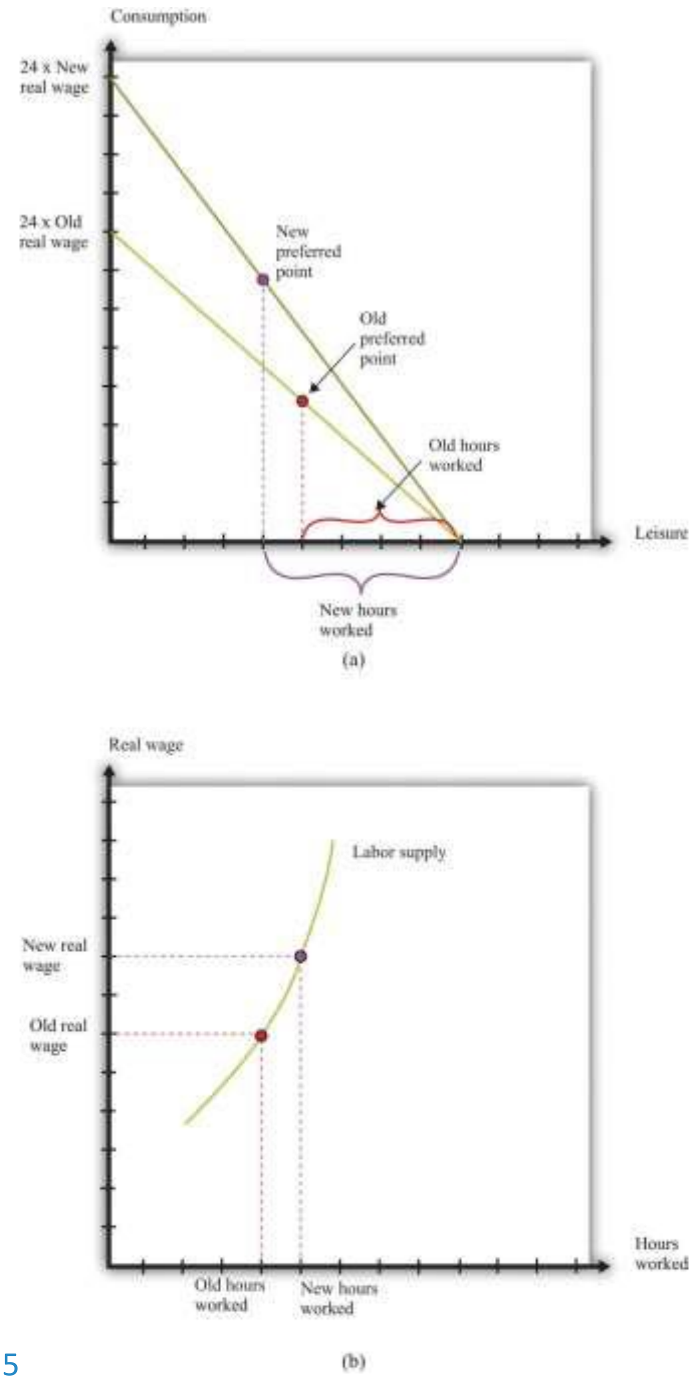
leisure hours × real wage + consumption = 24 × real wage.

As you move along the budget line, you trade hours of leisure for consumption goods. The slope of the budget line is the negative of the real wage. If you give up an hour of leisure, you obtain extra consumption equal to the real wage. Put differently, the opportunity cost of an hour of leisure is the amount of consumption you give up by not working. Once we have worked out how much leisure you consume, we have equivalently worked out how much labor you supply:

labor hours = 24 − leisure hours.

Part (a) of Figure 3.16 "The Effect of a Real Wage Increase on the Quantity of Labor Supplied" shows what happens when the real wage changes. When the real wage increases, the vertical intercept of the budget line is higher because the vertical intercept tells us how much consumption an individual could obtain if she worked for all 24 hours in the day. The horizontal intercept does not change as the real wage changes: if an individual does not work, then the level of consumption is zero regardless of wages. It follows that the budget line is steeper as the real wage increases. If an individual gives up an hour of leisure time, he or she gets more additional consumption when the real wage is higher. The opportunity cost of leisure in terms of forgone consumption is higher.

Figure 3.16 *The Effect of a Real Wage Increase on the Quantity of Labor Supplied*



[5]

(a) An increase in the real wage causes the budget line to rotate. Income and substitution effects are both at work: the income effect encourages more leisure (less work), while the substitution effect encourages more work. The substitution effect generally dominates, so higher real wages lead to more work. This means that the labor supply curve slopes upward, as shown in (b).

Part (b) of Figure 3.16 "The Effect of a Real Wage Increase on the Quantity of Labor Supplied" shows the individual labor supply curve that emerges from the labor-leisure choice.

Toolkit: Section 17.3 "The Labor Market"

The individual labor supply curve shows the number of hours that an individual chooses to work at each value of the real wage.

In fact, there are conflicting incentives at work here. As the real wage increases, the opportunity cost of leisure is higher, so you are tempted to work more. But at a higher real wage, you can enjoy the same amount of consumption with fewer hours of work, so this tempts you to work less. This is another example of substitution and income effects. The substitution effect says that when something gets more expensive, we buy less of it. When the real wage increases, leisure is more expensive. The income effect says that as the real wage increases, you can buy more of the things you like, including leisure. We know from our study of demand that, for normal goods, the income and substitution effects act in the same direction. In the case of supply, however, income and substitution effects point in *different* directions. Consistent with this, most economic studies find that, though the labor supply curve slopes upward, hours worked are not very responsive to changes in the real wage.

Some jobs do not give you any control over the number of hours that you must work. Labor supply then becomes a "unit supply" decision, analogous to the unit demand decisions we considered previously. Should you take a job at all and, if so, what job? To the extent that you can choose among different jobs that offer different hours of work, your decision about whether or not to work will still reflect a trade-off between leisure and consumption.

Individuals and Households

We have discussed almost everything in this chapter in terms of an individual's decision making. However, economists often think in terms of households rather than individuals. In part this is because—as we saw with the budget studies—much more economic data are collected for households than for individuals. Also, many of the decisions we have discussed are really made by a household as a whole, rather than by the individual members of that household.

For example, many households have two working adults. Their decisions about how much to work will usually be made jointly on the basis of the real wages they both face. To see some implications of this, consider a two-person household in which both are working. Now suppose that the real wage increases for one person in the household. One person will probably respond by increasing the number of hours worked. However, the other person may choose to work less. Imagine, for example, that there are household chores that either could do. By working less, one person can do more of these chores and thus compensate the other person for the extra hours worked. [1] Most of the time, though, we do not need to worry about the distinction between the individual and the household, and we often use the terms interchangeably.

Time Studies

Table 3.6 "Allocation of Hours in a Day" shows the allocation of time to certain activities for individuals in three countries: the United States, the United Kingdom, and Mexico. It shows the time allocated on average per day for each of four activities: work, study, personal care, and leisure.

Table 3.6 Allocation of Hours in a Day

Country	Age	Work	Study	Personal Care	Leisure
United States	15–24	2.65	2.2	9.95	5.46
United Kingdom	16–24	3.00	1.39	9.96	5.13
Mexico	20–29	4.49	0.72	10.37	3.1

For the United States and the United Kingdom, the average number of hours worked is between 2.65 and 3.00. This is an average: some people in this age group may work a full-time job, while others may be students who are not working for pay at all. The sample from Mexico differs from the US and UK samples. First, the group is slightly older. Second, Mexico is notably poorer than the United States and the United Kingdom. For these two reasons, individuals sampled in Mexico are more likely to be working and less likely to be studying and enjoying leisure—which is indeed what we see.

Combining Your Time and Spending Choices

So far we have looked at the allocation of your income separately from the allocation of your time. Yet these choices are linked. The allocation of your time influences the income you have to spend on goods and services. So a change in the wages you are paid will affect how you allocate your time *and* the goods and services you choose to buy.

In a similar fashion, the prices of goods and services you purchase will have an influence on your allocation of time. For example, if the price of a computer you want to buy decreases, you may respond by working a little more to earn extra income to purchase the computer. The reduction in the price of the computer raises your real wage, so you respond by working more.

Effects of Real Wages on Household Demand

If the real wage changes, there are changes in both consumption decisions and work choices. Figure 3.16 "The Effect of a Real Wage Increase on the Quantity of Labor Supplied" shows that an increase in the real wage means you can obtain more consumption for a given amount of work time. Further, as in Figure 3.6 "An Increase in Income", the budget set expands as income increases. Because an increase in the real wage will lead to an increase in hours worked (see Figure 3.16 "The Effect of a Real Wage Increase on the Quantity of Labor Supplied"), labor income will increase. So we can interpret the shift in the budget line inFigure 3.6 "An Increase in Income" as coming from this increase in labor income.

When income increases, you will generally consume more of all goods and services. An increase in the real wage leads to an outward shift in your demand curves for chocolate bars, downloads, and all other normal goods. Combining these figures, we can make the following predictions about the effects of an increase in the real wage:

- You will work more hours.
- You will have more income.
- You will consume more goods and services.
- You will be happier.

The last item in the list draws on Figure 3.6 "An Increase in Income" but is less direct than the other implications. As the real wage—and thus your income—increases, the set of bundles you can afford is larger. Moreover, every bundle you could afford when you had less income is still affordable now that you have more income. Thus we conclude that you will be happier. After all, you can always purchase the bundle you bought with lower income and still have some extra income to spend.

Effects of Prices on Time Allocation

Suppose the price of a chocolate bar increases. We saw from Figure 3.8 "A Decrease in the Price of a Chocolate Bar" that when this price increases, the budget set shrinks. We also saw from Figure 3.9 "The Demand Curve" that the demand for chocolate bars decreases when the price of a chocolate bar increases. But there are also implications for labor supply. Remember that the real wage is the nominal wage dividing by a price index representing a household's cost of purchasing a bundle of goods and services. So when the price of a chocolate bar increases, the cost of purchasing the bundle will increase and the real wage will decrease. From Figure 3.16 "The Effect of a Real Wage Increase on the Quantity of Labor Supplied", labor supply will decrease as the real wage decreases. This is a *movement along* the labor supply curve.

KEY TAKEAWAYS

- The time budget constraint states that the sum of the time spent on all activities each day must equal 24 hours.
- The opportunity cost of time spent on one activity is the time taken away from another.
- Decisions about how much to work depend on how much more you can purchase if you work a little more: that is, they depend on the real wage.
- The individual labor supply curve shows how much an individual will choose to work given the real wage.
- As the real wage increases, an individual will supply more labor if the substitution effect dominates the income effect.

CHECKING YOUR UNDERSTANDING

1. If you must sleep a minimum of five hours each day, how would you modifyFigure 3.14 "The Time Budget Constraint" to indicate this necessity of life?
2. If both the nominal wage and the price level double, what will happen to your allocation of time and consumption?

[1] This is an application of an important economic idea called *comparative advantage*, which we discuss in more detail in Chapter 5 "eBay and craigslist".

3.4 End-of-Chapter Material

In Conclusion

We all make economic decisions every day, often without giving them very much thought. In this chapter, we highlighted two fundamental decisions: allocating income and allocating time. You (and everyone else) generally make choices over income and time allocations in a manner that makes you happy. Thus we predict that you will not throw income away. Further, whatever combinations of goods and services and time allocations you choose, economic theory presumes that these are the best ones available to you. Remember also that even though it is usually easier to focus on one decision at a time, your spending and time allocation decisions are interconnected. Changes in the prices of goods and services affect how you spend your time, and changes in the real wage affect your consumption choices. [1]

Economics is often defined as the study of how we allocate scarce resources that have alternative uses. In this chapter, we saw this idea at work at the individual level. Your income is a scarce resource; you don't have enough income to buy everything that you would like. Your income has alternative uses because there are lots of things you might want to buy.

Perhaps the most fundamental idea of this chapter is that of opportunity cost. Given that you have limited income to allocate across goods and services, the opportunity cost of consuming one good or service is the amount of another good or service you give up. Given that you have limited time to allocate across activities, the opportunity cost of spending time on one activity is the value of the time you could have spent on another. The budget and time budget constraints are graphical representations of this central economic principle. And the interaction between these budget constraints and people's wants and desires is at the heart of the economic analysis of decision making.

Key Link

- *New York Times* Time Allocation Study:http://www.nytimes.com/interactive/2009/07/31/business/20080801-metrics-graphic.html

EXERCISES

1. Which statements are prescriptive? Which statements are descriptive?

a. The government should take care of the poor.

b. If the real wage increases, households will be willing to supply more labor.

c. When people's incomes decrease, they consume more cheap cuts of meat.

d. We ought to consume fewer resources to protect the planet.

e. Young people should purchase medical insurance because it is cheaper for young people than for old people.

Draw a budget line assuming disposable income equals $100, the price of a music download is $1, and the price of a chocolate bar is $5. On the same graph, draw another budget line assuming the same level of income but with the price of a music download equal to $2 and the price of a chocolate bar equal to $1. Explain how the budget sets differ. If you liked downloads but hated chocolate bars, which budget set would you prefer?

Consider a bundle consisting of 4 chocolate bars and 30 downloads (call it bundle A). Using the assumption that "more is better," what bundles can we say are definitely preferred to bundle A? What bundles are definitely worse than bundle A? Show your answers on our usual kind of diagram (that is, a budget set diagram with chocolate bars and downloads on the axes).

If we observe that a household buys bundle A but not bundle B and we know that bundle B has more of every good than does bundle A, what can we say about the household's preferences for bundle A and bundle B? What can we say about the household's budget set?

(Advanced) Look at Table 3.7 "Preferences over Downloads and Chocolate Bars". The top part of the table lists four different bundles of downloads and chocolate bars. The bottom part of the table shows which bundle is preferred when we compare any two bundles. Look at bundle 1. The top part of the table tells us it contains 0 downloads and 20 chocolate bars. The first row of the bottom part of the table shows how this bundle compares to the other bundles. So this individual prefers bundle 1 to bundle 2 but also prefers both bundle 3 and bundle 4 to bundle 1. Do these preferences satisfy "more is better"? Are they consistent or can you find some contradictions?

TABLE 3.7 PREFERENCES OVER DOWNLOADS AND CHOCOLATE BARS

Bundle	Downloads Consumption	Chocolate Bar Consumption
1	0	20
2	100	0

Bundle	Downloads Consumption		Chocolate Bar Consumption	
3	50		10	
4	110		30	
	Which Bundle Is Preferred When Comparing Bundles?			
Bundle	Bundle 1	Bundle 2	Bundle 3	Bundle 4
1	—	1	3	4
2	1	—	2	4
3	3	2	—	3
4	4	4	3	—

Suppose income increases by 10 percent, but the price of a chocolate bar and the price of downloads both increase by 5 percent. Will the budget line shift inward or outward? Will the slope of the budget line change?

We explained the household demand curve and the law of demand by focusing on how a change in the price of a chocolate bar influences the quantity of chocolate bars demanded. Redo this discussion and the figures to illustrate how a change in the price of downloads will affect the demand for downloads and the demand for chocolate bars.

In our example, we noted that it was not possible for both chocolate bars and music downloads to be inferior goods. Suppose there were three goods: chocolate bars, music downloads, and tuna sushi. Is it possible now that chocolate bars and music downloads are both inferior goods? Could all three be inferior goods?

(Advanced) Suppose the government imposes a tax on chocolate bars. Draw a diagram that shows what happens to the budget set. If chocolate bars and downloads are both normal goods, can you say whether the consumption of chocolate bars will increase or decrease? What about the consumption of downloads?

Explain why a price increase in movie tickets causes the demand curve for chocolate bars to shift.

Suppose you are thinking of buying chocolate bars. Your marginal valuation of the seventh chocolate bar is $3. The price of a chocolate bar is $4. Should you buy more or fewer than seven bars?

Explain how the law of demand works in the case of a unit demand curve.

Can preferences include altruism or a regard for fairness and still exhibit rationality?

Using the data in Table 3.5 "United Kingdom Budget Study", create a pie graph of expenditure shares. How might you explain the differences in spending between younger households in the United States and the United Kingdom? How might you explain the differences in spending in 2005 between younger households and average households?

(Advanced) In discussing labor supply, we did not allow an individual to decide not to work. Yet we observe many individuals who could work but choose not to. How would you have to amend the discussion to include the choice of working or not working?

If you face a big exam this week, how might this influence your time allocation choice in Figure 3.14 "The Time Budget Constraint"?

(Advanced) If there is a reduction in the price of a chocolate bar, what does our theory predict will happen to labor supply?

(Advanced) Suppose the government imposes a tax on labor. What will that tax do to the labor supply of a household and its demands for downloads and chocolate bars?

(Advanced) If one member of a two-person household gets a raise, what will that do to the hours worked by that person and to the other household member? Explain this using income and substitution effects. Could this raise cause the other household member not to work at all?

Economics Detective

1. Search the Internet to find the level of spending by Japanese households on food in a recent year. Convert this figure to dollars.
2. The data in Table 3.4 "Budget Shares in the United States" come from a survey. Who was surveyed? How frequently?
3. Go to the web page of the Office for National Statistics in the United Kingdom (http://www.statistics.gov.uk/CCI/Nscl.asp?ID=5407&Pos=1&ColRank=1&Rank=16UK) and create a version of Table 3.5 "United Kingdom Budget Study" for different income groups. What differences do you see in spending patterns across income groups? How would you explain the differences in spending patterns as income changes?
4. Go to http://www.bls.gov/tus. Pick two years. Prepare a table to illustrate how the allocation of time has changed for one of two age groups over these two years. How might you explain these changes?

Spreadsheet Exercise

1. (Advanced) Create a spreadsheet to reproduce the graph of a budget constraint with two goods (chocolate bars and downloads) in Figure 3.4 "The Budget Line". In column A, put the quantity of chocolate bars (from 0 to 20). In column B, put the price of a chocolate bar (that is, each cell should contain a 5). [2] In column C, put the price of downloads. In column D, put income. Then write an equation to enter in each cell of column E, based on the budget line. This equation should calculate the quantity of downloads in terms of the prices, income, and the quantity of chocolate bars. Make sure that you allow only nonnegative quantities of the goods. Use this to graph the budget line. Now try changing the prices and the level of income and make sure you can explain how the budget line shifts as income and prices change.

[1] Chapter 7 "Why Do Prices Change?" has more to say on the connections among different markets in the economy.

[2] This is the simplest but not the most elegant way to create this spreadsheet. If you are an experienced user of spreadsheets, you may know tricks that will allow you to create the spreadsheet in a more compact way.

Chapter 4
Life Decisions

Life Choices

Some economic decisions, like how to spend your money and your time, are everyday decisions. [1] There are also bigger and more difficult economic decisions that you confront only occasionally. In the months and years after graduation, you will face major life choices, such as the following:

- Upon graduation, which occupation should you choose?
- Should you go to graduate school?
- Should you purchase a new car?
- Should you purchase a house?
- How much of your income should you save?
- Should you purchase health insurance?
- Should you purchase insurance for your home or apartment?
- What should you do with the money you save?

These economic choices are more complicated than choosing how many chocolate bars to buy or how much time you should spend watching television today.

Two things make these decisions hard. First, there is the element of time—not the 24 hours in a day, but the fact that you must make decisions whose consequences will unfold over time. In choosing an occupation, deciding on graduate school, or picking a portfolio of financial assets, you must look ahead. Second, there is the element of uncertainty. Will you be healthy? Will you live to an old age? Will you succeed as a rock musician? The future is unknown, yet we cannot ignore it. The future is coming whether we like it or not.

We cannot tell you whether you should buy a new car or if you will be a rock star. But we can give you some tools that will help you when you are making decisions that involve time and uncertainty. In this chapter we tackle the following questions:

- How do we make decisions over time?

- How do we make plans for an uncertain future?

Road Map

The chapter is organized around the two themes of time and uncertainty. We begin with a brief review of the choice between two goods at a given time. [2]Then we look at choices over time. Economists typically assume that individuals are capable of choosing consistently among the bundles of goods and services they might wish to consume. The ability to make such a choice is perhaps not too onerous in the case of simple choices at a given time (such as whether to go to a movie or go to dinner). It is more difficult when we consider choices over a broad range of goods from now into the future.

- There are goods and services that will be available to you in the future that you cannot imagine today. When people chose among different types of handheld calculators 30 years ago, they could not imagine that today they would be choosing among different types of tablet computers. Many products that we now consume simply did not exist in any form until comparatively recently, and when we make choices now, we do so in ignorance of future consumption possibilities.
- Your tastes may change. When you are 20 years old, it is difficult to predict what goods and services you will want to buy when you are 30, 40, or 50 years old. Your future self might regret past decisions.

We tackle time and uncertainty separately. To begin with, we will suppose that the future is known with certainty. This allows us to focus on including time in our analysis of economic decision making. We begin with a discussion of the choice between consumption and saving and explain how this decision is affected by changes in interest rates. We then look at problems such as how to choose an occupation. A major part of this analysis is an explanation of how to compare income that we receive in different years.

We then turn to uncertainty. We explain the idea of risk and then discuss the kinds of risks you cannot avoid in life. We explain how insurance is a way to cope with these risks. We also discuss uncertainties that we *create* in our lives—through occupational choice, portfolio choice, and gambling.

[1] See Chapter 3 "Everyday Decisions" for more discussion.

[2] This decision is analyzed at length in Chapter 3 "Everyday Decisions".

4.1 Consumption and Saving

LEARNING OBJECTIVES

1. What is your lifetime budget constraint?
2. What factors influence your choice between consumption today and saving for the future?
3. What is the difference between real and nominal interest rates?
4. What are the effects of a change in the interest rate on consumption and saving?

Your choice at any given time between two goods—say, chocolate bars and music downloads—reflects the tension between your desires for chocolate bars and downloads and your income, as summarized by your **budget line**. The budget line shows us the bundles of goods and services that you can afford, given prices and your income, under the presumption that you do not throw any money away. For an individual choosing between two goods only (chocolate bars and music downloads), the budget line states that total spending is equal to spending on chocolate bars plus spending on downloads:

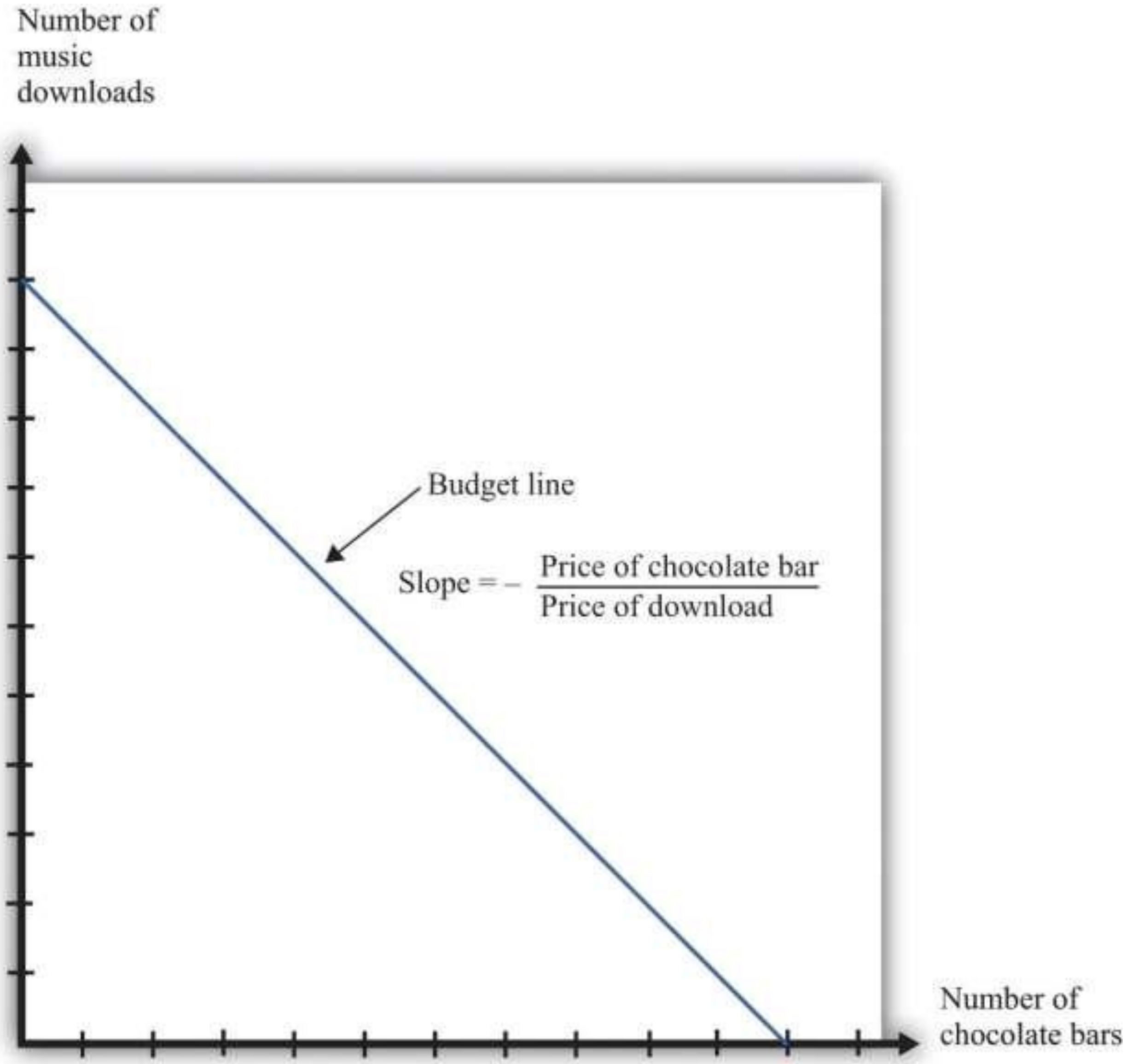
(number of chocolate bars × price of a chocolate bar) + (number of downloads × price of download) = disposable income.

Figure 4.1 "The Budget Line with Two Goods" displays the budget line for the choice between downloads and chocolate bars.

Toolkit: Section 17.1 "Individual Demand"

You can review the derivation and meaning of the budget line in the toolkit.

Figure 4.1 The Budget Line with Two Goods



This diagram shows the budget line for an individual choosing between chocolate bars and music downloads. The slope of the budget line reflects the rate at which an individual can trade off the two goods.

The slope of the budget line is

−price of a chocolate barprice of download.

As you move along the budget line, you are giving up downloads to get chocolate bars. If you give up one chocolate bar, you get an amount of money equal to the price of a chocolate bar. You can take this money and use it to buy music downloads. You have to divide this amount of money by the price of a download to determine how many downloads you can buy. The slope of the budget line reflects the opportunity cost of chocolate bars in terms of music downloads.

The Budget Line with Two Periods

Now think about what the budget line looks like when we are choosing between now and the future. Just as we find it easier to think about the choice between two goods rather than among 2,000 goods, so too is it easier to think about the choice over only two periods of time. We call these two periods "this year" and "next year." There is nothing special about the two-period example beyond the fact that it makes it easy to draw diagrams and see what is going on. The principles that we uncover for this case also apply to decisions made with more than two time periods in mind.

We also do not worry about all the different goods and services that are available, preferring instead to talk in general about consumption this year and consumption next year. We *aggregate* together all the different products that we consume. Thus "consumption" means the bundle of goods and services people consume. This consumption also has an associated price, which we call the price level. Think of this as the average price of goods and services in the economy. If you find it easier, imagine there is a single good, like chocolate bars, that you consume, and think of the price level as simply being the price of a chocolate bar.

Saving and the Nominal Interest Rate

If you choose not to spend all your income on consumption this year, you are saving. When you save, you can put your money into a financial institution and earn interest on it. Suppose you have $100 this year that you save by putting it in a bank. You are then *lending* to the bank—saving and lending are really the same thing. The bank acts as an intermediary, taking your $100 and giving it to someone else who borrows from the bank.

The bank offers you interest on this loan—for example, it may pay anominal interest rate of 5 percent per year. After a year, your bank account will contain your original $100, plus an extra 5 percent. Because 5 percent of $100 is $5, you earn $5 worth of interest. We talk about interest rates in percentage terms, but you should remember that a percentage is simply a number. For example, 5 percent is 0.05, and 20 percent is 0.2. The nominal interest rate is the interest rate at which individuals and firms in the economy can save or borrow. [1] It is called a *nominal* interest rate because it is measured in monetary terms. Most interest rates are quoted on an annual basis, meaning that they specify the amount earned per year.

Of course, if you put a $100 in the bank for one year, then next year you will still have the original $100 as well as the interest you earned. At an interest rate of 5 percent, $100 this year is worth $105 next year. To calculate the *total* amount of money that you can earn, we simply add one to the nominal interest rate, giving us the nominal interest factor:

nominal interest factor = 1 + nominal interest rate.

The nominal interest factor is used to convert dollars today into dollars next year.

Toolkit: Section 17.6 "The Credit Market"

The nominal interest rate is the rate at which individuals and firms in the economy can save or borrow. The nominal interest factor is 1 + the nominal interest rate.

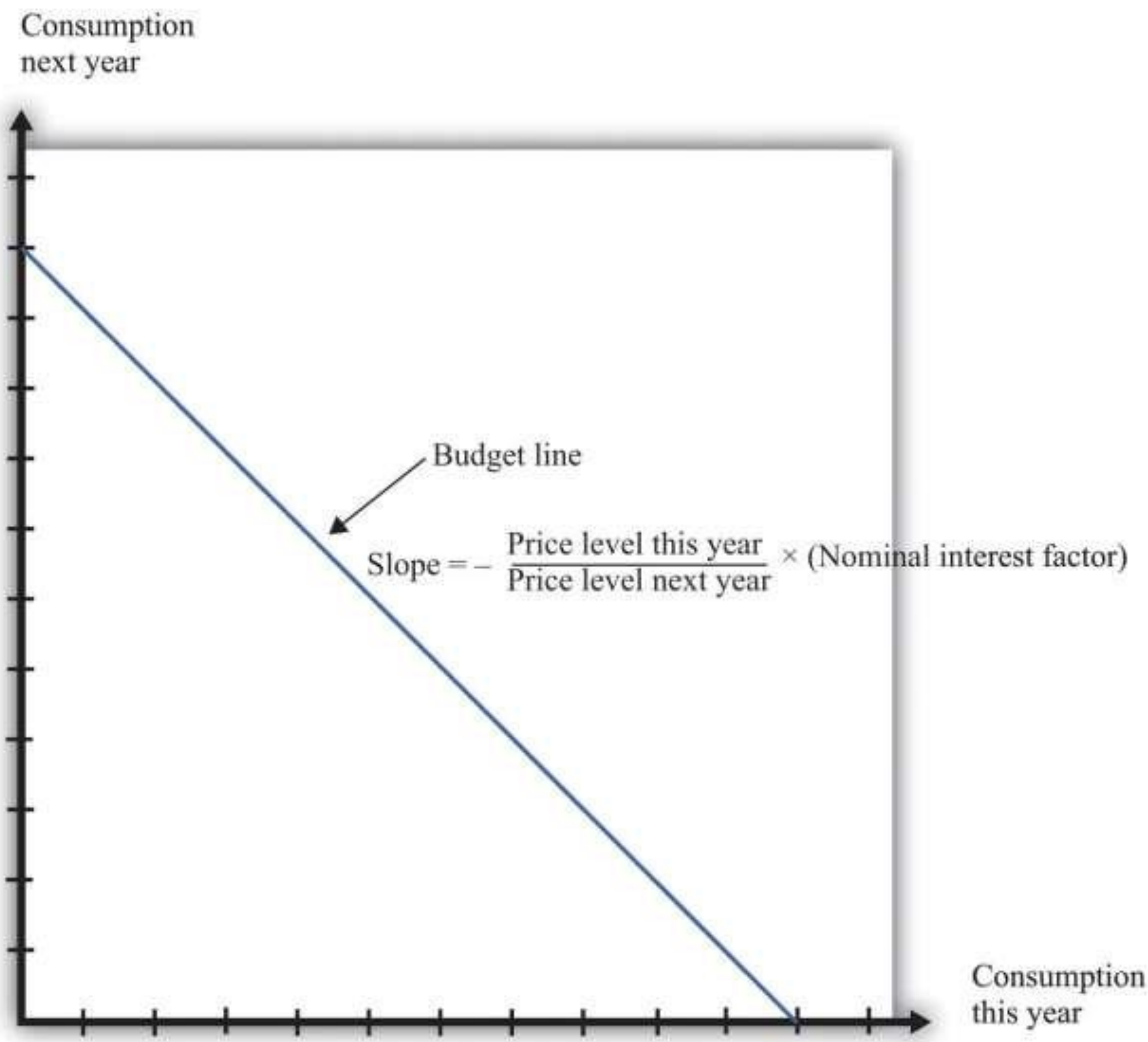
In this chapter, we typically use the nominal interest factor rather than the nominal interest rate because it makes the equations easier to understand. Just keep in mind that it is easy to move back and forth between the interest rate and the interest factor by adding or subtracting one.

If you have $100 today, then tomorrow it will be worth $100 × the nominal interest factor. In general,

z this year will be worth *z* × the nominal interest factor next year.

Armed with this idea of the nominal interest factor, we can graph the budget line for a two-period consumption-saving problem. Figure 4.2 "The Budget Line with Two Periods" shows consumption this year on the horizontal axis and consumption next year on the vertical axis. To discover what the budget line looks like, we first determine its slope and then its position.

Figure 4.2 The Budget Line with Two Periods



This diagram shows the budget line for an individual choosing consumption over time. The slope of the budget line depends on the price level this year, the price level next year, and the nominal interest factor (= 1 + nominal interest rate). Suppose the price level this year is $9, the price level next year is $10, and the nominal interest rate is 20 percent. Then the slope of the budget line is −(9/10) × 1.2 = −1.08. This means that if you give up 1 unit of consumption this year, you can get 1.08 units next year.

The Slope of the Budget Line

The slope of the budget line tells you how much extra consumption you will get next year if you give up a unit of consumption this year. So to determine the slope of the budget line, we use the following thought experiment.

1. If you give up one unit of consumption this year, you get an amount of dollars equal to the price of consumption this year (the price level).

2. You can then save those dollars, so next year you will have a number of dollars equal to the price level this year × the nominal interest factor.
3. Next year, you can take these funds and purchase

$$\frac{\text{price level this year} \times \text{nominal interest factor}}{\text{price level next year}}$$

   units of consumption.

So the slope of the budget line is as follows:

$$\text{slope of budget line} = -\frac{\text{price level this year}}{\text{price level next year}} \times \text{nominal interest factor}.$$

We show the budget line in Figure 4.2 "The Budget Line with Two Periods".

The budget line has a negative slope because—as with the choice between chocolate bars and downloads—you must give one thing up to get another. If you want to consume more in the future, you must be willing to consume less right now. If you want to consume more now, you will have to sacrifice consumption in the future.

The slope of the budget line depends not only on the nominal interest factor but also on prices this year and next year. Suppose the price of a unit of consumption this year is \$100 and next year it is \$110. Economists call the percentage increase in the price level the inflation rate; it is calculated as follows:

$$\text{inflation rate} = \frac{\text{price level next year}}{\text{price level this year}} - 1.$$

Put differently, it is the rate of growth of the price level. In our example, the inflation rate is 10 percent.

Now suppose the nominal interest rate is also 10 percent, which means that the nominal interest factor is 1.1. Then

$$\text{slope of budget line} = -\frac{100}{110} \times (1 + 0.1) = -\frac{1.1}{1.1} = -1.$$

In this case, the price level increased by 10 percent, from 100 to 110. But the nominal interest rate also increased by 10 percent, which offset the increase in prices. We see that the slope of the budget line depends on both the nominal interest factor and the rate of inflation. In fact, it depends on thereal interest factor:

slope of budget line = −(1 + real interest rate) = −real interest factor.

Toolkit: Section 17.6 "The Credit Market"

The real interest rate is the rate of interest adjusted for inflation. It tells you how much you will get next year, *in terms of goods and services*, if you give up a unit of goods and services this year. The real interest factor allows you to convert units of goods and services this year into units of goods and services next year. The real interest factor is 1 + the real interest rate.

As you move along the budget line in Figure 4.2 "The Budget Line with Two Periods", you are giving up chocolate (consumption) this year for chocolate next year. So the slope of the budget line must be a number, not a dollar amount. Because this year's price and next year's price are both denoted in dollars, their ratio is a number. Likewise, the interest rate is a number, so the slope of the budget line is indeed a number.

An example may help you understand the difference between the nominal interest rate and the real interest rate. Suppose you go to your bank and get a one-year, $20,000 loan to buy a car, with a nominal interest rate of 5 percent. Your contract with the bank thus stipulates that you must pay the bank $21,000 at the end of the year. If the inflation rate is zero, then the cost of borrowing measured in terms of real goods and services is $1,000, which is 5 percent of the amount that you borrowed. But if the inflation rate is 5 percent, then the $21,000 you pay to the bank at the end of the loan buys the same amount of goods and services that the $20,000 the bank lent to you. In this case, you are effectively able to borrow for free.

Good decisions about borrowing and lending are based on real interest rates rather than nominal interest rates. Your cost of borrowing to buy the car is not the monetary payments you make on the loan but rather the value of the goods and services you could have purchased with that money. So we need a way to convert from the commonly observed nominal interest rate to a measure of real interest rates. We do this by using a formula for the real interest rate that was discovered by a famous economist named Irving Fisher.

Toolkit: Section 17.8 "Correcting for Inflation"

The Fisher equation is a formula for converting from nominal interest rates to real interest rates, which is as follows:

real interest rate ≈ nominal interest rate – inflation rate.

Equivalently,

real interest factor ≈ nominal interest factor – inflation rate.

For example, suppose the nominal interest rate is 5 percent. If the rate of inflation is zero, then the real interest rate is 5 percent. But if the rate of inflation is 4 percent, the real interest rate is only 1 percent. The Fisher equation is a tool that tells us how to convert nominal interest rates—the interest rates you see in the newspapers and on television—into real interest rates, which are key for decision making. [2]

The Position of the Budget Line

Whereas the slope of your budget line depends on the real interest rate, the*position* of your budget line depends on how much income you have. When you have more income, the budget line is further away from the origin. One way to determine the position of the budget line is by looking at its intercepts. The horizontal intercept is the amount you can consume this year if you spend all of this year's income *and* borrow against your entire future income. The vertical intercept is the amount you can consume next year if you choose to consume nothing this year *and* save all of your current income.

It is easier and more instructive, however, to look at a different point on the budget line. Remember that the budget line is the bundles of consumption you can just afford. One bundle you can certainly afford is the bundle where you spend all of this year's income on consumption this year and all of next year's income on consumption next year. In other words, one available option is that you neither save nor borrow. In this case,

$$\text{consumption this year} = \text{real income this year} = \frac{\text{nominal income this year}}{\text{price level this year}}$$

and

consumption next year = real income next year =nominal income next yearprice level next year.

On the right-hand side of these equations, we divided dollar income by the price level to give us real income (that is, income measured in terms of purchasing power). We must do this to find out how much you can consume in terms of goods and services.

For example, suppose your nominal income this year is $23,000, and your nominal income next year is $24,200. Suppose the price level this year is $10 and the price level next year is $11. This means that

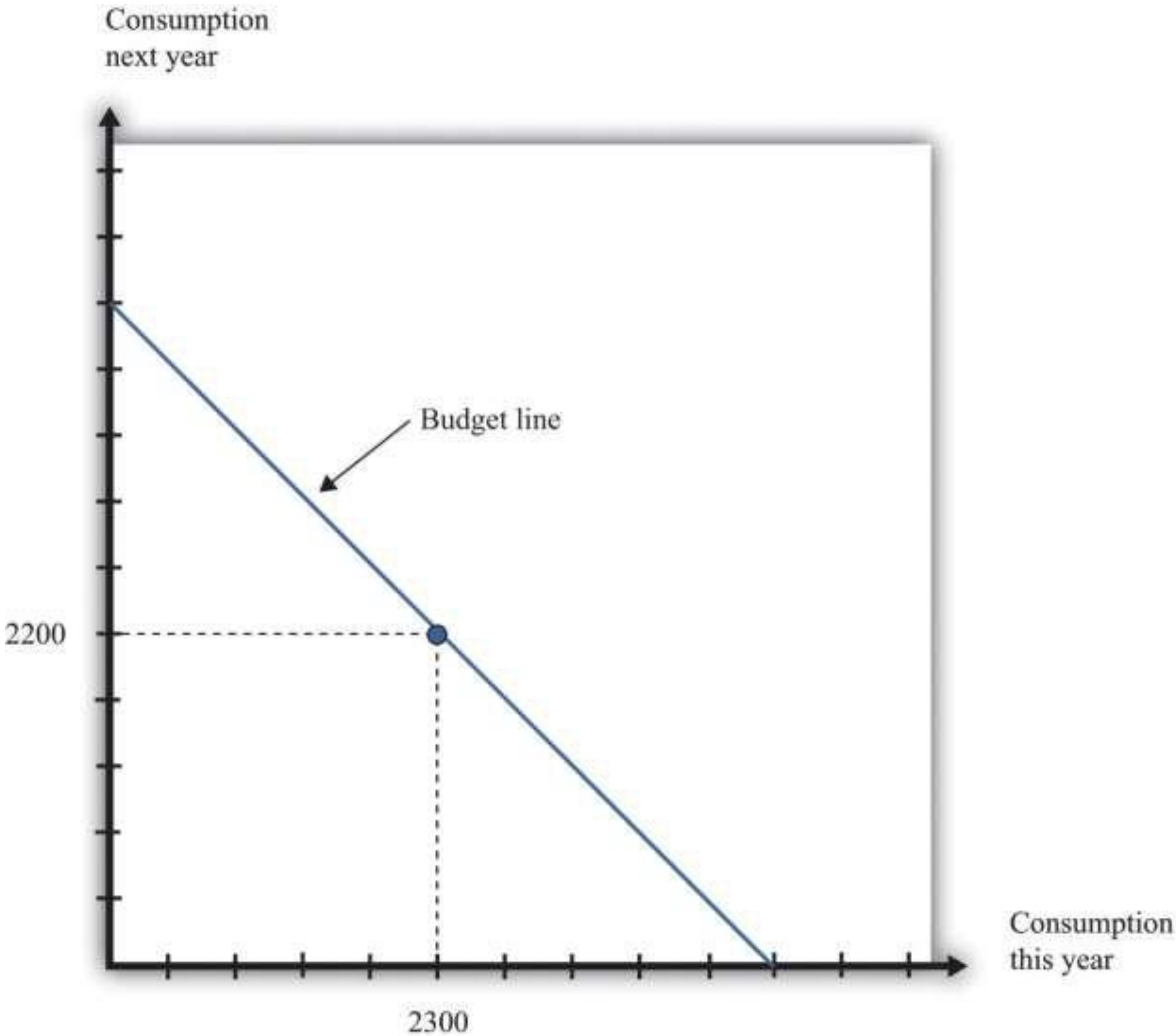
real income this year = $23,000$10

and

real income next year = $24,200$11.

So one possible consumption choice, as shown in Figure 4.3 "Determining the Position of the Budget Line", is 2,300 units of consumption this year and 2,200 units of consumption next year. In this case, you are neither borrowing nor saving. Of course, you might choose some different point on the budget line.Figure 4.3 "Determining the Position of the Budget Line" shows that your real income this year and next year does indeed pick out a point on the budget line. And because we already know the slope of the budget line, we are done; we can now draw the budget line.

Figure 4.3 Determining the Position of the Budget Line



The position of the budget line depends on income this year and next year. We know that one possible choice of consumption is where the consumer neither saves nor borrows. This means that the budget line must pass through this point.

Adding Income and Consumption Spending over Two Periods

Your budget line describes the condition that total spending equals total income. This is true for the choice about the consumption of downloads and chocolate bars, and it is also true for the choice over time. But once we move to two periods, we must be careful about measuring both total income and total spending.

Adding Nominal Income over Two Periods

Suppose you earn some income this year (say, $23,000) and will earn some more next year (say $24,200). What is your income for these two years together? Your first instinct is probably to add the income in the two years and say $47,200. Superficially this makes sense—after all, income is measured in dollars in both years. Unfortunately, this is not a very good way to add money over time. It is flawed because it views income in two different years as if they are the same thing. In fact, money this year and money next year are not the same.

Imagine that a friend asks to borrow $1,000 from you today, promising to pay you back the $1,000 twenty years from now. Would you be likely to agree to this? Even if you trust your friend completely, the answer is surely no. After all, you could take your $1,000 and put it in the bank for twenty years, and the bank will pay you interest on your money—that is, the bank is willing to pay you for the privilege of using your money. Over twenty years, you could earn quite a bit of interest. By contrast, your friend is asking for a *zero-interest loan* in which no interest is paid on the money that you lend.

Positive interest rates mean that a dollar today and a dollar in the future are not worth the same. Adding dollars in one year to dollars in another year makes no more sense than adding apples and oranges. We need to convert dollars next year into their value right now. Remember that

z this year will be worth $z \times$ the nominal interest factor next year.

We can turn this around. If the interest rate is 5 percent, then $105 next year will be worth only $100 this year. A dollar next year is worth $\frac{1}{\text{nominal interest factor}}$ dollars this year. This is the most you would be willing to give someone this year if he or she promised to give you a dollar next year. You would not give them more than this because you would lose money relative to the alternative of putting the dollar in the bank and earning interest. More generally,

$$z \text{ next year is worth } \frac{z}{\text{nominal interest factor}} \text{ this year.}$$

Toolkit: Section 17.4 "Choices over Time"

Unless the interest rate is zero, a dollar this year is not the same as a dollar a year from now. To avoid this problem, economists use discounted present value as a device for measuring flows that occur over time. Discounted present value tells you the value of something you will receive in the future, discounted back to the present.

For example, if we want to add income in dollars over two years, the discounted present value of such a two-year *flow of income* is given by the following formula:

$$\text{discounted present value of two-year flow of nominal income} = \text{nominal income this year} + \frac{\text{nominal income next year}}{\text{nominal interest factor}}.$$

This is the income term that we need for the budget line in our two-year example. Go back to our earlier example, where income this year is $23,000 and income next year is $24,200, and the nominal interest factor is 1.1. Then

$$\text{discounted present value of two-year flow of nominal income} = \$23{,}000 + \frac{\$24{,}200}{1.1} = \$45{,}000.$$

Even though income next year is higher in dollar terms, it is lower in terms of present value: $24,200 next year is worth only $22,000 today. Notice that when we measure the discounted value of a flow of nominal income, we still end up with a nominal value—the value of the income flow in terms of this year's dollars.

Table 4.1 "Discounted Present Value of Income" provides another illustration: it shows the calculation of the discounted present value of income when this year's income is $100 and next year's income is $200. You can see that, as the interest rate increases, the discounted present value of income decreases.

Table 4.1 Discounted Present Value of Income

Nominal Income This Year ($)	Nominal Income Next Year ($)	Discounted Present Value of Nominal Income Flow ($)		
		Nominal Interest Rate		
		0%	**5%**	**10%**
100	200	300.00	290.47	281.82
250	500			

Adding Nominal Consumption over Two Periods

When we want to add consumption spending this year and next year, measured in dollars, we use exactly the same logic as we did when adding income. Nominal consumption this year and next are given as follows:

nominal consumption this year = price level this year × consumption this year

and

nominal consumption next year = price level next year × consumption next year.

(Again, if you find it easier, just think of this as chocolate: total spending is the number of chocolate bars purchased times the price per bar. When we talk about "consumption," we mean something measured in real units, such as chocolate bars. When we talk about "nominal consumption," we are referring to a value measured in dollars.) Just as it is incorrect to add this year's and next year's income, so too should we not add together nominal consumption. Instead, we must calculate a discounted present value, exactly as we did before.

discounted present value of two-year flow of consumption spending = nominal consumption this year + nominal consumption next yearnominal interest factor.

As with income, the discounted present value of nominal consumption is measured in this year's dollars.

Which Interest Factor Should You Use?

Earlier, we emphasized that people think about the real interest factor when they are comparing this year and next year. Yet in calculating the discounted present value of income and consumption spending, we are using the nominal interest factor. What is going on?

The rule for determining which interest factor to use in a discounted present value calculation is simple. *If you are converting nominal values, then you should use the nominal interest factor. If you are converting real values, then you should use the real interest factor*. So if you want to know how much a given number of dollars in the future will be worth in dollars today, you should use the nominal interest factor. This is the normal case for most calculations that you would do. However, if you want to calculate a discounted present value for variables that have already been corrected for inflation, you must use the real interest factor. In this case, the answer you get is also a real quantity.

The Two-Period Budget Line Revisited

The tool of discounted present value gives us another way of thinking about the two-period budget line—the condition that

discounted present value of two-year flow of nominal consumption = discounted present value of two-year flow of nominal income.

Remember that both sides of this equation are measured in terms of this year's dollars. If we were to divide both sides of this equation by this year's price level, then we would get the equivalent expression in real terms:

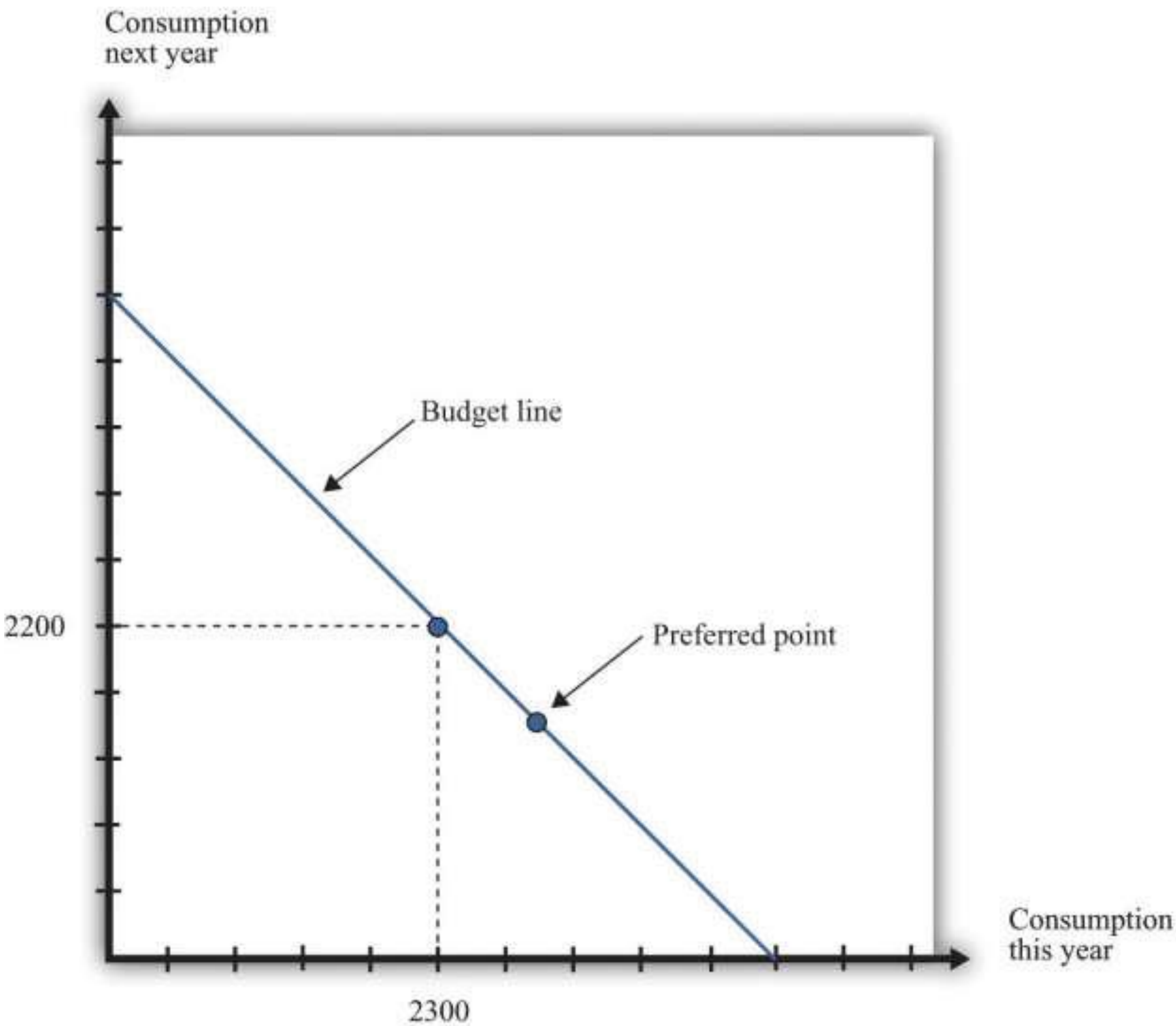
discounted present value of two-year flow of consumption = discounted present value of two-year flow of real income.

In this case, as we just explained, the discounting must be done using the real interest factor instead of the nominal interest factor.

Income, Consumption, and Saving

Given your budget line, we suppose you choose a combination of consumption this year and next year that makes you as well off as possible. An example of such a preferred point is indicated in Figure 4.4 "The Preferred Point".

Figure 4.4 The Preferred Point



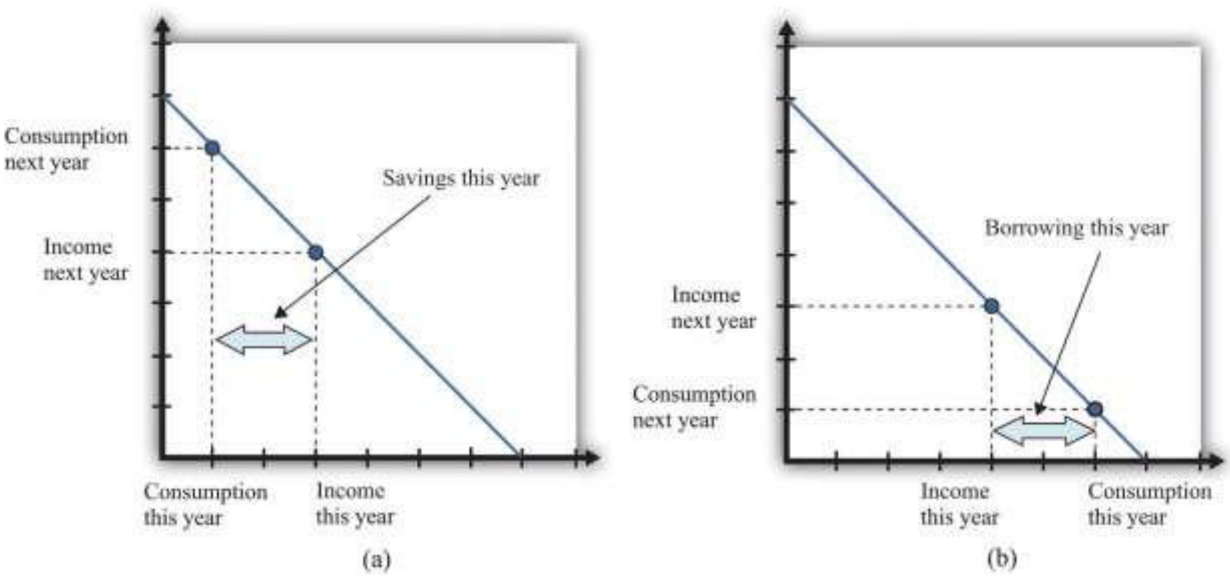
The consumer's preferred point must lie somewhere on the budget line. In this example, the consumer is choosing to consume in excess of his income this year. The consumer must borrow against future income, which means that consumption next year will be below next year's income.

The choice of a preferred point reflects two ideas. Whatever your tastes between consumption in the two years, we presume that you will not throw any income away. As a result, your choice will be on, not inside, the budget line. Further, if you choose well, according to your preferences, then you will pick the best combination of consumption; there is no other point on the budget line that you prefer.

Saving and Borrowing

Your preferred point implies a choice about how much saving or borrowing you do. Figure 4.5 "Consumption and Saving" shows two possible cases. In part (a) ofFigure 4.5 "Consumption and Saving", you are a saver: you are consuming less than your income this year. The difference between your income and your consumption is the amount of your savings. Those savings, plus interest, are available to you next year, so next year you can consume in excess of your income. In part (b) of Figure 4.5 "Consumption and Saving", you are a borrower: you are consuming more than your income this year. When you borrow this year, you must repay the loan with interest next year, so your consumption next year is less than your income.

Figure 4.5 Consumption and Saving



(a) The individual is a saver this year. (b) The individual is a borrower this year.

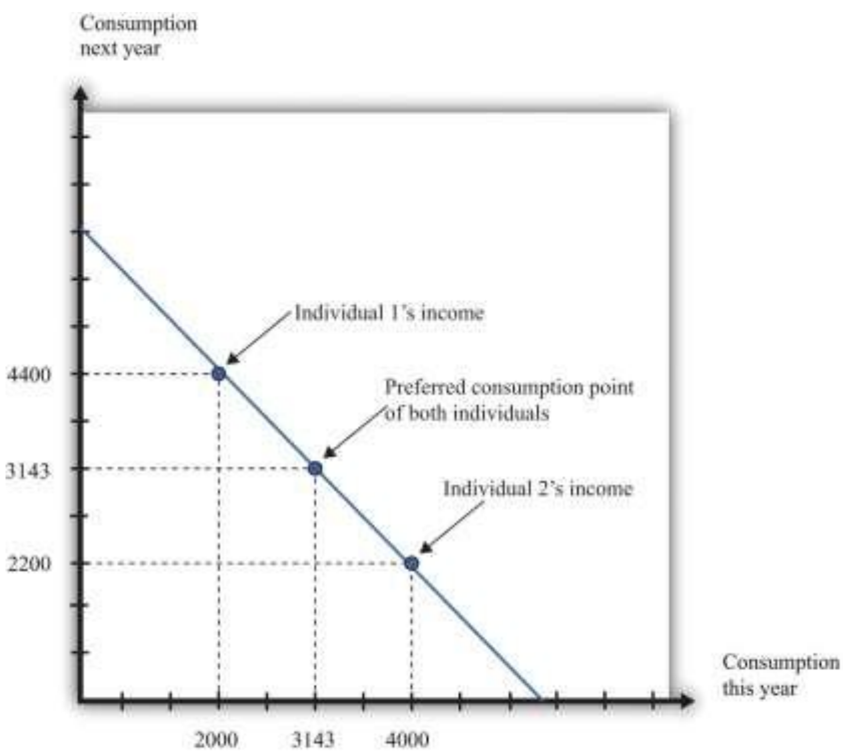
The budget line tells you the rate at which the *market* allows you to substitute goods between this year and next year. This is distinct from your personal tastes about consuming this year or next year. The saver in part (a) of Figure 4.5 "Consumption and Saving" is a relatively patient person: she is willing to give up a lot of consumption this year to be able to consume more next year. The borrower in part (b) of Figure 4.5 "Consumption and Saving" is a relatively impatient person: he wants to consume a lot this year and is willing to sacrifice a great deal of future consumption.

You will sometimes hear discussions of how much individuals like to "discount the future." This is a statement about their tastes. Someone who discounts the future a great deal is impatient. Such a person wants to consume right away, so he will give up a lot of future consumption to have more today. Someone who discounts the future only a little is patient. Such a person is willing to give up consumption today even if she gets only a little extra consumption in the future. Economists pass no judgment on whether it is better to be impatient or thrifty. These are matters of personal preference.

The Timing of Income

Interestingly, the timing of your income turns out not to matter for your choice of consumption, which is illustrated in Figure 4.6 "The Timing of Income".

Figure 4.6 The Timing of Income



The timing of income is irrelevant to the consumption choice. Here, one individual has low income this year and high income next year, while the opposite is true for the other individual. However, the discounted present value of income is the same in both cases. If they both have the same tastes, they will choose the same consumption point.

Suppose that the nominal interest rate is 10 percent and that the price level is $10 in both periods. This means that the inflation rate is zero, so—from the Fisher equation—the real interest rate is also 10 percent. Now imagine that there are two individuals who have identical tastes. One of them earns income of $40,000 this year and $22,000 next year, so real income (nominal income divided by the price level) is therefore 4,000 this year and 2,200 next year. The other person earns $20,000 this year and $44,000 next year, yielding a real income of 2,000 this year and 4,400 next year. Both of these individuals share the same budget line (see Figure 4.6 "The Timing of Income"). This is because the discounted present value of their nominal income is the same: $60,000. (Check to make sure you understand why this is true.) For example, suppose that the preferred point of both individuals is to consume the same amount in each year. Then they can both consume approximately 3,143 units of consumption in each period.

Let us see how this works. Because the price level is $10, this amount of consumption costs $31,430 in each period. The first individual takes her income of $40,000 and saves $8,570 by putting it in the bank. This saving earns 10 percent interest, so she gets an additional $857. She thus has income in the following year equal to $22,000 + $8,570 + $857 = $31,427. This allows her to buy 3,142.7 units of consumption goods. The second individual needs to borrow $11,430 to add to his income this year. Next year, he must repay this amount plus 10 percent interest (that is, another $1,143). So his income next year is $44,000 − $11,430 − $1,143 = $31,427. So one individual must save to reach her preferred consumption bundle, while the other must borrow to reach his. Yet because they have the same discounted present value of income and the same tastes, they will consume the same bundle of goods.

Keep in mind that our discussion so far ignores uncertainty. We assumed that both individuals know their current and future income with certainty. Just as importantly, we have supposed that a bank is confident that the borrower will have sufficient income next year to repay the loan. In a world of uncertainty, we do not know for sure how much money we will have next year, and lenders worry about the possibility that people might not make good on their loans. Later in the chapter, we explain more about decision making in an uncertain world.

Lifetime Budget Constraint

So far, we have worked everything out in terms of a two-period example. The two-period budget constraint tells us how income and consumption are linked over time. In reality, of course, you make these decisions with longer time horizons, and you can save or borrow for multiple years. But the same fundamental insight holds. If you save this year, then you will have extra resources to spend at some future date. If you borrow this year, then you will have to repay that loan sometime in the future, at which time you will have fewer resources to spend.

Toolkit: Section 17.5 "Discounted Present Value"

Individuals face a lifetime budget constraint. They can save in some periods of their lives and borrow (not save) in other periods. Over the course of any individual's lifetime, however, income and consumption spending must balance. (If you begin life with some assets [for example, a bequest], we count this as part of income. If you leave a bequest when you die, we count this as part of consumption.) The lifetime budget constraint is as follows:

discounted present value of lifetime consumption = discounted present value of lifetime income.

Again, it is important to be consistent in calculating the discounted present values in this expression. We have written the equation in terms of (real) consumption and (real) income, which means that the real interest factor must be used for discounting. An alternative is to measure both consumption and income in nominal terms and then use the nominal interest factor for discounting. There is a useful special case where real interest rates are zero, in which case it *is* legitimate simply to add income and consumption in different years. Thus the lifetime budget constraint becomes

total lifetime consumption = total lifetime income.

Although the principles of decision making are the same whether we are thinking about 2 months, 2 years, or an entire lifetime, it is obviously harder to make decisions over a 30-year horizon than over a 30-day horizon. One reason is that, over longer time horizons, we are more likely to face uncertainty. We don't know what our income will be 30 years from now, and we don't know our tastes. But even without that uncertainty, we may not always make good decisions.

In particular, economists and psychologists have discovered that we do not view choices involving the near future the same way as we view distant choices. For example, suppose an individual is given a choice between 1 cookie today or 2 cookies tomorrow. If he is impatient (or hungry), he is likely to choose the single cookie today. But if the same individual is given a choice between 1 cookie in 30 days or 2 cookies in 31 days, he or she may very well choose the 2-cookie option. Yet after 30 days have gone by, that person will be confronting the earlier decision, wishing that he or she could have the 2 cookies today.

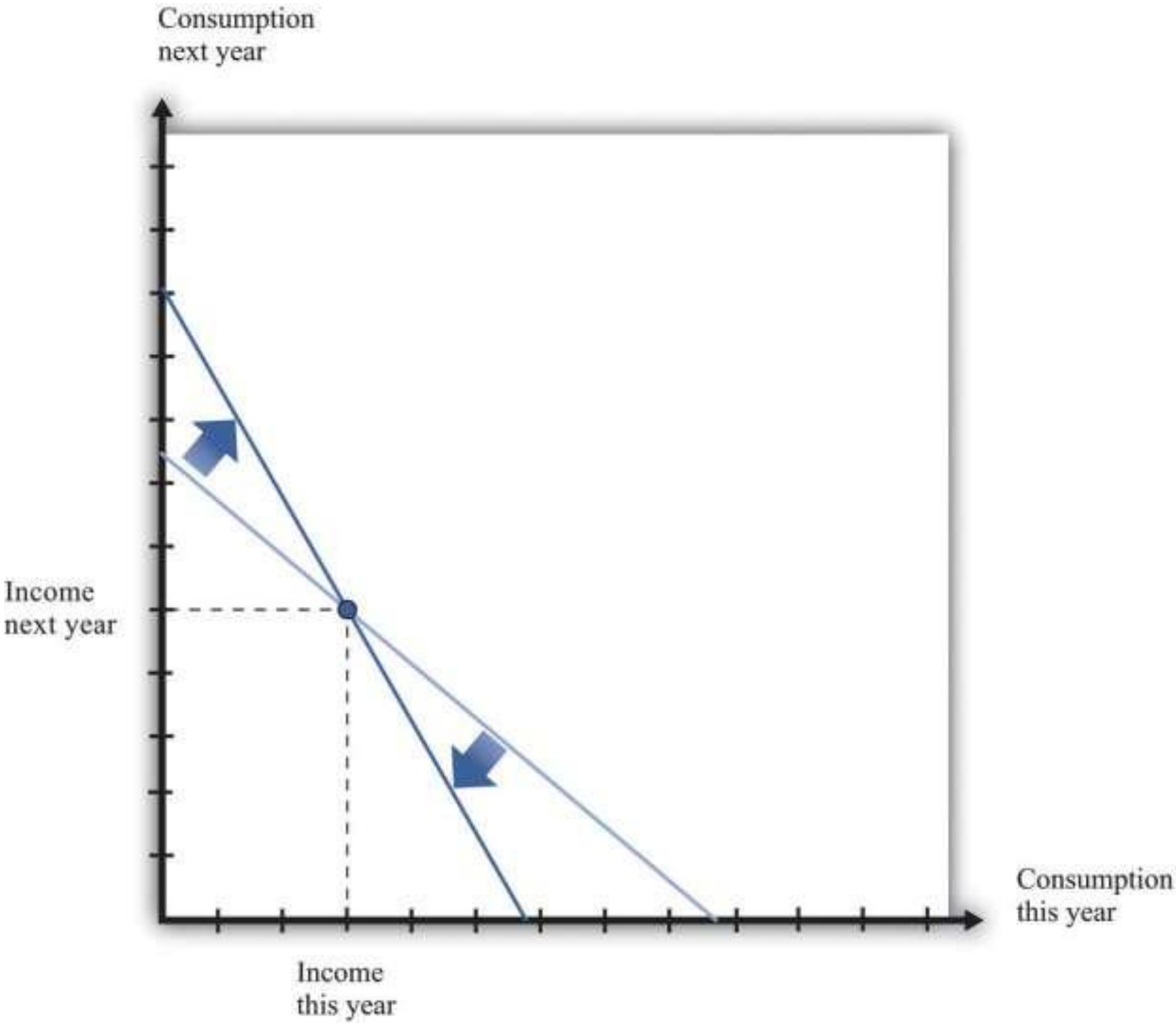
Another way of saying this is that our decisions are not always consistent over time: our future selves may wish that our current selves had displayed more self-control. For example, we may choose to consume a lot today—instead of saving—and then regret that decision when we are older. Indeed, people often engage in tricks to get around their lack of self-control. For example, some people have a separate bank account for their savings, so they are less tempted to spend that money. Governments also take actions that compensate for our lack of self-control. Social security is in some ways a "forced saving" scheme: the government takes money from us when we are working but pays us money when we are retired.

There is a new and exciting field of economics called "neuroeconomics" that tries to understand the processes in the brain that underlie economic decision making. This field, while still very much in its infancy, promises to help us understand why the economic theory of how we make choices often works well, and why it sometimes does not. Some recent research suggests that different brain processes may deliver conflicting messages when making choices over time. Some processes are deliberative, in line with the economic model, while others are more impulsive. It is likely that the next two decades will bring a much deeper understanding of how the brain makes decisions, perhaps leading to a richer theory of economic decision making.

Changes in Interest Rates

Whenever the real interest rate changes, then the *relative price* of consumption this year and next year changes. As we already know, changes in the real interest rate can come from two different sources: changes in the nominal interest rate and changes in the inflation rate. (Look back at the Fisher equation for a reminder of this.) Figure 4.7 "An Increase in the Real Interest Rate" shows the effect of an increase in the real interest rate on your budget line. The budget line becomes steeper because the opportunity cost of consumption this year increases. Notice, though, that the point at which you just consume your income in each period is still on the budget line. This is the point at which you are neither saving nor borrowing. Thus no matter what the interest rate, this point is always available to you.

Figure 4.7 An Increase in the Real Interest Rate



A change in the real interest rate changes the slope of the budget line. At anyreal interest rate, however, it is possible to consume exactly one's income. So the point corresponding to no saving or borrowing is always available, no matter what the real interest rate. An increase in the real interest rate therefore causes the budget line to become steeper and rotate through the income point.

Changes in relative prices lead to income and substitution effects. To understand the effect of an increase in the real interest rate, we must look at both effects.

- **Substitution effect.** An increase in the real interest rate makes consumption next year look more attractive relative to consumption this year. This encourages saving and discourages borrowing.
- **Income effect.** The income effect is somewhat more complicated. Look again at Figure 4.7 "An Increase in the Real Interest Rate". *Savers* are made better off by an increase in interest rates because, to the left of the income point, the new budget line lies outside the old budget set. This encourages savers to increase consumption this year and next year. Because income this year hasn't changed, but there is an incentive to consume more this year, we can see that the income effect discourages saving. *Borrowers*, meanwhile, are worse off by the increase in interest rates. To the right of the income point, the new budget line lies inside the old budget set. This encourages borrowers to decrease consumption this year and next year. This incentive to consume less tells us that the income effect discourages borrowing. In sum, the income effect gives an incentive for savers to save less and for borrowers to borrow less.

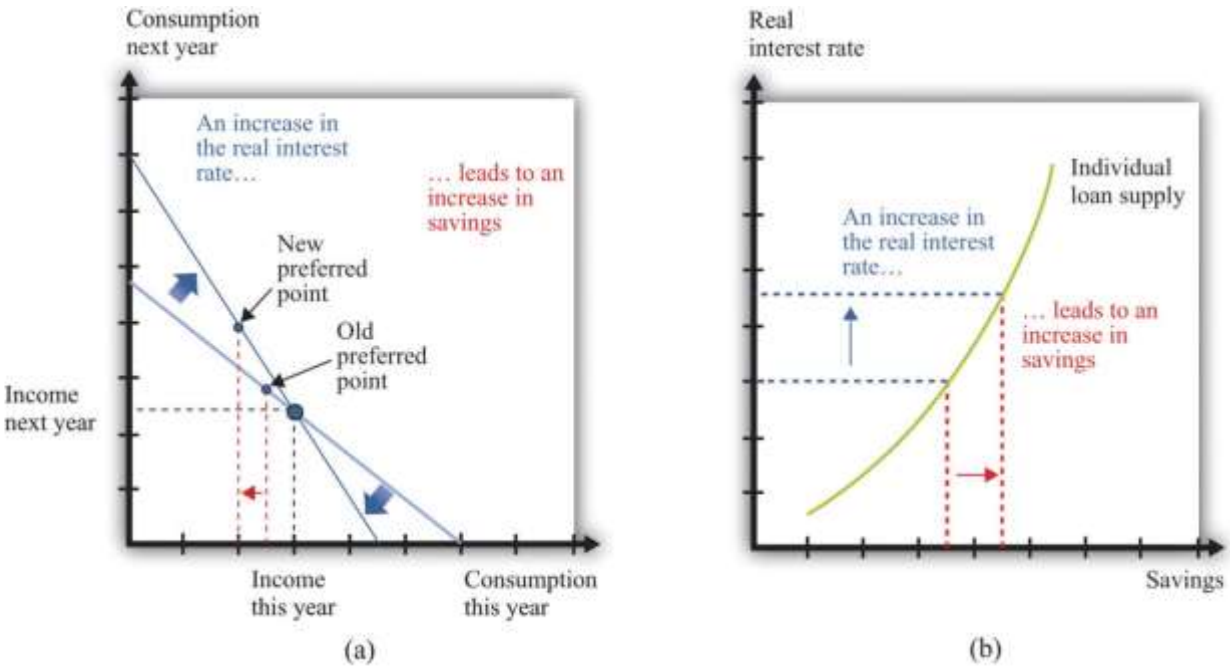
Combining the income and substitution effects and following an increase in the interest rate, borrowers have an incentive to borrow less. The substitution effect encourages saving, while the income effect discourages saving. The overall effect is ambiguous.

The evidence suggests that most people are like the individual in Figure 4.8 "Individual Loan Supply". For this person, the substitution effect dominates: the amount of saving increases as the real interest rate increases. Because an individual's savings represent funds that can be lent out to others in the economy, we call them the individual loan supply.

Toolkit: Section 17.6 "The Credit Market"

Individual loan supply is the amount of saving carried out by an individual at different values of the real interest rate. It is illustrated in a diagram with the real interest rate on the vertical axis and the supply of loans on the horizontal axis.

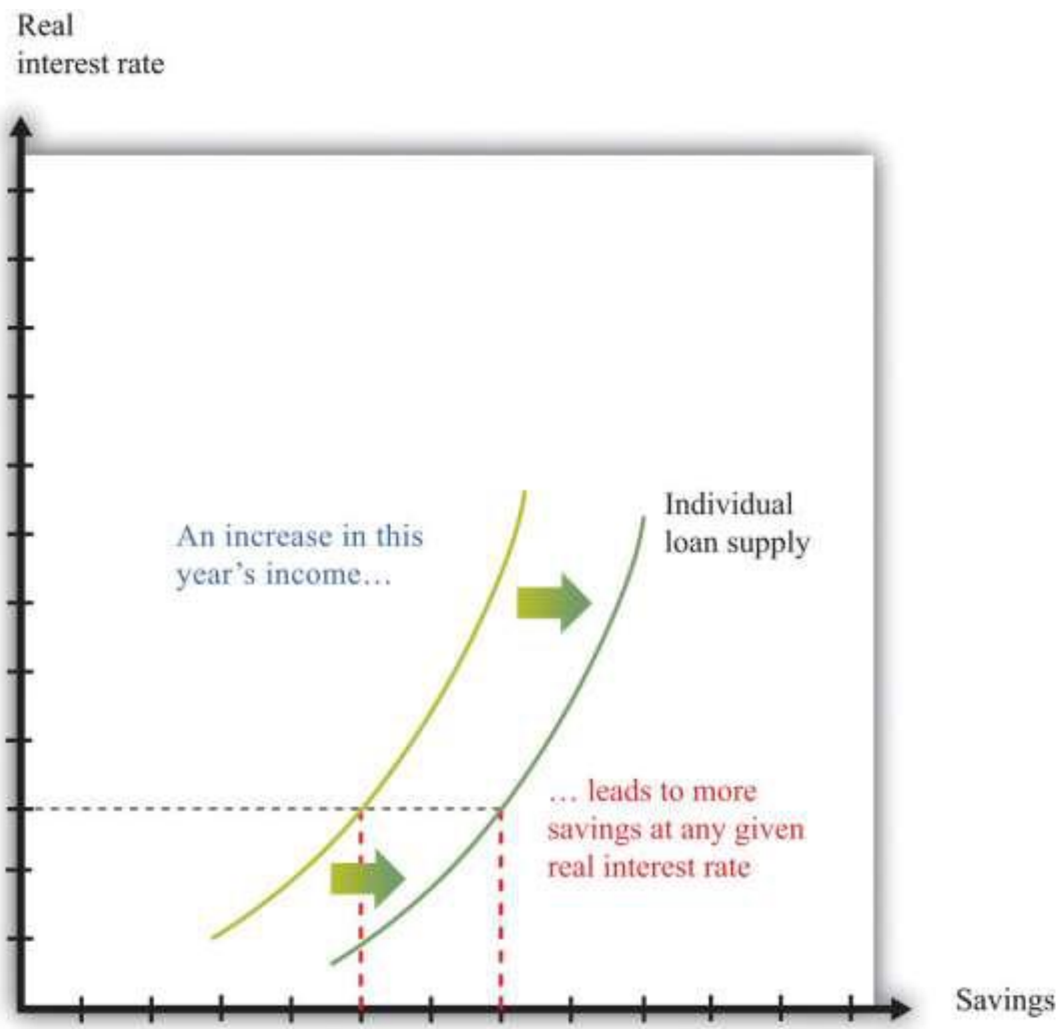
Figure 4.8 *Individual Loan Supply*



For savers in the economy, the effects of an increase in the real interest rate are ambiguous. The substitution effect encourages saving, but the income effect discourages saving. The evidence suggests that, on balance, the substitution effect dominates, so that savings increase. (a) In this two-period diagram, an increase in interest rates causes consumption this year to decrease. Because income this year is unchanged, savings increases. (b) The same diagram is applied to an individual supply of loans.

As the real interest rate changes, the response of individual saving is a*movement along* the loan supply curve. What might cause the whole curve to shift? If an individual has a higher income in the current year, this will cause the budget line to shift outward, and the person will consume more goods in the current year *and* more goods in the future. To consume more in the future, the person will have to save more. In this case, the supply of savings shifts outward as current income increases. This is shown in Figure 4.9 "A Shift in an Individual's Supply of Savings".

Figure 4.9 *A Shift in an Individual's Supply of Savings*



An increase in this year's income means that an individual will save more at any given interest rate. This means that the loan supply curve for the individual shifts outward.

KEY TAKEAWAYS

- Over the course of an individual's lifetime, the discounted present value of spending equals the discounted present value of income.
- Households save to consume more in the future.
- Unless the interest rate is zero, a dollar today does not have the same value as a dollar tomorrow.

- The nominal interest rate is expressed in dollar terms, while the real interest rate is expressed in terms of goods and services. Economists think that households and firms make decisions on the basis of real interest rates.

CHECKING YOUR UNDERSTANDING

1. Fill in the missing values in Table 4.1 "Discounted Present Value of Income".
2. If the interest rate increases, what will happen to the amount saved by a household? How does this answer depend on whether the household is a lender (saving is positive) or a borrower (saving is negative)?

[1] There are actually many different interest rates in an economy. Chapter 9 "Making and Losing Money on Wall Street" looks at some of these. Here, we simplify the process by supposing there is only one interest rate.

[2] The precise formula is as follows:

$$1 + \text{real interest rate} = \text{price this year}/\text{price next year} \times (1+\text{nominal interest rate}) = 1+\text{nominal interest rate}/1+\text{inflation rate}.$$

This equation is, to a very good approximation, the same as the one in the text.

4.2 Using Discounted Present Values

LEARNING OBJECTIVES

1. When should you use the tool of discounted present value?
2. How does an increase in the interest rate affect the discounted present value of a flow of income?

Section 4.1 "Consumption and Saving" introduced a valuable technique called discounted present value. You can use this technique whenever you need to compare flows of goods, services, or currencies (such as dollars) in different periods of time. In this section, we look at some of the big decisions you make during your life, both to illustrate discounted present value in action and to show how a good understanding of this idea can help you make better decisions.

Choosing a Career

A decision you typically make around the time that you graduate is your choice of a career. What makes the choice of a career so consequential is the fact that it can be very costly to switch from one career to another. For example, if you have trained as an engineer and then decide you want to be a lawyer, you will have to give up your engineering job (and give up your salary as well) and go to law school instead.

Suppose you are choosing among three careers: a lawyer, an insurance salesperson, or a barista. To make matters simple, we will work out an example with only two years. Table 4.2 "Which Career Should You Choose?" shows your earnings in each year at each occupation. In the first year in your career as a lawyer, we suppose that you work as a clerk, not earning very much. In the second year, you join a law firm and enjoy much higher pay. Selling insurance pays better than the legal career in the first year but worse in the second year. Working as a barista pays less than selling insurance in both years.

Table 4.2 Which Career Should You Choose?

Career	First-Year Income ($)	Second-Year Income ($)
Lawyer	5,000	60,000
Insurance salesperson	27,000	36,000
Barista	18,000	20,000

It is obvious that, if you care about the financial aspect of your career, you should not be a barista. (You would choose that career only if it had other benefits—such as flexible working hours and lack of stress—

that outweighed the financial penalty.) It is less obvious whether it is better financially to work as a lawyer or as an insurance salesperson. Over the two years, you earn $65,000 as a lawyer and $63,000 as an insurance seller. But as we have already explained, simply adding your income for the two years is incorrect. The high salary you earn as a lawyer comes mostly in the second year and must be discounted back to the present.

To properly compare these careers, you should use the tool of discounted present value. With this tool, you can compare the income flows from the different occupations. Table 4.3 "Comparing Discounted Present Values of Different Income Streams" shows the discounted present value of the two-year flow of income for each career, assuming a 5 percent interest rate (that is, an interest factor equal to 1.05). Look, for example, at the lawyer's income stream:

discounted present value of income as a lawyer = $5,000 + $60,0001.05=$62.143.

Similarly, the discounted present value of the income stream is $61,286 for the insurance salesperson and $37,048 for the barista. So if you are choosing your career on the basis of the discounted present value of your income stream, you should pick a career as a lawyer.

Table 4.3 Comparing Discounted Present Values of Different Income Streams

Career	First-Year Income ($)	Second-Year Income ($)	Discounted Present Value at 5% Interest Rate ($)
Lawyer	5,000	60,000	62,143
Insurance salesperson	27,000	36,000	61,286
Barista	18,000	20,000	37,048

This conclusion, however, depends on the interest rate used for discounting.Table 4.4 "Discounted Present Values with Different Interest Rates" adds another column, showing the discounted present values when the interest rate is 10 percent. You can see two things from this table: (1) The higher interest rate reduces the discounted present value for all three professions. If the interest rate increases, then future income is less valuable in present value terms. (2) The higher interest rate reverses our conclusion about which career is better. Selling insurance now looks better than being a lawyer because most of the lawyer's earnings come in the future, so the discounting has a bigger effect.

Table 4.4 Discounted Present Values with Different Interest Rates

Career	First-Year Income ($)	Second-Year Income($)	Discounted Present Value at 5% Interest Rate ($)	Discounted Present Value at 10% Interest Rate ($)
Lawyer	5,000	60,000	62,143	59,545
Insurance salesperson	27,000	36,000	61,286	59,727
Barista	18,000	20,000	37,048	36,182

Of course, what you might really like to do is to sell insurance for the first year and work as a lawyer in the second year. This evidently would have higher income. Sadly, it is not possible: it is almost impossible to qualify for a high-paying lawyer's job without investing a year as a law clerk first. Changing occupation can be very costly or even impossible if you don't have the right skills. So choosing a career path means you must look ahead.

Going to College

If you are like most readers of this book, you have already made at least one very important decision in your life. You have chosen to go to college rather than taking a job immediately after graduating from high school. Ignoring the pleasures of going to college—and there are many—there are direct financial costs and benefits of a college education.

Think back to when you were deciding whether to go to college or to start work immediately. To keep our example from being too complicated, we again look at a two-year decision. What if you could obtain a college degree in one year, at a tuition cost of $13,000, and the interest rate is 5 percent annually? Your earnings are presented in Table 4.5 "Income from Going to College versus Taking a Job". In your year at college, you would earn no income, and you have to pay the tuition fee. In the following year, imagine that you can earn $62,143 working as a lawyer. Alternatively, you could bypass college and go to work as a barista, earning $10,000 in the first year and $37,048 in the second year. (We are assuming, as before, that you know these figures with certainty when you are making your decision.)

Table 4.5 Income from Going to College versus Taking a Job

Career	Income in the Year at College ($)	Income in the Year after College ($)
College	−13,000	62,143

Career	Income in the Year at College ($)	Income in the Year after College ($)
Barista	10,000	37,048

Going to college is an example of an *investment decision*. You incur a cost in the year when you go to college, and then you get a benefit in the future. There are two costs of going to college: (1) the $13,000 tuition you must pay (this is what you probably think of first when considering the cost of going to college) and (2) the *opportunity cost* of the income you could have earned while working. In our example, this is $10,000. The explicit cost and the opportunity cost together total $23,000, which is what it costs you to go to college instead of working in the first year.

By the way, we do *not* think about living expenses as a cost of going to college. You have to pay for food and accommodation whether you are at college or working. Of course, if these living expenses are different under the two scenarios, then you should take this into account. For example, if your prospective college is in New York City and has higher rental costs than in the city where you would work, then the *difference* in the rent should be counted as another cost of college.

The benefit of going to college is the higher future income that you enjoy. In our example, you will earn $62,143 in the following year if you go to college, and $37,048 if you do not. The difference between these is the benefit of going to college: $62,143 − $37,048 = $25,095. Even though this is greater than the $23,000 cost of going to college, we cannot yet conclude that going to college is a good idea. We have to calculate the discounted present value of this benefit. Suppose, as before, that the interest rate is 5 percent. Then

discounted present value of gain from college = $25,0951.05=$23,900.

We can conclude that, with these numbers, going to college is a good investment. It is worth $900 more in discounted present value terms.

We could obtain this same conclusion another way. We could calculate the discounted value of the two-year income stream for the case of college versus barista, as in Table 4.6 "Income Streams from Going to College versus Taking a Job". We see that the discounted value of the income stream if you go to college is $46,184, compared to $45,284 if you work as a barista. The difference between these two is $900, just as before.

Table 4.6 Income Streams from Going to College versus Taking a Job

Career	Income in the Year at College ($)	Income in the Year after College ($)	Discounted Present Value at 5% Interest Rate ($)
College	−13,000	62,143	46,184
Barista	10,000	37,048	45,284

You might have noticed that the figures we chose as "income in the year after college" in Table 4.5 "Income from Going to College versus Taking a Job" are the same as the numbers that we calculated in Table 4.3 "Comparing Discounted Present Values of Different Income Streams". The numbers in Table 4.3 "Comparing Discounted Present Values of Different Income Streams" were themselves the result of a discounted present value calculation: they were the discounted present value of a two-year income stream. When we compare going to college with being a barista, we are therefore calculating a discounted present value of something that is already a discounted present value. What is going on?

To understand this, suppose you are deciding about whether to go to college in 2012. If you do go to college, then in 2013 you will decide whether to be a lawyer, an insurance salesman, or a barista. If you decide on the legal career, then you will be a law clerk in 2013, and you will earn the high legal salary in 2014. Our analysis in Table 4.3 "Comparing Discounted Present Values of Different Income Streams" is therefore about the choice you make in 2013, thinking about your income in 2013 and 2014. Table 4.3 "Comparing Discounted Present Values of Different Income Streams" gives us the discounted present value *in 2013* for each choice. If we then take those discounted present values and use them as "income in the year after college," as in Table 4.6 "Income Streams from Going to College versus Taking a Job", we are in fact calculating the discounted present value, in 2012, of the flow of income you receive in 2012, 2013, and 2014.

If you think carefully about this, you will realize that

$$\text{discounted present value of income if you go to college} = -\$13,000+\frac{\$5,000}{1.05}+\frac{\frac{\$60,000}{1.05}}{1.05}=-\$13,000+\$4,762+\frac{\$57,143}{1.05}=\$46,184.$$

This is the same answer that we got before. As this example suggests, you can calculate discounted present values of long streams of income, including income you will receive many years in the future. [1]

Economists have worked hard to measure the return on investment from schooling: "Alan B. Krueger, an economics professor at Princeton, says the evidence suggests that, up to a point, an additional year of schooling is likely to raise an individual's earnings about 10 percent. For someone earning the national median household income of $42,000, an extra year of training could provide an additional $4,200 a year. Over the span of a career, that could easily add up to $30,000 or $40,000 of present value. If the year's education costs less than that, there is a net gain." [2] Notice several things from this passage. First, the gains from education appear as an increase in earnings *each year*. So even if a 10 percent increase in earnings does not seem like a lot, it can be substantial once these gains are added over one's lifetime. Second, Krueger is careful to use the term *present value*. Third, the number given is an average. Some people will benefit more; others will benefit less. Equally, some forms of schooling will generate larger income gains than others. Fourth, Krueger correctly notes that the present value must be compared with the cost of education, but you should remember that the cost of education includes the opportunity cost of lost income.

Table 4.7 "Return on Education" provides some more information on the financial benefits of schooling. [3] The table shows average income in 2004. There is again evidence of a substantial benefit from schooling. Male college graduates, on average, earned more than $21,000 (68 percent) more than high school graduates, and female college graduates earned more than $16,000 (78 percent) more than high school graduates. The table shows that women are paid considerably less than men and also that the return on education is higher for women.

The presence of such apparently large gains from education helps explain why economists often suggest that education is one of the most important ingredients for the development of poorer countries. (In poorer countries, we are often talking not about the benefit of going to college but about the benefit of more years of high school education.) Moreover, the benefits from education typically go beyond the benefits to the individuals who go to school or college. There are benefits to society as a whole as well.

However, you should be careful when interpreting numbers such as these. We cannot conclude that if you randomly selected some high school graduates and sent them to college, then their income would increase by $17,000. As we all understand, individuals *decide* whether to go to college. These decisions reflect many things, including general intelligence, the ability to apply oneself to a task, and so on. People who have more of those abilities are more likely to attend—and complete—college.

One last point: we conducted this entire discussion "ignoring the pleasures of going to college." But those pleasures belong in the calculations. Economics is about not only money but also all the things that make us happy. This is why we occasionally see people 60 years old or older in college. They attend not as an investment but simply because of the pleasure of learning. This is not inconsistent with economic reasoning or our discussion here. It is simply a reminder that your calculations should not only be financial but also include the all the nonmonetary things you care about.

The Effects of Interest Rates on Labor Supply

After you have decided whether or not to go to college and have chosen your career, you will still have plenty of decisions to make involving discounted present value. Remember that you have a lifetime budget constraint, in which the discounted present value of your income, including labor income, equals the discounted present value of consumption.

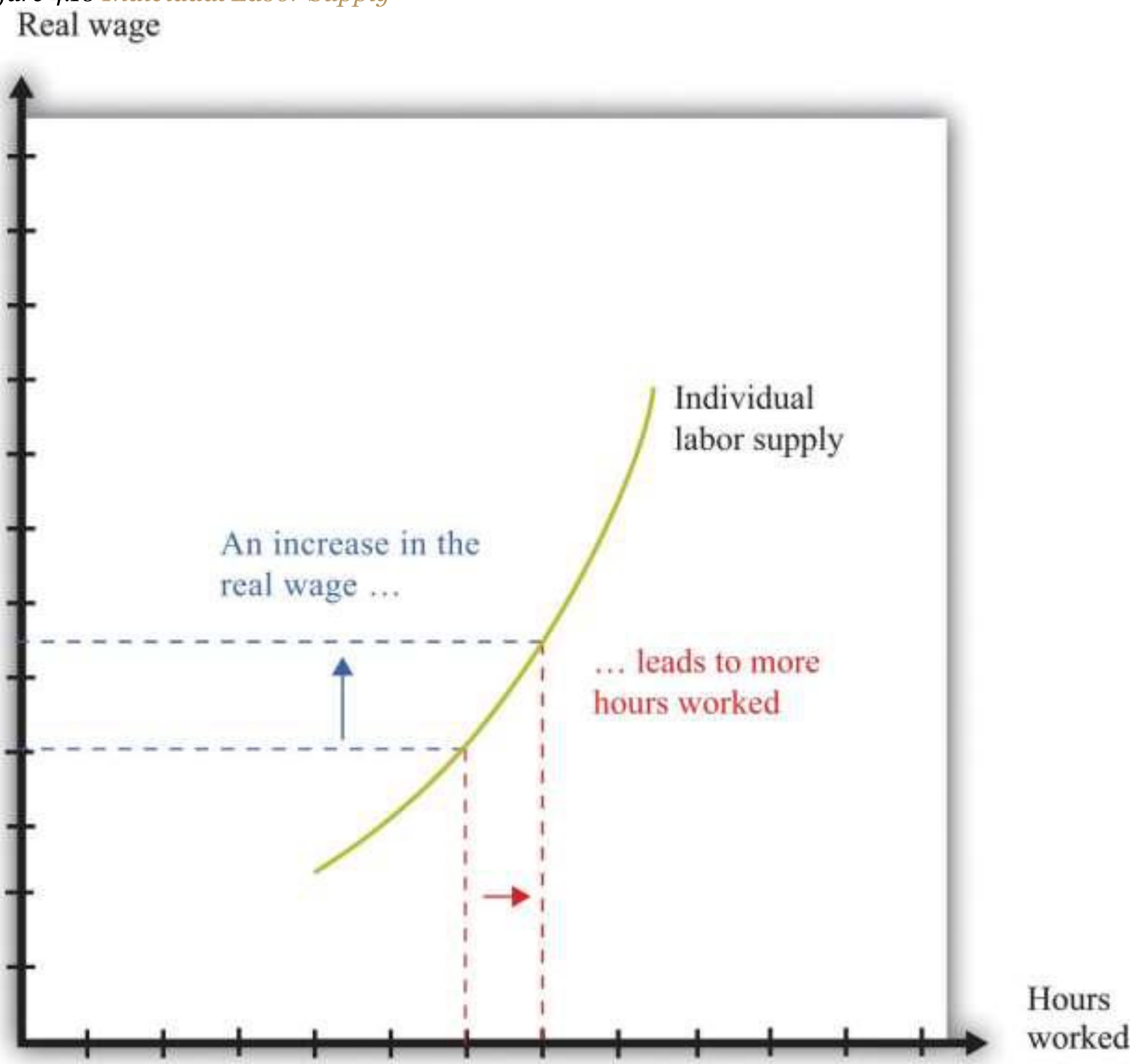
Your labor income is partly under your control. If you have some choice about how many hours to work, then your **individual labor supply curve** depends on the real wage, as shown in Figure 4.10 "Individual Labor Supply". [4] The labor supply curve illustrates the fact that as the real wage increases, you are likely to work more. Labor supply, like the supply of loans that we considered previously, is driven by substitution and income effects. As the real wage increases,

- the opportunity cost of leisure increases, so you have an incentive to work more (substitution effect); and
- you have more income to spend on everything, including leisure, so you have an incentive to consume more leisure and work less (income effect).

Toolkit: Section 17.3 "The Labor Market"

The toolkit contains more information if you want to review the labor supply curve.

Figure 4.10 Individual Labor Supply



An increase in the real wage encourages an individual to work more. The labor supply curve slopes upward.

When you are making a decision about how much to work over many periods of time, your choice is more complicated. How much you choose to work right now depends not only on the real wage today but also on the real wage in the future and on the real interest rate. This is because you work both today and in the future to earn the income that goes into your lifetime budget constraint. If you think it is likely to be harder to earn money in the future, then you will probably decide to work more today. If you think it will be easier to earn money in the future, then you might well decide to work less today.

It is easiest to see how this works with an example. Suppose you are a freelance construction worker in Florida in the aftermath of a hurricane. There is lots of work available, and construction firms are paying higher than usual wages. You realize that you can earn much more per hour of work right now compared to your likely wage a few months in the future. A natural response is to work harder now to take advantage of the unusually high wages.

We can understand your decision in terms of income and substitution effects. The higher wage leads to the usual substitution effect. But because the change in the wage is only temporary, and because you are thinking about your wages over your lifetime, it does not have a large income effect. In this case, therefore, we expect the substitution effect to strongly outweigh the income effect.

Interest rates may also influence your decision about how hard to work. If interest rates increase, then the gains from working today increase as well. If you save money at high interest rates, you can enjoy more consumption in the future. High interest rates, like temporarily high current wages, increase the return to working today compared to working in the future, so you are likely to work more today.

The Demand for Durable Goods

Some of the products we purchase, such as milk or a ticket to a football game, disappear as soon as they are consumed. Other goods last for a long time and are, in effect, consumed over and over. Some examples are a bicycle, a car, and a microwave oven.

Goods that last over many uses are called durable goods, while those that do not last very long are called nondurable goods. There is no hard-and-fast distinction between durable and nondurable goods. Many everyday items, such as plates, books, T-shirts, and downloaded music, are used multiple times. In economic statistics, however, the term *durable* is reserved for larger items that are bought only occasionally and that typically last for many years. Cars and kitchen appliances are classified as durable goods, but blue jeans and haircuts are not, even though they are not consumed all at once.

Because durable goods last a long time, making decisions about purchasing a durable good requires thinking about the future as well as the present. You must compare the benefits of the durable good *over its entire lifetime* relative to the cost you incur to pay for the durable now. A durable good purchase is typically a **unit demand** decision—you buy either a single unit or nothing. For unit demand, your decision rule is simple: buy if your **valuation** of the good exceeds the price of the good. (Remember that your valuation is the maximum you would be willing to pay for the good.) In the case of durable goods, there is an extra twist: your valuation needs to be a discounted present value.

Toolkit: Section 17.1 "Individual Demand"

You can review valuation and unit demand in the toolkit.

The idea is that you obtain a *flow of services* from a durable good. You need to place a valuation on that flow for the entire lifetime of the durable good. Then you need to calculate the discounted present value of that flow of services. If this discounted present value exceeds the price of the good, you should purchase it.

Suppose you are thinking of buying a new car that you expect to last for 10 years. You need to place a valuation on the flow of services that you get from the car each year: for example, you might decide that you are willing to pay $3,000 each year for the benefit of owning the car. To keep life really simple, let us think about a situation where the real interest rate is zero; this is the special case where it *is* legitimate to add these flows. So the car is worth $30,000 (= $3,000 per year × 10 years) to you now. This means you should be willing to buy the car if it costs less than $30,000, and you should not buy it otherwise. [5]

If real interest rates are not zero, then spending on durable goods will depend on interest rates. As interest rates increase, the future benefits of the durable good become smaller, in terms of discounted present value. This means that durable goods become more expensive relative to nondurable goods. Thus the demand for durable goods decreases as interest rates increase.

One way to understand this is to realize that it is often easy to defer the purchase of a durable good. New durable goods are frequently bought to replace old goods that are wearing out. People buy new cars to replace their old cars or new washing machines to replace their old ones. If interest rates are high, you can often postpone such replacement purchases; you decide whether you can manage another year with your old car or leaking washing machine. As a result, spending on durable goods tends to be very sensitive to changes in interest rates.

These examples of discounted present value illustrate one key point: whenever you are making economics decisions about the future—be it what career to follow, when it is best to work hard, or if you should buy a new car—your decisions depend on the rate of interest. Whenever the rate of interest is high, future costs and benefits are substantially discounted and are therefore worth less in present value terms. High interest rates, in other words, mean that you put a lot of weight on the present relative to the future. When the rate of interest decreases, the future should play a larger role in your decisions.

KEY TAKEAWAYS

- You should use discounted present value whenever you need to compare flows of income and expenses in different periods of time.
- The higher the interest rate, the lower the discounted present value of a flow.

CHECKING YOUR UNDERSTANDING

1. List five goods that are durable and five services that are nondurable. Is it possible to have a service that is durable?
2. Calculate the appropriate values in Table 4.3 "Comparing Discounted Present Values of Different Income Streams" assuming an interest rate of 8 percent.
3. If interest rates are lower, would you expect people to invest more or less in their health?

[1] The toolkit gives a more general formula for calculating the discounted present value. (See the more formal presentation of discounted present value at the end of Chapter 4 "Life Decisions", Section 4.4 "Embracing Risk".)

[2] Anna Bernasek, "What's the Return on Education?" Economic View, Business Section,*New York Times*, December 11, 2005, accessed February 24, 2011,http://www.nytimes.com/2005/12/11/business/yourmoney/11view.html.

[3] US Census Bureau, "Income, Earnings, and Poverty from the American Community Survey," 2010, table 5, accessed February 24, 2011,http://www.census.gov/hhes/www/poverty/publications/acs-01.pdf.

[4] Chapter 3 "Everyday Decisions" explains the individual supply of labor.

[5] What if you think you might sell the car before it wears out? In this case, the value of the car has two components: (1) the flow of services you obtain while you own it, and (2) the price you can expect to obtain when you sell it. When you buy a durable good, you are purchasing an **asset**: something that yields some flow of benefits over time and that you can buy and sell. In general, the value of an asset depends on both the benefits that it

provides and the price at which it can be traded. We examine these ideas in much more detail in Chapter 9 "Making and Losing Money on Wall Street".

4.3 Avoiding Risk

LEARNING OBJECTIVES

1. What is the definition of probability?
2. How can I calculate an expected value?
3. What is risk aversion?
4. How do individuals and firms deal with risk?
5. How does insurance work?

In life, there are many uncertainties. So far, we have ignored them all, but you will have to face them. In our various discussions of discounted present value, we pretended that you knew your future income—and your future tastes—with certainty. In real life, we must decide how much to save without knowing for sure what our future income will be. We must pick a career without knowing how much we will enjoy different jobs or how much they will pay. We must decide whether or not to go to college without knowing what kind of job we will be able to get, and so on. How can we deal with all these uncertainties?

Some of the uncertainties we face are forced on us with no choice of our own, such as the following:

- Accidents involving you, your automobile, your house, and so on
- Layoffs resulting in spells of unemployment
- Your health

As you know, one way to deal with these uncertain events is through insurance. Insurance is a way of trying to remove some of the risk that we face. We explain how it works later in this section.

Other risks are more under our control. We accept jobs that entail certain risks. We drive our cars even though we know that there is a risk of accident. We put our savings into risky stocks rather than safe assets. In these cases, we trade off these risks against other benefits. We drive faster, accepting the greater risk of accident to save time. Or we take a risky job because it pays well.

There are yet other kinds of risk that we actually seek out rather than avoid. We play poker or bet on sporting events. We climb mountains, go skydiving, and engage in extreme sports. In these cases, the risks are apparently something good that we seek out, rather than something bad that we avoid.

Risk and Uncertainty

Let us begin by making sure we understand what risk and uncertainty mean. (Here we will use the terms more or less interchangeably, although people sometimes reserve the term *uncertainty* for cases where it is hard to quantify the risks that we face.) Probably the simplest example of risk is familiar to us all: the toss of a coin. Imagine flipping a coin five times. Each time, the outcome will be either a head or a tail. Table 4.8 "Coin-Flipping Experiment" shows an example of such an experiment. In this experiment, the outcome was three heads and two tails. For each flip of the coin, there was uncertainty about the outcome. We did not know ahead of time whether there would be heads or tails. The outcome reported in Table 4.8 "Coin-Flipping Experiment" is only one example. If you were to carry out this experiment right now, you would almost certainly end up with a different outcome.

Table 4.8 Coin-Flipping Experiment

Flip	1	2	3	4	5
Outcome	Heads	Heads	Tails	Heads	Tails

Coin tosses are special because the flips of the coin are *independent* of each other (that is, the history of previous tosses has no effect on the current toss of a coin). In Table 4.8 "Coin-Flipping Experiment", the coin was not more likely to come up tails on the third toss because the previous tosses were both heads. Even if you have 100 heads in a row, this does not affect the outcome of the 101st toss of the coin. If you think that the coin is "fair," meaning that heads and tails are equally likely, then the 101st toss is still just as likely to be heads as tails. By contrast, the likelihood that it will be raining an hour from now is not independent of whether or not it is raining at this moment.

Financial Risk and Expected Value

Some of the risks that we confront are nonfinancial. An example of nonfinancial uncertainty is the risk that you might break your ankle playing basketball or the possibility that your favorite sporting team will win a big game and make you happy. Here, we will focus on financial uncertainty, by which we mean situations where there is money at stake. In other words, we are thinking about risks where you can measure the implications in monetary terms. An obvious example is the money you could win or lose from buying a lottery ticket or playing poker. Another is the money you would have to pay for repairs or medical expenses following a car accident. Another is the gains or losses from buying stocks, government bonds, or other financial assets. Another is the income you would lose if you were laid off from your job.

When we evaluate risky situations, we must have a way of describing the kinds of gambles that we confront. In general, we do this by listing all the possible outcomes together with the likelihood of each outcome. For example, Table 4.9 "Outcomes and Probabilities from a Coin Toss" lists the outcomes and the probability (that is, the likelihood of each outcome) for the experiment of tossing a coin one time.

Table 4.9 Outcomes and Probabilities from a Coin Toss

Outcome	Probability
Heads	12
Tails	12

Toolkit: Section 17.7 "Expected Value"

Probability is the percentage chance that something will occur. For example, there is a 50 percent chance that a tossed coin will come up heads. We say that the probability of getting the *outcome* "heads" is 0.5. There are five things to know about probability:

1. The list of outcomes must be complete.
2. The list of outcomes must not overlap.
3. If an outcome is certain to occur, it has probability of 1.
4. If an outcome is certain *not* to occur, it has probability of 0.
5. If we add the probabilities for all the possible outcomes, the total must equal 1.

Think about rolling a normal six-sided die one time and describing outcomes and probabilities.

- We must make sure that we include every outcome. We cannot list as possible outcomes "less than or equal to 2" and "greater than or equal to 4." Such a list ignores the possibility of rolling a 3.
- We cannot list as possible outcomes "less than or equal to 4" and "greater than or equal to 3." These categories overlap because a roll of a 3 or a 4 would show up in both categories.
- The outcome "less than or equal to 6" has a probability of 1 because it is certain.
- The outcome "9" has a probability of 0.

- Provided we have a complete list of outcomes (for example "less than or equal to 4" and "greater than or equal to 5"), the probabilities of all the outcomes will always sum to 1. (In this case, the probability of the first outcome is 2/3, and the probability of the second outcome is 1/3.)

Now suppose you are playing a gambling game based on a toss of a coin. If the coin comes up heads, you win $1. If it comes up tails, you win $0. When we look at a situation such as this, we are often interested in how much you would get, on average, if you played the game many times. In this example, it is easy to guess the answer. On average, you would expect to win half the time, so half the time you get $1, and half the time you get nothing. We say that the expected valueof each flip of the coin is 50 cents.

Toolkit: Section 17.7 "Expected Value"

The expected value of a situation with financial risk is the measure of how much you would expect to win (or lose) on average, if the situation were to be replayed a large number of times. Expected value is calculated as follows:

1. For each outcome, multiply the probability of that outcome by the amount you will receive.
2. Add these amounts over all the possible outcomes.

Table 4.10 "Outcomes and Probabilities from Investment in Internet Venture"gives another example of expected value. Suppose a friend is planning on establishing a small Internet business and asks you to invest $1,000. He tells you (and you believe him) that there is a 50 percent chance that the business will fail, so you will lose your money. There is a 40 percent chance that the business will just break even, so you will get your $1,000 back but nothing more. And there is a 10 percent chance that the business will be very successful, so you will earn $16,000.

Table 4.10 Outcomes and Probabilities from Investment in Internet Venture

Outcome	Probability	Amount You Will Receive ($)
Failure	0.5	0
Break even	0.4	1,000
Success	0.1	16,000

In this case, the expected value of the investment is given by the following:

expected value = (0.5 × \$0) + (0.4 × \$1,000) + (0.1 × \$16,000) = \$2,000.

Thus for your investment of \$1,000, you could expect to get \$2,000 back on average. This seems like a good investment. It is important to remember, though, what "on average" means. You will never actually *get* \$2,000. You will receive either \$16,000, \$1,000, or nothing. Even though this is a good investment on average, you might still decide that you don't want any part of it. Yes, you might get the big net gain of \$15,000. But there's also a 50 percent chance that you will be out \$1,000. The gamble might seem too risky for you.

Coin tosses are special because it is relatively easy to determine the probability of a head or a tail. This is not the case for all the types of uncertainty you might face. In some cases, financial instruments—such as the mortgage-backed securities that played a big role in the financial crisis of 2007–2009—are so complex that investors find it difficult to assess the probabilities of various outcomes.

We often do a bad job of *estimating* probabilities. One reason for this is because we are unduly influenced by things that we can easily bring to mind. Psychologists call this the "availability heuristic." For example, we tend to overestimate certain causes of death, such as car accidents, tornadoes, and homicides, and underestimate others, such as diabetes, stroke, and asthma. [1]We also often do a poor job at *using* probabilities; in particular, we often put too much emphasis on small probabilities. For example, consider two drugs that are equally effective in treating a disease, but suppose the older drug has a 1 in 10 million chance of having a certain side effect and the newer drug has a 1 in a 100 million chance of having the same side effect. Consumers might view the new drug as much more appealing, even though the side effect was already highly improbable with the older drug.

Diversification of Risk

In many cases, we would like to find some way of getting rid of—at least to some degree—the risks that we face. One way we eliminate risk is through insurance. Sometimes we purchase insurance on the market. Sometimes our employer provides us with insurance. Sometimes the government provides us with insurance. In the following subsections, we look at many different kinds of insurance, including property insurance, unemployment insurance, and deposit insurance. [2]

First, though, we need to understand how and why insurance works. Suppose you have a bicycle worth \$1,000, and (for some reason) you cannot purchase insurance. You think that, in any given year, there is about a 1 percent chance that your bike will have to be replaced (because it is either stolen or written off in an accident).

Now, in expected value terms, this may not look too bad. Your expected loss from an accident is \$0.01 × \$1,000 = \$10. So on average, you can expect to lose \$10 a year. But the problem is that, if you are unlucky, you are stuck with a very big expense. Most of us dislike this kind of risk.

You are complaining about this to a friend, and she sympathizes, saying that she faces exactly (and we mean exactly) the same problem. She also has a bike worth \$1,000 and thinks there is a 1 percent chance each year that she will need to replace it. And that's when you have the brilliant idea. *You can make an agreement that, if either one of you has to replace your bikes, you will share the costs*. So if you have to replace your bike, she will pay \$500 of your costs, and if she has to replace her bike, then you will pay \$500 of her costs. It is (almost exactly) twice as likely that you will have to pay *something*, but if you do, you will only have to pay half as much. With this scheme, your expected loss is unchanged. But you and your friend prefer this scheme because it is less risky; it is much less likely that you will have to make the big \$1,000 payout.

We are implicitly assuming here that your chances of having to replace your bike are independent of the chance that she will have to replace her bike. (If you are likely to crash into her, or both of your bikes are stolen, then it is a different story.) There is also still a chance that you will both experience the unlucky 1-in-100 chance, in which case you would both still have to pay \$1,000. But the likelihood of this happening is now tiny. (To be precise, the probability of both of you having an accident in the same year is 1 in 10,000 [that is, 0.0001]). This is because the probability that two independent events occur equals the probability of one multiplied by the probability of the other.)

But why stop here? If you can find two more friends with the same problem, then you can make it almost certain that you will have to pay out no more than \$250. It is true that you would be even more likely to have to make a payment because you will have to pay if you or one of your friends has to replace his or her bike. But because the payment is now being shared four ways, you will have to pay only 25 percent of the expenses. This is an example of diversification, which is the insight that underlies insurance: people share their risks, so it is less likely that any single individual will face a large loss.

Diversification and insurance don't prevent bad stuff from happening. We live in a world where bicycles are stolen; where houses are destroyed by floods, fires, or storms; where people have accidents or become ill; and so on. There is not a lot we can do about the fact that bad things happen. But we can make the consequences of these bad things easier to deal with. Insurance is a means of sharing—diversifying—these risks.

Continuing with our bicycle insurance example, suppose you could find thousands of friends who would agree to be part of this arrangement. As more and more people join the scheme, it becomes increasingly likely that you have to make a payment each year, but the amount you would have to pay becomes smaller and smaller. With a very large number of people, you would end up very close to a situation where you pay out $10 with certainty each year. Of course, organizing thousands of your friends into such a scheme would present all sorts of practical problems. This is where insurance companies come in.

Insurance companies charge you a premium (an annual payment). In return, they promise to pay you an indemnity in the event you suffer a loss. The indemnity is usually not the full amount of the loss. The part of the loss that is not covered is called the deductible. In our example, there is no deductible, and the indemnity is $1,000. An insurance company would charge you a premium equal to the expected loss of $10 plus a little extra. The extra payment is how the insurance company makes money. You and everyone else are willing to pay this extra amount in return for the removal of risk.

The idea of diversification can also be applied to investment. [3] Think back to our example of your friend with the Internet venture. You might not want to invest $1,000 in his scheme because it seems too risky. But if you had 100 friends with 100 similar (but independent) schemes, you might be willing to invest $10 in each. Again, you would be diversifying your risk.

Risk Aversion

The preceding discussion of insurance and diversification is based on the presumption that people typically wish to avoid risk whenever possible. In our example, you have a 1 percent chance of suffering a

$1,000 loss. Your expected loss is therefore $10. Now imagine we give you a choice between this gamble and a certain loss of $10. If you are just as happy in either case, then we say you are risk-neutral. But if you are like most people, then you would prefer a certain loss of $10 to the gamble whereby you have a 1 percent chance of losing $1,000. In that case, you are risk-averse.

Toolkit: Section 17.7 "Expected Value"

Suppose you are presented with the following gamble:

- Lose nothing (99 percent probability)
- Lose $1,000 (1 percent probability)

How much would you pay to avoid this gamble? If you are risk-neutral, you would be willing to pay only $10, which is the expected loss. If you are risk-averse, you would be willing to pay more than $10. The more risk-averse you are, the more you would be willing to pay.

It is risk aversion that allows insurance companies to make money. Risk-averse people prefer a sure thing to a gamble that has the same expected value. In fact, they will prefer the sure thing to a gamble with a slightly lower expected value. Because it can diversify risk, the insurance company cares only about the expected value. Thus an insurance company behaves as if it were risk-neutral.

Different Kinds of Insurance

There are many different kinds of insurance available to you. We briefly discuss some of the most important.

Property Insurance

Many forms of property are insured: houses, cars, boats, the contents of your apartment, and so on. Indeed, some insurance is often mandatory. People purchase insurance because there are risks associated with owning property. Houses burn down, cars are stolen, and boats are wrecked in storms. In an abstract sense, these risks are just like a coin flip: heads means nothing happens; tails means there is a fire, a robbery, or a storm.

Let us look at home insurance in more detail. Suppose you own a house that is worth $120,000. You might pay $1,000 per year as a premium for an insurance policy. If your house burns down, then the insurance company will pay you some money to recover *part* of the loss. If the deductible on the policy is $20,000, you would receive an indemnity of $100,000. You lose $20,000 when the house burns down because the insurance company does not fully cover your loss.

Thus, if your house burns down, the insurance company loses the indemnity minus the premium—a total of $99,000. You lose the deductible and the premium—a total of $21,000. Your joint loss is $120,000—the lost value of the house. This serves to remind us again that insurance is not some magic way of preventing bad things from happening. When the house does not burn down, the insurance company earns the $1,000 premium, and you pay the $1,000 premium. Your joint loss is zero in this case.

You may wonder why insurance companies typically insist on a deductible as part of an insurance contract. After all, you would probably prefer to be covered for the entire loss. Deductibles exist because insurance policies can have the effect of altering how people behave. We have assumed that the probability of a bad thing happening was completely random. But if you are fully insured, you might not be so careful about how you look after your house. You might worry less about turning off the stove, ensuring that you have put out the fire in the fireplace, falling asleep while smoking, and so on. Deductibles make sure that you still have a big incentive to take care of your property.

Unemployment Insurance

Not everyone who wants to work actually has a job. Some people are unemployed, meaning that they are actively looking for work but do not have jobs. The unemployment rate is the number of unemployed individuals divided by the sum of the number employed and the number unemployed.

Toolkit: Section 17.3 "The Labor Market"

If you want to learn more details about the definition and measurement of unemployment, refer to the toolkit.

Since 1960, the unemployment rate in the United States has averaged slightly under 6 percent. This means that for every 100 people in the labor force (either working or looking for a job), 94 of them are working, and the other 6 are looking for jobs. The labor market is fluid so that, over time, unemployed workers find jobs, while some employed workers lose jobs and become unemployed. The unemployed find jobs, and others lose them and go through spells of unemployment.

If you are laid off from your job and become unemployed, you obviously still need to spend money for food and rent. During a spell of unemployment, you have several possible sources of income. If you have an existing stock of accumulated savings, then you can draw on these. If you are a member of a union, you may receive some support from the union. You may receive some severance pay when you lose your job. You might be able to rely on the support of your family and friends. And, most relevant for this chapter, you may be eligible to receive income from the government, calledunemployment insurance.

Unemployment insurance is similar in some ways to health and property insurance. There is an unlucky event called unemployment, and the government provides insurance. Perhaps you think this is great news: after graduation, you can claim unemployment, collect from the government, and enjoy your leisure. Of course, life is not quite that good. First, to qualify for unemployment insurance, you have to hold a job for some period of time. The details of these regulations differ across countries and also across states within the United States. Second, unemployment benefits do not last forever, nor do they completely compensate for all of your lost income. Again, the details depend on the country or state in which you work.

Why is the government in the business of providing insurance? To answer this, look back at our example of home insurance. The typical insurance company will have many policies with many different households. Over the course of a year, some households will make a claim on their insurance, but most will not. As long as the insurance company has lots of policies in many locations, then, on average, the number of insurance claims will be nearly constant each year. Although individual households face risk, the insurance company is able to diversify almost all of this risk.

Unemployment is different. When the economy is doing well, unemployment is low, and few households need this form of insurance. When the economy is not doing well, then the unemployment rate can be very high. In such times, many people want to claim unemployment insurance *at the same time*. So unlike insurance policies for homeowners, there is no easy way to balance out the risks of unemployment. The risk of unemployment is not *independent* across all individuals.

If an individual insurance company tried to offer unemployment insurance, it might be unable to survive: during a period of low economic activity, the demands for insurance would be so severe that the insurance company might not be able to meet all the claims. The government has the ability to tax people and borrow as needed. This puts it in a much better position to offer unemployment insurance. So in many countries, the government raises revenue by taxing firms and workers and uses these funds to provide unemployment insurance.

Deposit Insurance

In the United States and in some other countries, deposits that you place in the bank are insured by the government. In the United States, the government provides insurance, up to $250,000 per deposit, to you in the event your bank closes. [4] Deposit insurance in the United States dates from the time of the Great Depression in the 1930s. In this period many banks had insufficient funds on hand to meet the demands of their depositors and so went bankrupt. When this occurred, depositors lost the money they had put in the bank. After the Great Depression, the US federal government instituted deposit insurance. Similar programs exist in most other countries.

The argument for why the government should provide deposit insurance is similar to the argument for government provision of unemployment insurance. During periods of financial turbulence, many banks are prone to failure. If there were a private insurance company providing deposit insurance, it would probably be unable to meet all the claims. In addition, there is considerable social value to deposit insurance. It gives people greater confidence in the bank and in the banking system, which in turn makes bank failures less likely. Because bank failures put a great deal of stress on the financial system, government has an interest in insuring deposits.

In the summer of 2007, the British bank Northern Rock entered a financial crisis. Savers who had put their money in this institution started to worry that the bank would go bust, in which case they would lose their money. The British government, like the US government, provides deposit insurance. However, the amount of this insurance was limited to a maximum of about $70,000, so some people were still concerned about their savings. As lines started to form outside Northern Rock branches, the British government—concerned that the possible failure of Northern Rock would put other banks at risk—ended up guaranteeing all of its deposits.

KEY TAKEAWAYS

- The probability of a particular outcome is the percentage chance that the outcome will occur.
- An expected value is calculated by multiplying the probability of each outcome by the value of that outcome and then adding these numbers for all outcomes.
- Risk aversion means a preference for a sure thing rather than a gamble with the same expected value.
- We face many types of risks in our lives, and we can often buy insurance as a way to deal with these risks.
- Insurance companies provide a way for individuals to diversify their individual risks.

CHECKING YOUR UNDERSTANDING

1. Imagine you have a die that is fair: the probability of rolling each number is ⅙. Each time you roll the die, it is independent of the previous roll. Suppose you roll the die 10 times, and you roll a 5 each time. What is the probability of getting a 5 on the next roll of the die? Relate your answer to the common analysis of sportscasters who say that a baseball hitter who normally bats 0.300 “is due” when that batter has no hits in his last 20 at bats.
2. List some risks that you face that are not fully covered by insurance.

[1] See Paul Slovic, Baruch Fischoff, and Sarah Lichtenstein, “Facts versus Fears: Understanding Perceived Risk,” in *Judgment under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (Cambridge, MA: Cambridge University Press, 1982), 463–89.

[2] We do not discuss health insurance here. In Chapter 15 "A Healthy Economy", we discuss the provision of health care and the problems of health insurance in detail.

[3] We discuss this in Chapter 9 "Making and Losing Money on Wall Street".

[4] You can find details at FDIC, “Your Insured Deposits,” accessed March 14, 2011,http://www.fdic.gov/deposit/deposits/insured/basics.html.

4.4 Embracing Risk

LEARNING OBJECTIVES

1. Why do people sometimes take on risks instead of avoiding them?
2. What are compensating wage differentials?

Insurance allows us to remove some risks from our lives, either partially or completely. There are other risks that we accept or even embrace. In some cases, we are willing to take on risks because we are compensated in some way for them. We discuss two examples here: job choice and investment in the stock market. In other cases, we actively seek risk, such as when we gamble, buy lottery tickets, or engage in extreme sports.

Job Choice

We have already discussed how your choice of career should be based on discounted present value, taking into account the fact that it can be costly to move from one job to another. Your choice of career also reflects how you feel about risk and uncertainty. There are numerous kinds of jobs that differ in many dimensions. When choosing a job, we pay particular attention to wages or salaries, vacation plans, health coverage, and other benefits. We also take into account the job's riskiness.

The most severe risk is, of course, the risk of death. Every year in the United States several thousand workers suffer fatal work injuries. For most of the last decade, this number has been between about 5,000 and 6,000, or the equivalent of roughly 4 deaths for every 100,000 workers. Figure 4.11 "Work Fatalities in the United States" shows data on work-related deaths in the United States for 2006–9. [1] The fatality rate has generally been declining over time. In the early 1990s, the fatality rate was in excess of 5 deaths for every 100,000 workers; in 2009, the fatality rate was 3.3.

The fatality rate in the United States is high relative to many other developed countries. For example, the corresponding rate in England in 2006 was 0.8 per 100,000 workers. [2] Figure 4.12 "Work Fatalities in Europe" shows the work-related fatalities for countries in the European Union. Most importantly for the current discussion, the fatality rate varies significantly across jobs. In the United States, "Agriculture and mining recorded the highest fatal work injury rates among the major industry sectors in 2005—32.5

fatalities per 100,000 workers for agriculture and 25.6 fatalities per 100,000 workers for mining." In the European Union, construction, agriculture, and transportation are the most dangerous sectors. [3]

The risk of death is not the only job risk. Jobs also differ in terms of their risk of injury. In some cases, these injuries can have a severe impact on a person's quality of life; in other cases, they may prevent an individual from working in the future. Professional athletes and other performers face the risk of injuries that can end their careers. These individuals sometimes buy insurance to help mitigate some of these risks. In 2006, for example, pop star Mariah Carey purchased a $1 billion insurance policy on her legs after signing up for an advertising campaign ("Legs of a Goddess") with the Gillette shaving company. Bruce Springsteen's voice is insured for $5.6 million. [4]

Riskier jobs generally pay more. A firm that exposes workers to more risk must compensate them for that risk. A compensating wage differential is any difference in pay received by identical workers doing different jobs. (By "identical" we mean workers with comparable education, skills, experience, and so on.) Jobs that are unpleasant or dangerous will typically pay higher wages to compensate workers for the negative aspects of their jobs.

To the extent workers do not like to face risks, jobs that are viewed as riskier tend to pay more on average. For example, a recent study found that nurses who were more likely to be exposed to the AIDS virus (HIV) received higher wages than comparable nurses who were less likely to be exposed. [5]

Asset Portfolio Choice

Young people's portfolios of assets are usually very simple: a typical college student might have only a checking account and a savings account. As you grow older, though, you will typically acquire a broader portfolio. Even if you do not directly purchase assets such as stocks and bonds, you may own them indirectly when you sign up for a pension plan. Because the return on stocks (and other assets) is uncertain, owning these assets is another type of risk you *choose* to take.

Owning a stock is somewhat like buying a lottery ticket. You pay some money to buy a share of the stock of some company. In return, you may be paid some dividends; at some time in the future, you may sell the stock. But at the time you buy the stock, you don't know the payments you will receive in the future and you don't know the future price of the stock. So by purchasing a stock, you are gambling. Whether the gamble is favorable or not depends on the price of the stock, the chance it will pay dividends in the future, and the future price. Choosing how to allocate the assets in a portfolio is a type of gamble we all make.

We cannot completely avoid this kind of gamble. Perhaps you think that putting cash in your mattress would be a way to avoid this risk, but that is not the case. Ultimately you care about what that money can buy in terms of goods and services. The real value of the money held in your mattress depends on the future prices of goods and services, which are not known to you today. The benefit of holding cash depends on the unknown inflation rate.

Buying a Lottery Ticket

Some people, at some times, are eager and willing to engage in risky activities. People engage in extreme sports, where the danger appears to be part of the attraction. People go to Las Vegas or to Monaco to gamble. In many countries, citizens can purchase a wide variety of lottery tickets sold by their governments. This is a form of gambling: you buy a ticket and if you have the lucky number, then you get a (sometimes large) prize. If you do not have the lucky number, your money is gone.

The existence of lotteries and other kinds of gambling seems like a puzzle. If people are risk-averse, then they are supposed to want to get rid of risk. The purchase of a lottery ticket is the exact opposite: you give up a sure thing (the price of the ticket) for an uncertain outcome. Unlike the purchase of insurance, which is a way to avoid risk, buying a lottery ticket is a demand for a gamble. Why do so many people buy lottery tickets? Why do governments sell them?

Do lottery tickets have an expected value that exceeds the cost of the ticket? If the difference is big enough, then even a risk-averse person might want to buy a ticket. Consider a very simply lottery. Suppose there is one fixed prize, and there is a probability that you win the prize. Then your expected gain is just the probability of winning times the prize:

expected gain = probability of winning × value of prize.

Using this equation, you can determine whether the price for a lottery ticket is high or low. If the price of the ticket exceeds the expected gain from buying the ticket, then the ticket is not a good deal. But if the price is low relative to the expected gain, then you may want to accept the risk and buy the ticket.

Let us look at an example. One US lottery is called Powerball. On February 18, 2006, the prize was worth $365 million to the winner. The chance of getting the jackpot was 1 in 988,172,368. [6] The expected value of a ticket at that time was the value of the prize times the probability of winning: $\$365\text{ million} \times \frac{1}{988{,}172{,}368} = \$0.37$—far less than the dollar price of the ticket. Despite the huge prize, the price of a ticket far exceeded its expected value.

Another perspective on the lottery is from the viewpoint of the government selling these tickets. Consider the Texas lottery. [7] The proceeds from the sale of tickets primarily support education. In 2005, about 60 percent of the income from the lottery went to payment for prizes and 28 percent went to a school fund. From the perspective of the Texas state government, selling lottery tickets is a way to fund programs. If the government is to make money on lottery tickets, then those buying the tickets must, on average, be losing money.

In fact, as this discussion suggests, the expected value of a lottery ticket is less—often substantially less—than the cost of the ticket. Why, then, do people buy lottery tickets? One possibility is that they simply enjoy gambling. This means that, at least with respect to these types of gambles, they are *risk-loving* rather than risk-averse. The pleasure of a lottery ticket is, among other things, the license to dream. Another possibility is that individuals overestimate their chances of winning.

KEY TAKEAWAYS

- One reason that people take on risks is because they enjoy gambling.
- Some jobs are riskier than others and pay more to compensate people for the risks they face.

CHECKING YOUR UNDERSTANDING

1. Suppose you take a job as an engineer at an oil company. There are two places where you can work. One is Houston, where you will live in the suburbs and commute to work. The other is on an oil rig in the North Sea (between Scotland and Norway), where you will spend most of your time on the rig. If the jobs are otherwise identical, which do you expect will pay more? Explain why.

[1] The data come from Occupational Safety and Health Administration, "Census of Fatal Occupational Injuries," accessed February 24, 2011,http://stats.bls.gov/iif/oshwc/cfoi/cfch0008.pdf. The Occupational Safety and Health Administration (OSHA) is the federal agency that monitors workplace safety.

[2] See Health and Safety Executive, "Fatal Injury Statistics," accessed March 14, 2011,http://www.hse.gov.uk/statistics/fatals.htm. Care must be taken when comparing international data because they are not always exactly comparable. Most importantly, European Union statistics exclude road traffic accidents, which account for about 43 percent of US deaths. Correcting for this, the US figures are similar to the European Union average.

[3] See US Department of Labor Bureau of Labor Statistics, "National Census of Fatal Occupational Injuries in 2005," accessed March 14, 2011,http://www.bls.gov/iif/oshwc/cfoi/cfnr0012.txt and Health and Safety Executive, "European Comparisons," accessed March 14, 2011,http://www.hse.gov.uk/statistics/european/fatal.htm.

[4] See starpulse.com, "Mariah Carey Takes Out $1 Billion Insurance Policy For Her Legs," accessed March 14, 2011,http://www.starpulse.com/news/index.php/2006/05/30/mariah_carey_takes_out_1_billion_insuran for an account of Ms. Carey's insurance policy; Paul Bannister, "World's Biggest Insurer Takes on All Risks," accessed March 14, 2011,http://www.bankrate.com/brm/news/insurance/old-lloyds1.asp, details a number of unusual insurance policies.

[5] Jeff DeSimone and Edward J. Schumacher, "Compensating Wage Differentials and AIDS Risk" (NBER Working Paper No. 10861, November 2004), accessed March 14, 2011,http://www.nber.org/papers/w10861.

[6] This is based on five balls selected ranging from 1 to 55 and a powerball ranging 1 to 42. See the Powerball site (http://www.usamega.com/powerball-howtoplay.htm) for details on this game. See also the odds calculator at CSGNetwork.com, "The Ultimate Lottery Games Odds Calculator," accessed March 14, 2011,http://www.csgnetwork.com/oddscalc.html.

[7] Texas Lottery home page, accessed March 14, 2011,http://www.txlottery.org/export/sites/default/index.html.

4.5 End-of-Chapter Material

In Conclusion

Many decisions involve a trade-off between now and the future. Whenever we invest our time or money, we are giving up something today to obtain something in the future. So by saving some of our income, we give up consumption now for consumption in the future. When we go to a university, we give up income and leisure time today to get more consumption (through higher income) in the future.

Once you start thinking about trade-offs over time, it is difficult to avoid the reality that many of these decisions are made in the face of great uncertainty. When we save, we are not certain of the return on our saving. When we go to school, we are not guaranteed a job in the future nor are we guaranteed a specific salary. We have provided some insights into the nature of these uncertainties and how to deal with them. Discounted present value and expected value are techniques that are worthwhile to master, as they will help you make better decisions throughout your life.

Key Links

- Unemployment insurance
    - New York State:http://www.labor.state.ny.us/workerprotection/laborstandards/workprot/ui.shtm
    - Texas: http://www.twc.state.tx.us/ui/bnfts/claimant1.html
- Deposit insurance: http://www.fdic.gov/deposit/index.html
- Powerball lottery: http://www.powerball.com
- Fun site to calculate probabilities:http://stattrek.com/Tools/ProbabilityCalculator.aspx

EXERCISES

1. Explain why an increase in the interest rate reduces the discounted present value of income.
2. What incentives exist for people to repay loans on education? On cars?
3. Suppose the nominal interest rate is 20 percent, the price level in the first year is 50, and the price level in the second year is 60. What is the real interest rate? How could you alter this example so that the real interest rate is 0?

4. Can the nominal interest rate ever be less than the real interest rate? If the real interest rate is negative, what happens to the slope of the budget line with two periods discussed in Chapter 4 "Life Decisions", Section 4.1 "Consumption and Saving"?
5. Do households sometimes borrow and lend simultaneously? Why might that happen? Is the interest rate they borrow at usually higher or lower than the interest rate they receive, say in the form of bank deposits?
6. In describing how changes in income influence the supply of loans, we assumed that the increase in income occurs this year. Suppose instead that the increase in income will occur next year even though everyone in the economy knew it would happen today. How would the news of a future increase in income influence the current loan supply curve?
7. When the government changes taxes, do you know if it is permanent or temporary?
8. (Advanced) One way that real wages received by workers can change is through a change in income taxes. Considering the information in this chapter, would you except temporary tax changes to have a bigger or a smaller impact on labor supply than a permanent tax change? What if the tax change is not through a change in the tax rate but rather through a fixed payment to households? What would that policy do to labor supply?
9. Look back at Table 4.2 "Which Career Should You Choose?" in the section "Choosing a Career." How would you edit the income entries in the table so that the insurance salesperson had a higher discounted present value than the lawyer even when the interest rate is 5 percent?
10. Look back at Table 4.2 "Which Career Should You Choose?" in the section "Choosing a Career." Explain why an increase in the interest rate makes it less attractive to be a lawyer.
11. Besides discounted present values of income, what other factors are important in choosing a career? How do you balance these with differences in the discounted present value of income?
12. What are the risks associated with choosing a particular career? How do those risks depend on whether the skills you learn at your job can be used in other jobs?
13. Show the calculation of the discounted present value from work in Table 4.5 "Income from Going to College versus Taking a Job". Redo the comparison of college and work assuming an interest rate of 20 percent.

14. (Advanced) Look back at Table 4.2 "Which Career Should You Choose?" in the section "Choosing a Career." Suppose the income of a lawyer increases by 20 percent each year after year 2 and the income of the insurance salesperson increases by $10,000 each year. Extend Table 4.2 "Which Career Should You Choose?" to 5 years. Which is a better profession?
15. Create your own version of Table 4.8 "Coin-Flipping Experiment" by flipping a coin 10 times. Imagine that each time the result is a head, you earn $1,000, and each time it is a tail, you lose $1,000. After you flip the coin 10 times, calculate how much you won (or lost). Now do this same experiment 20 times. Each time you flipped the coin 10 times, record how much you won (or lost), which will result in 20 numbers. What is the average of these 20 numbers? What is the expected value of how much money you will earn in each coin-flipping experiment??
16. Many products, such as computers, come with the option of an *extended warranty*. Suppose you are buying a computer with a one-year warranty. Thereafter, you can purchase an extended warranty for one more year, at a cost of $50. The warranty will repair or replace your computer in the event of breakdown. Suppose the average cost to the manufacturer of repairing or replacing the computer is $1,000. If the manufacturer is making no money from this warranty, what is the implied probability that the computer will need repair?
17. We wrote "As long as the insurance company has lots of policies in many locations, then, on average, the number of insurance claims will be nearly constant each year." Why did we include the statement about many locations?
18. Why is it difficult to diversify job risk? Is it possible to do some diversification within a family?
19. In the United States, the provision of unemployment insurance is partly at the state level and partly at the federal level. For your state, find out what the benefits are and what federally funded unemployment insurance might be available to you.

Economics Detective

1. Study the insurance policy you can buy when you purchase a new cell phone. Exactly what does this insurance protect you against? Given the price of the insurance and the coverage, what is the implied probability that you will make a claim for a new phone under the insurance? Is there a deductible? Why is it part of the policy?

2. Look at the insurance policy (if you have one) for the place where you are living. What is the deductible? List the ways in which you take actions to reduce the risk of fire where you live.
3. Our example of homeowners’ insurance did not use real numbers. Find a homeowners’ policy and determine the coverage, the premium, and the deductible.
4. One form of insurance occurs when you rent a car. Using the Internet or phoning local insurance agents, find out the kinds of insurance that are available when you rent a car. What is the cost per day? Exactly what risks do these policies protect you against? Given the price of the insurance and the coverage, what is the implied probability that you will have an accident and make a claim under the insurance? Does this probability seem reasonable to you?
5. What is the average price of a house in the United States? In your hometown?
6. What does it cost to insure a $100,000 house in your city? What does it cost to insure a $1,000,000 house in your city? Explain the differences in the insurance costs.
7. Pick a state in the Unites States. Suppose you work there and earn $2,000 each week as a manager. One day, the firm tells you that you are no longer needed. What unemployment insurance could you collect? Would you qualify for unemployment insurance? How much would the benefits be? How long would the benefits last?
8. Go to your local bank and see if there are any signs that indicate deposit insurance is provided. Ask about details of the program.
9. Use the Internet to find out about deposit insurance programs in the United States and in another country. How do these programs compare?
10. For the state in which you live, does the government sponsor a lottery? If so, how are the funds used?
11. In the financial crisis of 2008–9, was deposit insurance provided in the United States? In other countries?
12. In the state you live in, find out about the unemployment insurance program. How long do you receive benefits and how generous are the benefits?

Spreadsheet Exercises

1. Write a spreadsheet program to create a version of Table 4.8 "Coin-Flipping Experiment" for any combination of income flows and interest rates.
2. (Advanced) Create a spreadsheet program to simulate the flipping of a coin. Do *T* experiments with 5 flips per experiment. For each experiment, calculate the mean of the outcome. When you are finished, you

will have *T* means. What does the distribution of the *T* means look like? What is the mean of that distribution? What happens as *T* gets very large?

Chapter 5
eBay and craigslist
Buying on the Internet

eBay is one of the most famous sites on the Internet (http://www.ebay.com;Figure 5.1 "The eBay Home Page"). It was founded in 1995 and is now a very large company, with $60 billion in sales in 2009 and over 90 million active users worldwide. [1] One of the many ways in which the Internet is changing the world is that there are now ways of buying and selling—such as eBay—that were completely unavailable to people 20 years ago.

In or around the second and third centuries BCE, the island of Delos, in Greece, was a major center of trade—both of goods and slaves. At its height of activity, Delos, an island of five square kilometers, had a population of about 25,000 people. [2] This means Delos was about as densely populated as modern-day cities such as Istanbul, London, Chicago, Rio de Janeiro, or Vancouver. Visitors today see different things when they visit the ruins of Delos. Some see dusty pieces of shattered rock; others see the remains of a great culture. An economist sees the ruins of a trading center: a place where people such as the trader shown in Figure 5.2 "A Trader in Delos" were, in a sense, the eBay of their times.

Delos and eBay are separated by almost two and a half millennia of history, yet both are founded on a basic human activity: the trading of goods and services. How basic? Consider the following.

- Children learn to trade at an early age. First graders may be trading Pokémon cards before they have even learned the meaning of money.
- In prisoner-of-war camps during World War II, prisoners would receive occasional parcels from the Red Cross, containing items such as chocolate bars, cigarettes, jam, razor blades, writing paper, and so on. There was extensive trading of these items. In some cases, cigarettes even started to play the role of money. [3]
- In Nazi concentration camps, where people lived close to starvation in conditions of extreme danger and deprivation, the prisoners traded with each other. They traded scraps of bread, undergarments, spoons, basic medicines, and even tailoring services. [4]

Trade has played a central role in determining where many of us live today.

- In England, you will find the town of Market Harborough; in Germany, you can find Markt Isen; in Sweden, Lidköping. *Markt* and *köping* both mean market. These towns, and many others like them, date from the medieval period and, as their names suggest, owe their existence to the markets that were established there.
- Many of the great cities of the world developed in large part as ports, where goods were imported and exported. London, New Orleans, Hong Kong, Cape Town, Singapore, Amsterdam, and Montreal are a few examples.

Much of economics is about how we interact with each other. We are not alone in the economic world. We buy goods and services from firms, retailers, and each other. We likewise sell goods and services, most notably our labor time. In this chapter, we investigate different kinds of economic interactions and answer two of the most fundamental questions of economics:

1. How do we trade?
2. Why do we trade?

Road Map

The chapter falls naturally into two parts corresponding to these two questions. We begin by thinking about the ways in which individuals exchange goods and services.

In modern economies, most trade is highly *disintermediated*. You usually don't buy a good from its producer. Perhaps the producer sells the good to a retail store that then sells to you. Or perhaps the good is first sold to a wholesaler who then sells to a retailer who then sells to you. Goods are often bought and sold many times before you get the opportunity to buy them. [5]For the moment, however, we have a different emphasis. We do not yet get into the details of retailing in the economy but instead focus on trade among individuals—the kind of transaction that you can carry out on eBay and craigslist.

Specifically, we want to understand how potential buyers and sellers are matched up. We also want to know what determines the prices at which people exchange goods and services. Broadly speaking, prices can be established in the following ways.

- Some prices are the result of bargaining and negotiation. If you buy a car or a house, you will engage in a one-on-one negotiation with the seller in which there will typically be several rounds of offers and counteroffers before a final price is agreed on. Similarly, if you go to street markets in many countries in the world, you will not find posted prices but will have to bargain and haggle.
- Some prices are determined by auction, such as trinkets sold on eBay and antiques sold by Christie's. Auctions are a type of bargaining that must follow some preset rules.
- Some prices are chosen by the seller: she simply displays a price at which she is willing to sell a unit of the good or service. Even this is a very simple type of bargaining called the "take-it-or-leave-it offer." The seller posts a price (an "offer"), and prospective buyers then have a simple choice: either they buy at that price (they "take it") or they do not buy (they "leave it").

Take-it-or-leave-it offers are the most common form of price-setting in retail markets. The prices displayed in your local supermarket can be thought of as thousands of take-it-or-leave-it offers that the supermarket makes to you and other shoppers. Whenever you go to the supermarket, you reject most of these offers (meaning you don't buy most of the goods on display), but you accept some of them. Take-it-or-leave-it offers also occur when individuals trade. Classified advertisements in newspapers or on Internet sites like craigslist typically involve take-it-or-leave-it offers.

Once we understand how individuals trade with one another, we turn to an even more basic question: why do we trade? Whether we are talking about first graders swapping Pokémon cards, the purchase of a camera on eBay, the auction of a Renoir painting at Sotheby's, or traders in the Mediterranean islands over two millennia ago, there is one reason for trade: I have something you want, and you have something I want. (In many cases, one of these "somethings" is money. Keep in mind, though, that people don't want money for its own sake; they want money to buy goods and services.)

We therefore explain how differences in what we have and what we want provide a motive for trade and how such trade creates value in the economy. Then we go deeper. In a modern economy, trade is an essential part of life. We consume a large number of goods and services, but we play a role in the production of very few. Put differently, modern economies exhibit a great deal of specialization. We carefully investigate how specialization lies right at the heart of the gains from trade.

[1] See eBay Inc., “Who We Are,” accessed February 24, 2011,http://www.ebayinc.com/who.

[2] MyKonos Web, “History of Delos,” accessed January 25, 2011, http://www.mykonos-web.com/mykonos/delos_history.htm.

[3] For the classic discussion of trade in a prisoner-of-war camp, see R. A. Radford, “The Economic Organisation of a P.O.W. Camp,” *Economica* (*New Series)* 12, no. 48 (1945): 189–201.

[4] Jill Klein, “Calories for Dignity: Fashion in the Nazi Concentration Camp,” *Advances in Consumer Research* 15 (2003): 34–37.

[5] Such transactions are the focus in Chapter 6 "Where Do Prices Come From?" andChapter 7 "Why Do Prices Change?".

5.1 craigslist and the Gains from Trade

LEARNING OBJECTIVES

1. What are the gains from trade?
2. How does a market with take-it-or-leave-it offers work?
3. How are the gains from trade split between buyers and sellers?
4. What is economic efficiency?

To begin our investigation of why and how we trade, let us examine craigslist (http://www.craigslist.org), an Internet site devoted to exchange. The craigslist site is very similar to the classified advertisements in a newspaper except that the advertisements are online. It is local, in the sense that there is a different site for different places. You can find craigslist sites for cities and states throughout the United States, and—at the time of this writing—for 14 cities and 54 countries around the world. If you visit the craigslist website, you will see there are many types of goods and services listed. For now, we focus on the purchase of a good. Later, we will consider the purchase and even the exchange of services.

Pricing on craigslist commonly takes a take-it-or-leave-it form. The seller posts a price and then buyers and sellers communicate through (anonymous) e-mails. Of course, the buyer always has the option of trying to turn this take-it-or-leave-it scenario into back-and-forth bargaining by making a counteroffer. Once they have agreed to trade, the buyer and seller must find a way to consummate the transaction—delivering the good and making payment.

The Gains from Trade from a Single Transaction

Suppose you are interested in buying a car. You go to craigslist in your area and search through offers to sell cars. These offers typically provide lots of information about the product, usually including photos and a price. If you want to inquire about a particular car, you can contact the seller. If you want to buy the car, you can accept the seller's offer. If you want to negotiate, you can do so as well. To get at the heart of this kind of exchange, let us first take a simple case where there is a single seller and a single buyer.

Economists generally think that individuals make decisions in their own self-interests. If a seller is willing to sell a good at a given price, and a buyer is willing to buy at that price, our presumption is that this exchange makes them both better off. This deceptively simple idea is the very heart of economics:

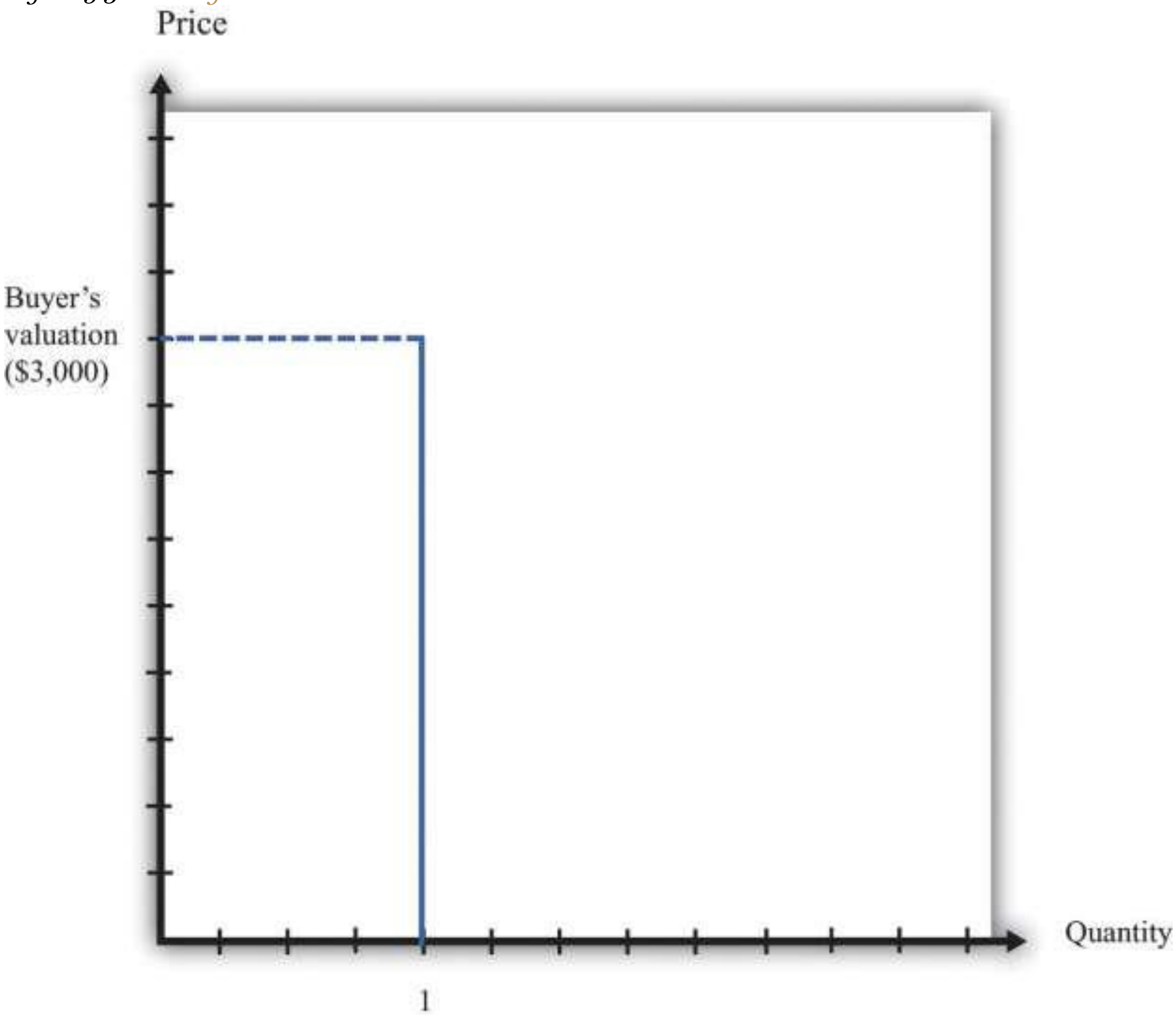
voluntary trade makes both participants better off. The word *voluntary* matters here. We are supposing that both people freely enter into this trade. If two people make a deal of their own free will and if they are rational, in the sense that they can make decisions in their own best interests, then the deal must make them both better off.

The demand for a car is an example of a **unit demand curve** because you are deciding whether or not to buy at all rather than how much you should buy. [1]The buyer has a **valuation** for the good, which represents the most he would be willing to pay for it. For example, suppose you see a used car on craigslist, and your valuation of this car is $3,000. This means that you would be equally happy either having the car and forfeiting $3,000 worth of other goods and services or not having the car. Figure 5.3 "The Buyer's Valuation" shows what your demand curve looks like in this case. You are choosing to buy either zero units or one unit, so if the price is above your valuation, you do not buy the good, whereas if the price is below your valuation, you buy the car.

Toolkit: Section 17.1 "Individual Demand"

You can review unit demand and valuation in the toolkit.

Figure 5.3 The Buyer's Valuation



The buyer follows the decision rule: "Buy if the price is less than the valuation."

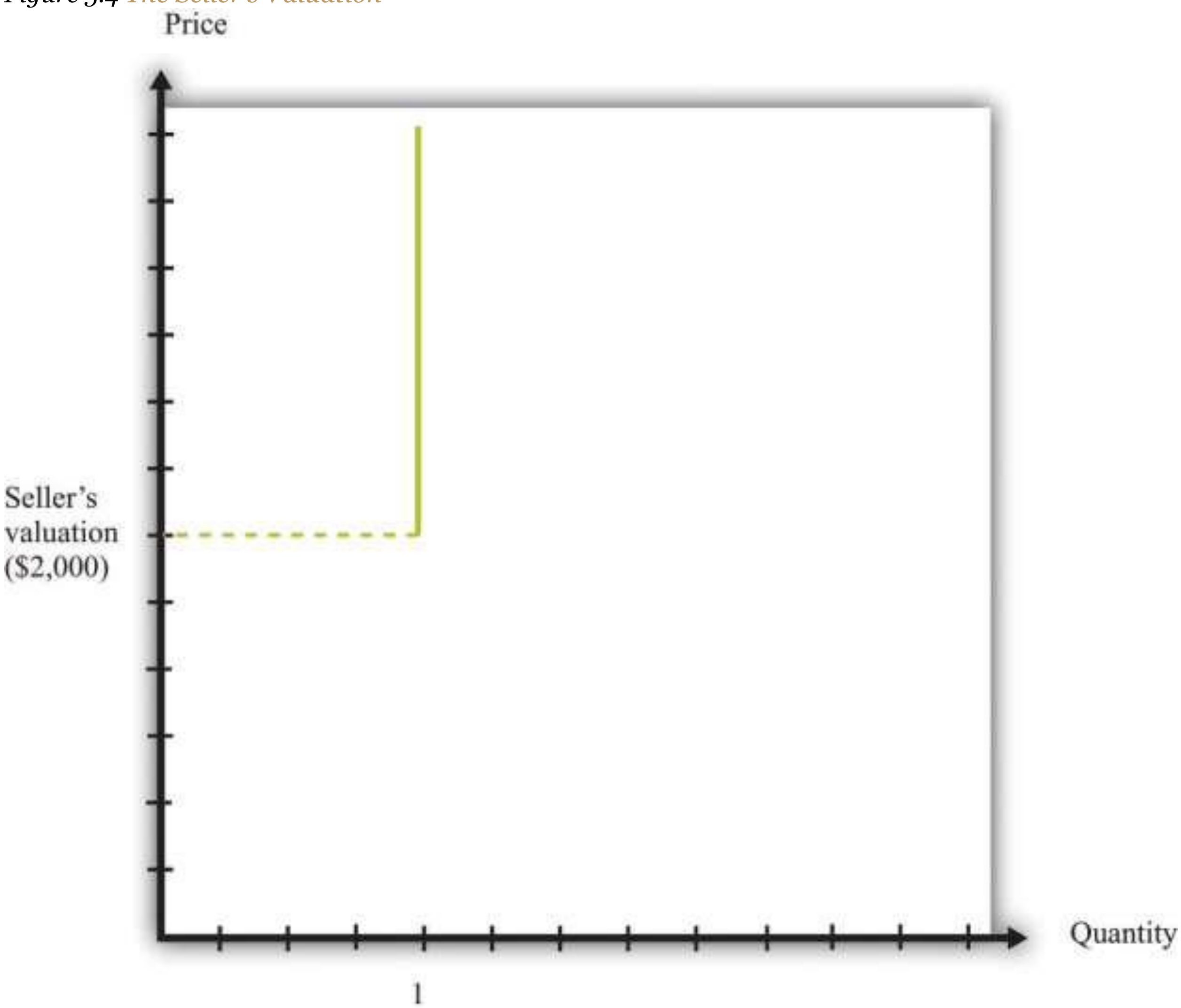
If your valuation were $3,000, then you would, of course, prefer to pay much less. If the car is for sale for $2,990, then it is true that you would be better off buying the car than not, but you won't get much out of the deal. You would be happier only to the tune of $10 (more precisely, $10 worth of goods and services). If the car is for sale for $2,400, then you will be happier by an amount equivalent to $600 worth of goods and services. On the other hand, if the car were for sale for $3,001, you definitely would not want to buy at that price. Buying the car would actually make you slightly less happy.

The seller also has a valuation of the car. The seller is not willing to sell it at any price. For example, if her valuation is $2,000, she is equally happy keeping the car or not having the car and having an extra

$2,000 worth of goods and services. If she can sell the car for more than $2,000, she will be better off. She won't sell the car for less than $2,000 because then she would be less happy than before. We can show the seller's willingness to sell in a way analogous to the buyer's willingness to buy. Figure 5.4 "The Seller's Valuation" shows that she will not sell the car at a price less than $2,000, but she will sell once the price is greater than $2,000.

By analogy to unit demand, we call this the unit supply curve. It tells us the price at which she is willing to sell. Below her valuation, she is unwilling to supply to the market. Above her valuation, she is willing to sell the good. Whereas the buyer's valuation is the *absolute maximum* that the buyer is willing to pay, the seller's valuation is the *absolute minimum* that the seller is willing to accept.

Figure 5.4 The Seller's Valuation



The seller follows the decision rule: "Sell if the price is greater than the valuation."

The buyer's valuation in our example is larger than the seller's valuation. This means it is possible to make both the buyer and the seller better off. The mere fact of transferring a good from someone who values it less to someone who values it more is an act that *creates value* in the economy. We say that there are gains from trade available here.

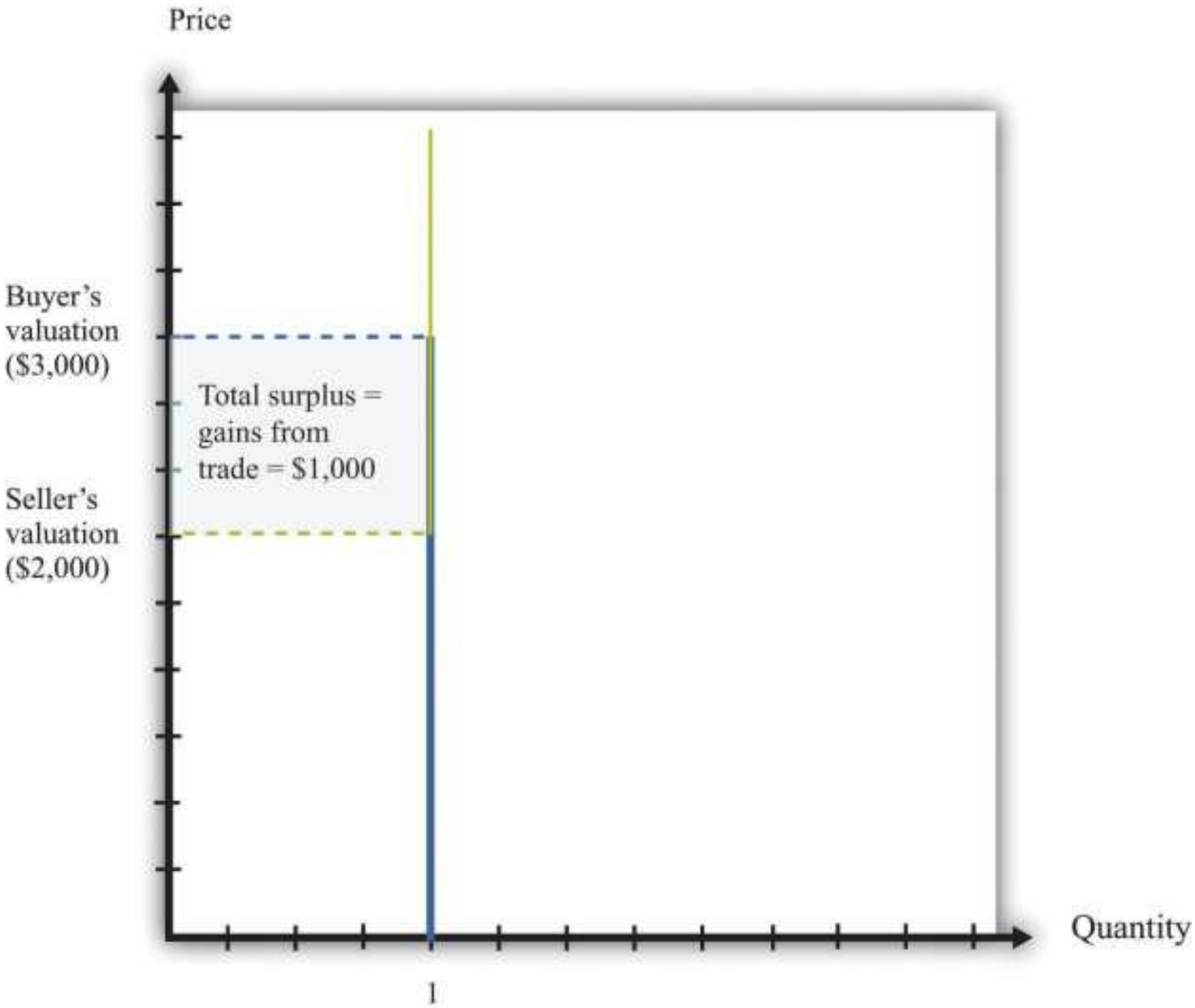
Toolkit: Section 17.10 "Buyer Surplus and Seller Surplus"

Total surplus is a measure of the gains from trade. In a single transaction,

total surplus = buyer's valuation − seller's valuation.

In this example, therefore, the total surplus is $1,000. This is the value created in the economy by the simple fact of transferring the car from a seller who values it less to a buyer who values it more. Figure 5.5 "Buyer and Seller Valuations" shows this graphically by combining the unit demand curve and the unit supply curve.

Figure 5.5 Buyer and Seller Valuations



The total surplus from a transaction is equal to the buyer's valuation minus the seller's valuation. Graphically, total surplus can be represented as a rectangle. The height of the rectangle is the difference in the valuations. The base of the rectangle is 1 because only one unit is being traded.

The buyer wants the price to be as low as possible, whereas the seller wants the price to be as high as possible. If both agree on a price of $2,100, for example, the buyer gets most of the surplus, and the seller does not get very much. If they agree on a price of $2,900, the situation is reversed: most of the benefit goes to the seller. The *distribution* of the value created depends on the price. Either way, though, they are both made better off by the trade, and in both cases the total surplus is the same (Figure 5.6 "The Distribution of Total Surplus").

Toolkit: Section 17.10 "Buyer Surplus and Seller Surplus"

The buyer surplus is a measure of how much the buyer gains from a transaction, and the seller surplus is a measure of how much the seller gains from a transaction:

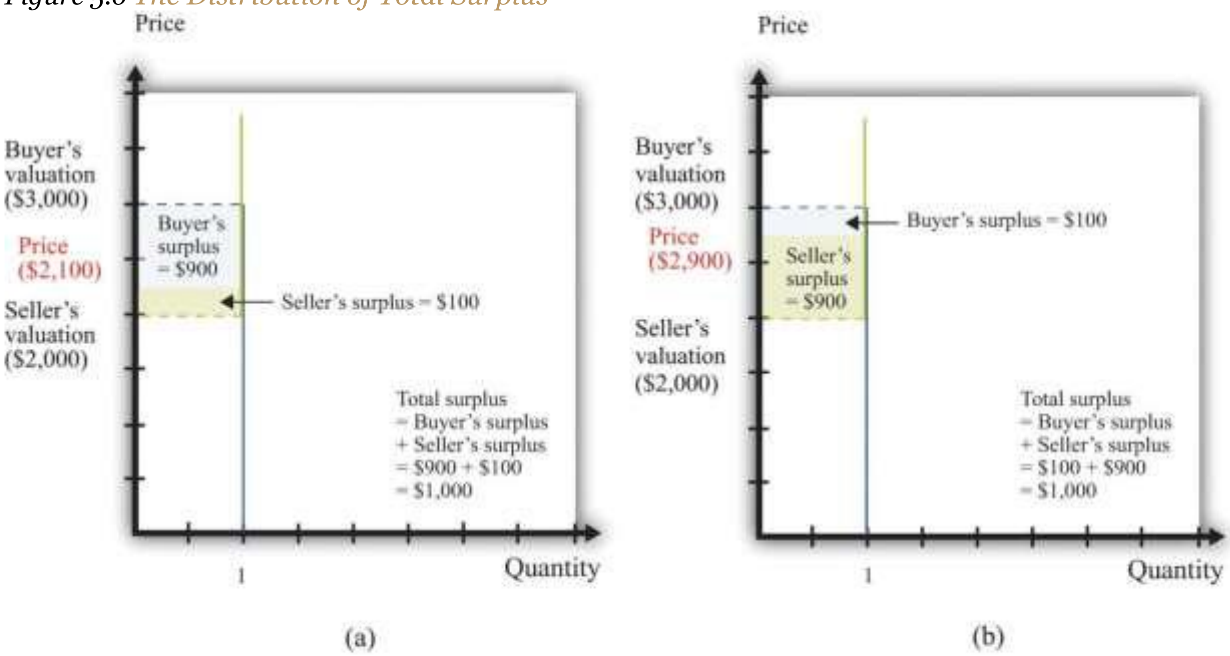
buyer surplus = buyer's valuation − price

and

seller surplus = price − seller's valuation.

The total surplus is the sum of the buyer surplus and the seller surplus.

Figure 5.6 The Distribution of Total Surplus



The distribution of surplus between the buyer and the seller depends on the price. A low price means that the buyer will get most of the surplus, while a high price means that the seller will get most of the surplus. The total surplus, however, is the same no matter what the price.

Economic Efficiency

When the buyer purchases the car from the seller, there is a reallocation of society's resources. Dollars have gone from the buyer to the seller, and the car has gone from the seller to the buyer. Economists have developed a specific criterion, called efficiency, for evaluating the way in which resources are allocated in a society.

It is actually easier to understand efficiency by looking at its opposite. Economists say that an allocation of resources is *inefficient* if there is some way to reallocate those resources that will make some people better off (that is, happier) without making anyone else worse off. For example, think about the situation where the buyer and seller have not traded the car. This allocation is inefficient. The buyer places a greater value on the car than does the seller, so it is inefficient for the car to remain with the seller. Any rearrangement of resources that makes some people better off without making anyone else worse off is *welfare improving*.

Toolkit: Section 17.11 "Efficiency and Deadweight Loss"

Efficiency is the basis that economists use for judging the allocation of resources in an economy. Resources are allocated efficiently if there is no way to reallocate them to make someone better off without making anybody else worse off.

Before the buyer and the seller trade, the allocation of resources is inefficient. However, there are many different trades that make both the buyer and the seller better off. In fact, any trade between a price of $2,000 and a price of $3,000 is welfare improving. The only thing that matters for economic efficiency is that a trade takes place, so the gains from trade can be realized. No matter how the surplus is distributed between the buyer and the seller, the outcome is efficient as long as the trade occurs.

Determining a Price on craigslist

We now know that as long as the buyer's valuation for a good exceeds the seller's valuation, there are potential gains from trade. We have not yet explored the mechanisms that allow trade to occur, nor have we explained what determines the price at which trades occur. To begin with, we ignore the possibility of bargaining. Then there are only two steps for selling an item on craigslist:

1. The seller of an item posts a price (makes a take-it-or-leave-it offer).
2. The buyer either accepts or rejects the offer.

If the buyer accepts the seller's offer, then an exchange is made. But what offer will the seller make? The answer depends on how much the seller knows about the *buyer's* valuation of the good. There are two cases to consider:

1. **The seller knows the buyer's valuation.** The seller would like her surplus to be as large as possible. If she knows the buyer's valuation, what should she do? To answer this question, she must put herself in the buyer's shoes. If she sets a price that is greater than the buyer's valuation, then the buyer will reject the offer. But as long as the price is less than the buyer's valuation, the buyer will accept the offer. With this in mind, the seller will set a price slightly less than the buyer's valuation to capture almost the entire surplus. In our example, the seller should put the car on sale for $2,999. The seller gets $999 worth of surplus, and the buyer gets $1. (Does it matter if the buyer also knows the seller's valuation? As a matter of pure economic theory, the answer is no. The buyer should be willing to buy provided he is getting some surplus—even if it is very little. After all, something is better than nothing. However, if the buyer knows that the seller is getting a lot of surplus, he may perceive this as unfair and might even choose not to buy out of spite. In reality, sellers often set a lower price, "giving away" some of the surplus to the buyer, to avoid this possibility. [2])
2. **The seller does not know the buyer's valuation.** This case is more likely and also much harder. The seller must trade off two different concerns. If she picks too high a price, then there is a risk of not making a sale at all. But the lower the price, the less surplus she gets in the event of a sale. There is no simple rule to know what price she should set. An economist looking from the outside finds this case more worrying than the first because it is possible that the gains from trade will be missed. If the seller offers a price that is greater than the buyer's valuation, then—under a take-it-or-leave-it offer—no trade takes place.

The knowledge of the buyer also matters. Suppose that the buyer knows the seller's valuation. Then he knows that there are possible gains from trade. In this case, it is natural to think that the buyer will try to negotiate with the seller, rather than just accept or reject the seller's offer. Indeed, if the buyer knows the seller's valuation, then we have the reverse of the first case. If the buyer offers a price slightly above the seller's valuation, then the buyer should be able to capture the entire surplus. We summarize this in Figure 5.7 "The Outcomes from a Take-It-or-Leave-It Offer".

In practice, the buyer is also likely to try to negotiate if the seller's price leaves the buyer with very little surplus. Thus even though craigslist is apparently based on take-it-or-leave-it offers, a great deal of bargaining does in fact take place.

Figure 5.7 The Outcomes from a Take-It-or-Leave-It Offer

The seller makes a take-it-or-leave-it offer		**Does the buyer know the seller's valuation?**	
		Yes	No
Does the seller know the buyer's valuation?	Yes	The seller may need to let the buyer have some surplus. Otherwise the buyer may say no out of spite!	The seller sets a price just below the buyer's valuation and gets almost all the surplus.
	No	There is a risk of the price being too high. However, if the buyer can make a counteroffer, he will offer a price just above the seller's valuation and get almost all the surplus.	There is a risk of the price being too high, in which case the gains from trade are missed.

KEY TAKEAWAYS

- If the seller's valuation of an object is less than the buyer's valuation of the same object, then there are gains to trade.
- One mechanism to reap the gains from trade when valuations are known is for the seller to post a price and the buyer to decide to purchase the good or not—that is, the seller makes a take-it-or-leave it offer.
- The way the gains to trade are split between the buyer and the seller depends on the way the bargaining occurs and the information the parties have about each other.
- An allocation is efficient if there is no way to make someone better off without also making somebody else worse off.

CHECKING YOUR UNDERSTANDING

1. In this discussion, we assumed that the seller's valuation was less than the buyer's valuation. What would happen if that were not true?
2. Suppose the seller's valuation is less than the buyer's but that the buyer, not the seller, sets the price. What price would the buyer set? Would there still be gains from trade?

[1] See Chapter 3 "Everyday Decisions" for more discussion.

[2] Technically, a take-it-or-leave-it offer is an example of an ultimatum game, which is discussed in Chapter 12 "Superstars".

5.2 eBay

LEARNING OBJECTIVES

1. What are the economics of an eBay auction?
2. How should I bid on eBay?
3. How are gains to trade determined and shared when there are multiple buyers?
4. What is the winner's curse?

So far we have supposed that there is only a single buyer and a single seller. If you are thinking about selling a good on craigslist, however, there are many potential buyers of your good. In addition, you probably don't know very much about the valuations of the different buyers. You might then like to find some way to make your buyers compete with each other. In other words, you might consider auctioning off the good instead.

You have probably at least visited the eBay site, and you may even have bought or sold an item on eBay. If so, you know it can be a convenient and efficient way to buy and sell goods. But what exactly is eBay? We answer this question by looking at the site from the perspective of participants. First we review how eBay works and look at it from the point of view of both a buyer and a seller. Then we bring some economic analysis to bear to better understand what is taking place on eBay and in other auctions.

Buying on eBay

Suppose you want to purchase something, such as a leather jacket, some cycling gloves, or a cell phone; the list of things that might interest you is endless. On the eBay page, you can search for the exact item you want to buy. Your search must be specific: if you search for "cell phone," you will find thousands of products. You need to know the exact model of phone you want, and even then you may find multiple items for sale.

Auctions on eBay have several characteristics, including the identity of the seller, the time limit on the auction, the acceptable means of payment, the means of delivery, and the reserve price.

- **Seller identity.** Unlike when you purchase from a store, you do not get to see the seller on eBay, and you cannot simply walk away with the good. This may concern you. After all, how do you know the seller will actually ship you the good after you have paid? How do you know if the product will work? This is a worry for you, but it is also a worry for any reputable seller. Sellers want you to trust them, so there are mechanisms to allow you to find out about sellers. On the eBay page, you can find detailed information about the seller, including a number—called a feedback score—that indicates the number of positively rated sales by that seller. If you dig a bit deeper, you can even find reviews of the performance of the seller.

- **Time remaining in the auction.** Online auctions have a fixed time limit. When you go to the auction site for a particular good, you will see the amount of time left in the auction. Much of the action in an auction often comes very near the end of the bidding.
- **Means of payment.** When you buy a good on eBay, you are not able to simply give the seller some cash. The means of payment accepted by the seller are indicated in the auction information. In many cases, sellers use an electronic payment system, such as PayPal.
- **Means of delivery.** The seller must have a way of shipping the good to you—perhaps via FedEx or another package delivery service. The auction specifies who pays the cost of shipping.
- **Reserve price.** The seller will frequently specify a minimum price, called a reserve price. As a potential buyer, however, you will not see the actual reserve price; the only information you will see is whether or not the reserve price has been reached. A natural reserve price for the seller is her valuation of the good. (More precisely, the reserve price would also include the fee that the seller must pay to eBay in the event of a sale. The only reason for a seller to set a higher valuation is if she thinks she might do better trying to auction the good again at some point in the future rather than settling for a low price in the current auction.)

You participate as a buyer in an eBay auction by placing a bid. For some products, you also have an option of clicking "Buy It Now," where you can purchase the good immediately. In other words, sellers sometimes make a take-it-or-leave it offer as well as offer an auction. To understand the details of the bidding process, look first at the description of how to bid on eBay:

Once you find an item you're interested in, it's easy to place a bid. Here's how:

1. *Once you're a registered eBay member, carefully look over the item listing. Be sure you really want to buy this item before you place a bid.*

2. *Enter your maximum bid in the box at the bottom of the page and then click the Place Bid button.*
3. *Enter your User ID and password and then click the Confirm Bid button. That's it! eBay will now bid on your behalf up to the maximum amount you're willing to pay for that item. You'll get an email confirming your bid. At the end of the listing, you'll receive another email indicating whether you've won the item with an explanation of next steps.* [1]

Because participants in an eBay auction are not all present to bid at the same time, eBay bids for you. All you have to do is to tell it how much you are willing to pay, and eBay takes over. This is known as "proxy bidding" or "automatic bidding."

The exact way in which eBay bids for you is not transparent from this description. It works as follows. Once you input your maximum bid, eBay compares this to the highest existing bid. If your maximum bid is higher than the existing highest bid, then eBay raises the bid by an increment on your behalf. Unless someone bids more, you will win the auction. If someone does bid more (the maximum bid exceeds the highest bid), then you, by proxy, will respond. In this way, the highest bid increases. This process ends with the item going to the bidder with the highest maximum bid. However, the buyer does not pay the amount of the maximum bid. The buyer pays the amount of the next highest bid, plus the increment.

Let us see how this works through an example. Suppose there are two buyers who put in maximum bids of $100 and $120 for a cell phone. Suppose that the increment is $1 and the bidding starts at $50. Because the maximum bids exceed $50, the highest bid will increase by increments of $1 until reaching $100. At this point, the higher of the two maximums, $120, will cause the highest bid to increase by another increment to $101. After this, there is no further action: the other bidder effectively drops out of the auction. The item goes to the buyer who bid $120, and he pays a price of $101 (provided this exceeds the seller's reserve price).

A Decision Rule for Bidding on eBay

Now that you understand how the auction works, you must decide how to bid. Suppose there is only one auction for the good you want (rather than multiple sellers of similar goods). In this case, there is a remarkably simple decision rule to guide your bidding.

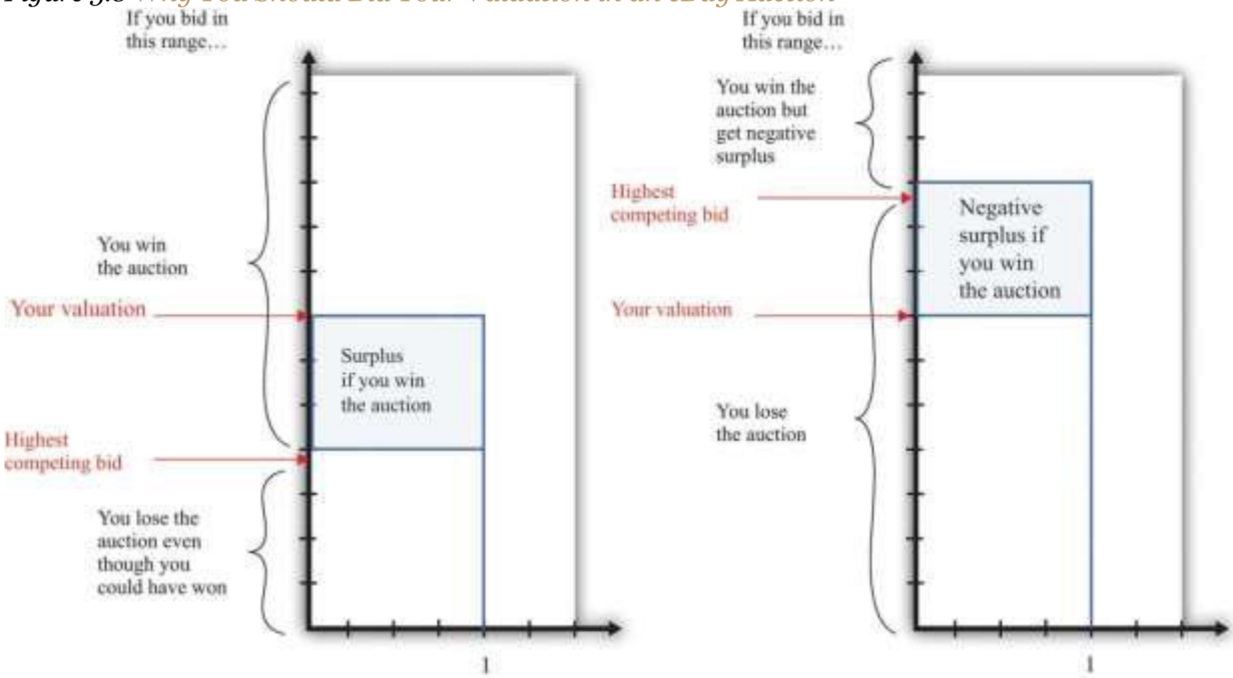
1. Decide on your valuation of the good—that is, the most you would be willing to pay for it.
2. Bid that amount.

This seems surprising. Your first reaction might well be that it is better to bid less than your valuation. But here is the key insight: the amount you actually pay if you win the auction doesn't depend on your bid. Your bid merely determines whether or not you win the auction.

If you pursue this strategy and win the auction, you will gain some surplus: the amount you pay will be less than the valuation you place on the good. If you don't win the auction, you get nothing. So winning the auction is better than not winning. If you bid more than your valuation, then there is a chance that you will have to pay more than your valuation. In particular, if the second-highest bidder puts in a bid that exceeds your valuation, then you will lose surplus. So this does not seem like a good strategy. Finally, if you bid less than your valuation, there is a chance that you won't win the auction even though you value the good more highly than anyone else. Therefore, you will lose the chance of getting some surplus. This is also not a good strategy.

Figure 5.8 "Why You Should Bid Your Valuation in an eBay Auction" illustrates one way of thinking about this. There are two possibilities: either your valuation is bigger than the highest competing bid or your valuation is smaller than the highest competing bid. We don't have to know anything about how other bidders are making their decisions. Part (a) of Figure 5.8 "Why You Should Bid Your Valuation in an eBay Auction" shows the first case. Here, there is a risk that you will lose out on surplus that you could have received. If you bid below the competing bid, you will lose the auction and hence lose out on the surplus. The surplus is the difference between your valuation and the competing bid, minus the increment.

Figure 5.8 Why You Should Bid Your Valuation in an eBay Auction



These illustrations show why you should always bid your valuation on eBay.

Part (b) of Figure 5.8 "Why You Should Bid Your Valuation in an eBay Auction"shows the case where your valuation is less than the competing bid. Here the risk is that you will win the auction and then regret it. If you bid an amount greater than the competing bid, then you will win the auction, but the amount you will have to pay exceeds your valuation. Your loss is the difference between the competing bid and your valuation, plus the increment.

Although automatic bidding by proxy sounds very fancy, the eBay auction is really the same as "English auctions" that are familiar from television and movies. In an English auction, an auctioneer stands at the front of the room and invites bids. Bidding increases in increments until all but one bidder drops out. The winning bidder pays the amount of his bid. The winning bidder therefore pays an amount equal to the highest competing bid, plus the increment, just as in the eBay auction. The amount that she wins is her valuation minus the price she pays, just as in the eBay auction.

The bidding in an English auction can be exciting to watch; you can also have the excitement of seeing how bids evolve on eBay (at least if you are willing to keep logging on and hitting the "refresh" button). But we could also imagine that eBay could do something simpler. It could carry out a "Vickrey auction," which is named after its inventor, the Nobel-prize-winning economist William Vickrey. In a Vickrey auction, the auctioneer (1) collects all the bids, (2) awards victory in the auction to the highest bidder, and (3) makes this person pay an amount equal to the *second-highest* bid.

The Vickrey auction sometimes goes by the more technical name of a second-price sealed-bid auction. Most people think this kind of auction sounds very odd when they first hear of it. Why make the winner pay the second-highest bid? Yet it is almost exactly the same as the eBay auction or the English auction. In those auctions, as in the Vickrey auction, the winner is the person with the highest bid, and the winner pays the amount of the second-highest bid (plus the increment). The only difference is that there is no increment in the Vickrey auction; because the increment is typically a very small sum of money, this is a minor detail.

Selling on eBay

Sellers on eBay typically provide information on the product being sold. This is often done by creating a small web page that describes the object and includes a photograph. Sellers also pay a listing fee to eBay for the right to sell products. They also specify the costs for shipping and handling. After the sale is completed, the buyer and the seller communicate about the shipping, and the buyer makes a payment. Then the seller ships the product, pays eBay for the right to sell, and pockets the remainder.

As we mentioned previously, the buyer can provide feedback on the transaction with the seller. This feedback is important to the seller because transactions require some trust. A seller who has built a reputation for honesty will be able to sell more items, potentially at a higher price.

An Economic Perspective on Auctions

As an individual participating in an auction, you have two concerns: (1) whether or not you win and (2) how much you have to pay. As economists observing from outside, there are other perspectives. Auctions play a very valuable role in the economy. They represent a leading way in which goods are bought and sold—that is, they represent a mechanism for trade.

As we have already stated, voluntary trade is a good thing because it creates value in the economy. Every transaction allows a good or service to be transferred from someone who values it less to someone who values it more. The English auction, such as on eBay, is attractive to economists because it does something more. It ensures that the good or the service goes to the person who values it the most—that is, it ensures that the outcome is efficient. It also has the fascinating feature that it induces people to reveal their valuations through their bids.

eBay is just one example of the many auctions you could participate in. There are auctions for all types of goods: treasury securities, art, houses, the right to broadcast in certain ranges of the electromagnetic spectrum, and countless others. These auctions differ not only in terms of the goods traded but also in their rules. For example, firms competing for a contract to improve a local road may submit sealed bids, with the contract going to the lowest bidder to minimize the costs of the project. Of course, other elements of the bid, including the reputation of the bidder, may also be taken into account.

Complications

The eBay auction sounds almost too good to be true. It is easy to understand, brings forth honest bids, and allocates the good to the person who values it the most. Are there any problems with this rosy picture?

Multiple Sellers

Suppose you have a video-game system for sale. You can put it up for auction on eBay, but you must be aware that many other people could be listing the identical item. What will happen?

First, your potential buyers will most likely look (and bid) at multiple auctions, not only your auction. Second, potential buyers will not be eager to bid in your auction. After all, if they don't win your auction, they can always hope for better luck in another auction. It follows that buyers may decide it is no longer such a good strategy to bid their valuations. Buyers who bid their valuations might end up paying a high price if they win an auction where someone else placed a relatively high bid. Such buyers might be more successful taking the chance of losing one auction and winning another in which the bidding is lower. As buyers monitor other auctions, they will also start to get a sense of how much other people are willing to pay and will adjust their bidding accordingly.

Unfortunately, we can't give you such simple advice about what to do as a buyer in these circumstances. It is not easy to develop the best bidding strategy. In fact, problems like this can be so hard that even expert auction theorists have not fully worked them out.

Tacit Collusion

Another concern is that bidders might want to find some way to collude. As a simple example, suppose there are three bidders for a good with an increment of $1. One bidder has a valuation of $50, one has a valuation of $99, and one has a valuation of $100. In an eBay auction, the winning bid would be $100, but the winner would end up with no surplus (because he would pay $99 plus the $1 increment). Now suppose that the two high-value bidders make an agreement. As soon as the third bidder drops out, they toss a coin. If it comes up heads, Mr. $99 drops out. If it comes up tails, Ms. $100 drops out.

This means that with 50 percent probability, Ms. $100 wins, pays $51, and gets a surplus of $49. With 50 percent probability Mr. $99 wins, pays $51, and gets a surplus of $48. Both buyers prefer this. It's certainly better for Mr. $99, who had no chance of winning before. It is also better for Ms. $100 because even though she may no longer win the auction, she stands to get some surplus if she does win. Of course, the seller wouldn't like this arrangement at all. And the dispassionate economist observing from afar doesn't like it either because sometimes the good may not go to the person who values it the most.

Explicit collusion of this type may very well be illegal, and it is also very hard to carry out. Yet it may be possible for buyers to collude indirectly, and there is speculation that such collusion is sometimes observed on eBay.

The Winner's Curse

We have been supposing throughout that potential buyers know their own valuations of the good being auctioned. In most circumstances, this seems reasonable. Valuations are typically a personal matter that depend on the tastes of the individual buyer.

Occasionally, however, a good with an objective monetary value that is*unknown* to potential buyers may be auctioned. A classic example is the drilling rights to an oilfield. There is a certain amount of oil in the ground, and it will earn a certain price on the market. However, bidders do not know these values in advance and must make their best guess.

It is easiest to see what can happen here with a numerical example. Suppose the true (but unknown) value of an oilfield is $100 million. Suppose there are five bidders, whose guesses as to the value of the oilfield are summarized in Table 5.1 "Valuations of Different Bidders in a Winner's Curse Auction". Notice that these bidders are right on average, but two overestimate the value of the field, and two underestimate it. Imagine that the bidders decide to follow the strategy that we recommended earlier and bid up to their best guess. Bidder E will win. He will have to pay the second-highest bid of $105 million, which is more than the oilfield is worth. He will lose $5 million.

Table 5.1 Valuations of Different Bidders in a Winner's Curse Auction

	Bidder A	**Bidder B**	**Bidder C**	**Bidder D**	**Bidder E**
Valuation ($ million)	90	95	100	105	110

The problem here is that the person who will win the auction is the person who makes the *worst overestimate* of the value of the field. Evidently it is not a good strategy in this auction to bid your best guess. You should recognize that your best guess may be inaccurate, and if you overestimate badly, you may win the auction but lose money. This phenomenon is known as the winner's curse. Your best strategy is therefore to bid less than you actually think the oilfield is worth. But how much less should you bid? That, unfortunately, is a very hard question for which there is no simple answer. It depends on how accurate you think your guess is likely to be *and* how accurate you think other bidders' guesses will be.

KEY TAKEAWAYS

- On eBay, the best strategy is to bid your true valuation of the object.
- Auctions, like eBay, serve to allocate goods from sellers to buyers.
- If the winner's curse if present, then you will want to bid less than your estimate of the value of the object.

CHECKING YOUR UNDERSTANDING

1. Suppose you bid less than your valuation on eBay. Explain how you could do better by bidding a little more.

2. Why didn't the winner's curse have an effect on your bidding in eBay?

[1] eBay Inc., "Help: How to Bid," accessed March 14, 2011,http://pages.ebay.com/help/new/contextual/bid.html.

5.3 Supply and Demand

LEARNING OBJECTIVES

1. What is the outcome of an auction with a large number of buyers and sellers?
2. What is the market demand curve?
3. What is the market supply curve?
4. What is the equilibrium of perfectly competitive markets?

An auction mechanism such as eBay is a natural thing for a seller to use if there are a large number of potential buyers for a good. But what happens if there is also a large number of potential sellers? In this section, we consider what might happen when we have a large number of potential buyers *and* a large number of potential sellers of a good.

We have already explained that it is very difficult to analyze what would happen on eBay when there are multiple buyers and sellers, but we can make a better guess about what will happen on a site like craigslist. As a buyer, you will look for the lowest price out there, bargain with sellers who post high prices, or both. As a seller, you would look at the prices posted by others and realize that you probably should set your price fairly close to those prices. In addition, we have some evidence that can help us understand the likely outcome in a world of many buyers and many sellers. It comes from looking at "double oral auctions." "Double" refers to the fact that there are a large number of buyers and sellers. "Oral" refers to the way in which the auction is conducted.

Double Oral Auctions

In a double oral auction, there is a large number of buyers, each of whom potentially has a different valuation of the good. There is also a large number of sellers, each of whom potentially has a different valuation of the good. Buyers and sellers negotiate with each other, one on one. If they cannot agree to a deal, either party can move on at any time and try to find someone else to bargain with.

Until quite recently, auctions such as this were common in many financial markets and commodity markets. These markets sometimes go by the name *pit markets* because buyers and sellers meet in a frenzy of activity in a trading area called the pit. Traders can hear and see the negotiations of others and often have access to the prices at which deals have been done. This means that both buyers and sellers have lots of information about what price is prevailing in the market.[1]

Economists have also conducted experiments in which they have put people in simulated pit markets to find out how they behave. The result is quite remarkable, but before we explain what happens, we need a framework to help us think about such markets.

Demand: Many Buyers

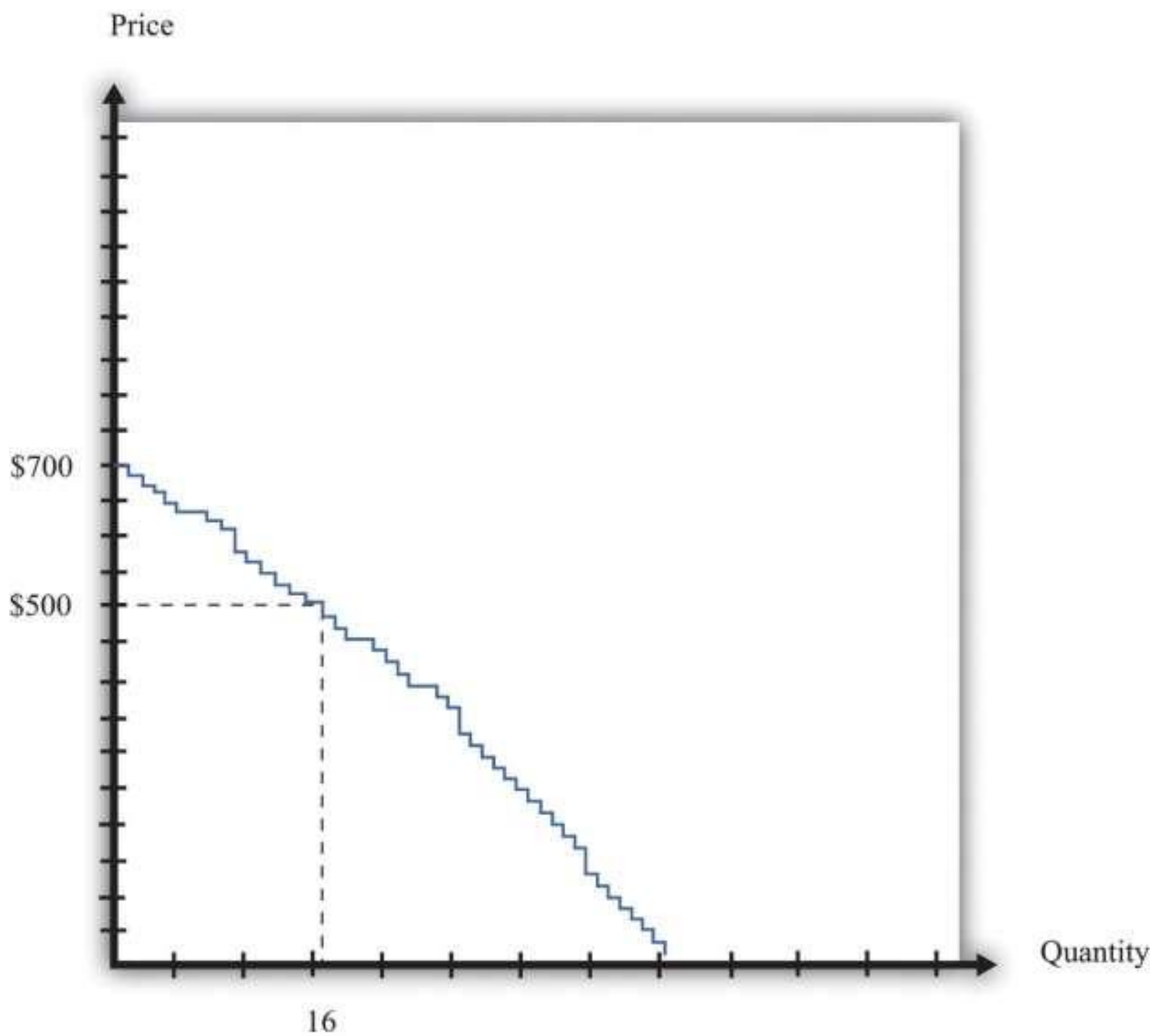
Suppose we are considering the purchase of a gaming console by a group of buyers. Each potential buyer has his own valuation. Some might be willing to pay as much as $700. Others might be willing to spend much less. After all, how much you are willing to pay for a gaming console depends on your income, how much you like playing, what equipment you currently own, and so on.

Each potential buyer has a unit demand curve like the one we saw in Figure 5.3 "The Buyer’s Valuation". We can add these unit demand curves together to get a picture of demand in the entire market: the market demand curve. For example, suppose only one person is willing to buy if the price is $700. However, suppose there is another buyer with a valuation of $660. If consoles were on sale for $660, then both individuals would want to purchase. At $660, in other words, the quantity demanded is 2. Perhaps the buyer with the next-highest valuation is willing to pay $640. If the price is $640, the quantity demanded is 3. Figure 5.10 "Obtaining the Market Demand Curve" shows what happens when we add together all these unit demand curves. The result is a downward sloping relationship that shows us how many units would be demanded at any given price.

Toolkit: Section 17.9 "Supply and Demand"

The market demand curve tells us how many units of a good or a service will be demanded at any given price. The market demand curve is obtained by adding together the individual demand curves in the economy and obeys the law of demand: as the price decreases, the quantity demanded increases.

Figure 5.10 Obtaining the Market Demand Curve



*We can add together the unit demand curves of different individuals in the economy to get the **market demand curve**.*

Supply: Many Sellers

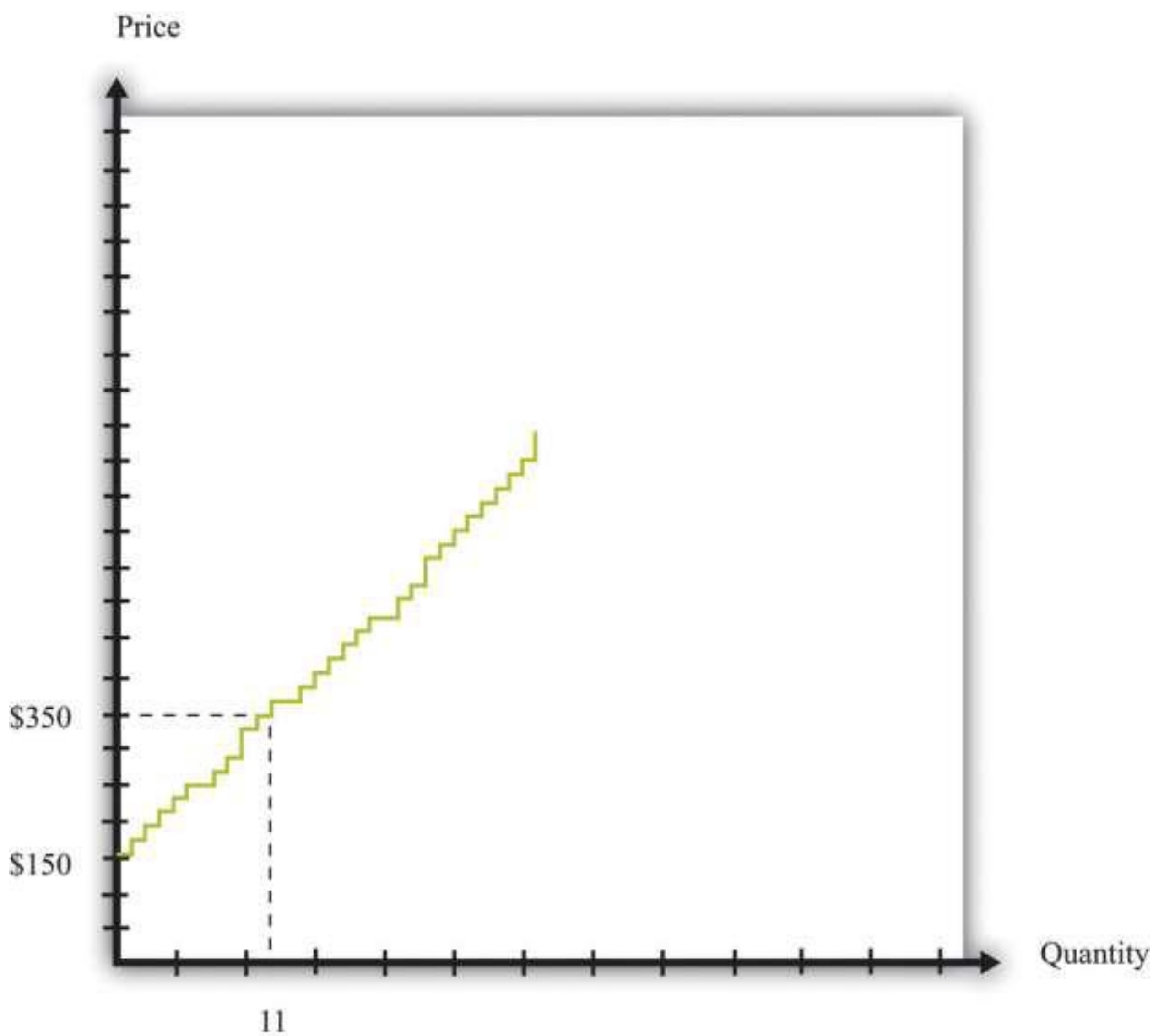
We saw earlier that each potential seller has a unit supply curve. If the price is less than a seller's valuation, she will not sell the good, but when the price becomes greater than her valuation, she will be willing to sell. Just as we added together the unit demand curves to get the market demand curve, so too can we add together the unit supply curves to get the market supply curve.

Toolkit: Section 17.9 "Supply and Demand"

The market supply curve tells us how many units of a good or a service will be supplied at any given price. The market supply curve is obtained by adding together the individual supply curves in the economy and typically slopes upward: as the price increases, the quantity supplied to the market increases.

In Figure 5.11 "Obtaining the Market Supply Curve", we see that the lowest valuation in the market is $150. There is one seller willing to sell a console at that price. As the price increases, more and more sellers will find the price attractive and will want to sell. For example, there are 11 potential sellers with a valuation less than $350. Thus, at this price, 11 consoles will be supplied to the market.

Figure 5.11 Obtaining the Market Supply Curve



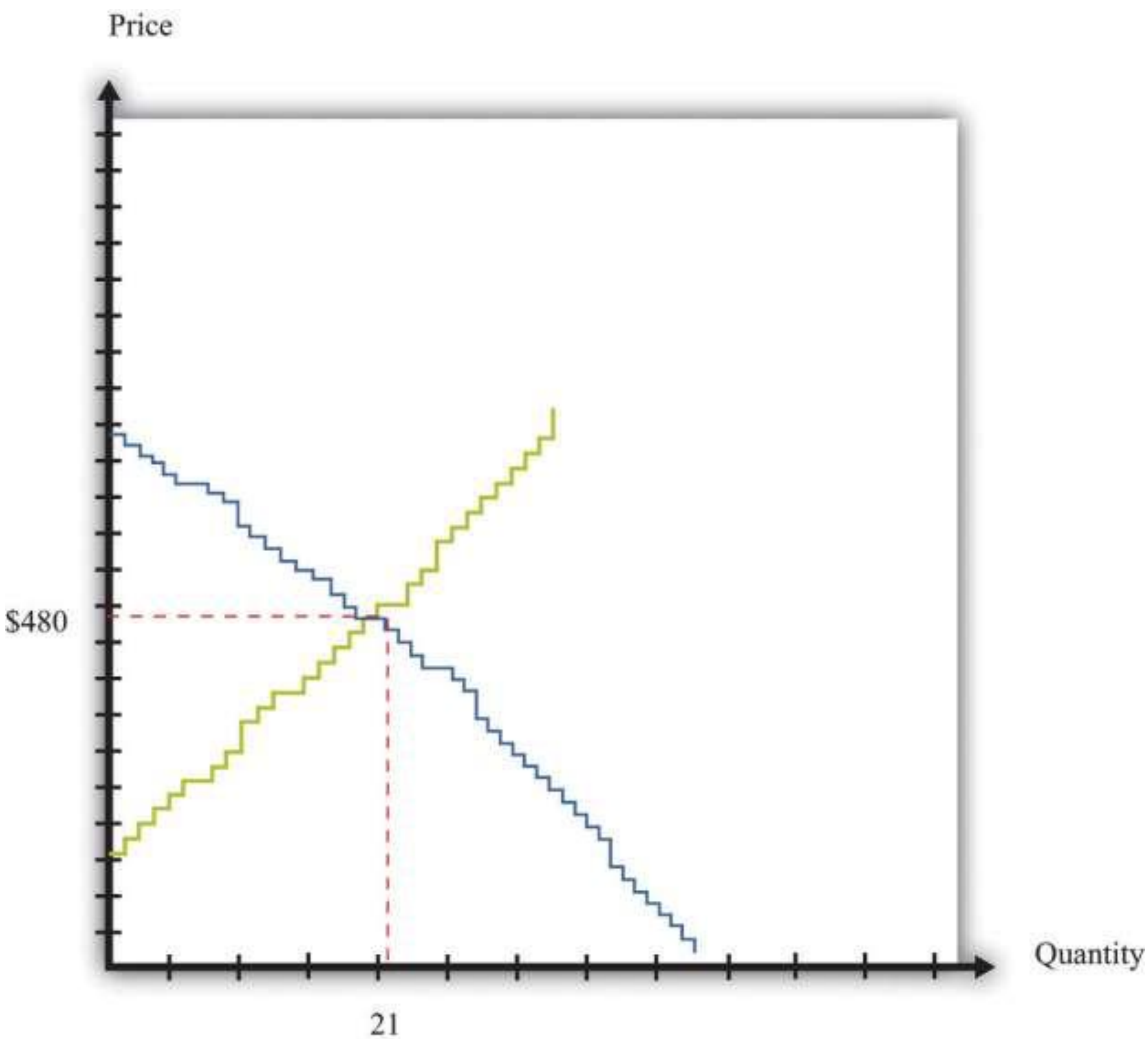
We can add together the unit supply curves of different individuals in an economy to get the ***market supply curve****.*

Equilibrium

Figure 5.10 "Obtaining the Market Demand Curve" and Figure 5.11 "Obtaining the Market Supply Curve" tell us the number of buyers willing to buy and the number of sellers willing to sell at each price.

Figure 5.12 "Market Equilibrium" shows what happens if we combine the demand curve and the supply curve on the same diagram. One point jumps out at us: the place where the demand and supply curves meet. In our example, this is at $480 and a quantity of 21 units. At this point, *the number of buyers with a valuation greater than the price is the number of sellers with a valuation less than the price*. If buyers and sellers were presented with this price, none would find themselves unable to transact. At this price, there is an exact match between the number of buyers and sellers.

Figure 5.12 Market Equilibrium



*In this figure, we combine the demand and supply curves to find the **equilibrium** price and quantity in the market.*

Toolkit: Section 17.9 "Supply and Demand"

Equilibrium in a market refers to an equilibrium price and an equilibrium quantity and has the following features:

- Given the equilibrium price, sellers supply the equilibrium quantity.
- Given the equilibrium price, buyers demand the equilibrium quantity.

Equilibrium is not only a point on a graph. It is a prediction about a possible outcome in a situation where a large number of buyers and sellers meet with the possibility of trading. It seems plausible that in a situation where a large number of buyers and sellers can meet and trade with each other, most will end up trading at or near the equilibrium price.

The equilibrium outcome is plausible because, at any other price, there will be a mismatch of buyers and sellers. Imagine, by contrast, that the buyers and sellers of our example are currently trading at $600, well above the equilibrium price of $480. At this high price, many more people want to sell than want to buy. Buyers would rapidly realize that they are in a strong bargaining position: if many sellers want your business, you can make them compete with each other and force price decreases. In fact, whenever the price is above equilibrium, the mismatch of buyers and sellers will tend to decrease prices.

By similar reasoning, a price of, say, $400 would also result in a mismatch between buyers and sellers. In this case, though, there are more people who want to buy than sell. Sellers can make buyers compete with each other, leading to price increases. At any price below the equilibrium price, prices will tend to increase.

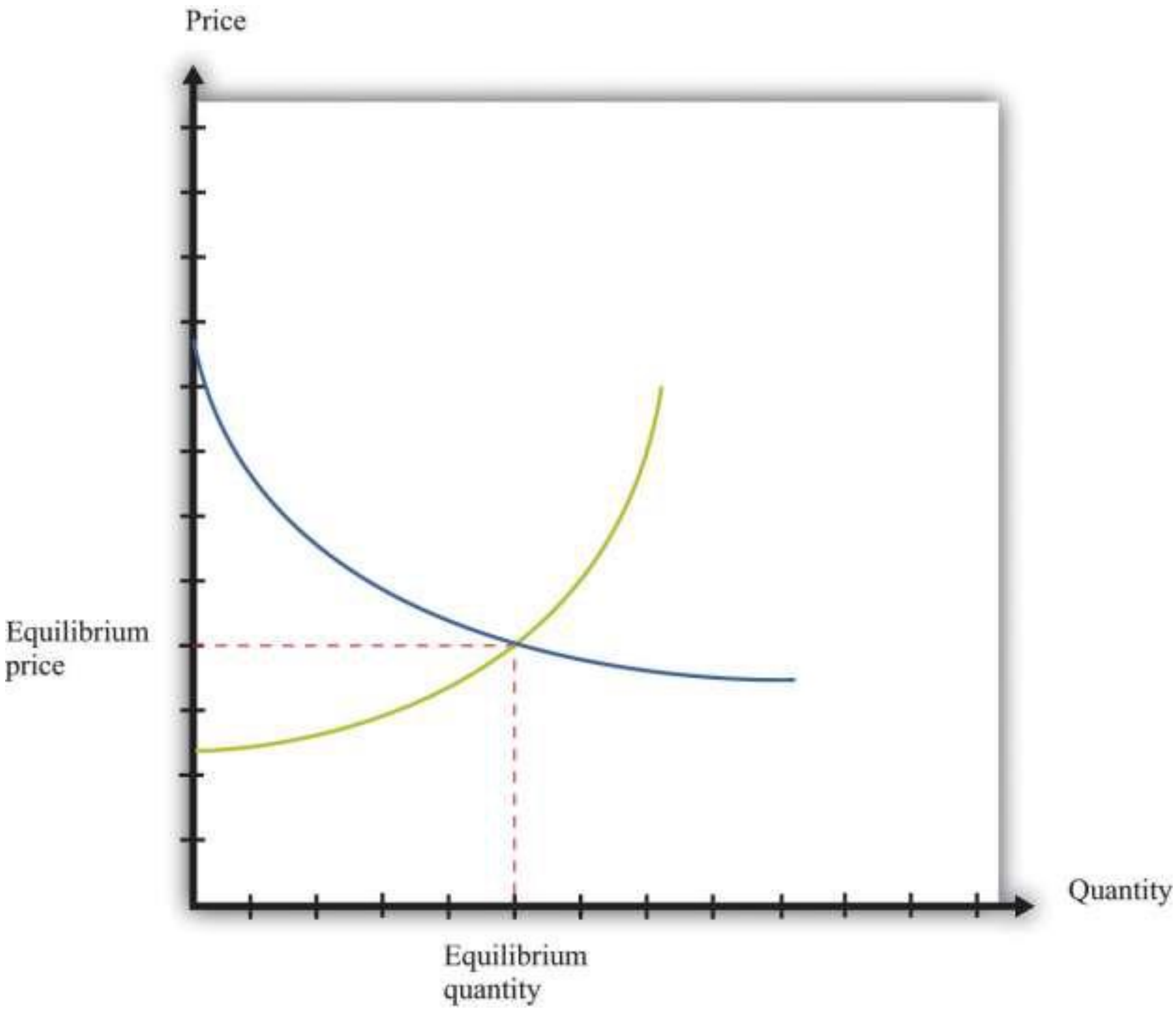
Perfectly Competitive Markets

Economists formalize the intuition we have just developed with the most famous framework in all of economics: supply and demand. [2]

Toolkit: Section 17.9 "Supply and Demand"

Supply and demand is a framework we use to explain and predict the equilibrium price and quantity of a good. This framework illustrates the willingness to sell (market supply) and buy (market demand) on a graph with price on the vertical axis and units of the good or the service on the horizontal axis. A point on the market supply curve shows the quantity that suppliers are willing to sell for a given price. A point on the market demand curve shows the quantity that demanders are willing to buy for a given price. The intersection of supply and demand determines the equilibrium price and quantity that will prevail in the market. A basic supply-and-demand framework is shown in Figure 5.13 "Supply and Demand".

Figure 5.13 Supply and Demand



When we have a large number of buyers and sellers of an identical good or service, the equilibrium price and quantity are determined by the intersection of the supply and demand curves.

The *position* of the demand curve depends on many things, such as income and the prices of other goods. A change in any of these will cause the entire demand curve to shift. Likewise, the *position* of the supply curve depends on factors such as a supplier's costs. A change in these will cause the entire supply curve to shift. When one (or both) of the curves shifts, the equilibrium price and quantity change.

Experience with double oral auctions, both in the laboratory and in actual pit markets, tells us that trading will typically settle down close to the equilibrium price within a relatively short period of time. In a situation where there is a large number of people buying and selling an identical good, we say that we have acompetitive market. We expect that most trades will take place at or close to the equilibrium price, and the quantity traded will be approximately equal to the equilibrium quantity. In fact, even when the number of participants in the auction is relatively small, we often find that a double oral auction still gets close to this equilibrium price and quantity. This is the remarkable finding that we mentioned earlier: in a double oral auction, the number of transactions and the prices of these transactions are usually very close to the equilibrium predicted by supply and demand.

Toolkit: Section 17.9 "Supply and Demand"

Suppose a market has the following two characteristics:

1. There are many buyers and many sellers, all of whom are small relative to the market.
2. The goods being traded are perfect substitutes.

In this case we say that we have a competitive market (sometimes called a perfectly competitive market). Buyers and sellers both take the price as given. This means that they think their actions have no effect on the price in the market, which in turn means we can employ the supply-and-demand framework.

The Gains from Trade in Equilibrium

Suppose all the transactions in Figure 5.12 "Market Equilibrium" take place at the equilibrium price of $480. What can we say about the surplus received by buyers and sellers? Each individual transaction looks like those we examined inChapter 5 "eBay and craigslist", Section 5.2 "eBay". The total surplus from any given transaction is equal to the difference between the buyer's valuation and the seller's valuation. The buyer surplus is the difference between his valuation and $480. The seller surplus is the difference between the price and her valuation. For example, Figure 5.14 "The Gains from Trade in a Single Transaction in Market Equilibrium" shows the gains from trade if a buyer with a valuation of $630 matches up with a seller whose valuation is $230:

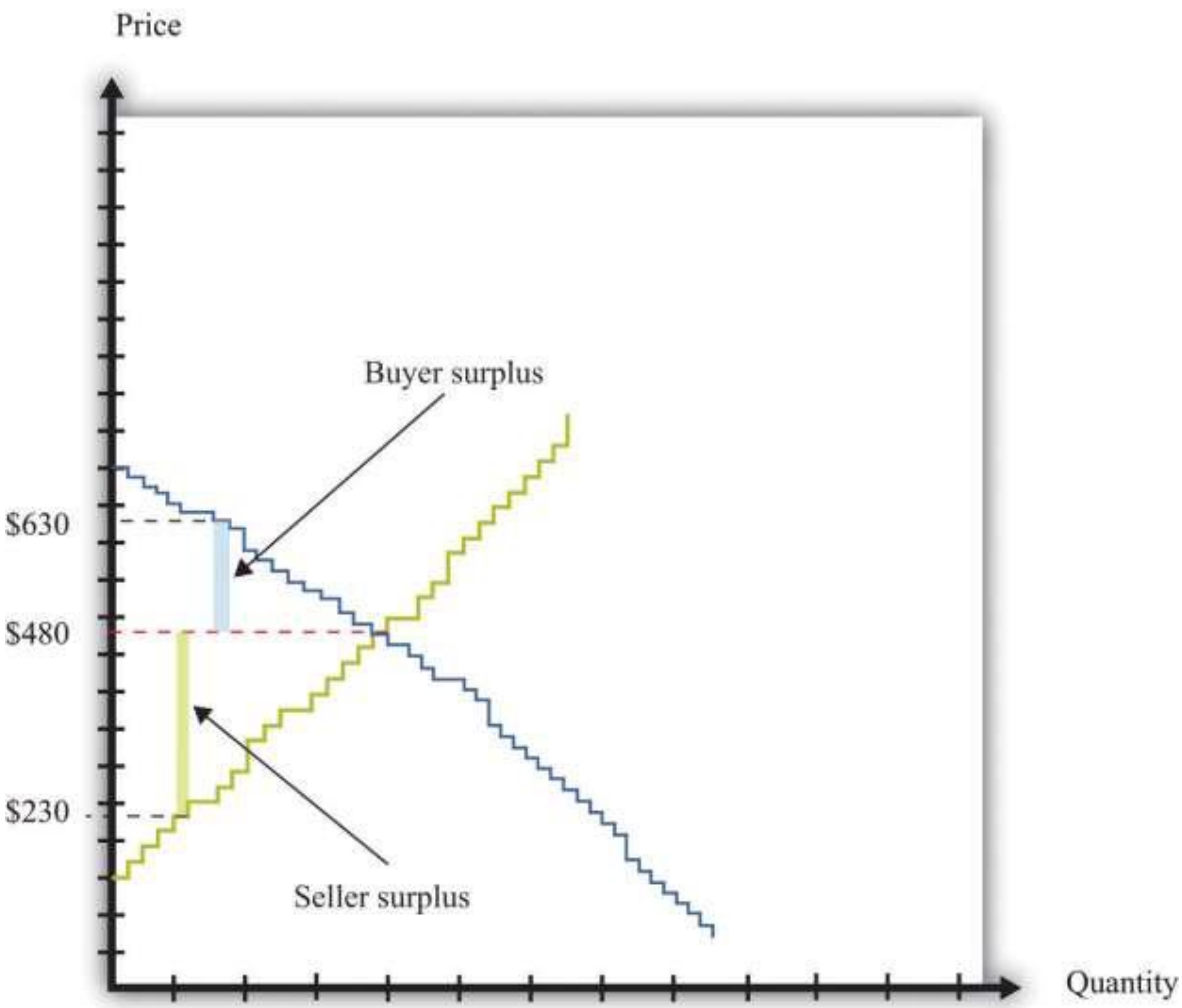
$$\text{buyer surplus} = \$630 - \$480 = \$150, \text{seller surplus} = \$480 - \$230 = \$250,$$

and

$$\text{total surplus} = \$150 + \$250 = \$400.$$

The transaction generates $400 worth of surplus: $150 goes to the buyer, and $250 goes to the seller.

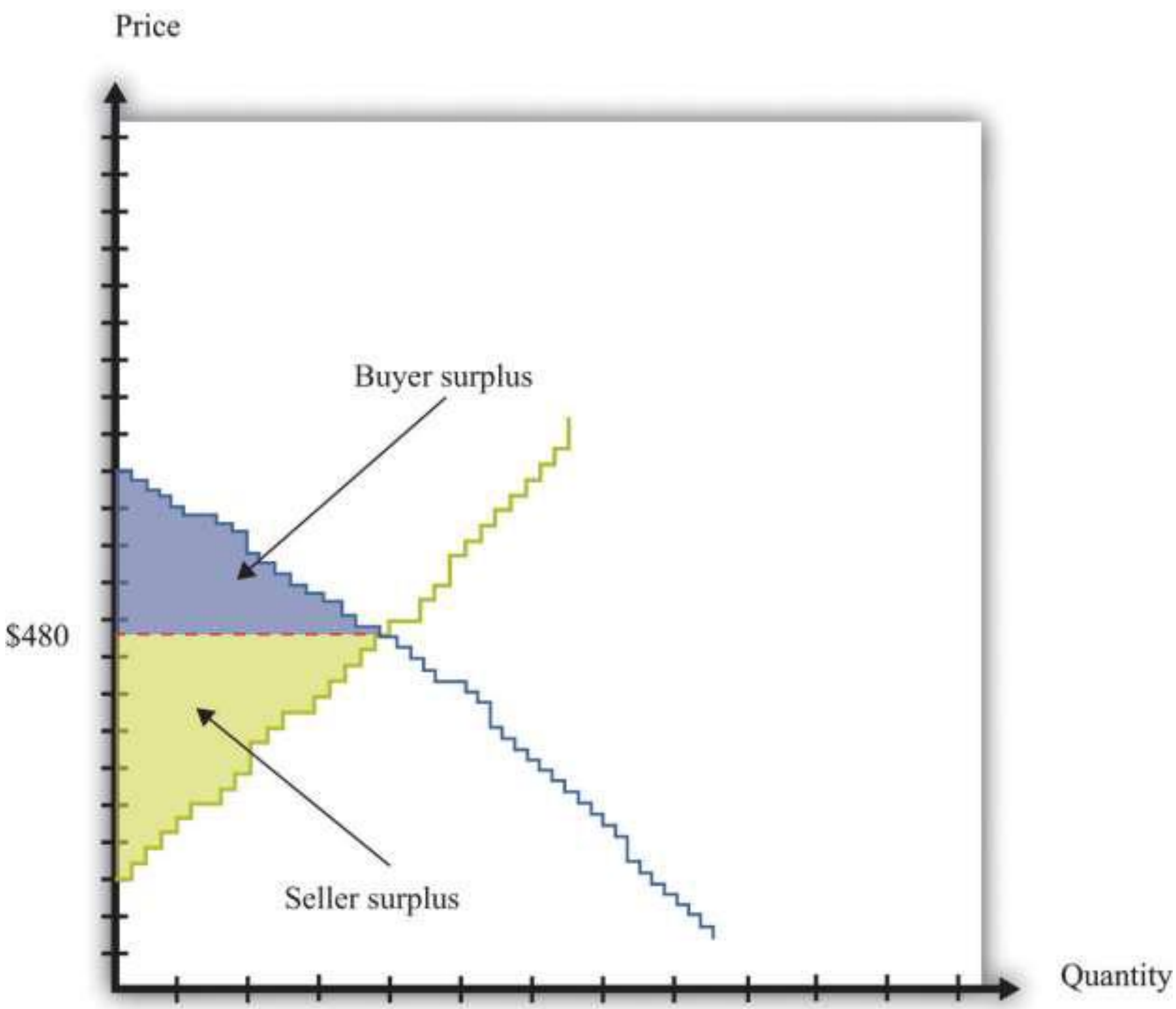
Figure 5.14 The Gains from Trade in a Single Transaction in Market Equilibrium



Each transaction in the market generates surplus.

We could draw exactly the same diagram for all 21 transactions in the market. If we combine them, we would end up with Figure 5.15 "Surplus in Equilibrium". The total surplus accruing to the buyers is equal to the area below the demand curve and above the price. The total surplus accruing to the sellers is equal to the area above the supply curve and below the price. The total surplus—that is, the total gains from trade in this market—is the sum of the buyer surplus in the market and the seller surplus in the market. The total surplus is therefore the area between the supply curve and the demand curve.

Figure 5.15 Surplus in Equilibrium



If we add the surplus from all trades in the market, supposing that they all take place at $480, we obtain the total surplus in the market.

If you look at Figure 5.15 "Surplus in Equilibrium", something else may become apparent to you. *All the gains from trade have been exhausted in the market*. If buyers and sellers trade at the market price, then they manage to achieve all the gains from trade that are possible in this market because

- every transaction that has been carried out has created surplus;
- any further transaction would generate negative surplus.

The first statement is true because all trades are voluntary. We can see that the second statement is true by imagining trying to match up another buyer and seller. All the buyers with valuations greater than $480 have now made a purchase. So every remaining potential buyer has a valuation less than $480. All the sellers with valuations less than $480 have now made a sale. So every potential seller has a valuation greater than $480. It follows that there is no mutually beneficial transaction to be carried out.

This is a truly remarkable result. A market where all potential buyers and sellers take as given the equilibrium price allows all the possible gains from trade to be realized. Thus a market is a very effective mechanism for generating an efficient allocation of resources. This is why economists place so much emphasis on markets and "market solutions" to economic problems. Markets allow buyers and sellers to come together to make mutually beneficial trades. Economists believe that, as far as possible, we should create circumstances in which people can meet and carry out voluntary transactions.

Although this argument for markets is very powerful, we must be careful. Buyers and sellers may benefit from trading, but sometimes other people not involved in the transaction may also be affected. For example, suppose you fill up your car with gas at your local gas station. Presumably, you benefit from this transaction—otherwise you wouldn't have bought the gas. Likewise, the gas station owner benefits from the transaction—otherwise the owner wouldn't have sold it to you. But your purchase will contribute to smog and air pollution when you drive the car, affecting other people in the vicinity. To the extent that you make a contribution to global climate change, your little transaction has the potential to have an effect—a very tiny effect but an effect nonetheless—on everyone else on the planet. As a more positive example, going to college is presumably a mutually beneficial transaction between you and your school. But many others may eventually benefit from your education as well. [3]

KEY TAKEAWAYS

- In a perfectly competitive market, buyers and sellers take the prices as given.
- In the equilibrium of a perfectly competitive market, there are no further gains to trade.
- The outcome of a double oral auction and the supply-and-demand framework are the same.

CHECKING YOUR UNDERSTANDING

1. Look at Figure 5.12 "Market Equilibrium". How could the equilibrium price be greater than $480?

2. Suppose there are two buyers. The first has a demand curve given by *quantity*= 5 − 0.5 × *price*. The second one has a demand curve of *quantity* = 15 − 1.5 ×*price*. What is the market demand at $1? Suppose there is a total supply of 10 units in this market. What is the equilibrium price? How is the surplus allocated?

[1] Chapter 9 "Making and Losing Money on Wall Street" has much more to say about these markets.

[2] The definition is repeated and discussed in more detail in Chapter 7 "Why Do Prices Change?"; we make extensive use of it in other chapters.

[3] In Chapter 13 "Cleaning Up the Air and Using Up the Oil", we consider such uncompensated costs and benefits in detail.

5.4 Production Possibilities

LEARNING OBJECTIVES

1. Where do the gains from trade come from?
2. What determines who produces which good?

So far we have discussed several different ways in which individuals trade with one another, including individual bargaining and Internet sites such as eBay and craigslist. We have considered situations with one seller and one buyer, one seller and many buyers, and many sellers and buyers. But why do we trade so much? Why is trade so central to our lives and indeed to the history of the human race?

On a typical craigslist website, many services are offered for sale. They are listed under categories such as financial, legal, computer, beauty, and so on. If you click on one of these headings and follow one of the offers, you typically find that someone is willing to provide a service, such as legal advice, in exchange for money. Sometimes there are offers to barter: to exchange a service for some other service or for some specific good. For example, we found the following offers listed on craigslist. [1]

> *Hello, I am looking for a dentist/oral surgeon who is willing to remove my two wisdom teeth in exchange for furniture repair and refinishing. If preferred, I'll come to your office and show you my teeth beforehand. Take a look at some of the work I have posted on my web page…quality professional furniture restoration. Bring new life to your antiques!*
>
> *We are new to the area and are looking for a babysitter for casual or part-time help with our three little girls. My husband is a chiropractor and offers adjustments, and I am a vegan and raw foods chef offering either culinary classes or prepared food in exchange for a few hours of babysitting each week.*
>
> *I have a web design company,…I figure I'd offer to barter in this slow economy. If you got something you'd be willing to trade for a website, let me know and maybe we can work something out!*

These offers provide a glimpse into why people trade. Some people are relatively more productive than others in the production of certain goods or services. Hence it makes sense that people should perform those tasks they are relatively good at and then in some way exchange goods and services. These offers reveal both a reason for trade and a mechanism for trade.

As individuals, we are involved in the production of a very small number of goods and services. The person who cuts your hair is probably not a financial advisor. It is unlikely that your economics professor also moonlights as a bouncer at a local nightclub. By contrast, we buy thousands of goods and services—many more goods and services than we produce. *We specialize in production* and *generalize in consumption*. One motivation for trade is this simple fact: we typically don't consume the goods we produce, and we certainly want to consume many more goods than we produce. Yet that prompts the question of why society is organized this way. Why do we live in such a specialized world?

To address this question, we leave our modern, complicated world—the world of eBay, craigslist, and the Internet—behind and study some very simple economies instead. In fact, we begin with an economy that has only one individual. This allows us to see what a world would look like without any trade at all. Then we can easily see the difference that trade makes.

Production Possibilities Frontier for a Single Individual

Inspired by the craigslist posts that we saw earlier, imagine an economy where people care about only two things: web pages and vegan meals. Our first economy has a single individual—we call him Julio—who has 8 hours a day to spend working. Julio can spend his time in two activities: web design and preparing vegan meals. To be concrete, suppose he can produce 1 web page per hour or 2 vegan meals per hour. Julio faces a time allocation problem: how should he divide his time between these activities? [2]

The answer depends on both Julio's productivity and his tastes. We start by looking at his ability to produce web pages and vegan meals in a number of different ways. Table 5.2 "Julio's Production Ability" shows the quantity of each good produced per hour of Julio's time. Julio can produce either 2 vegan meals or 1 web page in an hour. Put differently, it takes Julio half an hour to prepare a meal and 1 hour to produce a web page. These are the technologies—the ways of producing output from inputs—that are available to Julio.

Toolkit: Section 17.17 "Production Function"

A technology is a means of producing output from inputs.

Table 5.2 Julio's Production Ability

Vegan Meals per Hour	Web Pages per Hour
2	1

We could write these two technologies as equations:

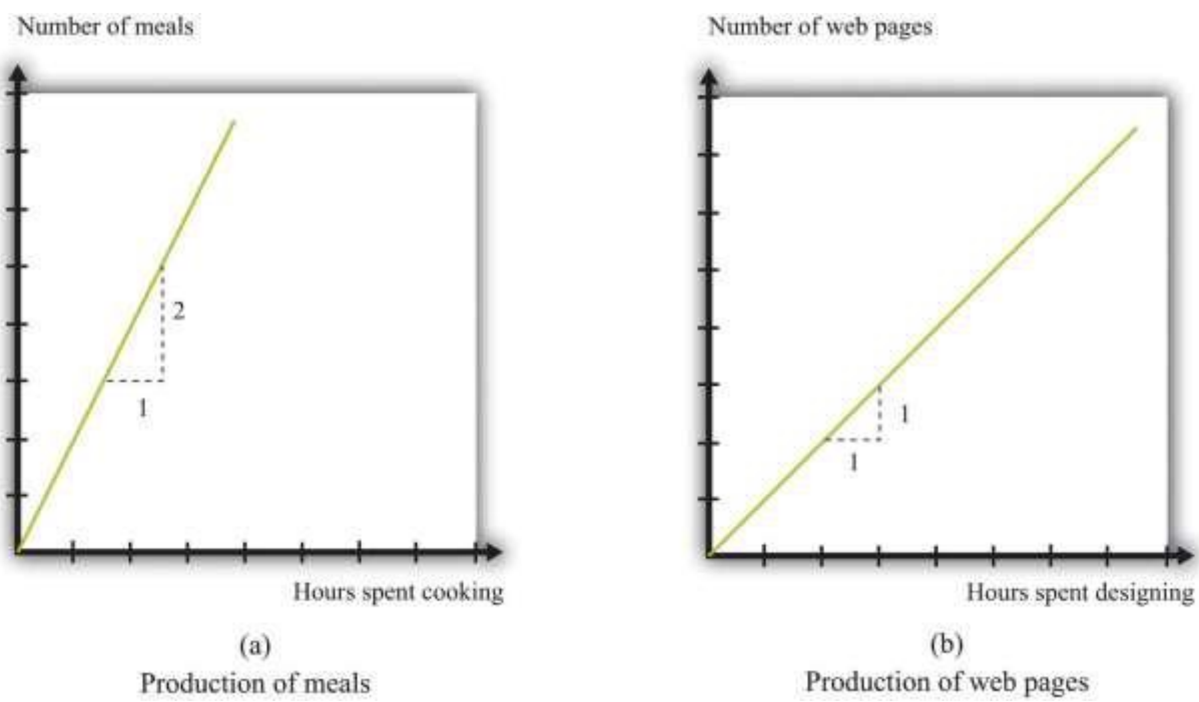
quantity of vegan meals = 2 × hours spent cooking

and

quantity of web pages = hours spent on web design.

Or we can draw these two technologies (Figure 5.16 "Julio's Production Ability"). The equations, the figure, and the table are three ways of showing exactly the same information.

Figure 5.16 Julio's Production Ability



These figures show Julio's technologies for producing vegan meals and producing web pages.

When we add in the further condition that Julio has 8 hours available each day, we can construct his different possible production choices—the combinations of web pages and vegan cuisine that he can produce given his abilities and the time available to him. The first two columns of Table 5.3 "Julio's Production Possibilities" describe five ways Julio might allocate his 8 hours of work time. In the first row, Julio allocates all 8 hours to preparing vegan meals. In the last row, he spends all of his time in web design. The other rows show what happens if he spends some time producing each service. Note that the total hours spent in the two activities is always 8 hours.

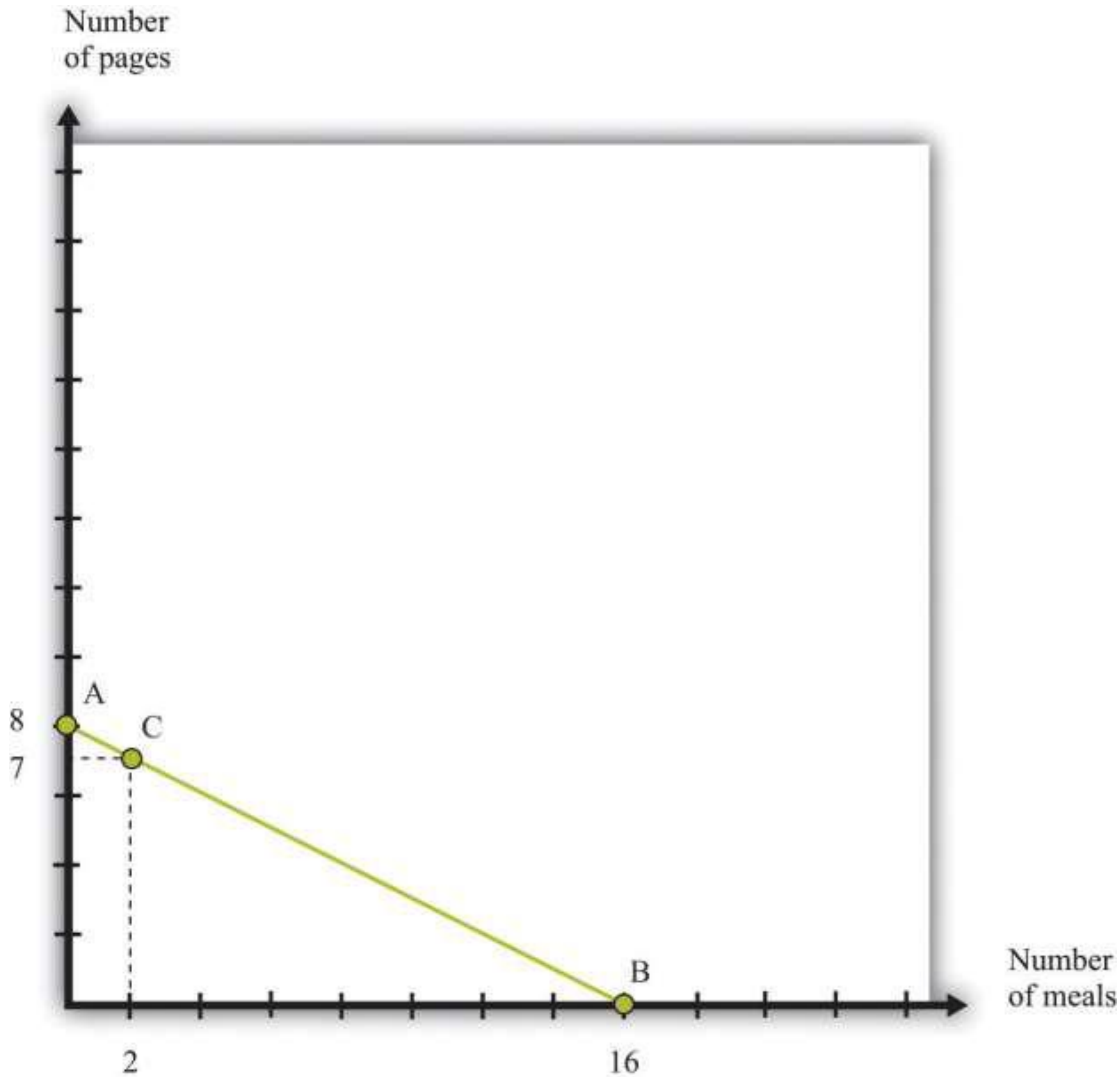
The third and fourth columns provide information on the number of vegan meals and web pages that Julio produces. Looking at the first row, if he works only on vegan meals, then he produces 16 meals and 0 web pages. If Julio spends all of his time designing web pages, then he produces 0 vegan meals and 8 web pages.

Table 5.3 Julio's Production Possibilities

Time Spent Producing		**Goods Produced**	
Vegan Meals	**Web Pages**	**Vegan Meals**	**Web Pages**
8	0	16	0
2	6	12	2
4	4		
6	2		
0	8	0	8

We can also illustrate this table in a single graph (Figure 5.17 "Julio's Production Possibilities") that summarizes Julio's production possibilities. The quantity of vegan meals is on the horizontal axis, and the quantity of web pages is on the vertical axis. To understand Figure 5.17 "Julio's Production Possibilities", first consider the vertical and horizontal intercepts. If Julio spends the entire 8 hours of his working day on web design, then he will produce 8 web pages and no vegan meals (point A). If Julio instead spends all his time cooking vegan meals and none on web design, then he can produce 16 vegan meals and 0 web pages (point B).

Figure 5.17 Julio's Production Possibilities



Julio's production possibilities frontier shows the combinations of meals and web pages that he can produce in an 8-hour day.

The slope of the graph is −1/2. To see why, start at the vertical intercept where Julio is producing only web pages. Suppose that he reduces web-page production by 1 page. This means he will produce only 7 web pages, which requires 7 hours of his time. The hour released from the production of web design can now be used to prepare vegan meals. This yields 2 vegan meals. The resulting combination of web pages and vegan meals is indicated as point C. Comparing points A and C, we can see why the slope is −1/2. A reduction of web-page production by 1 unit (the rise) yields an increase in vegan meals production of 2 (the run). The slope—rise divided by run—is −1/2.

Given his technology and 8 hours of working time, all the combinations of vegan meals and web pages that Julio can produce lie on the line connecting A and B. We call this the production possibilities frontier. Assuming that Julio equally likes both web design and vegan meals and is willing to work 8 hours, he will choose a point on this frontier. [3]

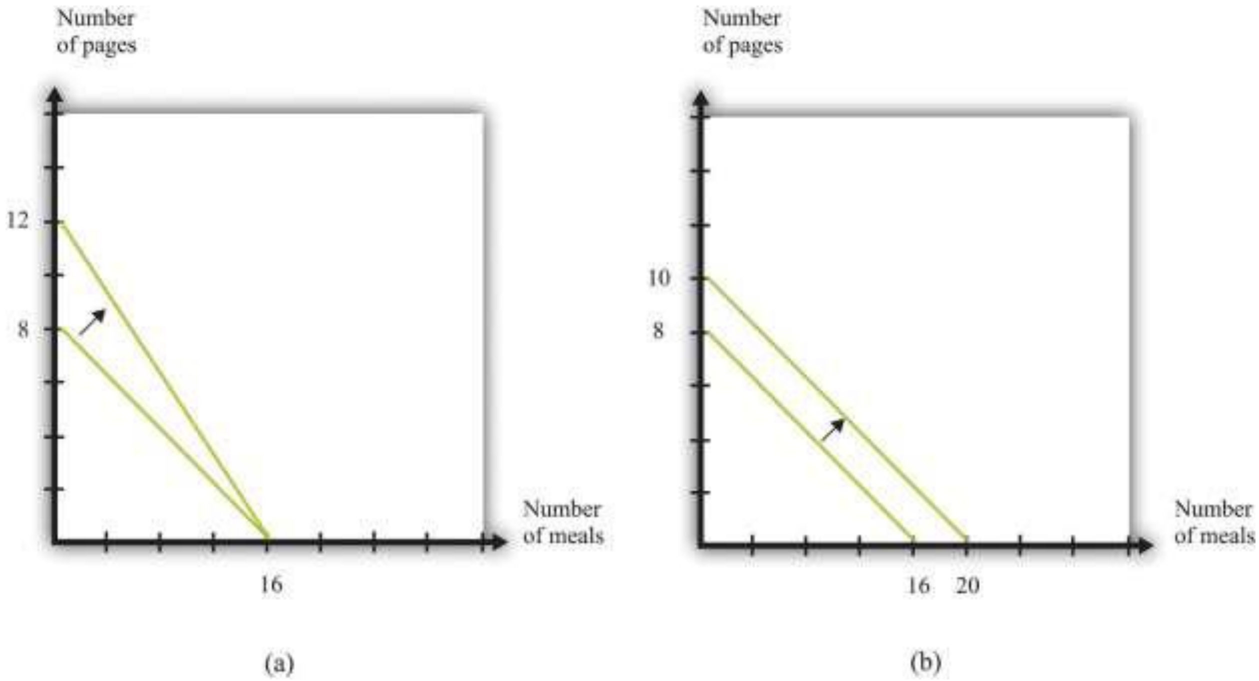
Toolkit: Section 17.12 "Production Possibilities Frontier"

The production possibilities frontier shows the combinations of goods that can be produced with available resources. It is generally illustrated for two goods.

What is the cost to Julio of cooking one more meal? To cook one more meal, Julio must take 30 minutes away from web design. Because it takes 30 minutes to produce the meal, and Julio produces 1 web page per hour, *the cost of producing an additional vegan meal is half of a web page*. This is his opportunity cost: to do one thing (produce more vegan meals), Julio must give up the opportunity to do something else (produce web pages). Turning this around, we can determine the opportunity cost of producing an extra web page in terms of vegan meals. Because Julio can produce 1 web page per hour or cook 2 meals per hour, *the opportunity cost of 1 web page is 2 vegan meals*. The fact that Julio must give up one good (for example, web pages) to get more of another (for example, vegan meals) is a direct consequence of the fact that Julio's time is scarce.

Could Julio somehow produce more web pages and more vegan meals? There are only two ways in which this could happen. First, his technology could improve. If Julio were able to become better at either web design or vegan meals, his production possibilities frontier would shift outward. For example, if he becomes more skilled at web design, he might be able to produce 3 (rather than 2) web pages in 2 hours. Then the new production possibilities frontier would be as shown in part (a) of Figure 5.18 "Two Ways of Shifting the Production Possibilities Frontier Outward".

Figure 5.18 *Two Ways of Shifting the Production Possibilities Frontier Outward*

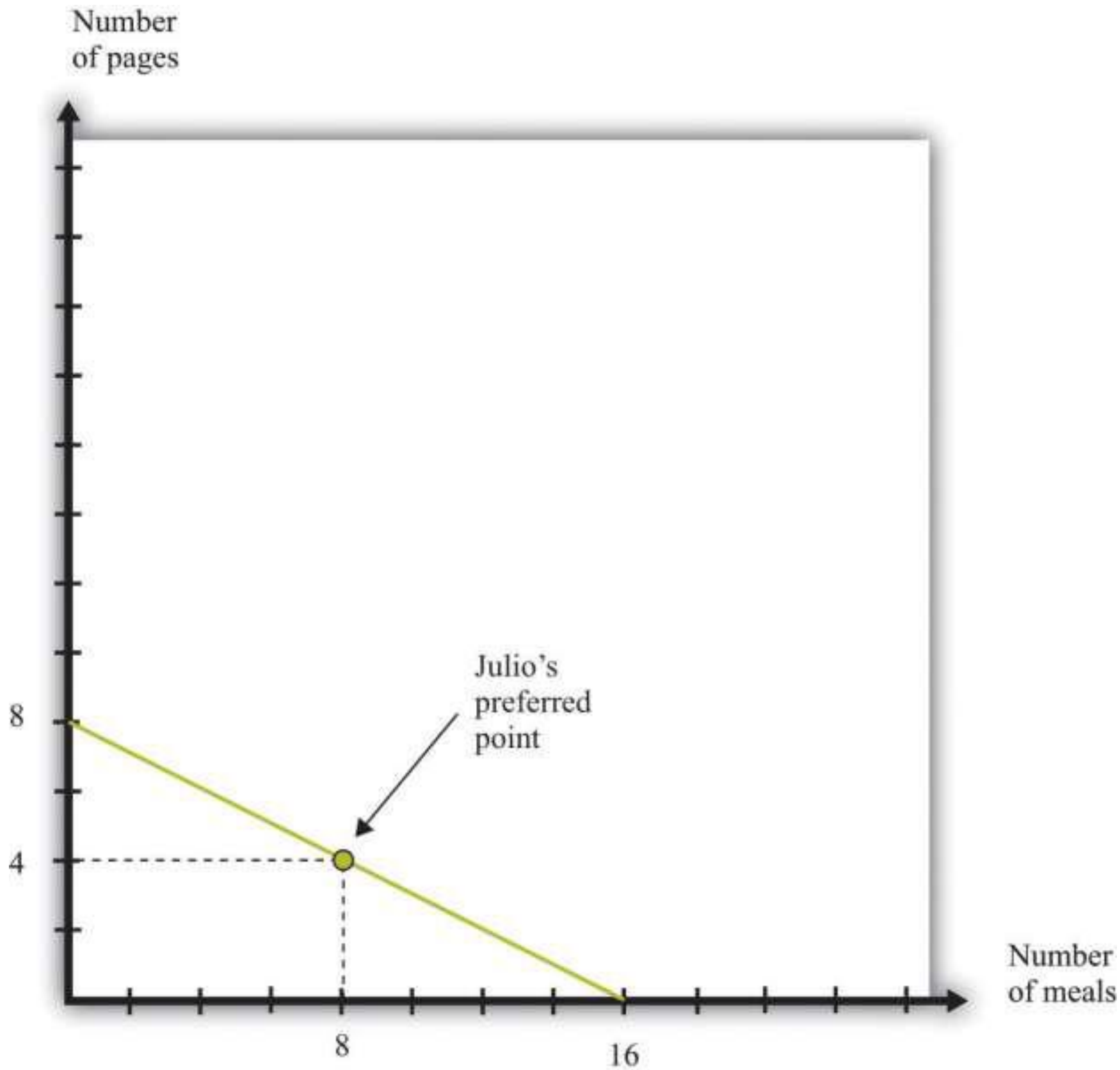


There are two ways in which Julio could produce more in a day: he could become more skilled, or he could work harder.

Alternatively, Julio could decide to work more. We have assumed that the amount of time that Julio works is fixed at 8 hours. Part (b) of Figure 5.18 "Two Ways of Shifting the Production Possibilities Frontier Outward" shows what the production possibilities frontier would look like if Julio worked 10 hours per day instead of 8 hours. This also has an opportunity cost. If Julio works longer, he has less time for his leisure activities.

We have not yet talked about where on the frontier Julio will choose to allocate his time. This depends on his tastes. For example, he might like to have 2 vegan meals for each web page. Then he would consume 4 web pages and 8 meals, as inFigure 5.19 "Julio's Allocation of Time to Cooking Meals and Producing Web Pages".

Figure 5.19 Julio's Allocation of Time to Cooking Meals and Producing Web Pages



If Julio likes to consume pages and meals in fixed proportions (2 vegan dishes for every web page), he will allocate his time to achieve the point shown in the figure.

The Production Possibilities Frontier with Two People

If this were the end of the story, we would not have seen the advertisements on craigslist to trade vegan meals or web pages. Things become more interesting and somewhat more realistic when we add another person to our economy.

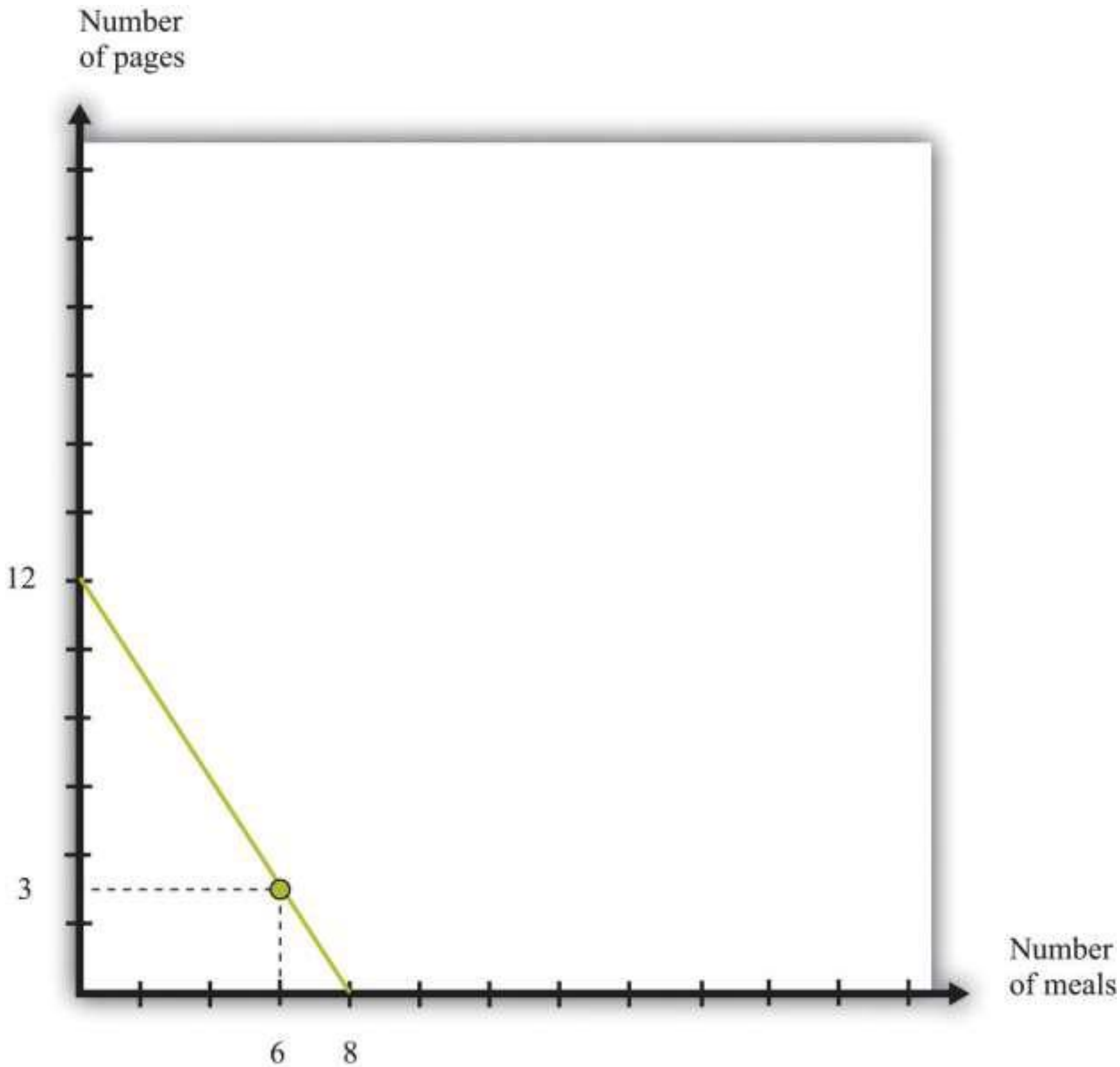
Hannah has production possibilities that are summarized in Table 5.4 "Production Possibilities for Julio and Hannah". She can produce 1 vegan meal in an hour or produce 1.5 web pages in an hour. Hannah, like Julio, has 8 hours per day to allocate to production activities. In Table 5.4 "Production Possibilities for Julio and Hannah" we have also included their respective opportunity costs of producing web pages and vegan meals.

Table 5.4 Production Possibilities for Julio and Hannah

	Vegan Meals per Hour	Web Pages per Hour	Opportunity Cost of Vegan Meals (in Web Pages)	Opportunity Cost of Web Pages (in Vegan Meals)
Julio	2	1	1/2	2
Hannah	1	1.5	3/2	2/3

Table 5.4 "Production Possibilities for Julio and Hannah" reveals that Hannah is more productive than Julio in the production of web pages. By contrast, Julio is more productive in vegan meals. Hannah's production possibilities frontier is illustrated in Figure 5.20 "Hannah's Production Possibilities Frontier". It is steeper than Julio's because the opportunity cost of vegan meals is higher for Hannah than it is for Julio.

Figure 5.20 Hannah's Production Possibilities Frontier



Hannah's production possibilities frontier is steeper than Julio's.

Economists use the ideas of absolute advantage andcomparative advantage to compare the productive abilities of Hannah and Julio.

Toolkit: Section 17.13 "Comparative Advantage"

Comparative advantage and absolute advantage are used to compare the productivity of people (or firms or countries) in the production of a good or a service. A person has an absolute advantage in the production of a good if that person can produce more of that good in a unit of time than another person. A person has a comparative advantage in the production of one good if the opportunity cost, measured by the lost output of the other good, is lower for that person than for another person.

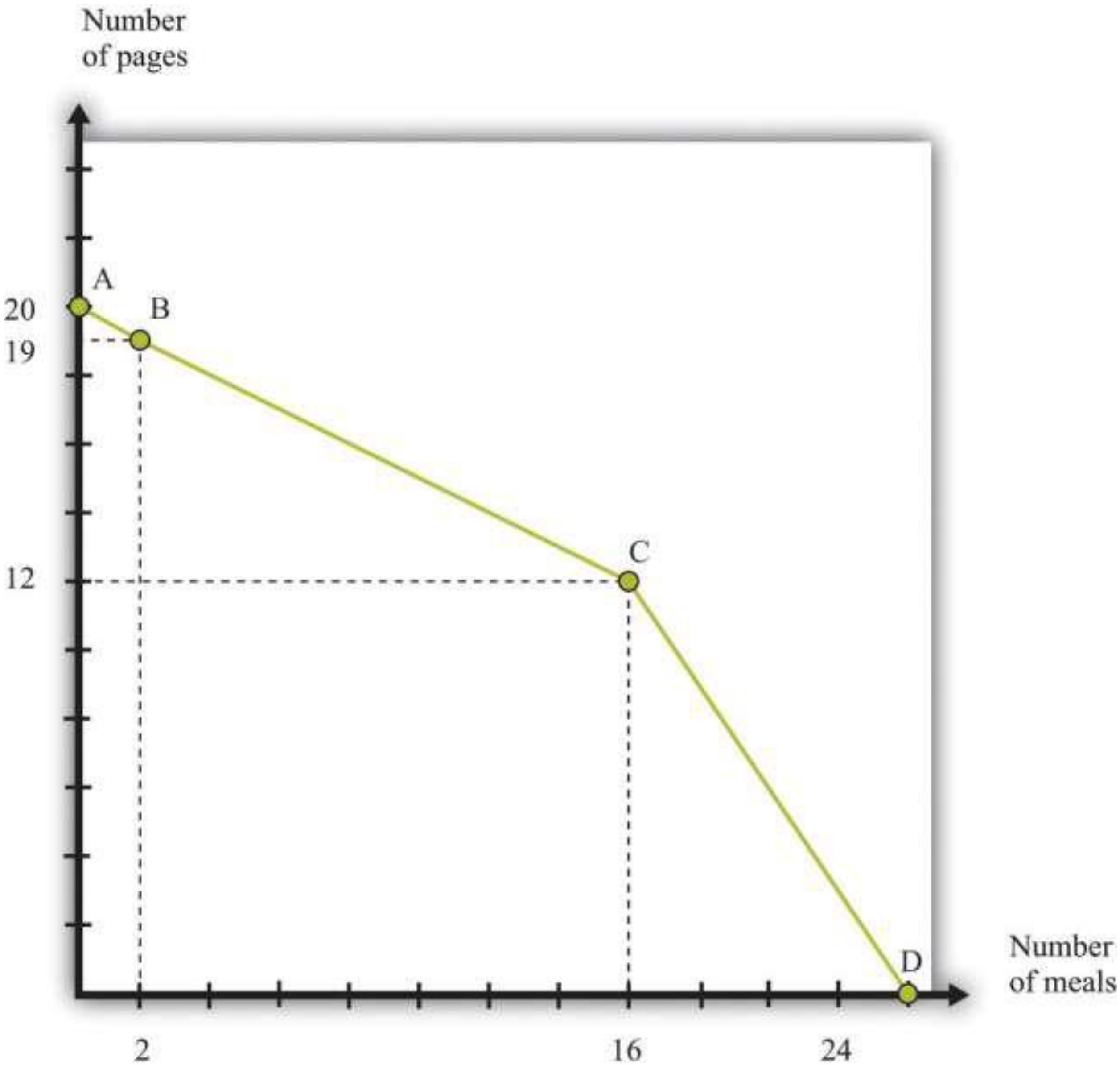
Both absolute and comparative advantage are relative concepts because they compare two people. In the case of absolute advantage, we compare the productivity of two people for a given good. In the case of comparative advantage, we compare two people and two goods because opportunity cost is defined across two goods (web pages and vegan meals in our example). Comparing Hannah and Julio, we see that Hannah has an absolute advantage in the production of web pages. She is better at producing web pages than Julio. By contrast, Julio has an absolute advantage in the production of vegan meals. Therefore, it is not surprising that Hannah also has a comparative advantage in the production of web pages—the opportunity cost of web pages is lower for her than it is for Julio—whereas Julio has a comparative advantage in vegan meals.

It is entirely possible for one person to have an absolute advantage in the production of both goods. For example, were Hannah's productivity in both activities to double, she would be better both at both web design and vegan meals. However, her opportunity cost of web pages would be unchanged. Julio would still have a comparative advantage in vegan meals. In general, one person always has a comparative advantage in one activity, while the other person has a comparative advantage in the other activity (the only exception is the case where the two individuals have exactly the same opportunity costs).

Suppose that Hannah's tastes for vegan meals and web pages are the same as Julio's: like Julio, Hannah wants to consume 2 meals for every web page. Acting alone, she would work for 2 hours in web design and spend 6 hours cooking, ending up with 3 web pages and 6 meals. But there is something very odd going on here. Julio, who is good at preparing meals, spends half his time on web design. Hannah, who is good at web design, spends three-quarters of her time cooking. Each has to spend a lot of time doing an activity at which he or she is unproductive.

This is where we see the possibility of gains from trade. Imagine that Julio and Hannah join together and become a team. What is their *joint* production possibilities frontier? If both Julio and Hannah devote their 8 hours of time to the production of web pages, then the economy can produce 20 web pages (8 from Julio and 12 from Hannah). At the other extreme, if both Julio and Hannah devote their 8 hours to cooking vegan meals, then the economy can produce 24 meals (16 from Julio and 8 from Hannah). These two points, which represent specialization of their two-person economy in one good, are indicated by point A and point D in Figure 5.21 "Julio and Hannah's Joint Production Possibilities Frontier".

Figure 5.21 Julio and Hannah's Joint Production Possibilities Frontier



Julio and Hannah's joint production possibilities frontier.

To fill in the rest of their joint production possibilities frontier, start from the vertical intercept, where 20 web pages are being produced. Suppose Julio and Hannah jointly decide that they would prefer to give up

1 web page to have some vegan meals. If Hannah were to produce only 11 web pages, she could free up 2/3 of an hour for vegan meals. She would produce 2/3 of a meal. Conversely, if Julio produced 1 fewer web page, that would free an hour of his time (because he is less efficient than Hannah at web design), and he could create 2 vegan meals. Evidently, it makes much more sense for Julio to shift from web design to cooking (see point B in Figure 5.21 "Julio and Hannah's Joint Production Possibilities Frontier").

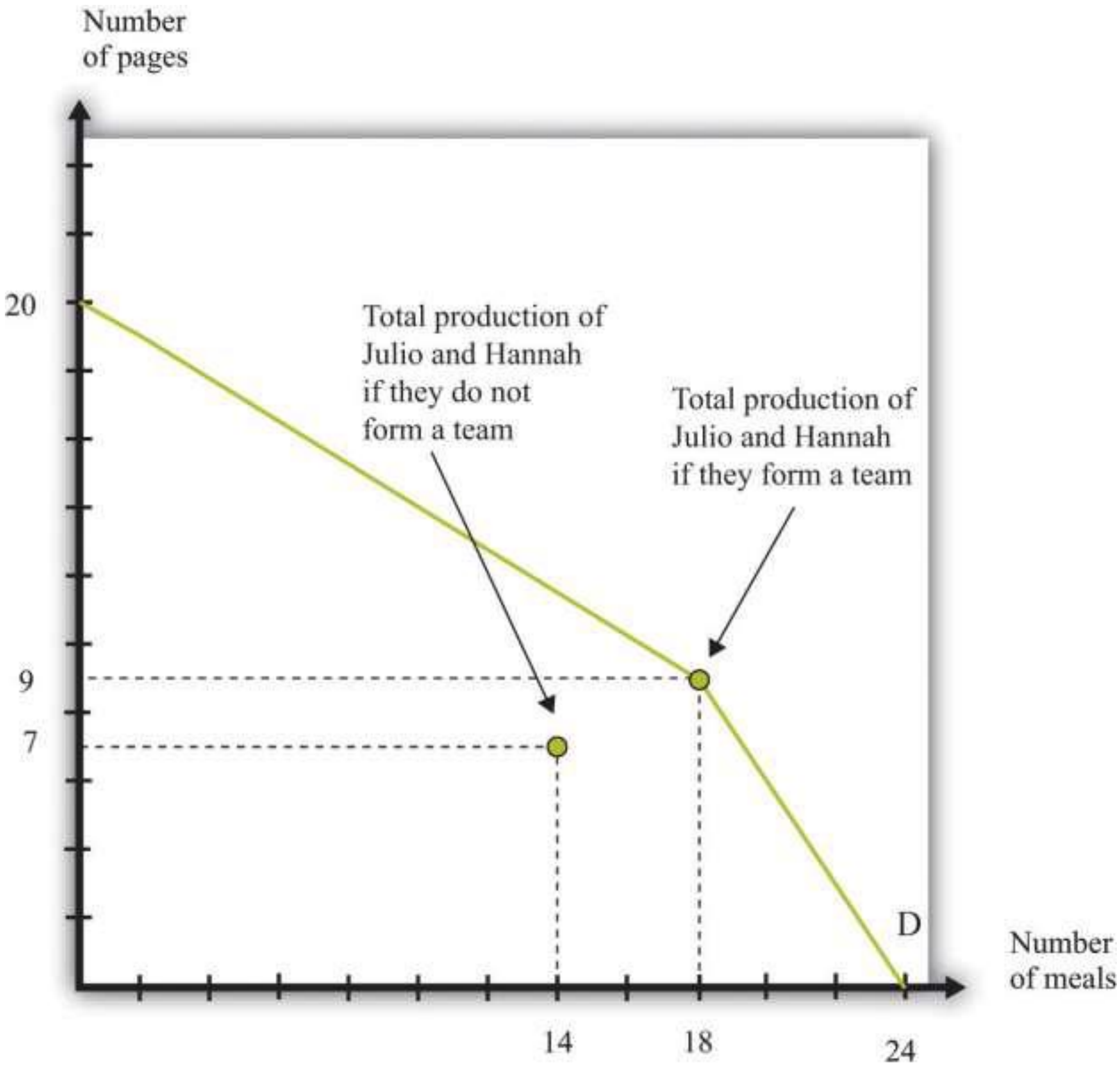
The most efficient way to substitute web design production for vegan meals production, starting at point A, is to have Julio switch from producing web pages to producing vegan meals. Julio should switch because he has a comparative advantage in cooking. As we move along the production possibilities frontier from point A to point B to point C, Julio continues to substitute from web pages to meals. For this segment, the slope of the production possibilities frontier is −1/2, which is Julio's opportunity cost of web pages.

At point C, both individuals are completely specialized. Julio spends all 8 hours on vegan meals and produces 16 meals. Hannah spends all 8 hours on web design and produces 12 web pages. If they would like to have still more vegan meals, it is necessary for Hannah to start producing that service. Because she is less efficient at cooking and more efficient at web design than Julio, the cost of extra vegan meals increases. Between point A and point C, the cost of vegan meals was 1/2 a webpage. Between point C and point D, the cost of a unit of vegan meals is 1.5 web pages. The production possibilities frontier becomes much steeper. If you look carefully at Figure 5.19 "Julio's Allocation of Time to Cooking Meals and Producing Web Pages", Figure 5.20 "Hannah's Production Possibilities Frontier", and Figure 5.21 "Julio and Hannah's Joint Production Possibilities Frontier", the joint production possibilities frontier is composed of the two individual frontiers joined together at point C.

Gains from Trade Once Again

Figure 5.22 "Julio and Hannah's Preferred Point" again shows the production possibilities frontier for the Julio-Hannah team: all the combinations of web pages and vegan meals that they can produce in one day, using the technologies available to them.

Figure 5.22 Julio and Hannah's Preferred Point



Julio and Hannah can benefit from joining forces.

Julio and Hannah have a great deal to gain by joining forces. We know that Julio, acting alone, produces 4 web pages and 8 vegan meals. Hannah, acting alone, produces 3 web pages and 6 vegan meals. The joint total is 7 web pages and 14 vegan meals. In Figure 5.22 "Julio and Hannah's Preferred Point", we have labeled this as "Total production of Julio and Hannah if they do not form a team." But look at what they can achieve if they work together: they can produce 9 web pages and 18 vegan meals.

Evidently they both can be better off when they work together. For example, each could get an additional web page and 2 vegan meals. Julio could have 5 web pages and 10 vegan meals (instead of 4 and 8, respectively), and Hannah could have 4 web pages and 8 vegan meals (instead of 3 and 6, respectively).

How do they do this? Julio specializes completely in vegan meals. He spends all 8 hours of his day cooking, producing 16 vegan meals. Hannah, meanwhile, gets to spend most of her time doing what she does best: designing web pages. She spends 6 hours on web design, producing 9 web pages, and 2 hours cooking, producing 2 vegan meals.

The key to this improvement is that we are no longer requiring that Julio and Hannah consume only what they can individually produce. Instead, they can produce according to their comparative advantage. They each specialize in the production of the good that they produce best and then trade to get a consumption bundle that they are happy with. The gains from trade come from the ability to specialize.

It is exactly such gains from trade that people are looking for when they place advertisements on craigslist. For example, the first ad we quoted was from someone with a comparative advantage in fixing furniture looking to trade with someone who had a comparative advantage in dental work. Comparative advantage is one of the most fundamental reasons why people trade, and sites like craigslist allow people to benefit from trade. Of course, in modern economies, most trade does not occur through individual barter; stores, wholesalers, and other intermediaries mediate trade. Although there are many*mechanisms* for trade, comparative advantage is a key *motivation* for trade.

Specialization and the History of the World

Economics famously teaches us that there is no such thing as a free lunch: everything has an opportunity cost. Paradoxically, economics also teaches us the secret of how we can make everyone better off than before simply by allowing them to trade—and if that isn't a free lunch, then what is?

This idea is also the story of why the world is so much richer today than it was 100 years, 1,000 years, or 10,000 years ago. The ability to specialize and trade is a key to prosperity. In the modern world, almost everybody is highly specialized in their production, carrying out a very small number of very narrow tasks. Specialization permits people to become skilled and efficient workers. (This is true, by the way, even if people have similar innate abilities. People with identical abilities will still usually be more efficient at producing one good rather than two.) Trade means that even though people specialize in production, they can still generalize in consumption. At least in the developed world, we enjoy lives of luxury that were unimaginable even a couple of centuries ago. This luxury would be impossible without the ability to specialize and trade.

The story of Julio and Hannah is therefore much more than a textbook exercise. One of the first steps on the ladder of human progress was the shift from a world where people looked after themselves to a world where people started producing in hunter-gatherer teams. Humans figured out that they could be more productive if some people hunted and others gathered. They also started to learn the benefits of team production—hunting was more efficient if a group of hunters worked together, encircling the prey so it could not escape. Such hunting teams are perhaps the first example of something that looks like a firm: a group of individuals engaged jointly in production.

Production Possibilities Frontier for a Country

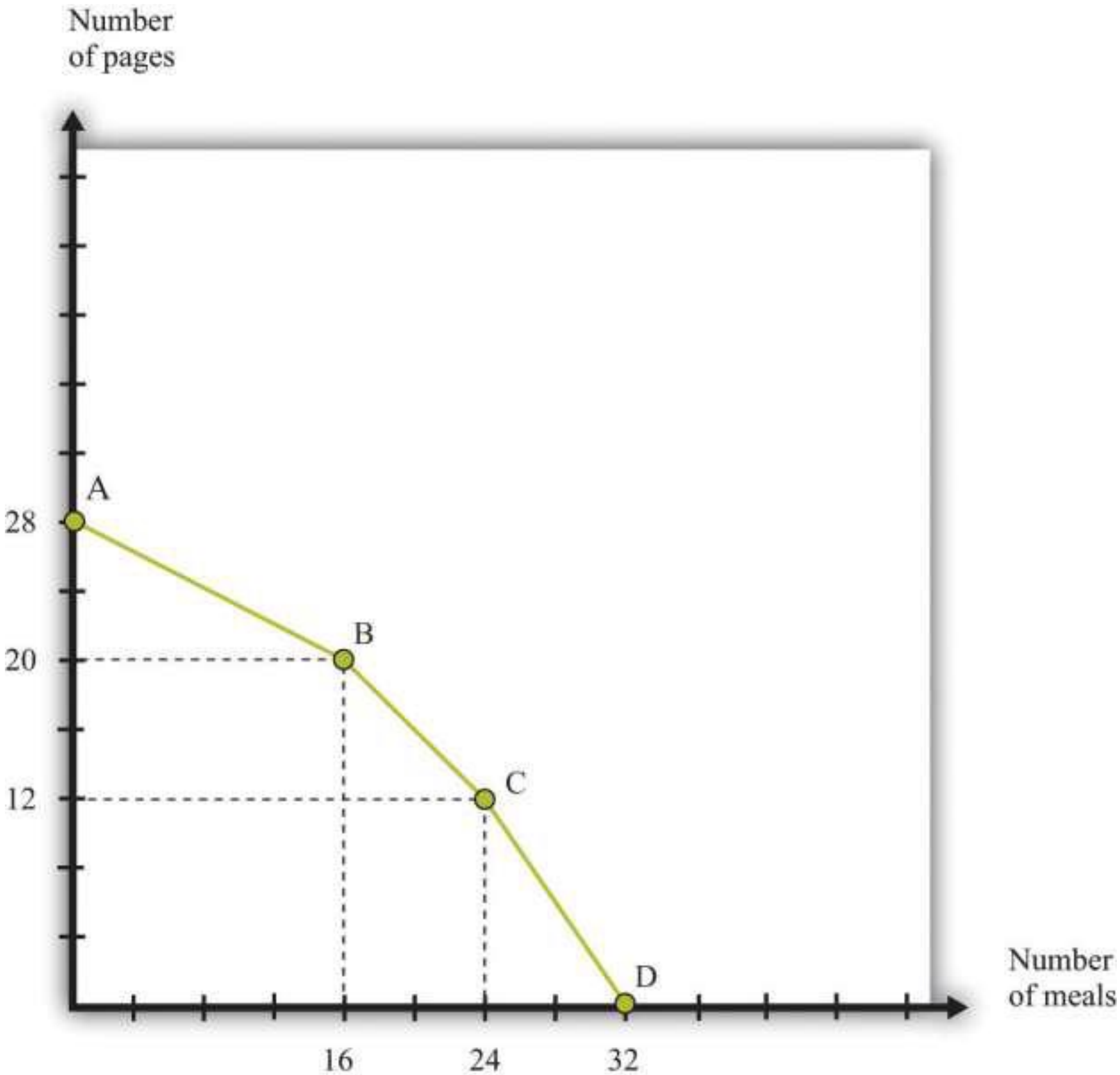
Before we finish with this story, let us try to get a sense of how we can expand it to an entire economy. We begin by adding a third individual to our story: Sergio. Sergio is less efficient than both Julio and Hannah. He has no absolute advantage in anything; he is no better at web design than Hannah, and he is no better preparing vegan meals than Julio. Remarkably, Julio and Hannah will still want to trade with him.

Table 5.5 Production Possibilities for Julio, Hannah, and Sergio

	Vegan Meals per Hour	Web Pages per Hour	Opportunity Cost of Web Pages (in Vegan Meals)	Opportunity Cost of Vegan Meals (in Web Pages)
Julio	2	1	1/2	2
Hannah	1	1.5	3/2	2/3
Sergio	1	1	1	1

We begin by constructing the production possibilities frontier for these three individuals. The logic is the same as before. Start from the position where the economy produces nothing but web pages (see Figure 5.23 "The Production Possibilities Frontier with Three People", point A). Together, Julio, Hannah, and Sergio can produce a total of 28 web pages in a day. Then we first shift Julio to cooking because he has the lowest opportunity cost of that activity. As before, the slope of this first part of the frontier is −1/2, up to point B. At point B, Julio is producing nothing but vegan meals, and Hannah and Sergio are devoting all of their time to producing web pages. At point B, the economy produces 20 web pages and 16 vegan meals.

Figure 5.23 *The Production Possibilities Frontier with Three People*



Here we show a production possibilities frontier with three individuals. Notice that it is smoother than the production possibilities frontier for two people.

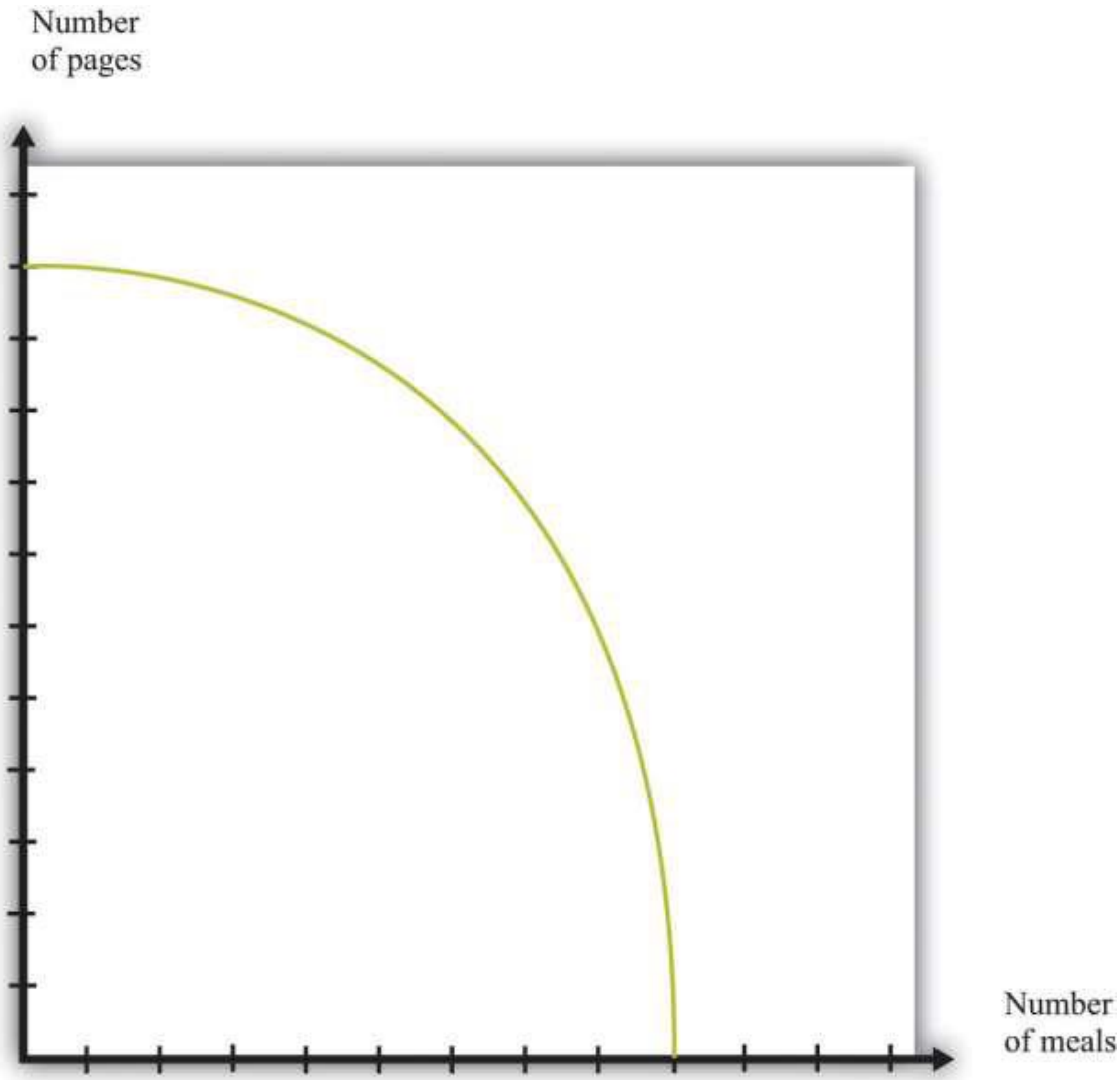
If they want more vegan meals, who should switch next? The answer is Sergio because his opportunity cost of vegan meals is lower than Hannah's. As Sergio starts shifting from web design to vegan meals, the frontier has slope −1, and we move from point B to point C. At point C, Sergio and Julio cook meals, and Hannah produces web pages. The number of web pages produced is 12, and there are 24 vegan meals. Finally, the last segment of the frontier has slope −3/2, as Hannah also shifts from web design to vegan meals. At point D, they all cook, with total production equal to 32 meals.

Let us suppose that Sergio, like the others, consumes in the ratio of 2 vegan meals for every web page. If the economy is at point C, the economy can produce 12 web pages and 24 vegan meals. Earlier, we saw that Hannah and Julio could together produce 9 web pages and 18 vegan meals, so bringing in Sergio allows for an extra 3 web pages and 6 vegan meals, which is more than Sergio could produce on his own. Even though Sergio is less efficient than both Julio and Hannah, there are still some gains from trade. The easiest way to see this is to note that it would take Sergio 9 hours to produce 3 web pages and 6 vegan meals. In 8 hours he could produce almost 3 web pages and over 5 vegan meals.

Where do the gains from trade come from? They come from the fact that, relative to Hannah, Sergio has a comparative advantage in vegan meals. Previously, Hannah was devoting some of her time to vegan meals, which meant she had to divert time from web design. This was costly because she is good at web design. By letting Sergio do the vegan meals, Hannah can specialize in what she does best. The end result is that there are extra web pages and vegan meals for them all to share.

Comparing Figure 5.22 "Julio and Hannah's Preferred Point" and Figure 5.23 "The Production Possibilities Frontier with Three People", you can see that the frontier becomes "smoother" when we add Sergio to the picture. Now imagine that we add more and more people to the economy, each with different technologies, and then construct the frontier in the same way. We would get a smoother and smoother production possibilities frontier. In the end, we might end up with something like Figure 5.24 "The Production Possibilities Frontier with a Large Number of People".

Figure 5.24 The Production Possibilities Frontier with a Large Number of People



As we add more and more people to the economy, the production possibilities frontier will become smoother.

It is easy enough to imagine that Julio, Hannah, and Sergio could all get together, agree to produce according to the principle of comparative advantage, and then share the goods that they have produced in a way that makes them all better off than they would be individually. Exactly how the goods would be shared would involve some kind of negotiation and bargaining among them. Once we imagine an economy with a large number of people in it, however, it is less clear how they would divide up the goods after they were produced. And that brings us full circle in the chapter. It is not enough that potential gains from trade exist. There must also be mechanisms, such as auctions and markets, that allow people to come together and realize these gains from trade.

KEY TAKEAWAYS

- Gains in trade partly come from the fact that individuals specialize in production and generalize in consumption.
- The efficient way to organize production is by looking at comparative advantage.
- Gains to trade emerge when individuals produce according to comparative advantage and then trade goods and services with one another.

CHECKING YOUR UNDERSTANDING

1. Fill in the missing values in Table 5.3 "Julio's Production Possibilities".

[1] These are actual offers that we found on craigslist, edited slightly for clarity.

[2] We study the time allocation problem in Chapter 3 "Everyday Decisions".

[3] All of this may seem quite familiar. The production possibility frontier for a single individual is the same as the time budget line for an individual. See Chapter 3 "Everyday Decisions" for more information.

5.5 End-of-Chapter Material

In Conclusion

This chapter is our first look at how individuals exchange. We have emphasized two points:

1. *How* individuals trade. You have seen that some very familiar things, such as eBay and craigslist, provide mechanisms to facilitate trade.
2. *Why* individuals trade. These gains may simply arise from differences in how people value items, as in Chapter 5 "eBay and craigslist", Section 5.2 "eBay". Or, as in Chapter 5 "eBay and craigslist", Section 5.5 "End-of-Chapter Material", these gains may reflect the fact that people differ in their abilities to produce different goods and services.

In reality, individuals differ across these two dimensions and more. [1]

Auctions such as eBay, newspaper classified advertisements, and sites such as craigslist are all means by which *individuals* in an economy can trade with one another. Of course, these are not the only forms of trade. Our discussion, by design, has ignored other common forms of trade in the economy, such as individuals buying goods and services from a firm (perhaps through a retailer) and individuals selling their labor services to firms. [2]

The biggest insight you should take away from this chapter is the fact that exchange is a means of creating value. When a seller sells a good or a service to a buyer, there is a presumption that both become better off. We have such a presumption because people enter into trades voluntarily: nobody forces a buyer to buy; nobody forces a seller to sell. The fact that voluntary exchange creates value is one of the most powerful ideas in economics.

Key Links

- eBay: http://www.ebay.com
- craigslist: http://www.craigslist.org

EXERCISES

1. Would you expect to get an item for less by buying it through craigslist rather than a regular store? If so, why?

2. (Advanced) Suppose the government imposed a tax on trading through craigslist, so the seller had to pay 5 percent of the price to the government. What might be the impact on trade in this market? What might happen to prices?
3. If you are a seller on craigslist, what would be the cost of setting a very high price?
4. If the price of an item traded increases, can both the surplus to the buyer and the surplus of the seller increase simultaneously?
5. If the owner of a car values it at $5,000 and there is a prospective buyer who is willing to pay $7,000 for that car, what does efficiency dictate about the price the buyer should pay for the car?
6. What are the differences between buying an item on eBay and buying that same item on craigslist?
7. In what settings do you have to be aware of the winner's curse?
8. In what way does a double oral auction differ from craigslist? From eBay?
9. If more people come into an auction, how should that affect your bidding in a winner's curse situation? Should you bid more or less? Why?
10. Suppose that instead of producing 2 vegan meals each hour, Julio can produce 3 vegan meals each hour. Draw his production possibilities frontier. What is his opportunity cost of web pages in terms of vegan meals with this alternative technology?
11. Suppose that Julio can produce 3 vegan meals each hour but requires 2 hours to design a web page. Draw his production possibilities frontier.
12. If Julio had a choice between the technology in Table 5.2 "Julio's Production Ability" and the one described in question 10, which would he prefer? Explain why.
13. What would the production possibilities frontier look like if, starting from point A in Figure 5.8 "Why You Should Bid Your Valuation in an eBay Auction", we first shifted Hannah rather than Julio to vegan meal production?
14. Show how the production possibilities frontier shifts if Hannah becomes more productive in producing web pages.
15. (Advanced) Suppose that both Julio and Hannah like each of the goods in a ratio of 6 vegan meals to 1 web page. Show that there are still gains to trade using the technologies described in Table 5.1 "Valuations of Different Bidders in a Winner's Curse Auction".

16. Explain what it means to "specialize in production and generalize in consumption." How many jobs do people usually have at a point in time? How many items does a food shopper usually have in his or her basket at the store?
17. Explain the connection between opportunity cost and comparative advantage.

Economics Detective

1. Find an auction to buy or sell the following items: a house, a car, a government bond, and licenses for the electromagnetic spectrum. What do you have to do to become a bidder at one of these auctions? How is the auction conducted?
2. Suppose you want to purchase a painting at Sotheby's, a famous English auction house. How would you do so? How would the auction operate? In what ways would it differ from buying art on eBay or craigslist?

Spreadsheet Exercise

1. (Advanced) Create a spreadsheet to input data like that in the first two columns of Table 5.4 "Production Possibilities for Julio and Hannah". Suppose there are two people who can produce two goods. Enter into the spreadsheet how much of each good they can produce in an hour.
   a. Calculate the opportunity costs, as in the last two columns ofTable 5.4 "Production Possibilities for Julio and Hannah".
   b. Assuming each has 8 hours a day to work, use the spreadsheet to calculate the total amount of each good each individual could produce if they produced only that good.
   c. Use this information to graph the production possibilities frontier for each person.
   d. Use this information to graph the production possibilities frontier for the two people combined.
   e. As you input different levels of output per hour per person, watch how these graphs change.
   f. Where do you see comparative advantage coming into play?

[1] Chapter 9 "Making and Losing Money on Wall Street" explores two further reasons for trade: differences in information and differences in attitudes toward risk.

[2] Such exchanges are discussed in Chapter 6 "Where Do Prices Come From?" and Chapter 7 "Why Do Prices Change?".

Chapter 6
Where Do Prices Come From?
The Price of a Pill

If you walk down the aisles of a supermarket, you will see thousands of different goods for sale. Each one will have a price displayed, telling you how much money you must give up if you want the good in question. On the Internet, you can find out how much it would cost you to stay in a hotel in Lima, Peru, or how much you would have to pay to rent a four-wheel drive vehicle in Nairobi, Kenya. On your television every evening, you can see the price that you would have to pay to buy a share of Microsoft Corporation or other companies.

Prices don't appear by magic. Every price posted in the supermarket or on the Internet is the result of a decision made by one or more individuals. In the future, you may find yourself trying to make exactly such a decision. Many students of economics have jobs in the marketing departments of firms or work for consulting companies that provide advice on what prices firms should charge. To learn about how managers make such decisions, we look at a real-life pricing decision.

In 2003, a major pharmaceutical company was evaluating the performance of one of its most important drugs—a medication for treating high blood pressure—in a Southeast Asian country. (For reasons of confidentiality, we do not reveal the name of the company or the country; other than simplifying the numbers slightly, the story is true.) Its product was known as one of the best in the market and was being sold for $0.50 per pill. The company had good market share and income in the country. There was one major competing drug in the market that was selling at a higher price and a few less important drugs.

In pharmaceutical companies, one individual often leads the team for each major drug that the company sells. In this company, the head of the product team—we will call her Ellie—was happy with the performance of the drug. Nonetheless, she wondered whether her company could make higher profits by setting a higher or lower price. In many countries, the prices of pharmaceutical products are heavily regulated. In this particular country, however, pharmaceutical companies were largely free to set whatever price they chose. Together with her team, therefore, Ellie decided to review the pricing strategy for her product. In this chapter, we therefore tackle the following question:

How should a firm set its price?

Road Map

Price-setting in retail markets typically takes the form of a take-it-or-leave-it offer. The seller posts a price, and prospective customers either buy or don't buy at that price. The prices you encounter every day in a supermarket, a coffee shop, or a fast-food restaurant, for example, are all take-it-or-leave-it offers that the retailer makes to you and other customers. [1]

In this chapter, we put you in the place of a marketing manager who has been given the job of determining the price that a firm should charge for its product. We first discuss the goals of this manager: what is she trying to achieve? We then show what information she needs to make a good decision. Finally, we derive some principles that allow her to set the right price. The chapter is built around two ideas: [2]

1. **The law of demand.** Each firm faces a demand curve for its product. This demand curve obeys the law of demand: if a firm sets a higher price, it must be willing to sell a smaller quantity; if a firm wishes to sell a larger quantity, it must set a lower price.
2. **Profit accounting.** Firms earn income from selling their goods and services, but they also incur costs from producing those goods and services. These costs include the costs of raw materials, the wages paid to the firm's workers, and so on. The difference between a firm's revenues and its costs is the firm's profits.

   The choice of price, via the demand curve, determines the amount of output a firm sells. The amount of output determines a firm's revenues and costs. Together, revenues and costs determine the profits of a firm.

[1] Chapter 5 "eBay and craigslist" has more discussion.

[2] This chapter and Chapter 5 "eBay and craigslist" are linked because they are both about mechanisms that allocate goods and services. In Chapter 5 "eBay and craigslist" we explain how eBay, craigslist, and newspapers are ways in which individuals exchange goods and services. In this chapter, we study how goods and services are allocated from firms to households. At the end of this chapter, we show that the supply-and-demand framework

introduced in Chapter 5 "eBay and craigslist" is also a useful framework *when the same product is produced by a large number of firms*. In particular, we show that our ideas about pricing also allow us to understand the foundations of supply.

6.1 The Goal of a Firm

LEARNING OBJECTIVE

1. What is the goal of a firm?

Firms devote substantial resources to their decisions about pricing. Large firms often have individuals or even entire departments whose main job is to make pricing decisions. Consulting firms specialize in providing advice to firms about the prices that they should charge. Some companies, such as airlines, have dedicated software to help them make these decisions. It isn't hard to understand why firms pay so much attention to the prices they charge. More than anything else, price determines the profits that a firm earns.

Economists are prone to talk about the decisions and objectives of a firm, and we often use the same shorthand. A firm, though, is just a legal creation—a collection of individuals who use some kind of technology. A firm takes labor, raw materials, and other inputs and turns them into products that people want to buy. Some of the people in a firm—the managers—decide how many workers it should hire, what prices it should set, and so on.

To understand pricing, we begin with the goal of a firm (that is, its managers). If a firm's managers are doing their jobs well, they should be making decisions to serve the interests of the owners of that firm. The owners of a firm are its shareholders. If you buy a share in a firm, then you own a fraction (your *share*) of the firm, which gives you the right to a fraction of the firm's earnings. Shareholders, for the most part, have one reason for buying and owning shares: to earn income. So the managers, if they are doing their jobs well, want a firm to make as much money as possible. We need to be careful, though. What matters is not the total amount of money received by a firm, but how much is available to be distributed to its owners. The owners of a firm hope to earn as high a return as possible on their shares.

Toolkit: Section 17.15 "Pricing with Market Power"

The money that is available for distribution to the shareholders of a firm is called a firm's profits. A firm pays money for raw materials, energy, and other supplies, and it pays wages to its workers. These expenses are a firm's costs of production. When it sells the product(s) it has produced, a firm earns revenues. Accountants analyze these revenues and costs in more detail, but in the end all the monies that flow in and out of a firm can be classified as either revenues or costs. Thus

profits = revenues − costs.

Consider, then, a marketing manager who wants to set the best price for a product—such as Ellie choosing the price for her company's blood pressure medication. She wants to find the price that will yield the most profits to her company. In an ideal world, a marketing manager might have access to a spreadsheet table, such as Figure 6.1 "A Spreadsheet That Would Make Pricing Decisions Easy", which displays a firm's monthly profits for different possible prices that it might set. Then Ellie's job would be easy: she would just have to look at the table, find the cell in column B with the highest number, and set the corresponding price. In this case, she would set a price of $15.

Figure 6.1 A Spreadsheet That Would Make Pricing Decisions Easy

Monthly Profits.xls

Price ($)	Monthly Profit ('000 $)
1	-6840
2	-5760
3	-4760
4	-3840
5	-3000
6	-2240
7	-1560
8	-960
9	-440
10	0
11	360
12	640
13	840
14	960
15	1000
16	960
17	840
18	640
19	360
20	0
21	-440
22	-960
23	-1560
24	-2240
25	-3000

But the reality of business is different. It is very difficult and expensive—perhaps even impossible—to gather information such as that in Figure 6.1 "A Spreadsheet That Would Make Pricing Decisions Easy". You might imagine that a firm could experiment, trying different prices and seeing what profits it earned. Unfortunately, this would be very costly because most of the time a firm would earn much lower profits than it could. Experimenting might even generate losses. For example, suppose that, one September, Ellie chose to try a price of \$2 per pill. The firm would lose nearly \$6 million—the equivalent of about six months' profits even at the very best price. Ellie would rapidly find herself looking for another job.

It is clear that trial and error—choosing different prices at random and seeing how much profit you get—could lead to costly mistakes, and there is no guarantee that you would ever find the best price. By adding some structure to a trial-and-error process, though, there is a simple strategy for finding the best price:

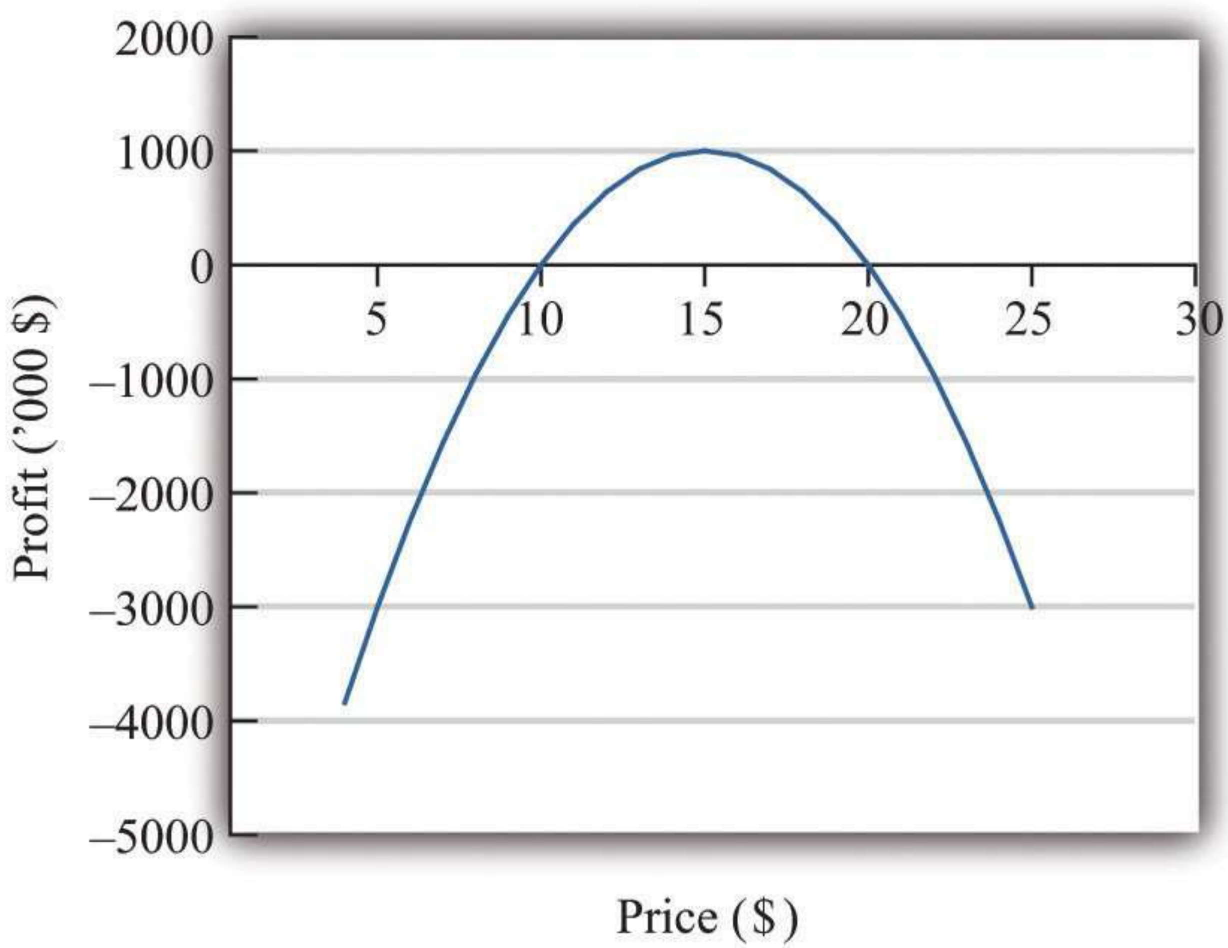
begin by slightly raising the firm's price. If profits increase, then you are on the right track. Keep raising the price, little by little, until profits stop increasing. On the other hand, if profits decrease when you raise the price, then you should try lowering the price instead. If profits increase, then keep lowering the price little by little.

Figure 6.2 "A Change in Price Leads to a Change in Profits" shows how a change in price translates into a change in profits. A change in a firm's price leads to a change in the quantity demanded. As a result, the revenues and costs of a firm change, as do its profits. Figure 6.3 "The Profits of a Firm" shows the profits a firm will earn at different prices. Our pricing strategy simply says the following. You are trying to get to the highest point of the profit hill in Figure 6.3 "The Profits of a Firm", and you will get there eventually if you always walk uphill. At the very top of the hill, the change in profits is zero.

Figure 6.2 A Change in Price Leads to a Change in Profits

If a firm changes its price, then there will be a change in demand. This then leads to changes in revenues and costs, which changes in the profits of a firm.

Figure 6.3 The Profits of a Firm



We could end the chapter right here. But we want to dig deeper and uncover some principles that tell us more about how pricing works. Then we can learn what information Ellie and other managers like her need to make better pricing decisions—and how they can make these decisions effectively. Our starting point is our earlier observation that

profits = revenues − costs.

KEY TAKEAWAY

- The objective of a firm is to maximize its profits, defined as revenues minus costs.

CHECKING YOUR UNDERSTANDING

1. If the manager of a firm chose a price to maximize sales, what would that price be? What would profits be at that price?
2. Explain in words why the profit function has the shape shown in Figure 6.3 "The Profits of a Firm".

6.2 The Revenues of a Firm

LEARNING OBJECTIVES

1. What is the demand curve faced by a firm?
2. What is the elasticity of demand? How is it calculated?
3. What is marginal revenue?

A firm's revenues are the money that it earns from selling its product. Revenues equal the number of units that a firm sells times the price at which it sells each unit:

revenues = price × quantity.

For example, think about a music store selling CDs. Suppose that the firm sells 25,000 CDs in a month at $15 each. Then its total monthly revenues are as follows:

revenues = 15 × 25,000 = $375,000.

There are two ways in which firms can obtain higher revenues: sell more products or sell at a higher price. So if a firm wants to make a lot of revenue, it should sell a lot of its product at a high price. Then again, you probably do not need to study economics to figure that out. The problem for a manager is that her ability to sell a product is limited by what the market will bear. Typically, we expect that if she sets a higher price, she will not be able to sell as much of the product:

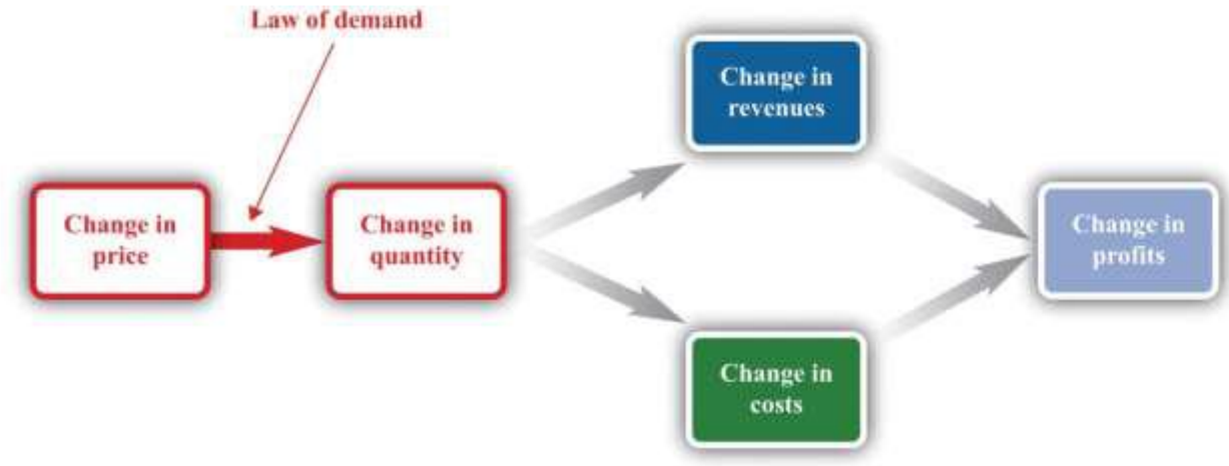
$$\uparrow \text{price} \rightarrow \downarrow \text{quantity}$$

Equivalently, if she wants to sell a larger quantity of product, she will need to drop the price:

$$\uparrow \text{quantity} \rightarrow \downarrow \text{price}$$

This is the law of demand in operation (Figure 6.4 "A Change in the Price Leads to a Change in Demand").

Figure 6.4 A Change in the Price Leads to a Change in Demand



An increase in price leads to a decrease in demand. A decrease in price leads to an increase in demand.

The Demand Curve Facing a Firm

There will typically be more than one firm that serves a market. This means that the overall demand for a product is divided among the different firms in the market. We have said nothing yet about the kind of "market structure" in which a firm is operating—for example, does it have a lot of competitors or only a few competitors? Without delving into details, we cannot know exactly how the market demand curve will be divided among the firms in the market. Fortunately, we can put this problem aside—at least for this chapter. [1] For the moment, it is enough just to know that each firm faces a demand curve for its own product.

When the price of a product increases, individual customers are less likely to think it is good value and are more likely to spend their income on other things instead. As a result—for almost all products—a higher price leads to lower sales.

Toolkit: Section 17.15 "Pricing with Market Power"

The demand curve facing a firm tells us the price that a firm can expect to receive for any given amount of output that it brings to market or the amount it can expect to sell for any price that it chooses to set. It represents the market opportunities of the firm.

An example of such a demand curve is

quantity demanded = 100 − (5 × price).

Table 6.1 "Example of the Demand Curve Faced by a Firm" calculates the quantity associated with different prices. For example, with this demand curve, if a manager sets the price at $10, the firm will sell 50 units because 100 − (5 × 10) = 50. If a manager sets the price at $16, the firm will sell only 20 units: 100 − (5 × 16) = 20. For every $1 increase in the price, output decreases by 5 units. (We have chosen a demand curve with numbers that are easy to work with. If you think that this makes the numbers unrealistically small, think of the quantity as being measured in, say, thousands of units, so a quantity of 3 in this equation means that the firm is selling 3,000 units. Our analysis would be unchanged.)

Table 6.1 Example of the Demand Curve Faced by a Firm

Price ($)	Quantity
0	100
2	90
4	80
6	70
8	60
10	50
12	40
14	30
16	20
18	10
20	0

Equivalently, we could think about a manager choosing the quantity that the firm should produce, in which case she would have to accept the price implied by the demand curve. To write the demand curve this way, first divide both sides of the equation by 5 to obtain

$$\frac{\text{quantity demanded}}{5} = 20 - \text{price}.$$

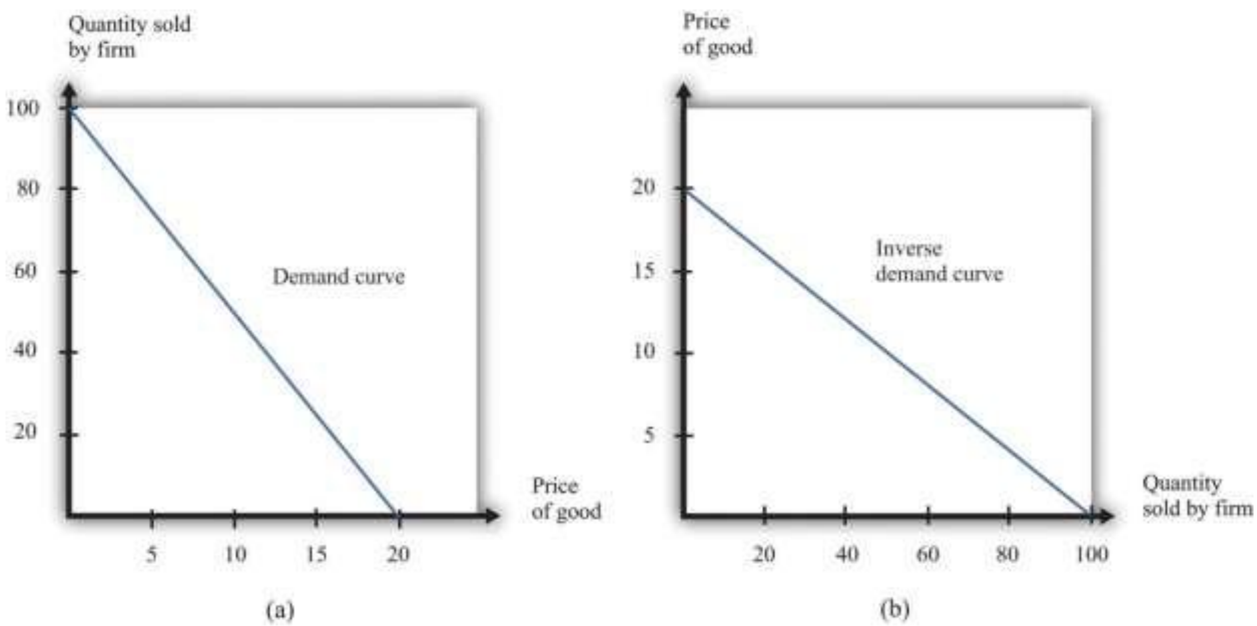
Now add "price" to each side and subtract "$\frac{\text{quantity demanded}}{5}$" from each side:

$$\text{price} = 20 - \frac{\text{quantity demanded}}{5}.$$

For example, if the manager wants to sell 70 units, she will need a price of $6 (because 20 − 70/5 = 6). For every unit increase in quantity, the price decreases by 20 cents.

Either way of looking at the demand curve is perfectly correct. Figure 6.5 "Two Views of the Demand Curve" shows the demand curve in these two ways. Look carefully at the two parts of this figure and convince yourself that they are really the same—all we have done is switch the axes.

Figure 6.5 Two Views of the Demand Curve



*There are two ways that we can draw a **demand curve**, both of which are perfectly correct. (a) The demand curve has price on the horizontal axis and quantity demanded on the vertical axis (b). The demand curve has price on the vertical axis, which is how we normally draw the demand curve in economics.*

The firm faces a trade-off: it can set a high price, such as $18, but it will be able to sell only a relatively small quantity (10). Alternatively, the firm can sell a large quantity (for example, 80), but only if it is willing to accept a low price ($4). The hard choice embodied in the demand curve is perhaps the most fundamental trade-off in the world of business. Of course, if the firm sets its price too high, it won't sell anything at all. The choke price is the price above which no units of the good will be sold. In our example, the choke price is $20; look at the vertical axis in part (b) of Figure 6.5 "Two Views of the Demand Curve". [2]

Every firm in the economy faces some kind of demand curve. Knowing the demand for your product is one of the most fundamental necessities of successful business. We therefore turn next to the problem Ellie learned about the demand curve for her company's drug.

The Elasticity of Demand: How Price Sensitive Are Consumers?

Marketing managers understand the law of demand. They know that if they set a higher price, they can expect to sell less output. But this is not enough information for good decision making. Managers need to know whether their customers' demand is very sensitive or relatively insensitive to changes in the price. Put differently, they need to know if the demand curve is steep (a change in price will lead to a small change in output) or flat (a change in price will lead to a big change in output). We measure this sensitivity by the own-price elasticity of demand.

Toolkit: Section 17.2 "Elasticity"

The own-price elasticity of demand (often simply called the *elasticity of demand*) measures the response of quantity demanded of a good to a change in the price of that good. Formally, it is the percentage change in the quantity demanded divided by the percentage change in the price:

$$\text{elasticity of demand} = \frac{\text{percentage change in quantity}}{\text{percentage change in price}}.$$

When price increases (the change in the price is positive), quantity decreases (the change in the quantity is negative). The price elasticity of demand is a negative number. It is easy to get confused with negative numbers, so we instead use

$$-(\text{elasticity of demand}) = \frac{\text{percentage change in quantity}}{\text{percentage change in price}},$$

which is always a positive number.

- If −(elasticity of demand) is a large number, then quantity demanded is sensitive to price: increases in price lead to big decreases in demand.
- If −(elasticity of demand) is a small number, then quantity demanded is insensitive to price: increases in price lead to small decreases in demand.

Throughout the remainder of this chapter, you will often see −(elasticity of demand). Just remember that this expression always refers to a positive number.

Calculating the Elasticity of Demand: An Example

Go back to our earlier example:

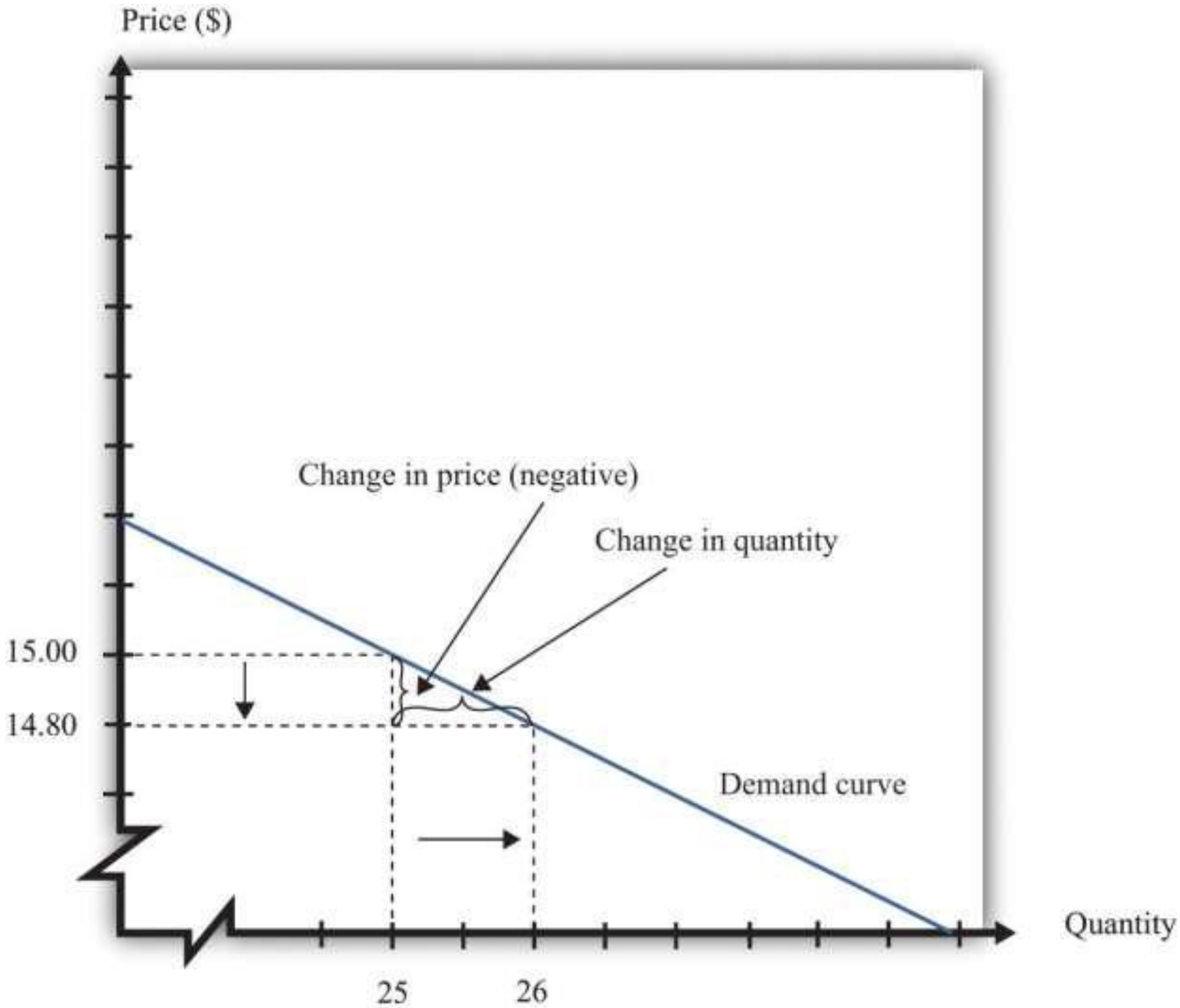
quantity demanded = 100 − 5 × price.

Suppose a firm sets a price of $15 and sells 25 units. What is the elasticity of demand if we think of a change in price from $15 to $14.80? In this case, the change in the price is −0.2, and the change in the quantity is 1. Thus we calculate the elasticity of demand as follows:

1. The percentage change in the quantity is 125 (4 percent).
2. The percentage change in the price is −0.215=175 (approximately −1.3 percent).
3. −(elasticity of demand) is 1/251/75=3.

The interpretation of this elasticity is as follows: when price decreases by 1 percent, quantity demanded increases by 3 percent. This is illustrated in Figure 6.6 "The Elasticity of Demand".

Figure 6.6 *The Elasticity of Demand*



When the price is decreased from $15.00 to $14.80, sales increase from 25 to 26. The percentage change in price is −1.3 percent. The percentage change in the quantity sold is 4. So −(elasticity of demand) is 3.

One very useful feature of the elasticity of demand is that it does not change when the number of units changes. Suppose that instead of measuring prices in dollars, we measure them in cents. In that case our demand curve becomes

quantity demanded = 100 − 500 × price.

Make sure you understand that this is exactly the same demand curve as before. Here the slope of the demand curve is −500 instead of −5. Looking back at the formula for elasticity, you see that the change in the price is 100 times greater, but the price itself is 100 times greater as well. The percentage change is unaffected, as is elasticity.

Market Power

The elasticity of demand is very useful because it is a measure of themarket power that a firm possesses. In some cases, some firms produce a good that consumers want very much—a good in which few substitutes are available. For example, De Beers controls much of the world's market for diamonds, and other firms are not easily able to provide substitutes. Thus the demand for De Beers' diamonds tends to be insensitive to price. We say that De Beers has a lot of market power. By contrast, a fast-food restaurant in a mall food court possesses very little market power: if the fast-food Chinese restaurant were to try to charge significantly higher prices, most of its potential customers would choose to go to the other Chinese restaurant down the aisle or even to eat sushi, pizza, or burritos instead.

Ellie's company had significant market power. There were a relatively small number of drugs available in the country to treat high blood pressure, and not all drugs were identical in terms of their efficacy and side effects. Some doctors were loyal to her product and would almost always prescribe it. Some doctors were not very well informed about the price because doctors don't pay for the medication. For all these reasons, Ellie had reason to suspect that the demand for her drug was not very sensitive to price.

The Elasticity of Demand for a Linear Demand Curve

The elasticity of demand is generally different at different points on the demand curve. In other words, the market power of a firm is not constant: it depends on the price that a firm has chosen to set. To illustrate, remember that we found −(elasticity of demand) = 3 for our demand curve when the price is \$15. Suppose we calculate the elasticity for this same demand curve at \$4. Thus imagine that that we are originally at the point where the price is \$4 and sales are 80 units and then suppose we again decrease the price by 20 cents. Sales will increase by 1 unit:

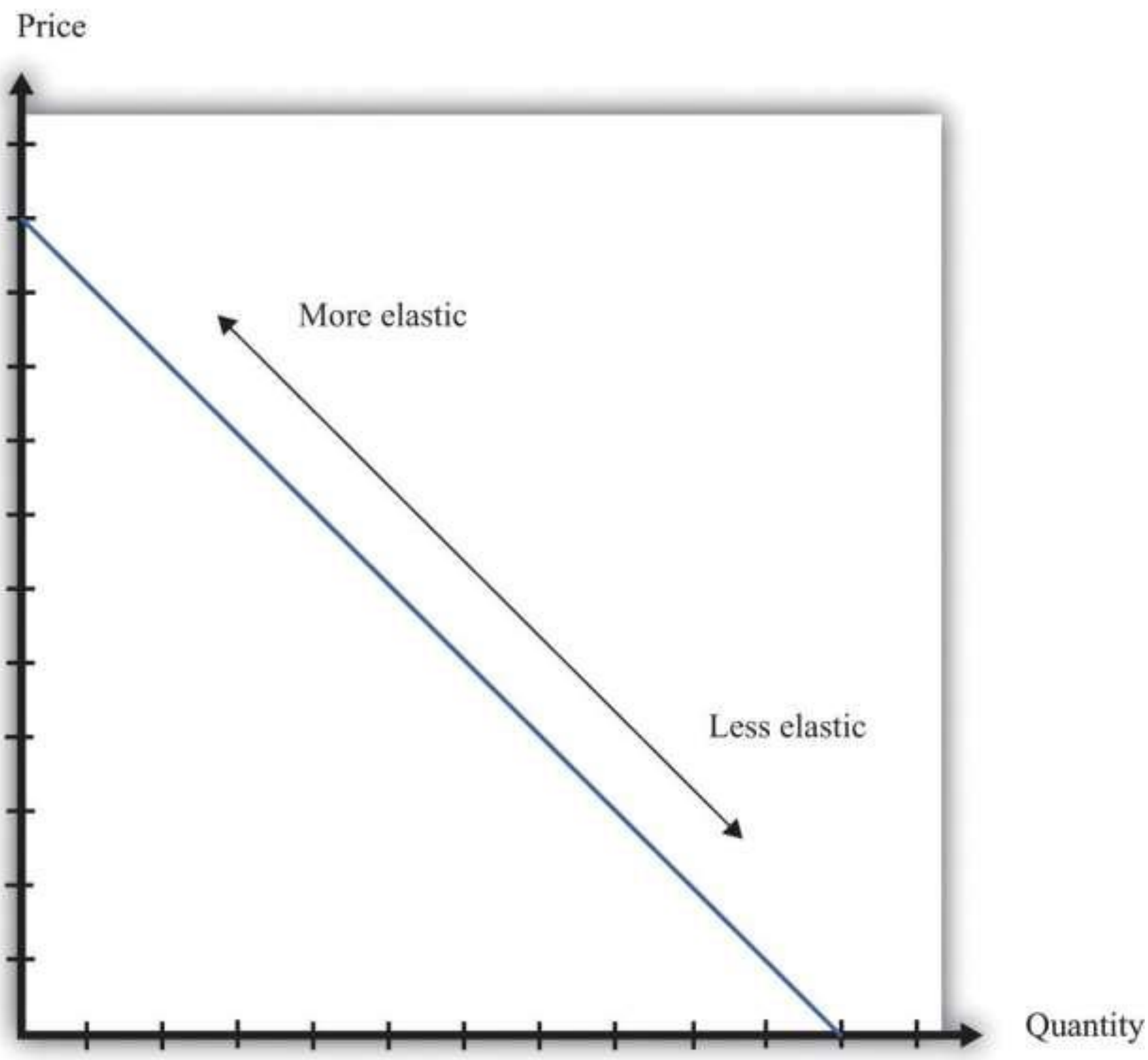
1. The percentage change in the quantity is 180 (1.25 percent).

2. The percentage change in the price is $-0.24=-120$ (5 percent).

3. −(elasticity of demand) is $1/801/20=0.25$.

The elasticity of demand is different because we are at a different point on the demand curve. When −(elasticity of demand) increases, we say that demand is becoming more elastic. When −(elasticity of demand) decreases, we say that demand is becoming less elastic. As we move down a linear demand curve, −(elasticity of demand) becomes smaller, as shown in Figure 6.7 "The Elasticity of Demand When the Demand Curve Is Linear".

Figure 6.7 The Elasticity of Demand When the Demand Curve Is Linear



The elasticity of demand is generally different at different points on a demand curve. In the case of a linear demand curve, –(elasticity of demand) becomes smaller as we move down the demand curve.

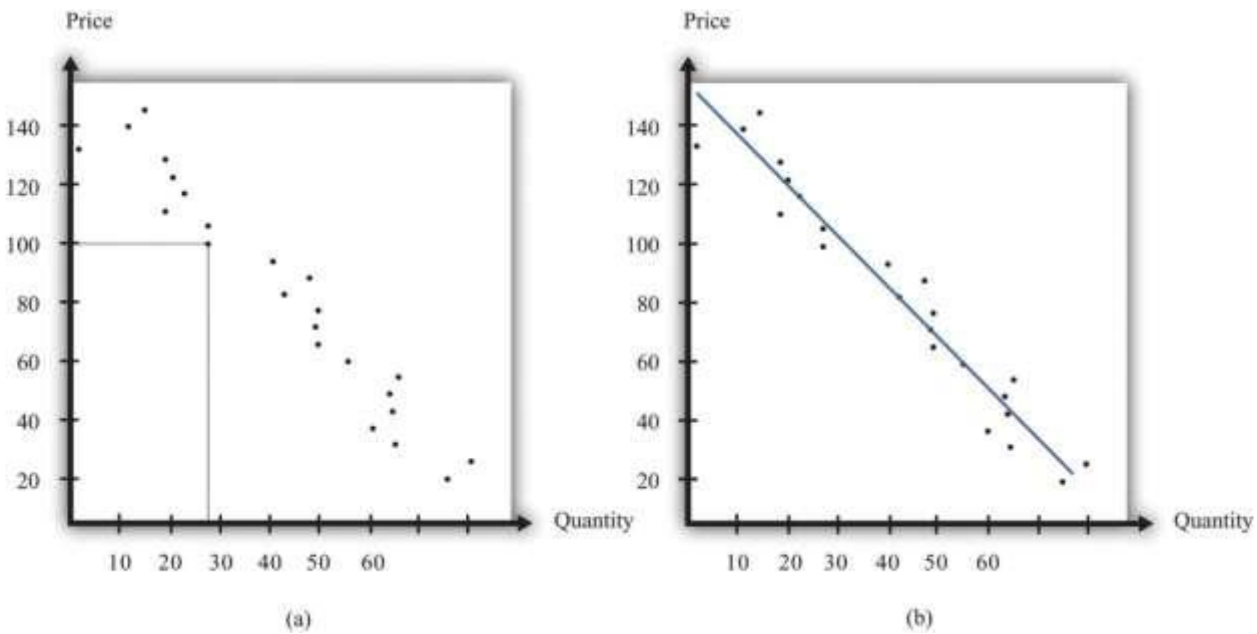
Measuring the Elasticity of Demand

To evaluate the effects of her decisions on revenues, Ellie needs to know about the demand curve facing her firm. In particular, she needs to know whether the quantity demanded by buyers is very sensitive to the price that she sets. We now know that the elasticity of demand is a useful measure of this sensitivity. How can managers such as Ellie gather information on the elasticity of demand?

At an informal level, people working in marketing and sales are likely to have some idea of whether their customers are very price sensitive. Marketing and sales personnel—if they are any good at their jobs—spend time talking to actual and potential customers and should have some idea of how much these customers care about prices. Similarly, these employees should have a good sense of the overall market and the other factors that might affect customers' choices. For example, they will usually know whether there are other firms in the market offering similar products, and, if so, what prices these firms are charging. Such knowledge is much better than nothing, but it does not provide very concrete evidence on the demand curve or the elasticity of demand.

A firm may be able to make use of existing sales data to develop a more concrete measure of the elasticity of demand. For example, a firm might have past sales data that show how much they managed to sell at different prices, or a firm might have sales data from different cities where different prices were charged. Suppose a pricing manager discovers data for prices and quantities like those in part (a) of Figure 6.8 "Finding the Demand Curve". Here, each dot marks an observation—for example, we can see that in one case, when the price was $100, the quantity demanded was 28.

Figure 6.8 *Finding the Demand Curve*



(a) This is an example of data that a manager might have obtained for prices and quantities. (b) A line is fit to the data that represents a best guess at the underlying demand curve facing a firm.

The straight-line demand curves that appear in this and other books are a convenient fiction of economists and textbook writers, but no one has actually seen one in captivity. In the real world of business, demand curves—if they are available at all—are only a best guess from a collection of data. Economists and statisticians have developed statistical techniques for these guesses. The underlying idea of these techniques is that they *fit a line* to the data. (The exact details do not concern us here; you can learn about them in more advanced courses in economics and statistics.) Part (b) of Figure 6.8 "Finding the Demand Curve" shows an example. It represents our best prediction, based on available data, of how much people will buy at different prices.

If a firm does not have access to reliable existing data, a third option is for it to generate its own data. For example, suppose a retailer wanted to know how sensitive customer demand for milk is to changes in the price of milk. It could try setting a different price every week and observe its sales. It could then plot them in a diagram like Figure 6.8 "Finding the Demand Curve" and use techniques like those we just discussed to fit a line. In effect, the store could conduct its own experiment to find out what its demand curve looks like. For a firm that sells over the Internet, this kind of experiment is particularly attractive because it can randomly offer different prices to people coming to its website.

Finally, firms can conduct market research either on their own or by hiring a professional market research firm. Market researchers use questionnaires and surveys to try to discover the likely purchasing behavior of consumers. The simplest questionnaire might ask, “How much would you be willing to pay for product x?” Market researchers have found such questions are not very useful because consumers do not answer them very honestly. As a result, research firms use more subtle questions and other more complicated techniques to uncover consumers’ willingness to pay for goods and services.

Ellie decided that she should conduct market research to help with the pricing decision. She hired a market research firm to ask doctors about how they currently prescribed different high blood pressure medications. Specifically, the doctors were asked what percentage of their prescriptions went to each of the drugs on the market. Then they were asked the effect of different prices on those percentages. Based on this research, the market research firm found that a good description of the demand curve was as follows:

quantity demanded = 252 − 300 × price.

Remember that the drug was currently being sold for $0.50 a pill, so

quantity demanded = 252 − 300 × 0.5 = 102.

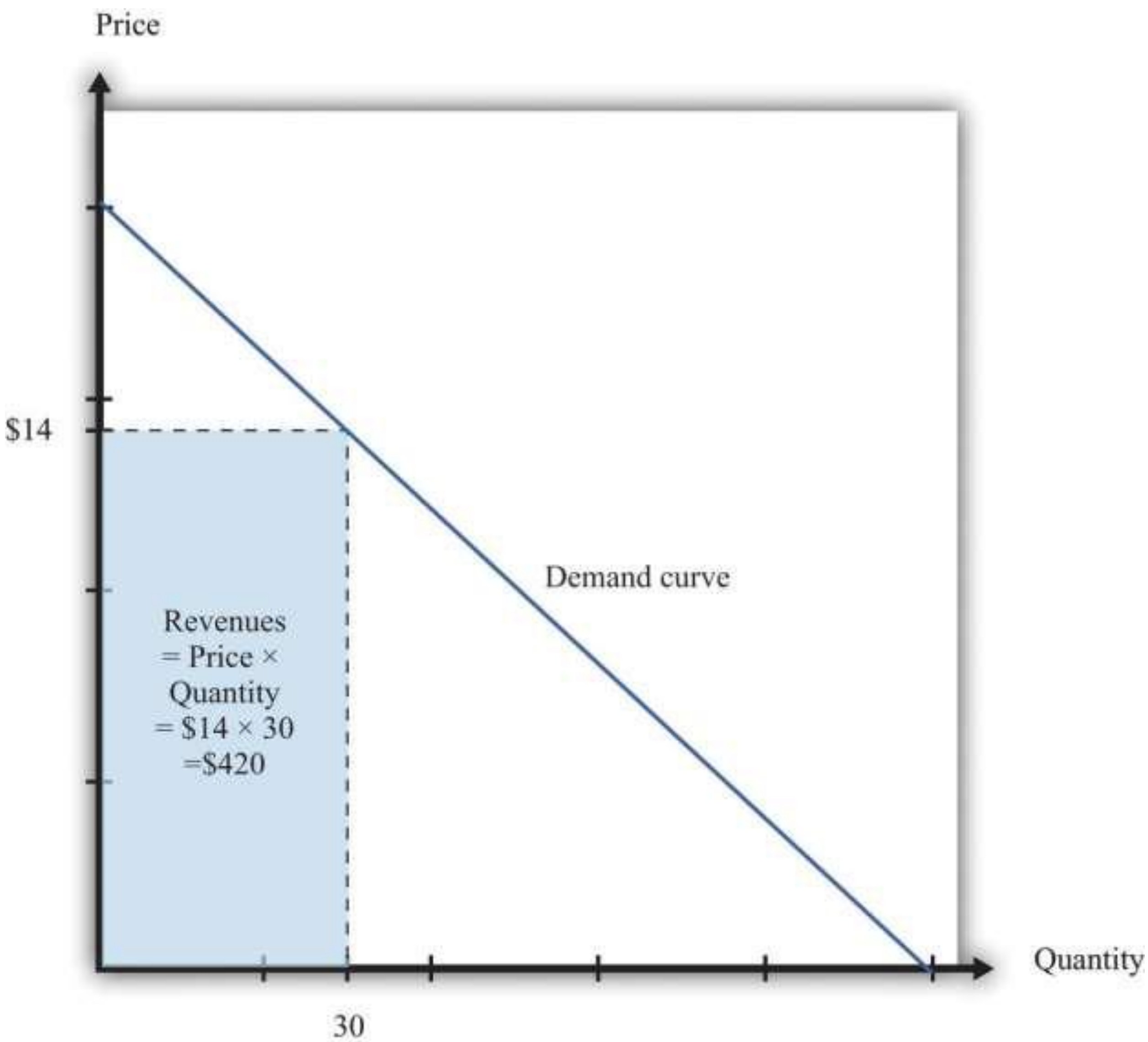
The demand curve also told Ellie that if she increased the price by 10 percent to $0.55, the quantity demanded would decrease to 87 (252 − 300 × 0.55 = 87). Therefore, the percentage change in quantity is $\frac{87-102}{102}=-14.7.$ From this, the market research firm discovered that the elasticity of demand at the current price was

$$-(\text{elasticity of demand})=\frac{\text{percentage change in quantity}}{\text{percentage change in price}}=\frac{14.7}{10}=1.47.$$

How Do Revenues Depend on Price?

The next step is to understand how to use the demand curve when setting prices. The elasticity of demand describes how quantity demanded depends on price, but what a manager really wants to know is how *revenues* are affected by price. Revenues equal price times quantity, so we know immediately that a firm earns $0 if the price is $0. (It doesn't matter how much you give away, you still get no money.) We also know that, at the choke price, the quantity demanded is 0 units, so its revenues are likewise $0. (If you sell 0 units, it doesn't matter how high a price you sell them for.) At prices between $0 and the choke price, however, the firm sells a positive amount at a positive price, thus earning positive revenues. Figure 6.9 "Revenues" is a graphical representation of the revenues of a firm. Revenues equal price times quantity, which is the area of the rectangle under the demand curve. For example, at $14 and 30 units, revenues are $420.

Figure 6.9 Revenues



*The **revenues of a firm** are equal to the area of the rectangle under the demand curve.*

We can use the information in Table 6.1 "Example of the Demand Curve Faced by a Firm" to calculate the revenues of a firm at different quantities and prices (this is easy to do with a spreadsheet). Table 6.2 "Calculating Revenues" shows that if we start at a price of zero and increase the price, the firm's revenues also increase. Above a certain point, however (in this example, $10), revenues start to decrease again.

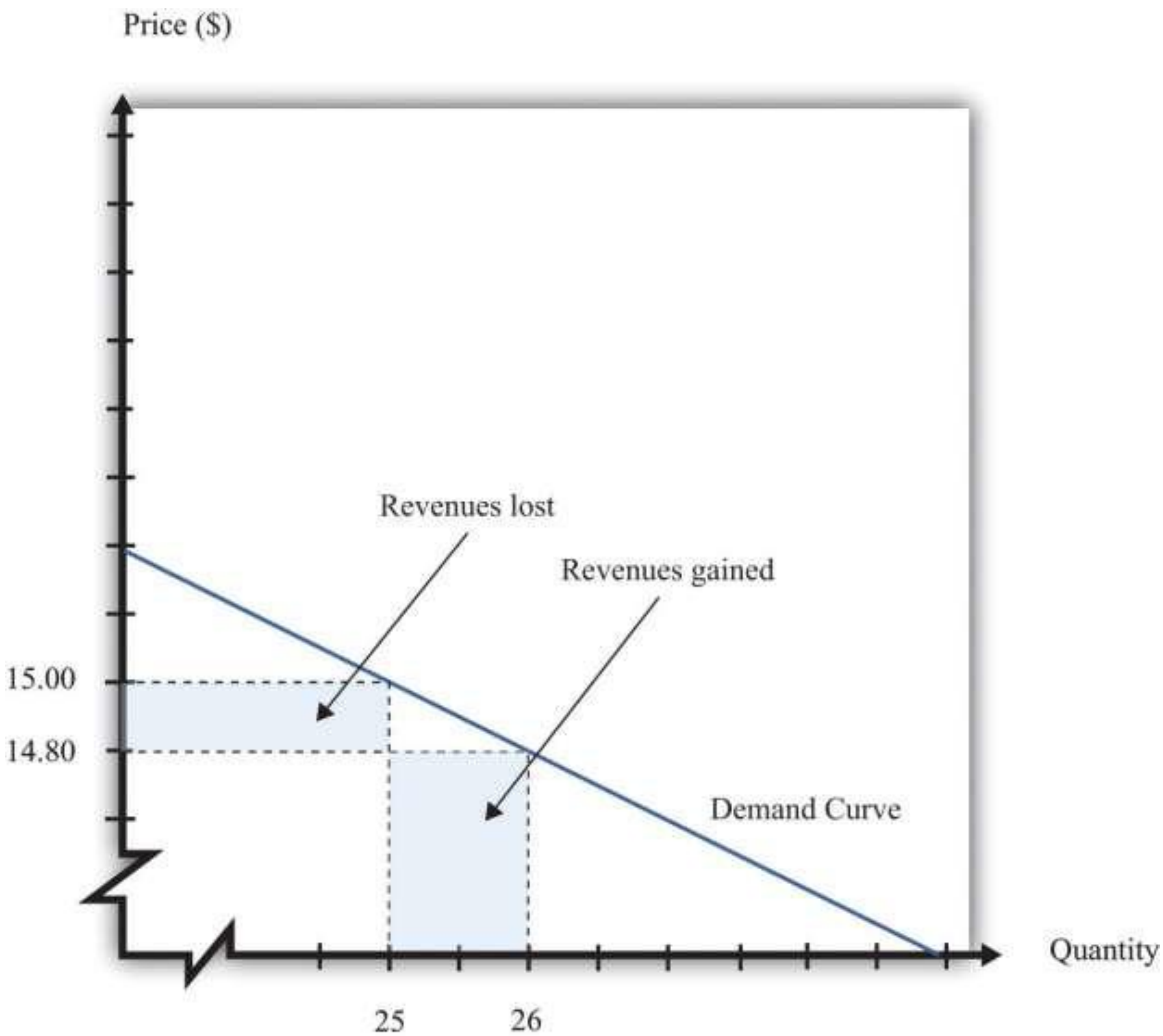
Table 6.2 Calculating Revenues

Price($)	Quantity	Revenues ($)
0	100	0
2	90	180
4	80	320
6	70	420
8	60	480
10	50	500
12	40	480
14	30	420
16	20	320
18	10	180
20	0	0

Marginal Revenue

Earlier we suggested that a good strategy for pricing is to experiment with small changes in price. So how do small changes in price affect the revenue of a firm? Suppose, for example, that a firm has set the price at $15 and sells 25 units, but the manager contemplates decreasing the price to $14.80. We can see the effect that this has on the firm's revenues in Figure 6.10 "Revenues Gained and Lost".

Figure 6.10 Revenues Gained and Lost

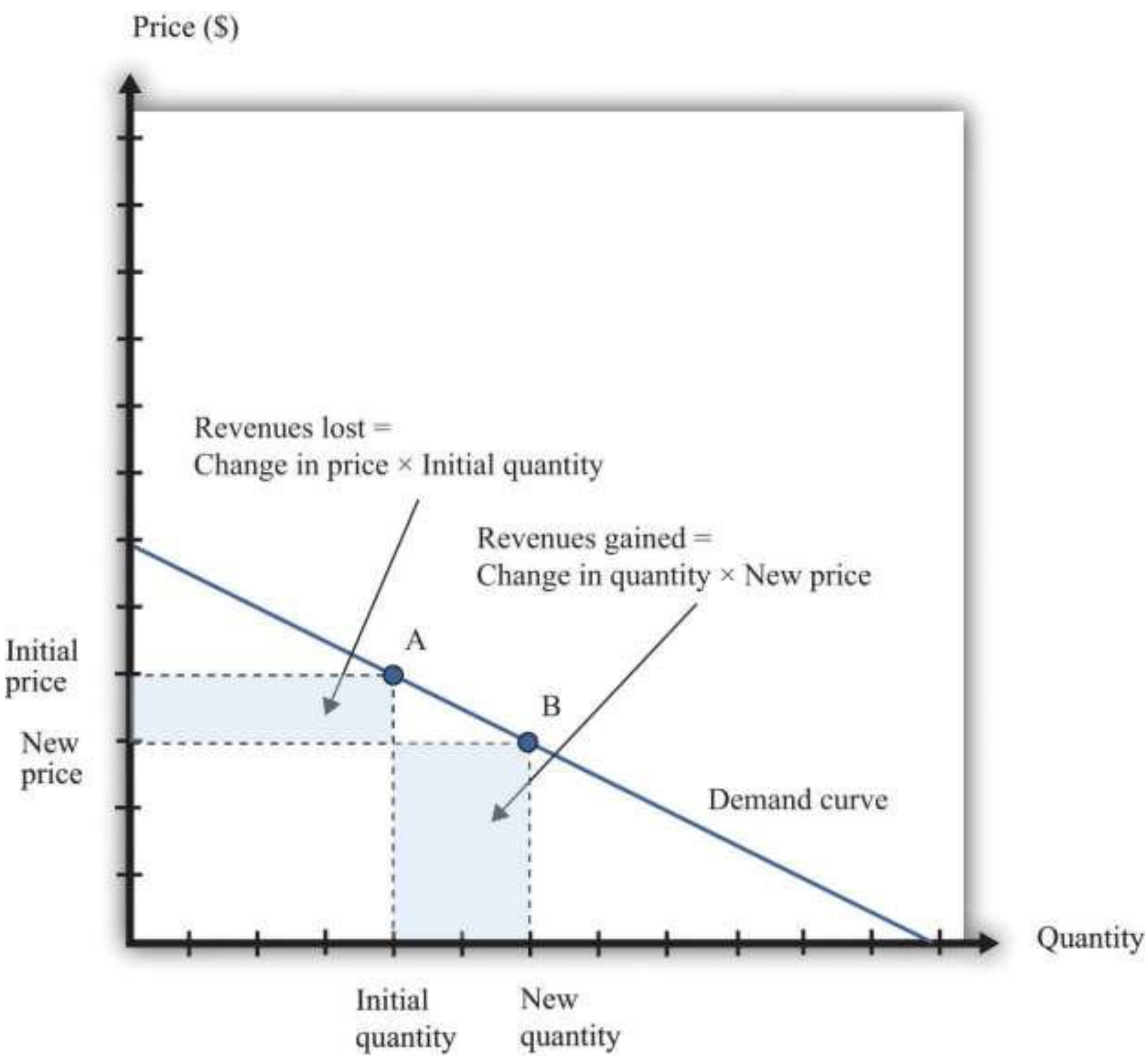


If a firm cuts its price, it sells more of its product, which increases revenues, but sells each unit at a lower price, which decreases revenues.

The firm will lose 20 cents on each unit it sells, so it will lose $5 in revenue. This is shown in the figure as the rectangle labeled "revenues lost." But the firm will sell more units: from the demand curve, we know that when the firm decreases its price by $0.20, it sells another unit. That means that the firm gains $14.80, as shown in the shaded area labeled "revenues gained." The overall change in the firm's revenues is equal to $14.80 – $5.00 = $9.80. Decreasing the price from $15.00 to $14.80 will increase its revenues by $9.80.

Look carefully at Figure 6.10 "Revenues Gained and Lost" and make sure you understand the experiment. *We presume throughout this chapter that a firm must sell every unit at the same price*. When we talk about moving from $15.00 to $14.80, we are *not* supposing that a firm sells 25 units for $15 and then drops its price to $14.80 to sell the additional unit. We are saying that the manager is choosing between selling 25 units for $15.00 or 26 units for $14.80.

Figure 6.11 Calculating the Change in Revenues



If a manager has an idea about how much quantity demanded will decrease for a given increase in price, she can calculate the likely effect on revenues.

Figure 6.11 "Calculating the Change in Revenues" explains this idea more generally. Suppose a firm is originally at point A on the demand curve. Now imagine that a manager decreases the price. At the new, lower price, the firm sells a new, higher quantity (point B). The change in the quantity is the new quantity minus the initial quantity. The change in the price is the new price minus the initial price (remember that this is a negative number). The change in the firm's revenues is given by

change in revenues = (change in quantity × new price) + (change in price × initial quantity).

The first term is positive: it is the extra revenue from selling the extra output. The second term is negative: it is the revenue lost because the price has been decreased. Together these give the effect of a change in price on revenues, which we call a firm's marginal revenue.

Toolkit: Section 17.15 "Pricing with Market Power"

Marginal revenue is the *change* in revenue associated with a *change* in quantity of output sold:

$$\text{marginal revenue} = \frac{\text{change in revenue}}{\text{change in quantity}}.$$

We can write this as [3]

$$\text{marginal revenue} = \text{price} \times \left(1 + \frac{\text{percentage change in price}}{\text{percentage change in quantity}}\right).$$

Marginal Revenue and the Elasticity of Demand

Given the definitions of marginal revenue and the elasticity of demand, we can write

$$\text{marginal revenue} = \text{price} \times \left(1 - \frac{1}{-(\text{elasticity of demand})}\right).$$

It may look odd to write this expression with two minus signs. We do this because it is easier to deal with the positive number: –(elasticity of demand). We see three things:

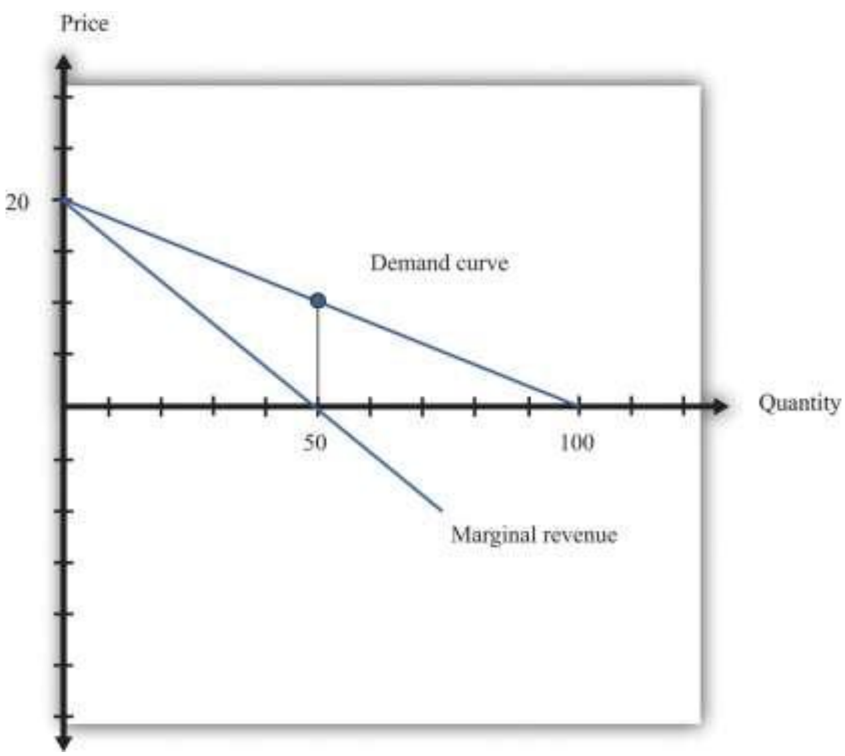
1. Marginal revenue is always less than the price. Mathematically, $\left(1 - \frac{1}{-(\text{elasticity of demand})}\right) < 1$. Suppose a firm sells an extra unit. If the price stays the same, then the extra revenue would just equal the price. But the price does not stay the same: it decreases, meaning the firm gets less for every unit that it sells.
2. Marginal revenue can be negative. If –(elasticity of demand) < 1, then $\frac{1}{-(\text{elasticity of demand})} > 1$, and $\left(1 - \frac{1}{-(\text{elasticity of demand})}\right) < 0$. When marginal revenue is negative, increased

production results in *lower* revenues for a firm. The firm sells more output but loses more from the lower price than it gains from the higher sales.

3. The gap between marginal revenue and price depends on the elasticity of demand. When demand is more elastic, meaning −(elasticity of demand) is a bigger number, the gap between marginal revenue and price becomes smaller.

These three ideas are illustrated in Figure 6.12 "Marginal Revenue and Demand". The demand curve shows us the price at any given quantity. The marginal revenue curve lies below the demand curve because of our first observation: at any quantity, marginal revenue is less than price. [4] The marginal revenue curve intersects the horizontal axis at 50 units: when output is less than 50 units, marginal revenue is positive; when output exceeds 50, marginal revenue is negative. We explained earlier that a linear demand curve becomes more inelastic as you move down it. When the demand curve goes from being relatively elastic to relatively inelastic, marginal revenue goes from being positive to being negative.

Figure 6.12 Marginal Revenue and Demand



The marginal revenue curve lies below the demand curve because at any quantity, marginal revenue is less than price.

Earlier, we showed that when a firm sets the price at $15, −(elasticity of demand) = 3. Thus we can calculate marginal revenue at this price:

$$\text{marginal revenue} = \text{price} \times \left(1 - \frac{1}{-(\text{elasticity of demand})}\right) = 15\left(1 - \frac{1}{3}\right) = 15 \times \frac{2}{3} = 10.$$

What does this mean? Starting at $15, it means that if a firm decreases its price—and hence increase its output—by a small amount, there would be an increase in the firm's revenues.

When revenues are at their maximum, marginal revenue is zero. We can confirm this by calculating the elasticity of demand at $10. Consider a 10 percent increase in price, so the price increases to $11. At $10, sales equal 50 units. At $11, sales equal 45 units. In other words, sales decrease by 5 units, so the decrease in sales is 10 percent. It follows that

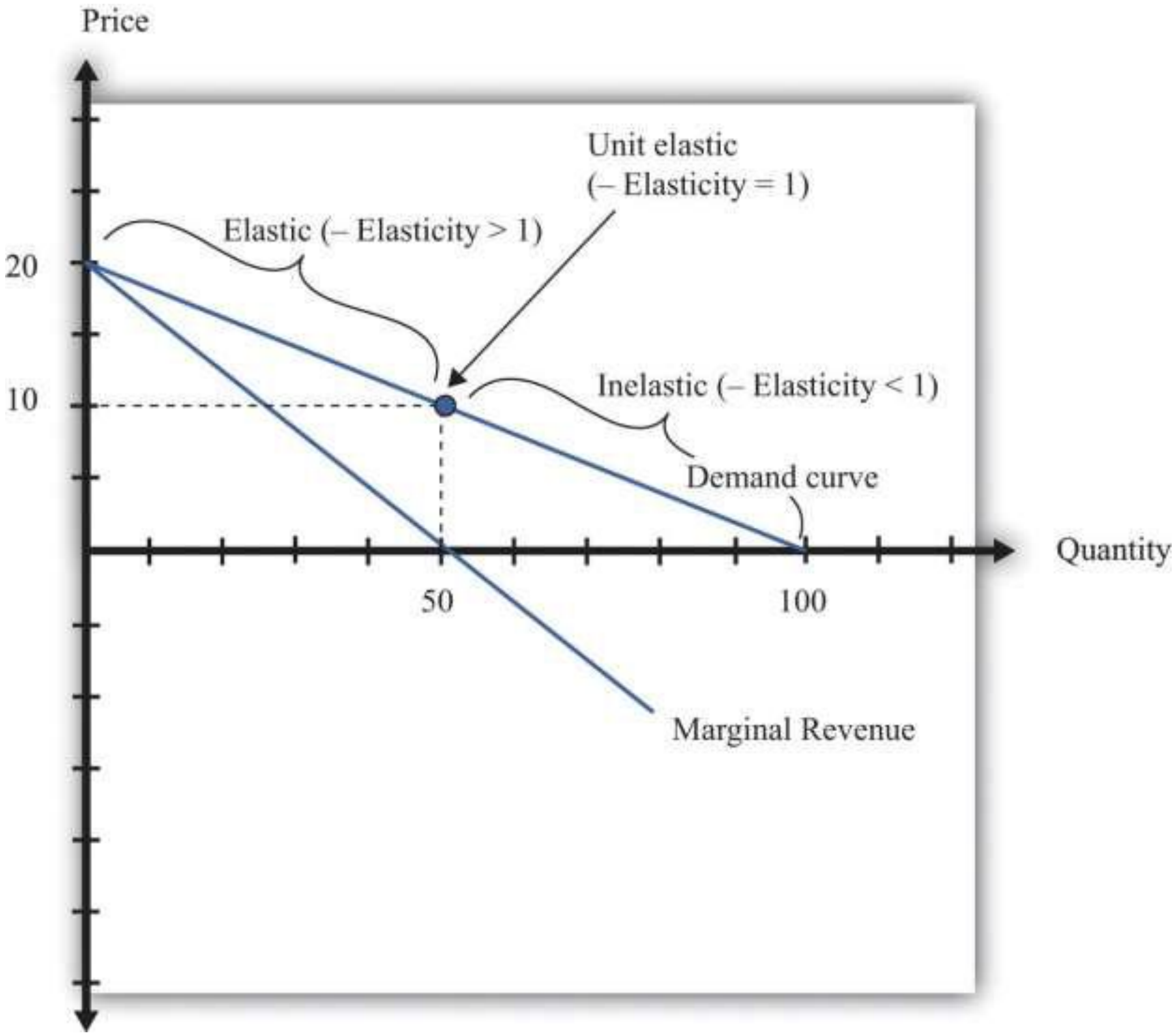
$$-(\text{elasticity of demand}) = -\frac{\text{percentage change in quantity}}{\text{percentage change in price}} = 1.$$

Plugging this into our expression for marginal revenue, we confirm that

$$\text{marginal revenue} = \text{price} \times \left(1 - \frac{1}{-(\text{elasticity of demand})}\right) = 10\left(1 - \frac{1}{1}\right) = 10 \times 0 = 0.$$

At $10, a small change in price leads to no change in revenue. The benefit from selling extra output is exactly offset by the loss from charging a lower price.

Figure 6.13 Marginal Revenue and the Elasticity of Demand



The demand curve can be divided into two parts: at low quantities and high prices, marginal revenue is positive and the demand curve is elastic; at high quantities and low prices, marginal revenue is negative and the demand curve is inelastic.

We can thus divide the demand curve into two parts, as in Figure 6.13 "Marginal Revenue and the Elasticity of Demand". At low quantities and high prices, a firm can increase its revenues by moving down the demand curve—to lower prices and higher output. Marginal revenue is positive. In this region, −(elasticity of demand) is a relatively large number (specifically, it is between 1 and infinity) and we say that the demand curve is relatively elastic. Conversely, at high quantities and low prices, a decrease in price will decrease a firm's revenues. Marginal revenue is negative. In this region, −(elasticity of demand) is between 0 and 1, and we say that the demand curve is inelastic. Table 6.3 represents this schematically.

Table 6.3

−(Elasticity of Demand)	Demand	Marginal Revenue	Effect of a Small Price Decrease
• > −(elasticity of demand) > 1	Relatively elastic	Positive	Increase revenues
−(elasticity of demand) = 1	Unit elastic	Zero	Have no effect on revenues
1 > −(elasticity of demand) > 0	Relatively inelastic	Negative	Decrease revenues

Maximizing Revenues

The market research company advising Ellie made a presentation to her team. The company told them that if they increased their price, they could expect to see a decrease in revenue. At their current price, in other words, marginal revenue was positive. If Ellie's team wanted to maximize revenue, they would need to recommend a reduction in price: down to the point where marginal revenue is $0—equivalently, where −(elasticity of demand) = 1.

Some members of Ellie's team therefore argued that they should try to decrease the price of the product so that they could increase their market share and earn more revenues from the sale of the drug. Ellie reminded them, though, that their goal wasn't to have as much revenue as possible. It was to have as large a *profit*as possible. Before they could decide what to do about price, they needed to learn more about the costs of producing the drug.

KEY TAKEAWAYS

- The demand curve tells a firm how much output it can sell at different prices.
- The elasticity of demand is the percentage change in quantity divided by the percentage change in the price.
- Marginal revenue is the change in total revenue from a change in the quantity sold.

CHECKING YOUR UNDERSTANDING

1. Earlier, we saw that the demand curve was

   quantity demanded = 252 − 300 × price.

a. Suppose Ellie sets the price at $0.42. What is the quantity demanded?

b. Suppose Ellie sets a price that is 10 percent higher ($0.462). What is the quantity demanded?

c. Confirm that −(elasticity of demand) = 1 when the price is $0.42.

If a firm's manager wants to choose a price to maximize revenue, is this the same price that would maximize profits?

If a demand curve has the same elasticity at every point, does it also have a constant slope?

[1] We look at market structure in Chapter 14 "Busting Up Monopolies".

[2] A small mathematical technicality: the equation for the demand curve applies only if both the price and the quantity are nonnegative. At any price greater than the choke price, the quantity demanded is zero, so the demand curve runs along the vertical axis. A negative price would mean a firm was paying consumers to take the product away.

[3] For the derivation of this expression, see the toolkit.

[4] When a demand curve is a straight line, the marginal revenue curve is also a straight line with the same intercept, but it is twice as steep.

6.3 The Costs of a Firm

LEARNING OBJECTIVE

1. What is marginal cost?
2. What costs matter for a firm's pricing decision?

The goods and services that firms put up for sale don't appear from nowhere. Firms produce these goods and incur costs as a result. When a marketing manager is thinking about the price that she sets, she must take into account that different prices lead to different levels of production and hence to different costs for the firm.

Your typical image of a firm probably involves a manufacturing process. This could be very simple indeed. For example, there are firms in Malaysia that produce palm oil. A firm in this context is little more than a big piece of machinery in the middle of a jungle of palm trees. The production process is hot, noisy, and very straightforward: (1) laborers harvest palm nuts from the trees surrounding the factory; (2) these palm nuts are crushed, heated, and pressed to extract the oil; and (3) the oil is placed into barrels and then sold. If you wanted to run a palm oil production business in Malaysia, you would need to purchase the following:

- Machine for extracting the oil
- Truck to transport the oil to market
- Generator to power the machinery
- Fuel to power the generator
- Gasoline for the truck
- Labor time from workers—to harvest the nuts, run the machinery, and transport the oil to market

That's it. It is not difficult to become a palm oil entrepreneur! In this case, it is quite easy to list the main costs of production for the firm.

In other businesses, however, it is much more difficult. Imagine trying to make a similar list for Apple Computer, with all its different products, production plants in different countries, canteens for their workers, pension plans, and so on. Of course, Apple's accountants still need to develop a list of Apple's expenses, but they keep their jobs manageable by grouping Apple's expenditures into various categories.

If this were an accounting textbook, we would discuss these categories in detail. Our task here is simpler: we only need to determine how these costs matter for pricing decisions.

Marginal Cost

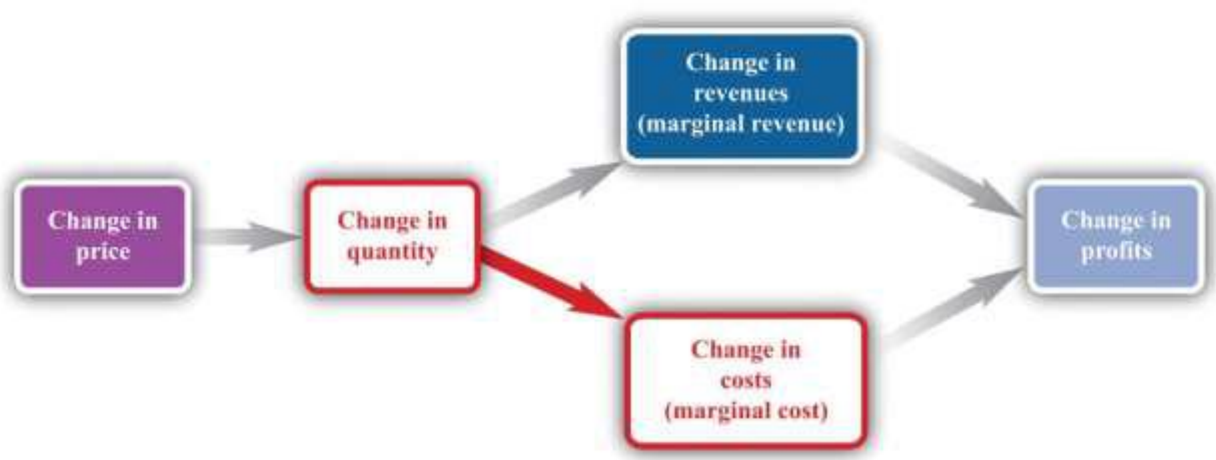
Earlier, we showed how a firm's revenues change when there is a change in quantity that the firm produces. If we also know how a firm's costs change when there is a change in output, we have all the information we need for good pricing decisions. As with revenues, we scale this change by the size of the change in quantity. Figure 6.14 "Marginal Cost" shows how marginal cost fits into our road map for the chapter.

Toolkit: Section 17.14 "Costs of Production"

Marginal cost is the *change* in cost associated with a *change* in quantity of output produced:

$$\text{marginal revenue} = \frac{\text{change in cost}}{\text{change in quantity}}.$$

Figure 6.14 Marginal Cost



When a firm sets a higher price, it sells a smaller quantity and its costs of production decrease.

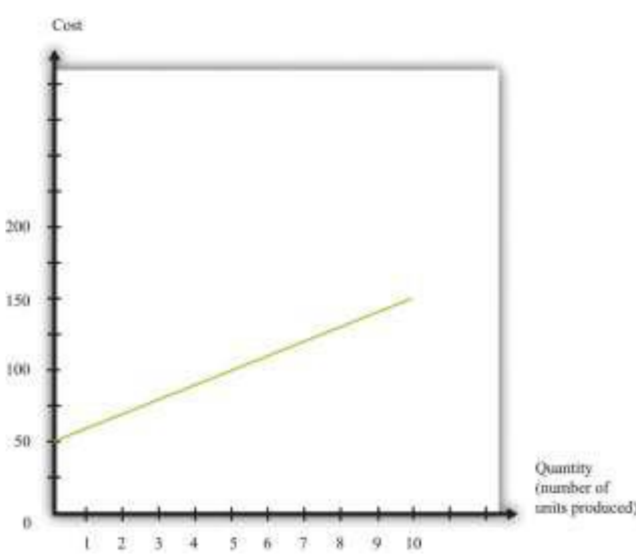
Table 6.4 Marginal Cost

Output	Total Costs	Marginal Cost ($)
0	50	
		10

Output	Total Costs	Marginal Cost ($)
1	60	
		10
2	70	
		10
3	80	
		10
4	90	
		10
5	100	
…		

Table 6.4 "Marginal Cost" shows an example of a firm's costs. It calculates marginal cost in the last column. We have presented this table with marginal cost on separate rows to emphasize that marginal cost is the cost of going from one level of output to the next. In our example, marginal cost—the cost of producing one more unit—is $10. If you want to produce one unit, it will cost you $60. If you want to produce two units, your must pay an *additional* $10 in costs, for a total of $70. If you want to produce three units, you must pay the $70 to produce the first two units, plus the additional marginal cost of $10, for a total cost of $80, and so on. Graphically, marginal cost is the *slope* of the cost line, as shown in Figure 6.15 "An Example of a Cost Function".

Figure 6.15 An Example of a Cost Function



This graph illustrates the cost function for a firm. In this case, the cost function has the equation cost = 50 + 10 × quantity. The cost of producing each additional unit (the marginal cost) is $10.

To emphasize again, *only those costs that change matter for a firm's pricing decision*. When a firm considers producing extra output, many of its costs do not change. We can completely ignore these costs when thinking about optimal pricing. This is not to say that other costs don't matter; quite the contrary. They are critical for a different decision—whether the firm should be in business at all.[1] But as long as we are interested in pricing, we can ignore them.

The costs for developing pharmaceutical products are typically quite high. The drug that Ellie was responsible for was first developed in research laboratories, then tested on animals, and then run through a number of clinical studies on human patients. These studies were needed before the Food and Drug Administration in the United States, and equivalent drug safety organizations in other countries, would approve the drug for sale. But these development costs have no effect on marginal cost because they were all incurred before a single pill could be sold.

Surprisingly, this means that even though the drug was very expensive to develop, Ellie's team—quite correctly—paid no attention to that fact. In determining what price to set, they looked at the price sensitivity of their customers and at marginal cost. They understood that the development costs of the drug were not relevant to the pricing decision. In fact, Ellie's team had only a very vague idea how much the drug had cost to develop: after all, that development had been carried out by a completely different arm of the company in other parts of the world.

You will often hear the opposite argument. It is common for people to say that pharmaceutical companies charge high prices because it costs so much to develop their drugs. This argument is superficially appealing, but it is completely backward. Pharmaceutical companies don't charge high prices because they incur large development costs. They are willing to incur large development costs because they can charge high prices.

Estimating Marginal Cost

Estimating marginal cost is generally much easier than estimating demand because the cost side of the business is largely under the control of the firm. The firm's costs depend on its technology and the decisions made about using that technology. In most medium- or large-sized firms, there is an "operations department" that takes care of the production process. The marketing manager ought to be able to consult with her colleagues in operations and learn about the costs of the firm. Most importantly, even if it is unreasonable to expect an operations manager to know the firm's entire cost function, the operations manager should have a good idea about marginal cost (that is, should know how much it would cost per unit to scale up operations by a small amount). And that is the information the marketing manager needs for her pricing decisions.

KEY TAKEAWAYS

- Marginal cost measures the additional costs from producing an extra unit of output.
- It is only the change in costs—marginal cost—that matter for a firm's pricing decision.

CHECKING YOUR UNDERSTANDING

1. If the marginal cost in Table 6.4 "Marginal Cost" were $20, what would be the cost of producing 10 units of the good?

[1] See Chapter 8 "Growing Jobs".

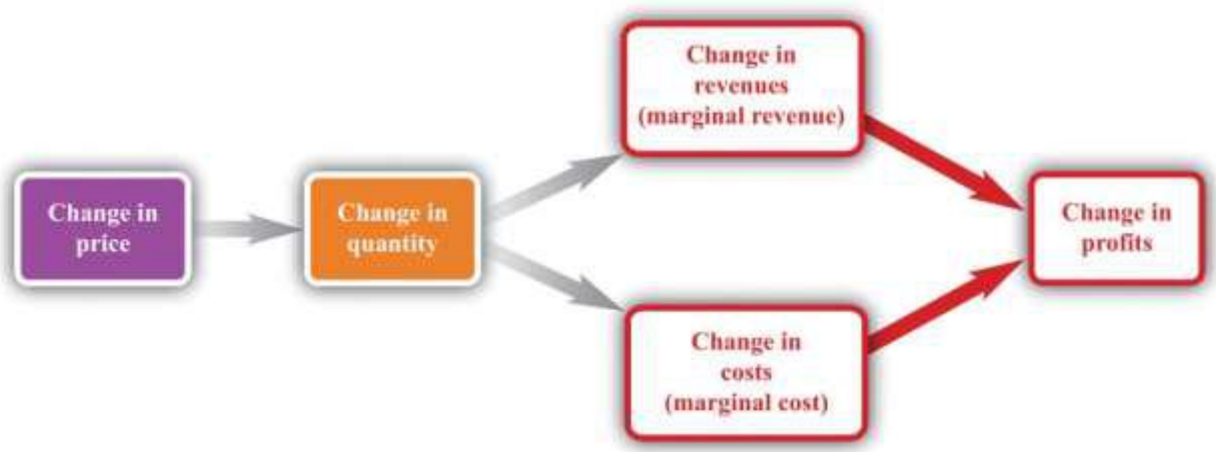
6.4 Markup Pricing: Combining Marginal Revenue and Marginal Cost

LEARNING OBJECTIVES

1. What is the optimal price for a firm?
2. What is markup?
3. What is the relationship between the elasticity of demand and markup?

Let us review the ideas we have developed in this chapter. We know that changes in output lead to changes in both revenues and costs. Changes in revenues and costs lead to changes in profits (see Figure 6.16 "Changes in Revenues and Costs Lead to Changes in Profits"). We have a measure of how much revenues change if output is increased—called marginal revenue, which you can calculate if you know price and the elasticity of demand. We also have a measure of how much costs change if output is increased—this is called marginal cost. Given information on current marginal revenue and marginal cost, a marketing manager can then decide if a firm should change its price. In this section, we derive a rule that tells us how a manager should make this decision.

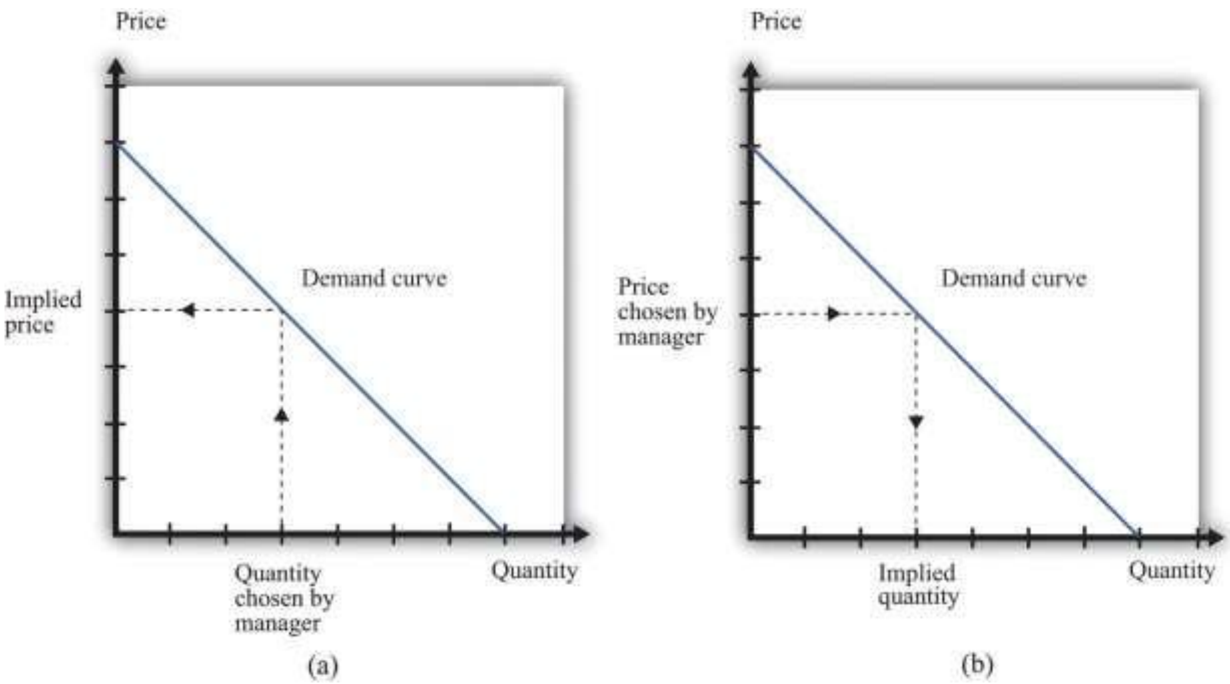
Figure 6.16 Changes in Revenues and Costs Lead to Changes in Profits



When a firm changes its price, this leads to changes in revenues and costs. The change in a firm's profit is equal to the change in revenue minus the change in cost—that is, the change in profit is marginal revenue minus marginal cost. When marginal revenue equals marginal cost, the change in profit is zero, so a firm is at the top of the profit hill.

In the real world of business, firms almost always choose the price they set rather than the quantity they produce. Yet the pricing decision is easier to analyze if we think about it the other way round: a firm choosing what quantity to produce and then accepting the price implied by the demand curve. This is just a matter of convenience: a firm chooses a point on the demand curve, and it doesn't matter if we think about it choosing the price and accepting the implied quantity or choosing the quantity and accepting the implied price (Figure 6.17 "Setting the Price or Setting the Quantity").

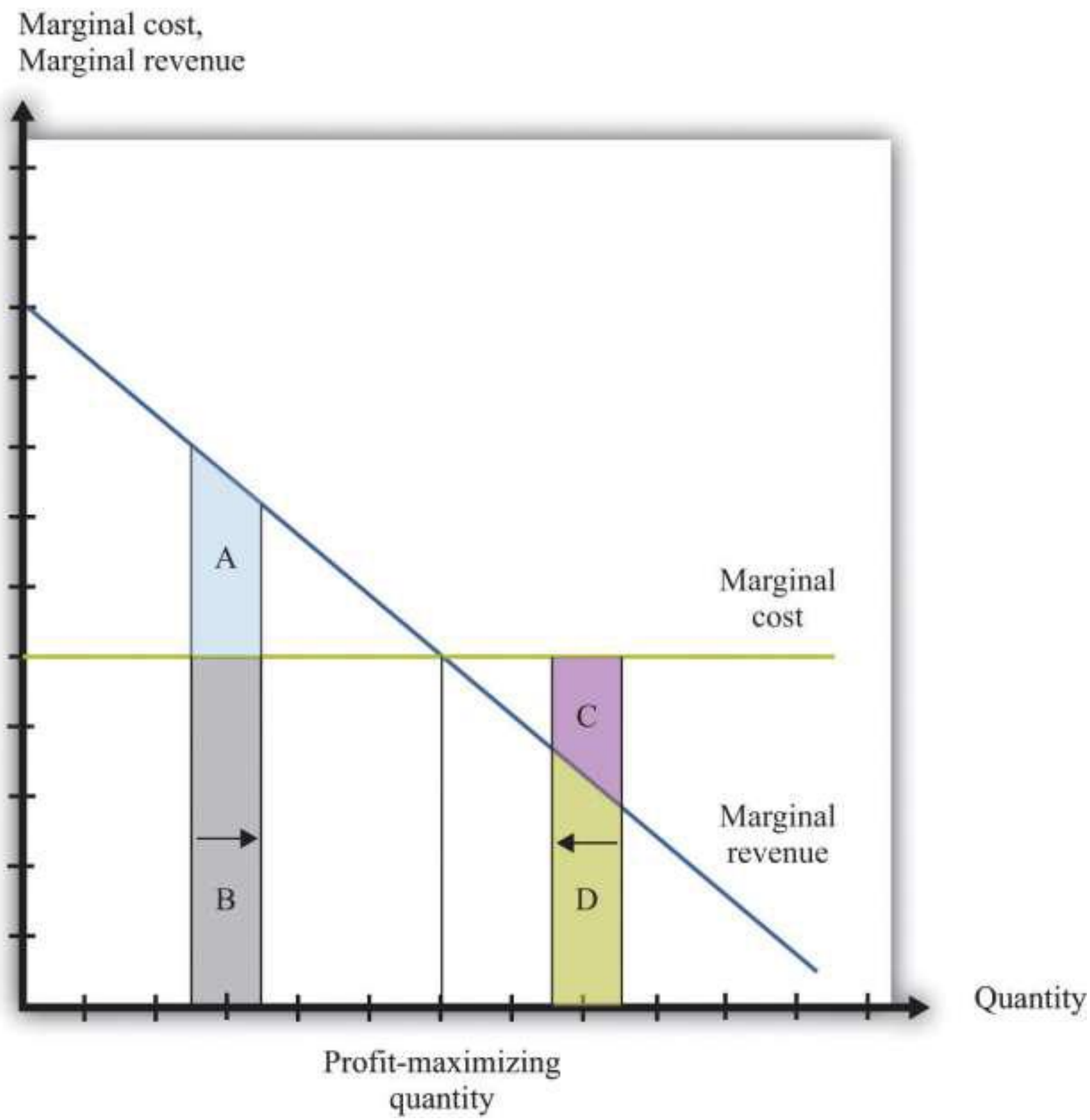
Figure 6.17 Setting the Price or Setting the Quantity



It doesn't matter if we think about choosing the price and accepting the implied quantity or choosing the quantity and accepting the implied price

Suppose that a marketing manager has estimated the elasticity of demand, looked at the current price, and used the marginal revenue formula to discover that the marginal revenue is $5. This means that if the firm increases output by one unit, its revenues will increase by $5. The marketing manager has also spoken to her counterpart in operations, who has told her that the marginal cost is $3. This means it would cost an additional $3 to produce one more unit. From these two pieces of information, the marketing manager knows that an increase in output would be a good idea. An increase in output leads to a bigger increase in revenues than in costs. As a result, it leads to an increase in profits: specifically, profits will increase by $2. This tells the marketing manager that it is a good idea to increase output. From the law of demand, she should think about decreasing the price.

Figure 6.18 Optimal Pricing



To the left of the point marked "profit-maximizing quantity," marginal revenue exceeds marginal cost so increasing output is a good idea. The opposite is true to the right of that point.

Figure 6.18 "Optimal Pricing" shows this idea graphically. To the left of the point marked “profit-maximizing quantity,” marginal revenue exceeds marginal cost. Suppose a firm is producing below this level. If it increases its output, the extra revenue it obtains will exceed the extra cost. We see that an increase in output yields extra revenue equal to the areas $A + B$ and extra costs equal to B. The increase in output yields extra profit, which is equal to A. Increasing its output is thus a good idea. Conversely, to the right of the profit-maximizing point, marginal revenue is less than marginal cost. If a firm reduces its output, the decrease in costs ($C + D$) exceeds the decrease in revenue (D). Decreases in output lead to increases in profit.

Profits are greatest when

marginal revenue = marginal cost.

This is the point where a change in price leads to no change in profits, so we are at the very top of the profit hill that we drew in Figure 6.3 "The Profits of a Firm". See also Figure 6.17 "Setting the Price or Setting the Quantity".

The Markup Pricing Formula

Think about Ellie’s company. If it became more expensive for the company to produce each pill, it seems likely they would respond by raising their costs. Also, we said earlier that their customers are not very sensitive to changes in the price, which should allow them to set a relatively high price. In other words, the profit-maximizing price is related to the elasticity of demand and to marginal cost. These are the two critical ingredients of the pricing decision.

Toolkit: Section 17.15 "Pricing with Market Power"

Firms should set the price as a markup over marginal cost: [1]

$$\text{price}=(1+\text{markup})\times\text{marginal cost}$$

and

$$\text{markup}=1-(\text{elasticity of demand})-1.$$

There are three facts about markup:

1. Markup is greater than or equal to zero—that is, the firm never sets a price below marginal cost.
2. Markup is smaller when demand is more elastic.
3. Markup is zero when the demand curve is perfectly elastic: −(elasticity of demand) = •.

   Ellie's team looked at their numbers. At the current price, −(elasticity of demand) = 1.47. They learned that the marginal cost was $0.28 per pill, and they were charging $0.50 per pill. Their current markup, in other words, was about 79 percent: 0.5 = (1+ 0.79) × 0.28. But if they applied the markup pricing formula based on the *current* elasticity of demand, they could charge a markup of 1/0.47 = 2.12—that is, more than a 200 percent markup, leading to a price of $0.87. It was clear that they could do better by increasing their price

A Pricing Algorithm

To summarize, a manager needs two key pieces of information when determining price:

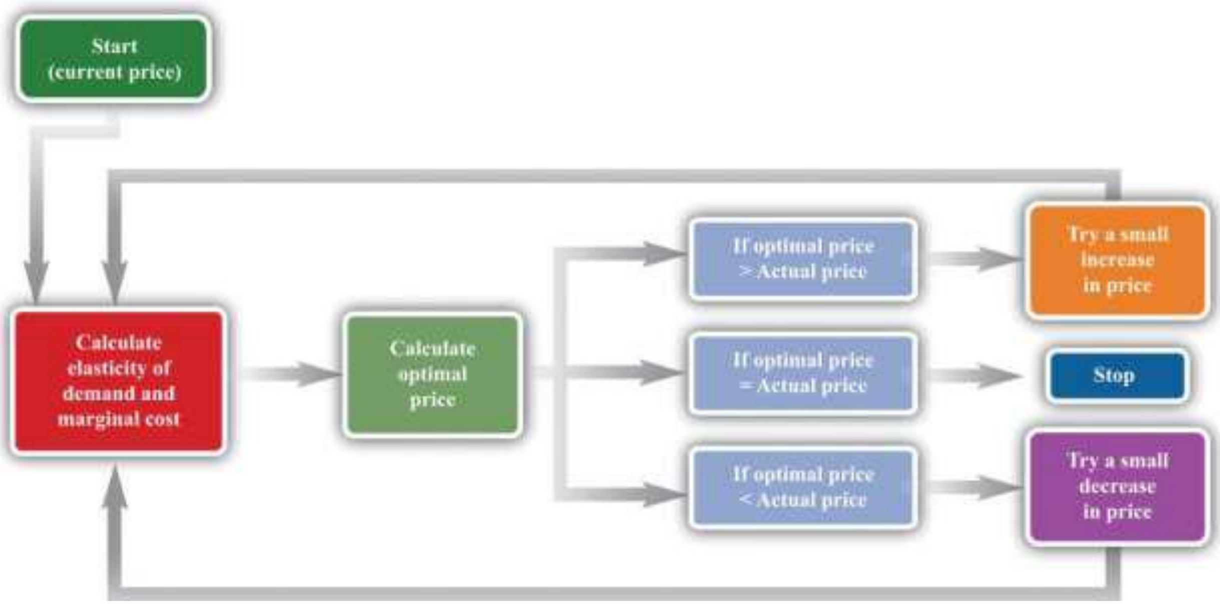
1. **Marginal cost.** We have shown that the profit-maximizing price is a markup over the marginal cost of production. If a manager does not know the magnitude of marginal cost, she is missing a critical piece of information for the pricing decision.
2. **Elasticity of demand.** Once a manager knows marginal cost, she should then set the price as a markup over marginal cost. But this should not be done in an ad hoc manner; the markup must be based on information about the elasticity of demand.

Given these two pieces of information, a manager can then use the markup formula to determine the optimal price. Be careful, though. The markup formula looks deceptively simple, as if it can be used in a "plug-and-play" manner—given marginal cost and the elasticity of demand, plug them into the formula and calculate the optimum price. But if you change the price, both marginal cost and the elasticity of demand are also likely to change. A more reliable way of using this formula is in the algorithm shown in Figure 6.19 "A Pricing Algorithm", which is based on our earlier idea that you should find your way to the top of the profit hill. The five steps are as follows:

1. At your current price, estimate marginal cost and the elasticity of demand.
2. Calculate the optimal price based on those values.
3. If the optimal price is greater than your actual price, increase your price. Then estimate marginal cost and the elasticity again and repeat the process.

4. If the optimal price is less than your actual price, decrease your price. Then estimate marginal cost and the elasticity again and repeat the process.
5. If the current price is equal to this optimal price, leave your price unchanged.

Figure 6.19 A Pricing Algorithm



*This **pricing algorithm** shows how to get the best price for a product.*

Ellie's team members were aware that, even though demand for the drug was apparently not very sensitive to price, they should not immediately jump to a much higher markup. They had found that *based on current marginal cost and elasticity*, the price could be raised. But as they raised the price, they knew that the elasticity of demand would probably also change. Looking more closely at their market research data, they found that at a price of $0.56 (a 100 percent markup), the elasticity of demand would increase to about 2. An elasticity of 2 means that the markup should be 100 percent to maximize profits. Thus—at least if their market research data were reliable—they knew that a price of $0.56 would maximize profits. Ellie recommended to senior management that the price of the drug be raised by slightly over 10 percent, from $0.50 per pill to $0.56 per pill.

Shifts in the Demand Curve Facing a Firm

So far we have looked only at *movements along* the demand curve—that is, we have looked at how changes in price lead to changes in the quantity that customers will buy. Firms also need to understand what factors might cause their demand curve to shift. Among the most important are the following:

- **Changes in household tastes.** Starting around 2004 or so, low-carbohydrate diets started to become very popular in the United States and elsewhere. For some companies, this was a boon; for others it was a problem. For example, companies like Einstein Bros. Bagels or Dunkin' Donuts sell products that are relatively high in carbohydrates. As more and more customers started looking for low-carb alternatives, these firms saw their demand curve shift inward.
- **Business cycle.** Consider Lexus, a manufacturer of high-end automobiles. When the economy is booming, sales are likely to be very good. In boom times, people feel richer and more secure and are more likely to purchase a luxury car. But if the economy goes into recession, potential car buyers will start looking at cheaper cars or may decide to defer their purchase altogether. Many companies sell products that are sensitive to the state of the business cycle. Their demand curves shift as the economy moves from boom to recession.
- **Changes in competitors' prices.** In a business setting, this is a critical concern. If a competitor decreases its price, this means that the demand curve you face will shift inward. For example, suppose that British Airways decides to decrease its price for flights from New York to London. American Airlines will find that its demand curve for that route has shifted inward. Ellie certainly has to worry about this because her company's product has only a small number of competitors. A change in price of a competing blood pressure drug might make a big difference in the sales and profits of Ellie's product.

If the demand curve shifts, should a firm change its price? The answer is yes *if the shift in the demand curve also leads to a change in the elasticity of demand*. In practice, this is likely to be the case, although it is certainly possible for a demand curve to shift without a change in the elasticity of demand. The correct response to a shift in the demand curve is to reestimate the elasticity of demand and then decide if a change in price is appropriate.

Complications

Pricing is a difficult and delicate job, and there are many factors that we have not yet considered: [2]

- By far the most important problem that we have neglected is as follows: When making pricing decisions, *firms may need to take into account how other firms will respond to their decisions*. For example, a manager might estimate her firm's elasticity of demand and marginal cost and determine that she could make more money by decreasing price. That calculation presumes that competing firms keep their prices unchanged. In markets with a small number of competitors, it is instead quite likely that other firms would respond by decreasing their prices. This would cause a firm's demand curve to shift inward and probably leave it worse off than before.
- We have assumed throughout that a firm has to charge the same price for every unit that it sells. In many cases, this is an accurate description of pricing behavior. When a grocery store posts a price, that price holds for every unit on the shelf. But sometimes firms charge different prices for different units—by either charging different prices to different customers or offering individual units at different prices to the same customer. You have undoubtedly encountered examples. Firms sometimes offer quantity discounts, so the price is lower if you buy more units. Sometimes they offer discounts to certain groups of customers, such as cheap movie tickets for students. We could easily fill an entire chapter with other examples—some of which are remarkably sophisticated.
- Firms can have pricing strategies that call for the price to change over time. For example, firms sometimes engage in a strategy known as penetration pricing, whereby they start off by charging a low price in an attempt to develop or expand the market. Imagine that Kellogg's develops a new breakfast cereal. It might decide to offer the cereal at a low price to induce people to try the product. Only after it has developed a group of loyal customers would it start setting their prices according to the markup principle.
- Pricing plays a role in the overall marketing and branding strategy of a firm. Some firms position themselves in the marketplace as suppliers of high-end offerings. They may choose to set high prices for their products to ensure that customers perceive them appropriately. Consider a luxury hotel that is contemplating setting a very low price in the off-season. Even though such a strategy might make sense in terms of its profits at that time, it might do long-term damage to the hotel's

reputation. For various reasons, customers often use the price of a product as an indicator of that product's quality, so a low price can adversely affect a firm's image.

- Psychologists who study marketing have found that demand is sensitive at certain price points. For example, if a firm increases the price of a product from $99.98 to $99.99, there might be very little effect on demand. But if the price increases from $99.99 to $100.00, there might be a much bigger effect because $100.00 is a psychological barrier. Such consumer behavior does not seem completely rational, but there is little doubt that it is a real phenomenon.
- Throughout this chapter, we have said that there is no difference between a firm choosing its price and taking as given the implied quantity or choosing its quantity and taking as given the implied price. Either way, the firm is picking a point on the demand curve. This is true, but there is a footnote that we should add. A firm's demand curve depends on what its competitors are doing and, oddly enough, it *does* make a difference if those competitors are choosing quantities or prices. [3]
- We have focused our attention on the market power of firms as sellers, as reflected in the downward-sloping demand curves they face. Firms can also have market power as buyers. Walmart is such an important customer for many of its suppliers that it can use its position to negotiate lower prices for the goods it buys. Governments are also often powerful buyers and may be able to influence the prices they pay for goods and services. For example, government-run health-care systems may be able to negotiate favorable prices with pharmaceutical companies.

KEY TAKEAWAYS

- At the profit-maximizing price, marginal revenue equals marginal cost.
- Markup is the difference between price and marginal cost, as a percentage of marginal cost.
- The more elastic the demand curve faced by a firm, the smaller the markup.

CHECKING YOUR UNDERSTANDING

1. We said that markup is always greater than zero. Look at the formula for markup. If markup is greater than zero, what must be true about −(elasticity of demand)? Can you see why this must be true? Look back at Figure 6.13 "Marginal Revenue and the Elasticity of Demand" for a hint.

2. If price is a markup over marginal cost, then how does marginal revenue influence the pricing decision of a firm?
3. Starting at the profit-maximizing price, if a firm increases its price, could revenue increase?

[1] This expression comes from combining the formula for marginal revenue and the condition that marginal revenue equals marginal cost. See the toolkit for more details.

[2] We address some of them in other chapters of the book; others are topics for more advanced classes in economics and business strategy.

[3] See Chapter 14 "Busting Up Monopolies" for discussion of this. We should also note that firms often do not know their demand curves with complete certainty. Suppose, for example, that the true demand curve for a firm's product is actually further outward than a firm expects. If the firm sets the price, it will end up with an unexpectedly large quantity being demanded. If the firm sets the quantity, it will end up with an unexpectedly high price.

6.5 The Supply Curve of a Competitive Firm

LEARNING OBJECTIVES

1. What is a perfectly competitive market?
2. In a perfectly competitive market, what does the demand curve faced by a firm look like?
3. What happens to the pricing decision of a firm in a perfectly competitive market?

In this chapter, we have paid a great deal of attention to demand, but we have not spoken of supply. There is a good reason for this: a firm with market power does not have a supply curve. A supply curve for a firm tells us how much output the firm is willing to bring to market at different prices. But a firm with market power looks at the demand curve that it faces and then chooses a point on that curve (a price and a quantity). Price, in this chapter, is something that a firm chooses, not something that it takes as given. What is the connection between our analysis in this chapter and a **market supply curve**?

Perfectly Competitive Markets

If you produce a good for which there are few close substitutes, you have a great deal of market power. Your demand curve is not very elastic: even if you charge a high price, people will be willing to buy the good. On the other hand, if you are the producer of a good that is very similar to other products on the market, then your demand curve will be very elastic. If you increase your price even a little, the demand for your product will decrease a lot.

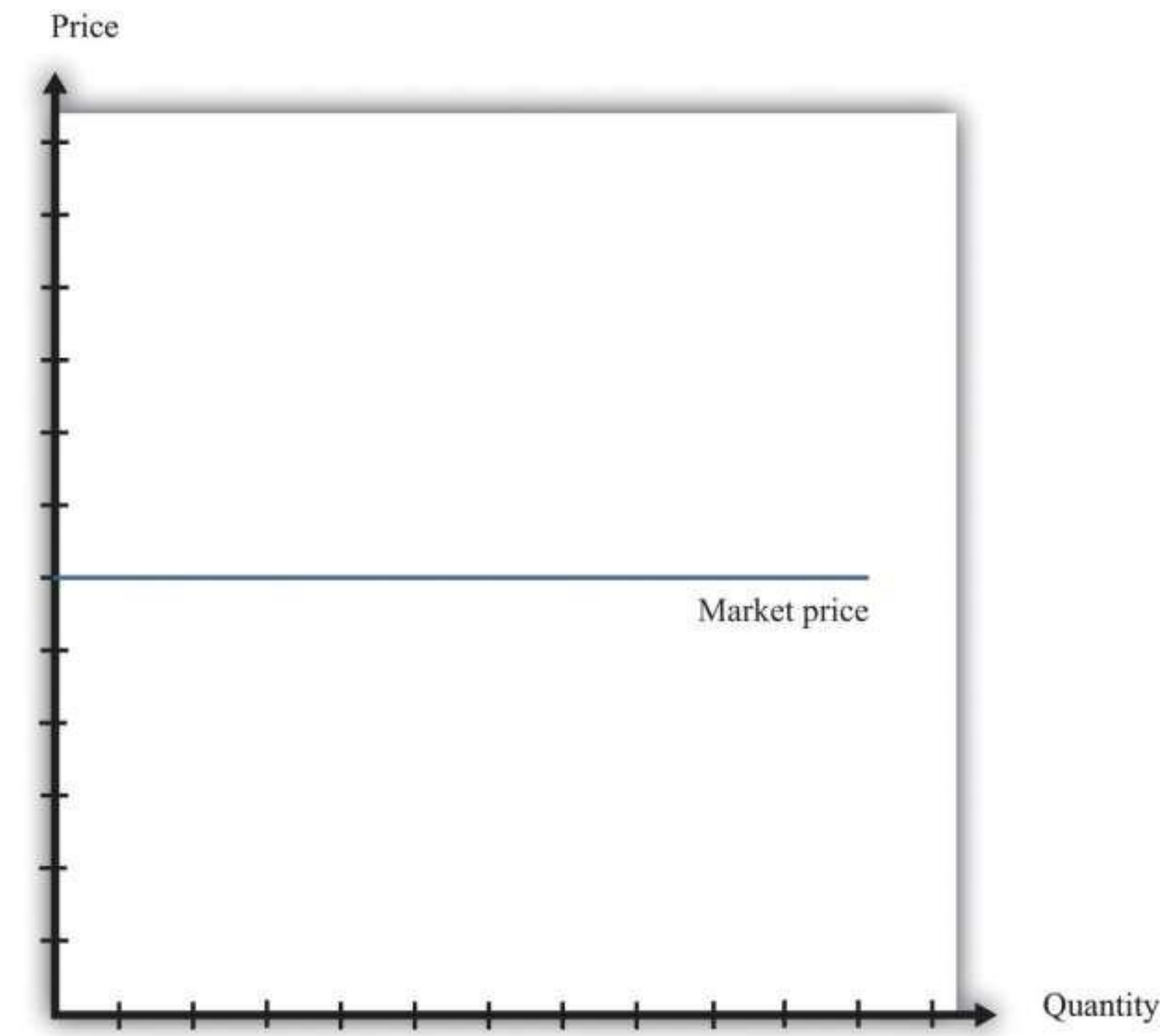
The extreme case is called a perfectly **competitive market**. In a perfectly competitive market, there are numerous buyers and sellers of exactly the same good. The standard examples of perfectly competitive markets are those for commodities, such as copper, sugar, wheat, or coffee. One bushel of wheat is the same as another, there are many producers of wheat in the world, and there are many buyers. Markets for financial assets may also be competitive. One euro is a perfect substitute for another, one three-month US treasury bill is a perfect substitute for another, and there are many institutions willing to buy and sell such assets.

Toolkit: Section 17.9 "Supply and Demand"

You can review the market supply curve and the definition of a perfectly competitive market in the toolkit.

An individual seller in a competitive market has no control over price. If the seller tries to set a price above the going market price, the quantity demanded falls to zero. However, the seller can sell as much as desired at the market price. When there are many sellers producing the same good, the output of a single seller is tiny relative to the whole market, and so the seller's supply choices have no effect on the market price. This is what we mean by saying that the seller is "small." It follows that a seller in a perfectly competitive market faces a demand curve that is a horizontal line at the market price, as shown in Figure 6.20 "The Demand Curve Facing a Firm in a Perfectly Competitive Market". This demand curve is *infinitely elastic*: −(elasticity of demand) = ∞. Be sure you understand this demand curve. As elsewhere in the chapter, it is the *demand faced by an individual firm*. In the background, there is a market demand curve that is downward sloping in the usual way; the market demand and market supply curves together determine the market price. But an individual producer does not experience the market demand curve. The producer confronts an infinitely elastic demand for its product.

Figure 6.20 The Demand Curve Facing a Firm in a Perfectly Competitive Market



The demand curve faced by a firm in a perfectly competitive market is infinitely elastic. Graphically, this means that it is a horizontal line at the market price.

Everything we have shown in this chapter applies to a firm facing such a demand curve. The seller still picks the best point on the demand curve. But because the price is the same everywhere on the demand curve, picking the best point means picking the best quantity. To see this, go back to the markup formula. When demand is infinitely elastic, the markup is zero:

$$\text{markup}=1-(\text{elasticity of demand})-1=1\infty=0,$$

so price equals marginal cost:

price = (1 + markup) × marginal cost = marginal cost.

This makes sense. The ability to set a price above marginal cost comes from market power. If you have no market power, you cannot set a price in excess of marginal cost. A perfectly competitive firm chooses its level of output so that its marginal cost of production equals the market price.

We could equally get this conclusion by remembering that

marginal revenue = marginal cost

and that when −(elasticity of demand) is infinite, marginal revenue equals price. If a competitive firm wants to sell one more unit, it does not have to decrease its price to do so. The amount it gets for selling one more unit is therefore the market price of the product, and the condition that marginal revenue equals marginal cost becomes

price = marginal cost.

For the goods and services that we purchase regularly, there are few markets that are truly perfectly competitive. Often there are many sellers of goods that may be very close substitutes but not absolutely identical. Still, many markets are close to being perfectly competitive, in which case markup is very small and perfect competition is a good approximation.

The Supply Curve of a Firm

Table 6.5 "Costs of Production: Increasing Marginal Cost" shows the costs of producing for a firm. In contrast to Table 6.4 "Marginal Cost", where we supposed marginal cost was constant, this example has higher marginal costs of production when the level of output is greater. [1]

Table 6.5 Costs of Production: Increasing Marginal Cost

Output	Total Costs ($)	Marginal Cost ($)
1	22	12
2	38	16
3	58	20
4	82	24
5	110	28

Output	Total Costs ($)	Marginal Cost ($)
…		

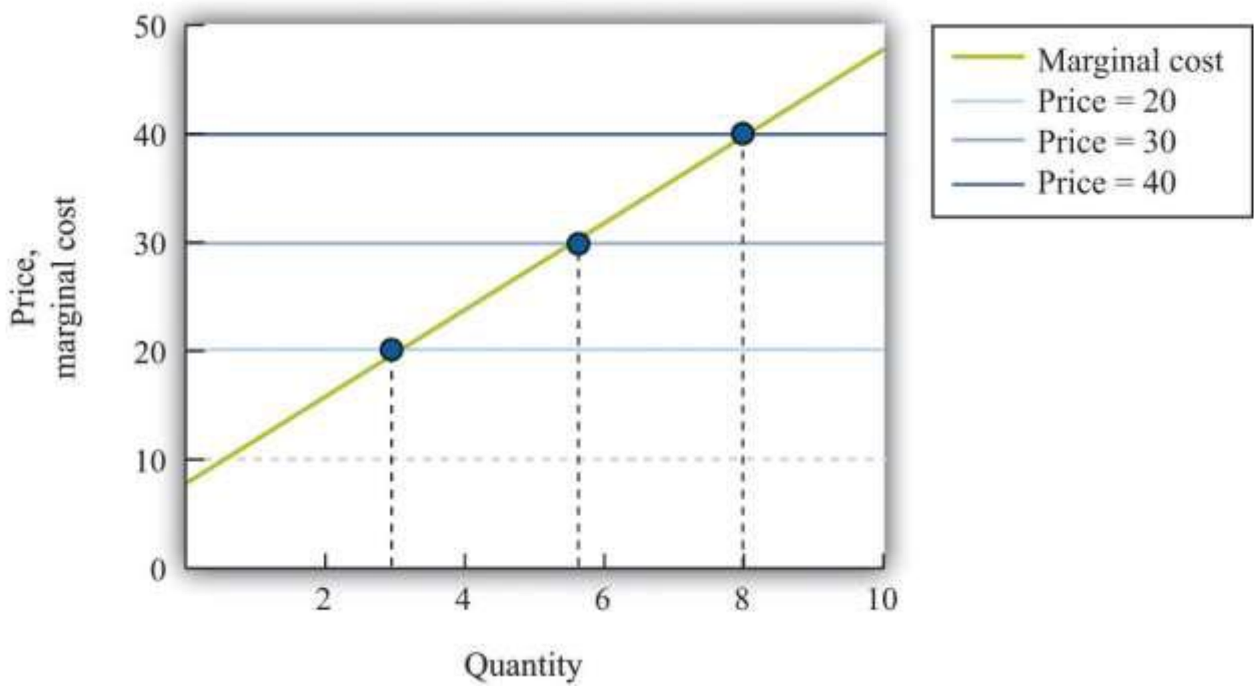
Figure 6.21 "The Supply Curve of an Individual Firm" shows how we derive the supply curve of an individual firm given such data on costs. The supply curve tells us how much the firm will produce at different prices. Suppose, for example, that the price is $20. At this price, we draw a horizontal line until we reach the marginal cost curve. At that point, we draw a vertical line to the quantity axis. In this way, you can find the level of output such that marginal cost equals price. Looking at the figure, we see that the firm should produce 3 units because the marginal cost of producing the third unit is $20. When the price is $30, setting marginal cost equal to price requires the firm to produce 5.5 units. When the price is $40, setting marginal cost equal to price requires the firm to produce 8 units.

The supply curve shows us the quantity that a firm will produce at different prices. Figure 6.21 "The Supply Curve of an Individual Firm" reveals something remarkable: *the* individual supply curve *of the firm is the marginal cost curve*. They are the same thing. As the price a firm faces increases, it will produce more. Note carefully how this is worded. We are *not* saying that if a firm produces more, it will charge a higher price. Firms in a competitive market must take the price as given. Instead, we think about the response of a firm to a change in the price. [2]

Toolkit: Section 17.9 "Supply and Demand"

The individual supply curve shows how much output a firm in a perfectly competitive market will supply at any given price. Provided that a firm is producing output, the supply curve is the same as marginal cost curve.

Figure 6.21 The Supply Curve of an Individual Firm



The firm chooses its quantity such that price equals marginal cost, which implies that the marginal cost curve of the firm is the supply curve of the firm.

KEY TAKEAWAYS

- A perfectly competitive market has a large number of buyers and sellers of exactly the same good.
- In a perfectly competitive market, an individual firm faces a demand curve with infinite elasticity.
- In a perfectly competitive market, the firm does not set a price but chooses a level of output such that marginal cost equals the market price.

CHECKING YOUR UNDERSTANDING

1. Explain why the demand curve a firm faces in a perfectly competitive market is horizontal even though the market demand curve is not horizontal.
2. Why is the cost of one unit $60 in Table 6.4 "Marginal Cost" but only $22 inTable 6.5 "Costs of Production: Increasing Marginal Cost"?

[1] Total cost in Table 6.5 "Costs of Production: Increasing Marginal Cost" is 50 + 10 × quantity + 2 × quantity^2.

[2] The individual firm's supply curve is an exact counterpart to something we show inChapter 3 "Everyday Decisions", where we derive the demand curve for an individual. We show that an individual buys a good up to the point where marginal valuation equals price. From this we can conclude that the demand curve for an individual is

the same as the individual's marginal valuation curve. In Chapter 7 "Why Do Prices Change?", we use an individual firm's supply curve as the basis for the **market supply curve**. Likewise, we use the individual demand curve as the basis for the **market demand curve**. By combining these curves, we obtain the **supply**-**and**-**demand** framework, which we can use to understand changing prices in an economy.

6.6 End-of-Chapter Material

In Conclusion

Choosing the right price is one of the hardest problems that a manager faces. It is also one of the most consequential: few other decisions have such immediate impact on the health and success of a firm. It is hardly surprising that firms devote considerable resources to deciding on the price to set. Even though firms operate in many different market settings, our analysis of markup pricing is very general: it applies to firms in all sorts of different circumstances. It is thus a powerful tool for understanding the behavior of firms in an economy.

One goal of this textbook is to help you make good decisions, both in your everyday life and in your future careers. In this chapter, we have set out the principles of how prices *should* be set, assuming that the goal of a manager is to make as much profit for a firm as possible. It does not necessarily follow, however, that this is how managers actually behave in real life. Does this chapter just describe a make-believe world of economists or does it also describe how prices are set in the real business world?

The answer is a bit of both. Managers must think carefully about costs and demand when setting prices. Market research firms routinely investigate consumers' price sensitivities and estimate elasticities. At the same time, pricing decisions are sometimes more haphazard than this chapter might suggest. In practice, managers often use rules of thumb or standard markups that are not necessarily solidly based on the elasticity of demand.

There is one reason to think that managers do not stray too far from the prices that maximize their firms' profits, however. Business is a competitive affair, and firms that make poor decisions will often not survive in the marketplace. If a firm consistently sets the wrong price, it will make less money than its competitors and will probably be forced out of business or taken over by another firm that can do a better job of management. The marketplace imposes a harsh discipline on badly managed firms, but the end result is—usually at least—a more efficient and better-functioning economy.

Key Link

- College pricing: http://www.collegeboard.com/student/pay/add-it-up/4494.html

EXERCISES

1. We suggested that a grocery store could conduct an experiment to find a demand curve by charging a different price each week for some product.
    a. Do you think that technique would be more accurate for a perishable good, such as milk, or for a nonperishable good, such as canned tuna? Why?
    b. Do you think the technique would be more accurate if the firm announced the price each week in advance or if it just let customers discover the different prices when they came to the store? Why?

Extend Table 6.5 "Costs of Production: Increasing Marginal Cost" for output levels 6–10. What does Table 6.5 "Costs of Production: Increasing Marginal Cost" look like if the fixed cost is $100?

Suppose your company is selling a product that is an inferior good. What do you think will happen to the demand curve facing your firm when the economy goes into recession?

Suppose you are a producer of DVDs and imagine that producers of DVD players decrease their prices. What do you think will happen to the demand curve you face?

If you were running a fast-food restaurant, what factors would you take into account in setting a price for burgers?

Suppose a monopolist could produce an extra unit at zero marginal cost and, at the current price, faces a demand curve with an −(elasticity of demand) of 2. Should the monopolist raise or lower its price to make more profit?

Suppose that instead of maximizing profit, the monopolist in Question 6 wants to maximize revenues. Would it behave any differently? What if the marginal cost was positive?

If the price of steak is $25.00 a pound and the −(elasticity of demand) is 2, what decrease in price would lead the quantity sold to increase by 4 percent?

Explain why marginal revenue must be less than the price when a firm faces a downward-sloping demand curve.

A monopolist is maximizing profit. Perhaps due to an innovation in some other product line, he finds that the elasticity of demand for his product is lower. What will this change in the elasticity of demand due to the profit of the monopolist? How will the monopolist respond to this change?

The following is an excerpt from an article in the Singaporean newspaper, the *Straits Times*:

Singaporeans with a sweet tooth could soon find themselves paying more for their favourite treats, as bakers and confectioners buckle under soaring sugar prices.

Since March last year, the price of white sugar has shot up by 70 per cent, according to the New York Board of Trade. As if that didn't make life difficult enough for bakers, butter and cheese prices have also risen, by 31 per cent and 17 per cent respectively.

The increases have been caused by various factors: a steep drop in Thailand's sugarcane production due to drought, higher sea freight charges, increasing demand from China's consumers for dairy products and the strong Australian and New Zealand dollar.

For the consumer in Singapore, what this may eventually boil down to is a more expensive bag of cookies, with prices at some bakeries expected to rise between 10 and 20 per cent.

[The owner of a Singapore bakery, Mr. Leong Meng Pock], said that he intends to raise prices possibly as early as next month. A sugared doughnut at his shop sells for 50 cents [about US$0.30] and a slice of Black Forest cake for $1.80 [about US$1.13], prices that have remained unchanged since 1990. Next month, the doughnuts may go up to 60 cents and the Black Forest cake to $2.

Said Mr Leong: "In Singapore, you have bread and cake prices that are at least 10 years old. This is especially true for the HDB [government-subsidized housing] neighborhoods, where customers are very price-sensitive." [1]

a. Do you think bakers face a demand curve that is relatively elastic or relatively inelastic? Why?
b. What has happened to their marginal cost?

Explain the difference between a shift in the demand curve and movement along a demand curve.

If you observe the price of a product, then you can infer the marginal cost of the product if and only if the market is competitive. Explain.

Spreadsheet Exercise

1. Suppose that the cost function for a product is given by total costs = 100 + 2,000 × quantity. Create a spreadsheet to calculate the costs for different levels of output and use it to produce a graph like Figure 6.16 "Changes in Revenues and Costs Lead to Changes in Profits".

Economics Detective

1. What prices are your local gas stations currently charging for gas? Do the stations generally have the same price for gas? If not, what would explain the differences in prices they set? Do the stations charge the same price all the time or does the price change? When the price changes, what might be the reason for that change?
2. Think about the college you are attending. What determines the profit of the college—what are its revenues and what are its costs? What is tuition at your college? Would you advocate an increase or a decrease in tuition rates to increase revenue at your college? What factors determine the elasticity of demand faced by your college?

[1] See http://straitstimes.asia1.com.sg.

Chapter 7
Why Do Prices Change?

Prices in the News

Here are two recent headlines. The first discusses beer prices in England.

Price of a Pint "Could Rise 60%"

The average price of a pint of beer could hit £4 [about $8]...

Scottish & Newcastle today forecast "material price increases" next year. The brewer, which sells three of the top 10 beer brands in Europe including Kronenbourg and Foster's, is also reviewing its supply chain in a bid to cut costs.

Industry experts say the cost of an average pint will rise by at least 15p, although some are now predicting rises of up to 60%....

"It is a bleak time for everyone," said Iain Lowe, research and information manager at Camra [The Campaign for Real Ale]. "These price rises have been predicted for a long time." [1]

The second concerns sales of baseball merchandise in Detroit at the start of the 2010 season.

Tigers' Merchandise Off to Roaring Start

Opening Day is still a day away for Detroit's baseball fans, but its impending arrival already is generating its share of Detroit Tiger retail hits.

Thousands of fans have flooded Comerica Park's pro shop and other Metro Detroit sporting outlets in anticipation of Friday afternoon's home game against the Cleveland Indians, snapping up Tigers jerseys, T-shirts and hats bearing the surnames Cabrera, Verlander and Damon, even Granderson—Tigers old and new.

...

Though it's too soon to tell which Tigers will prove most popular at the checkout line, former players have been relegated to the discount bin.

"Retailers take a pretty aggressive stand," Powell said. Most shops have marked down jerseys and T-shirts branded with ex-Tigers between 25 and 50 percent.

"Granderson and Polanco—we discount the price," said Brian King, owner of Sports Authentics in Rochester Hills. "Unfortunately, you can't take the names off the back." [2]

We could have picked thousands of other examples. If you search Google's news aggregator on any day with a string such as "an increase in the price of" you will find dozens, perhaps hundreds, of recent news articles that contain this phrase. Our task in this chapter is to see where all these price changes come from and what they imply for other economic variables, such as the quantity of these goods traded.

To see how good you are at this, think about these two stories. Can you explain why the price of beer increased? Can you explain why "Granderson" T-shirts are being sold at discounted prices? What do you think happened to the quantity of beer sold as the price increased? What do you think happened to the quantity of T-shirts sold as the price decreased?

Understanding the sources and consequences of changing prices in the economy is one of the most important tasks of an economist. There is an almost endless list of such analyses in economics. In fact, most of the applications in this textbook ultimately come down to understanding, explaining, and predicting changes in prices. The question that motivates this chapter is so important that we have chosen it as the title:

Why do prices change?

Road Map

All prices in the economy are ultimately chosen by someone. Sometimes they are chosen by marketing or pricing managers in big companies. Sometimes they are chosen by bidders in an auction. Sometimes they are agreed on by the buyer and the seller after bargaining. [3] Yet we can often make good predictions about prices without looking closely at the individual decision making of buyers and sellers by summarizing their decisions with demand curves and supply curves. Building on the ideas of the individual demand curve and a firm's supply curve for a good or service, we develop the ideas of supply and demand for an entire market. [4] In this chapter, we look at the trade that occurs between firms and households or among different firms in the economy. In the business world, these are called business-to-consumer (B2C) and business-to-business (B2B) trade, respectively. The market demand and supply curves that we derive allow us to predict what will happen to prices and quantities traded when there are changes that influence the market. [5]

An old joke says that you can ask an economist any question, and he will always give the same answer: supply and demand. Yet—strictly speaking—we are supposed to use the supply-and-demand framework only when we are talking about a competitive market—a market in which a homogeneous good is traded by a large number of buyers and sellers. In practice, economists and others use the framework all the time in settings where these assumptions do not hold. Perhaps surprisingly, this can be a completely reasonable thing to do, and we explain why.

Once we understand why prices change, we consider the implications of these price changes for the functioning of the economy. Prices convey information to both producers and consumers. When the price of a good or a service increases, it encourages consumers to consume less and producers to produce more. As we will see, this means that prices play a crucial role in allocating resources in the economy.

We finish this chapter by looking at three very significant markets in the economy: the labor market, the credit market, and the foreign exchange market. Understanding how these three markets work is necessary for a good understanding of both microeconomics and macroeconomics.

[1] Teena Lyons, "The Price of a Pint Could Rise 60%," *Guardian Unlimited*, November 20, 2007, accessed January 27, 2011,http://www.guardian.co.uk/business/2007/nov/20/fooddrinks.foodanddrink?gusrc=rss&feed=networkfront.

[2] "Tigers' Merchandise off to Roaring Start," *The Detroit News*, April 8, 2010, accessed January 27, 2011, http://www.detnews.com/article/20100408/BIZ/4080349/1129/Tigers-merchandise-off-to-roaring-start.

[3] We discuss these choices in Chapter 5 "eBay and craigslist" and Chapter 6 "Where Do Prices Come From?".

[4] Individual demand and supply curves are introduced in Chapter 3 "Everyday Decisions" and Chapter 6 "Where Do Prices Come From?".

[5] Chapter 5 "eBay and craigslist" also looks at supply and demand in the context of trade between individuals.

7.1 Market Supply and Market Demand

LEARNING OBJECTIVES

1. How is the market demand curve derived?
2. What is the slope of the market demand curve?
3. How is the market supply curve derived?
4. What is the slope of the market supply curve?
5. What is the equilibrium of a perfectly competitive market?

We begin the chapter with the **individual demand curve**—sometimes also called the household demand curve—that is based on an individual's choice among different goods. (In this chapter, we use the terms *individual* and *household* interchangeably.) We show how to build the market demand curve from these individual demand curves. Then we do the same thing for supply, showing how to build a market supply curve from the supply curves of individual firms. Finally, we put them together to obtain the market equilibrium.

Market Demand

Figure 7.1 "The Demand Curve of an Individual Household" is an example of a household's demand for chocolate bars each month. Taking the price of a chocolate bar as given, as well as its income and all other prices, the household decides how many chocolate bars to buy. Its choice is represented as a point on the household's demand curve. For example, at $5, the household wishes to consume five chocolate bars each month. The remainder of the household income—which is its total income minus the $25 it spends on chocolate—is spent on other goods and services. If the price decreases to $3, the household buys eight bars every month. In other words, the quantity demanded by the household increases. Equally, if the price of a chocolate bar increases, the quantity demanded decreases. This is the law of demand in operation.

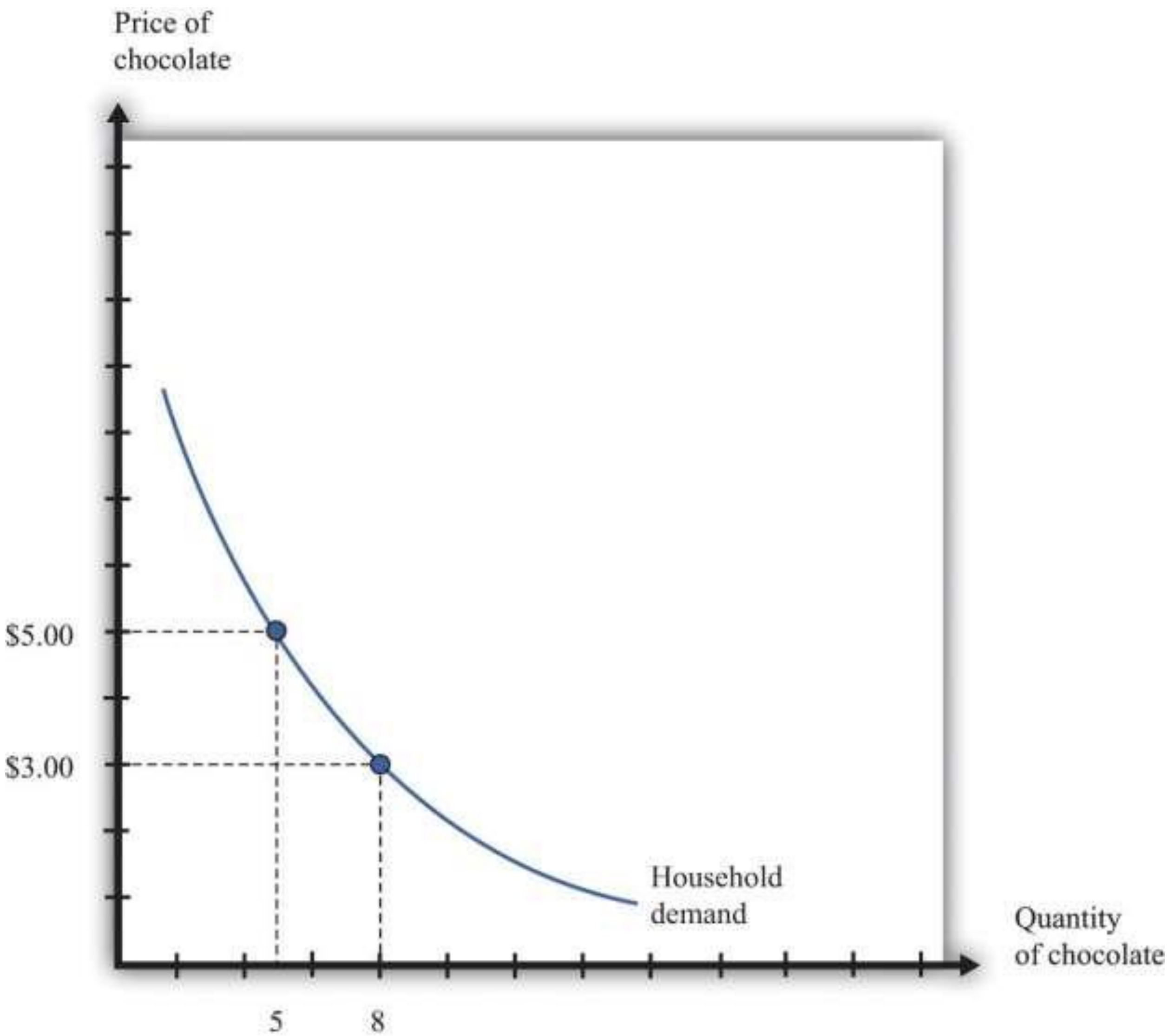
One way to summarize this behavior is to say that the household compares its**marginal valuation** from one more chocolate bar to price. The marginal valuation is a measure of how much the household would like one more chocolate bar. The household will keep buying chocolate bars up to the point where

marginal valuation = price.

Toolkit: Section 17.1 "Individual Demand"

You can review the foundations of individual demand and the idea of marginal valuation in the toolkit.

Figure 7.1 The Demand Curve of an Individual Household



The household demand curve shows the quantity of chocolate bars demanded by an individual household at each price. It has a negative slope: higher prices lead people to consume fewer chocolate bars.

Table 7.1 Individual and Market Demand

Price ($)	Household 1 Demand	Household 2 Demand	Market Demand
1	17	10	27

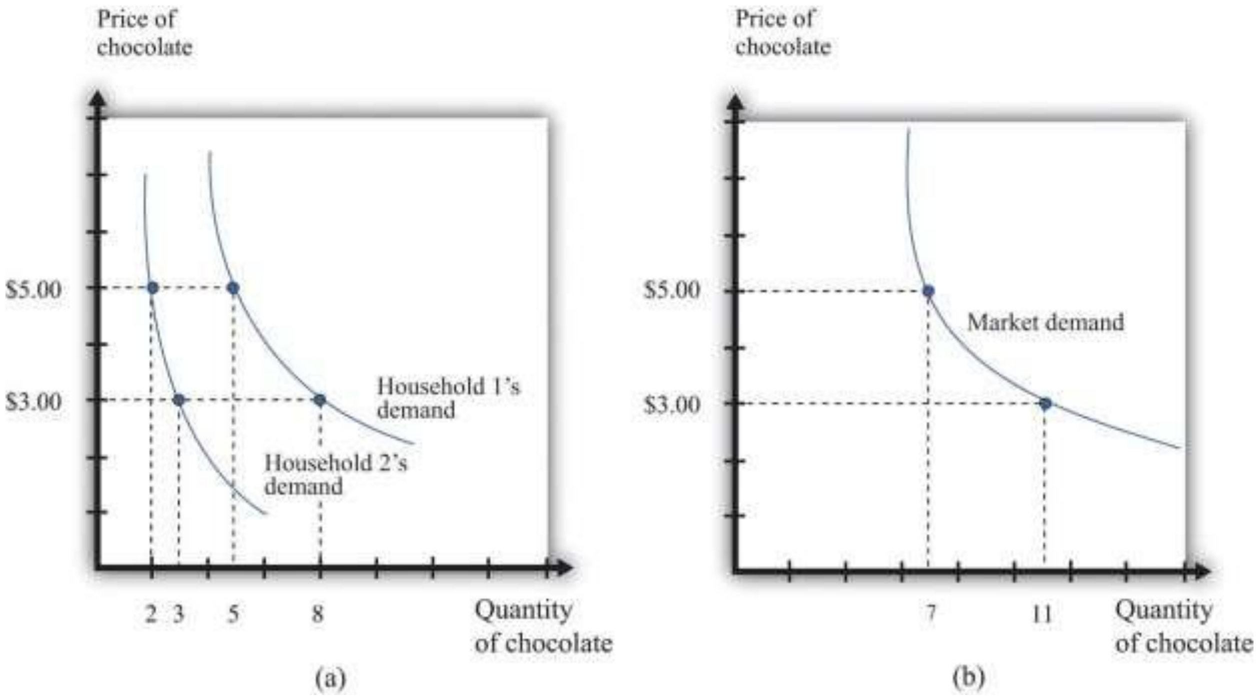
Price ($)	Household 1 Demand	Household 2 Demand	Market Demand
3	8	3	11
5	5	2	7
7	4	1.5	5.5

In most markets, many households purchase the good or the service traded. We need to add together all the demand curves of the individual households to obtain the **market demand curve**. To see how this works, look at Table 7.1 "Individual and Market Demand" and Figure 7.2 "Market Demand". Suppose that there are two households. Part (a) of Figure 7.2 "Market Demand" shows their individual demand curves. Household 1 has the demand curve from Figure 7.1 "The Demand Curve of an Individual Household". Household 2 demands fewer chocolate bars at every price. For example, at $5, household 2 buys 2 bars per month; at $3, it buys 3 bars per month. To get the market demand, we simply add together the demands of the two households at each price. For example, when the price is $5, the market demand is 7 chocolate bars (5 demanded by household 1 and 2 demanded by household 2). When the price is $3, the market demand is 11 chocolate bars (8 demanded by household 1 and 3 demanded by household 2). When we carry out the same calculation at every price, we get the market demand curve shown in part (b) of Figure 7.2 "Market Demand".

Toolkit: Section 17.9 "Supply and Demand"

You can review the market demand curve in the toolkit.

Figure 7.2 Market Demand



Market demand is obtained by adding together the individual demands of all the households in the economy.

Because the individual demand curves are downward sloping, the market demand curve is also downward sloping: the law of demand carries across to the market demand curve. As the price decreases, each household chooses to buy more of the product. Thus the quantity demanded increases as the price decreases. Although we used two households in this example, the same idea applies if there are 200 households or 20,000 households. In principle, we could add together the quantities demanded at each price and arrive at a market demand curve.

There is a second reason why demand curves slope down when we combine individual demand curves into a market demand curve. Think about the situation where each household has a **unit demand curve**: that is, each individual buys at most one unit of the product. As the price decreases, the number of individuals electing to buy increases, so the market demand curve slopes down. [1] In general, both mechanisms come into play.

- As price decreases, some households decide to enter the market; that is, these households buy some positive quantity other than zero.

- As price decreases, households increase the quantity that they wish to purchase.

  When the price decreases, there are more buyers, and each buyer buys more.

Market Supply

In a competitive market, a single firm is only one of the many sellers producing and selling exactly the same product. The demand curve facing a firm exhibits **perfectly elastic demand**, which means that it sets its price equal to the price prevailing in the market, and it chooses its output such that this price equals its marginal cost of production. [2] If it were to try to set a higher price, it could not sell any output at all. If it were to set a lower price, it would be throwing away profits. Thus, for a competitive firm, the quantity produced satisfies this condition:

price = marginal cost.

Toolkit: Section 17.2 "Elasticity"

For more information on elasticity, see the toolkit.

We typically expect that marginal cost will increase as a firm produces more output. Marginal cost is the cost of producing one *extra* unit of output. The cost of producing an additional unit of output generally increases as firms produce a larger and larger quantity. In part, this is because firms start to hit constraints in their capacities to produce more product. For example, a factory might be able to produce more output only by running extra shifts at night, which require paying higher wages.

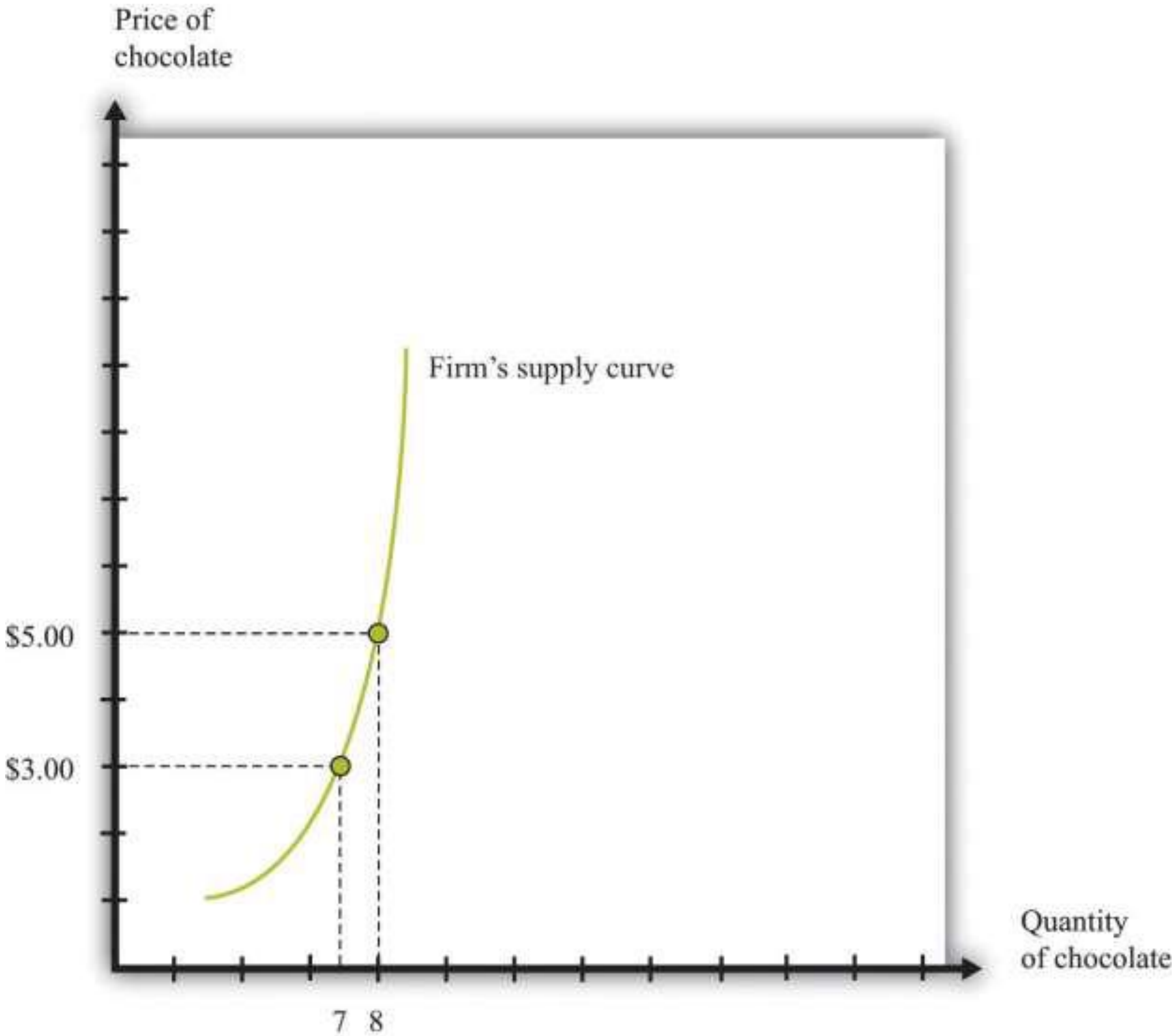
If marginal cost is increasing, then we know the following:

- Given a price, there is only one level of output such that price equals marginal cost.
- As the price increases, a firm will produce more.

Indeed, the supply curve of an individual firm is the same as its marginal cost curve.

Figure 7.3 "The Supply Curve of an Individual Firm" illustrates the supply curve for a firm. A firm supplies seven chocolate bars at $3 and eight chocolate bars at $5. From this we can deduce that the marginal cost of producing the seventh chocolate bar is $3. Similarly, the marginal cost of producing the eighth chocolate bar is $5.

Figure 7.3 The Supply Curve of an Individual Firm

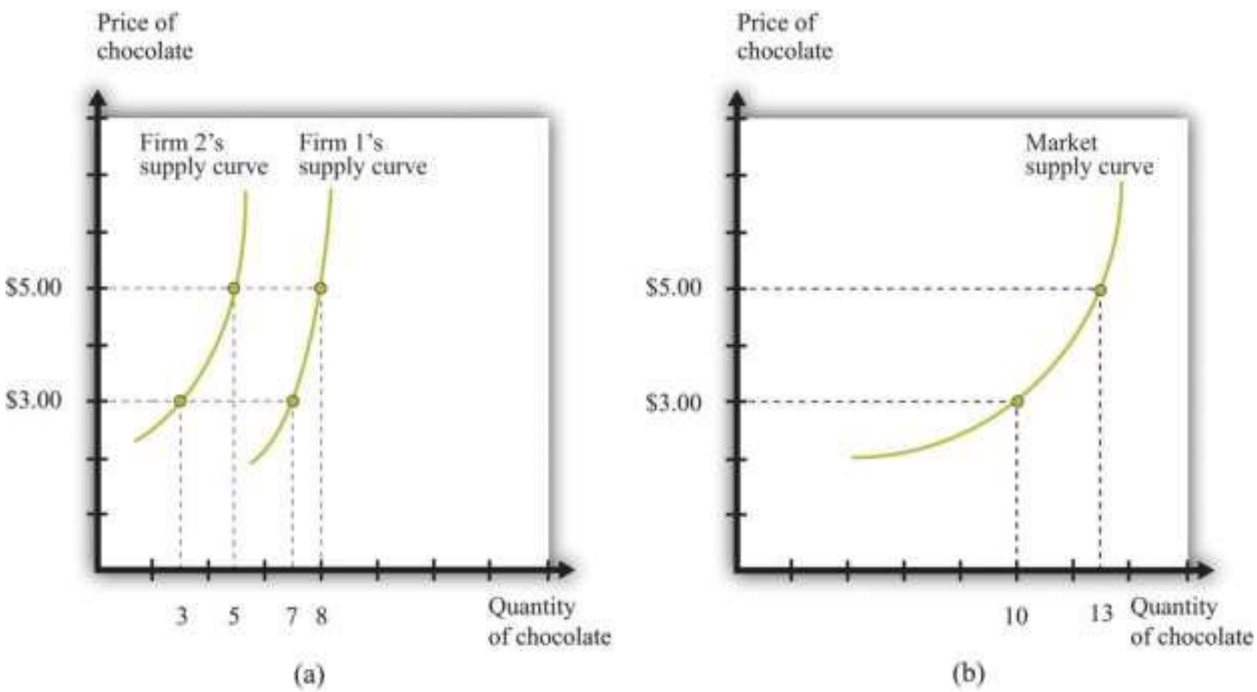


A firm's supply curve, which is the same as its marginal cost curve, shows the quantity of chocolate bars it is willing to supply at each price.

Just as the market demand curve tells us the total amount demanded at each price, the **market supply curve** tells us the total amount supplied at each price. It is obtained analogously to the market demand curve: at each price we add together the quantity supplied by each firm to obtain the total quantity supplied at that price. If we perform this calculation for every price, then we get the market supply curve. Figure 7.4 "Market Supply" shows an example with two firms. At $3, firm 1 produces 7 bars, and firm 2 produces 3 bars. Thus the total supply at this price is 10 chocolate bars. At $5, firm 1 produces 8 bars, and firm 2 produces 5 bars. Thus the total supply at this price is 13 chocolate bars.

The market supply curve is increasing in price. As price increases, each firm in the market finds it profitable to increase output to ensure that price equals marginal cost. Moreover, as price increases, firms who choose not to produce and sell a product may be induced to enter into the market. [3]

Figure 7.4 Market Supply



Market supply is obtained by adding together the individual supplies of all the firms in the economy.

In general, both mechanisms come into play. The market supply curve slopes up for two reasons:

1. As the price increases, more firms decide to enter the market—that is, these firms produce some positive quantity other than zero.
2. As the price increases, firms increase the quantity that they wish to produce.

   When the price increases, there are more firms in the market, and each firm produces more.

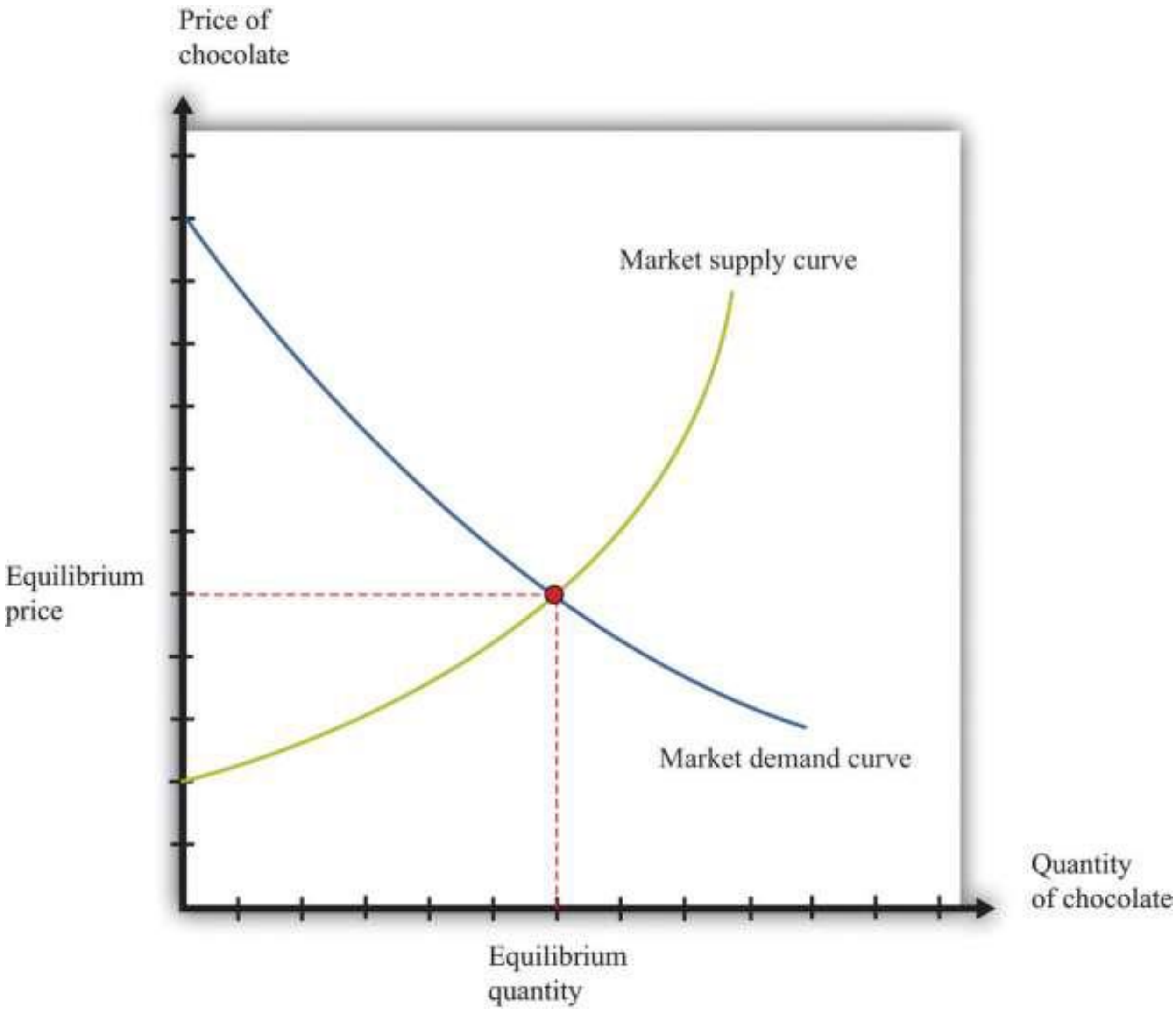
Market Equilibrium

In a perfectly competitive market, we combine the market demand and supply curves to obtain the **supply-and-demand** framework shown in Figure 7.5 "Market Equilibrium". The point where the curves cross is the market **equilibrium**. [4] At this point, there is a perfect match between the amount that buyers want to buy and the amount that sellers want to sell. The term *equilibrium* refers to the balancing of the forces of supply and demand in the market. At the equilibrium price, the suppliers of a good can sell as much as they wish, and demanders of a good can buy as much of the good as they wish. There are no disappointed buyers or sellers.

Toolkit: Section 17.9 "Supply and Demand"

You can review the definition and meaning of equilibrium in the supply-and-demand framework in the toolkit.

Figure 7.5 Market Equilibrium



In a competitive market, the equilibrium price and the equilibrium quantity are determined by the intersection of the supply and demand curves.

Because the demand curve has a negative slope and the supply curve has a positive slope, supply and demand will cross once. Both the equilibrium price and the equilibrium quantity will be positive. (More precisely, this is true as long as the vertical intercept of the demand curve is larger than the vertical intercept of the supply curve. If this is not the case, then the most that any buyer is willing to pay is less than the least any seller is willing to accept and there is no trade in the market.)

Table 7.2 Market Equilibrium: An Example

Price ($)	**Market Supply**	**Market Demand**

Price ($)	Market Supply	Market Demand
1	5	95
5	25	75
10	50	50
20	100	0

Table 7.2 "Market Equilibrium: An Example" shows an example of market equilibrium with market supply and market demand at four different prices. The equilibrium occurs at $10 and a quantity of 50 units. The table is based on the following equations:

market demand = 100 − 5 × price

and

market supply = 5 × price.

Equations such as these and diagrams such as Figure 7.5 "Market Equilibrium"are useful to economists who want to understand how the market works. Keep in mind, though, that firms and households in the market do not need any of this information. This is one of the beauties of the market. An individual firm or household needs to know only the price that is prevailing in the market.

Reaching the Market Equilibrium

Economists typically believe that a perfectly competitive market is likely to reach equilibrium for several reasons.

- If the prevailing price is different from the equilibrium price, then there will be an imbalance between demand and supply, which gives buyers and sellers an incentive to behave differently. For example, if the prevailing price is less than the equilibrium price, demand will exceed supply. Disappointed buyers might start bidding the price up, or sellers might realize they could charge a higher price. The opposite is true if the prevailing price is too high: suppliers might be tempted to try decreasing prices, and buyers might look for better deals. These are informal stories because the supply and demand curves are based on the idea that firms and consumers take prices as given. Still, the idea that there will be pressure on prices away from equilibrium is a plausible one.

- There is strong support for market predictions in the evidence from experimental markets. [5]
- The supply-and-demand framework generally provides reliable predictions about the movement of prices.

KEY TAKEAWAYS

- The market demand curve is obtained by adding together the demand curves of the individual households in an economy.
- As the price increases, household demand decreases, so market demand is downward sloping.
- The market supply curve is obtained by adding together the individual supply curves of all firms in an economy.
- As the price increases, the quantity supplied by every firm increases, so market supply is upward sloping.
- A perfectly competitive market is in equilibrium at the price where demand equals supply.

CHECKING YOUR UNDERSTANDING

1. In Table 7.2 "Market Equilibrium: An Example", market supply was equal to 5 × price. Suppose instead that market supply = 15 × price. Would the equilibrium price still be $10? If not, construct a new column in the table and find the new equilibrium price.
2. Explain why supply and demand cross only once. Do they always cross at a positive price?

[1] See Chapter 3 "Everyday Decisions" and Chapter 5 "eBay and craigslist" for discussions of unit demand.

[2] At the end of Chapter 6 "Where Do Prices Come From?", we derive the supply curve of a firm in a competitive market.

[3] A similar idea is in Chapter 5 "eBay and craigslist", where we show how to add together unit supply curves to obtain a market supply curve.

[4] The definition of equilibrium is also presented in Chapter 5 "eBay and craigslist".

[5] In Chapter 5 "eBay and craigslist", we explain that a double oral auction, in which buyers and sellers meet individually and bargain over prices, typically yields results very close to the market outcome in Figure 7.5 "Market Equilibrium".

7.2 Using the Supply-and-Demand Framework

LEARNING OBJECTIVES

1. Why do market prices increase and decrease?
2. How can I predict what is going to happen to prices?

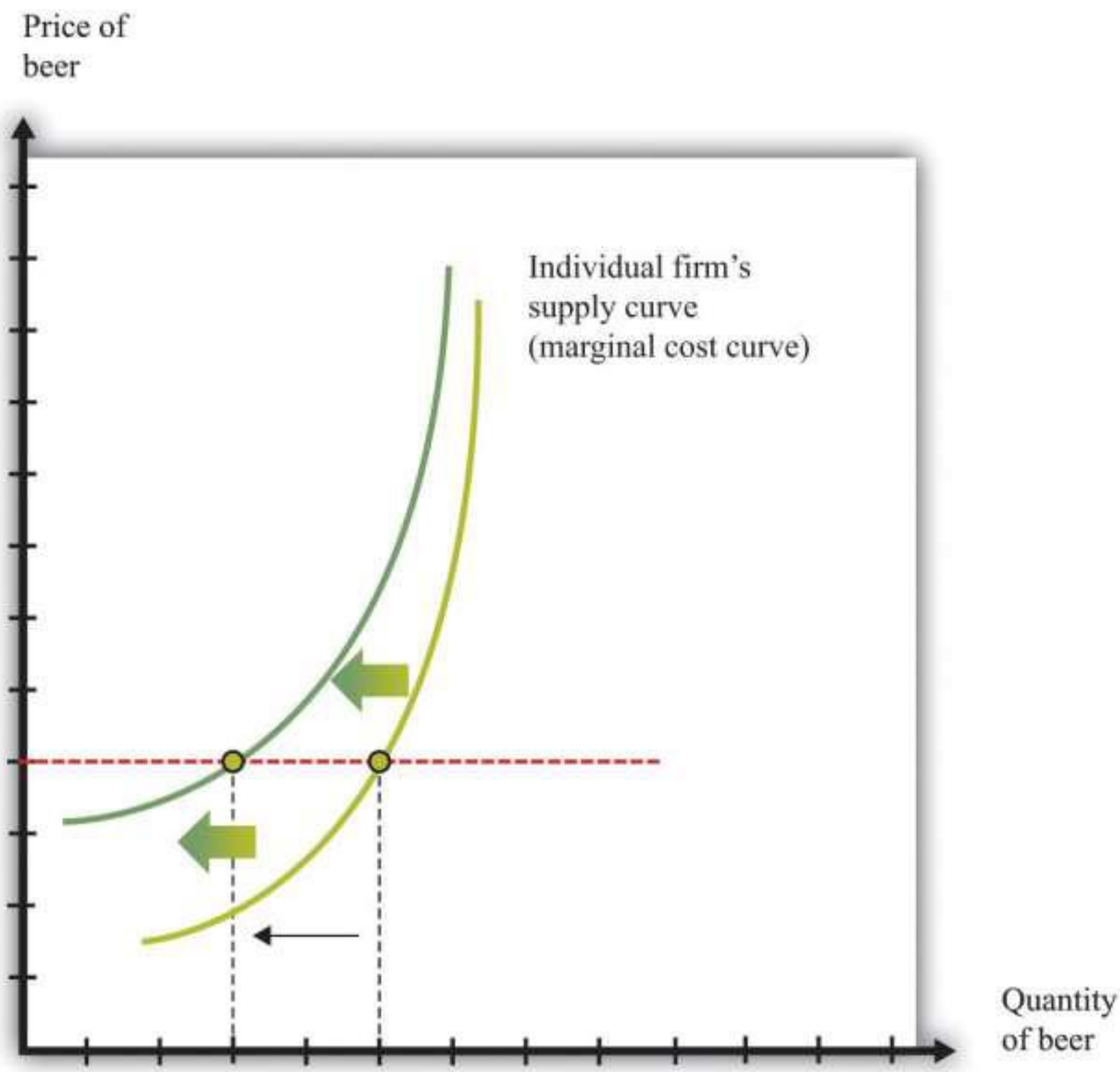
Economists are often asked to make predictions about the effects of events on economic outcomes. They do so by using the supply-and-demand framework. To use this framework, we must first distinguish between those things that we take as given (exogenous variables) and those that we seek to explain (endogenous variables).

Toolkit: Section 17.16 "Comparative Statics"

An exogenous variable is something that comes from outside a model and is not explained in our analysis. An endogenous variable is one that is explained within our analysis. When using the supply-and-demand framework, price and quantity are endogenous variables; everything else is exogenous.

A Shift in Supply: Beer Prices in Britain

Figure 7.6 A Shift in the Supply Curve of an Individual Firm



An increase in marginal cost leads a firm to produce less output at any given price. This means that a firm's supply curve shifts upward and to the left.

When we quoted the British newspaper article about beer prices in the chapter introduction, we omitted some sentences. The first sentence reads, in full: "The average price of a pint of beer could hit £4 [pounds sterling] after poor weather forced up the price of hops." A few sentences later the article states: "Hop farmers have not seen any price rises for years, but the appalling summer has finally forced the prices up." According to the article, the price of beer is increasing because the price of hops has increased.

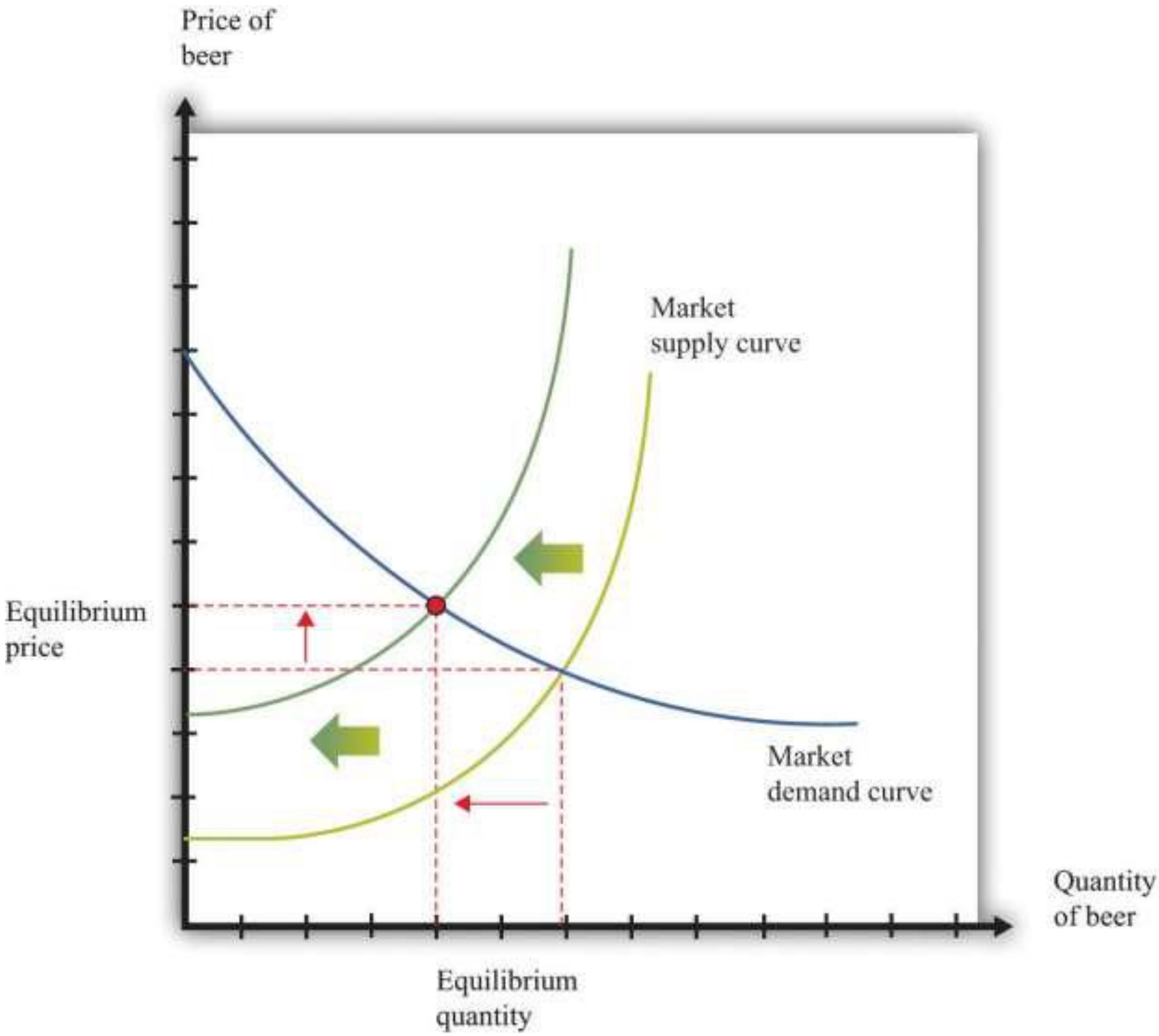
This makes intuitive sense, but it is worth understanding the exact chain of reasoning here. Hops are a key ingredient in the production of beer. An increase in the price of hops means an increase in the cost of producing beer. More precisely, the marginal cost of producing beer increases. The typical beer producer decides how much to produce by observing this decision rule:

price = marginal cost.

If marginal cost increases, then, at the existing price, the producer will find that price is now less than marginal cost. To bring price back in line with marginal cost, the producer will have to produce a smaller quantity. In fact, at any given price, an increase in marginal cost leads to a reduction in output (Figure 7.6 "A Shift in the Supply Curve of an Individual Firm"). The supply curve of an individual firm shifts to the left.

The increase in the price of hops affects all firms in the market. Each firm sees an increase in its marginal cost of production, so each firm produces less output at a given price: the shift in supply shown in Figure 7.6 "A Shift in the Supply Curve of an Individual Firm" applies to all firms in the market. Figure 7.7 "A Shift in Market Supply" shows the outcome in the market. Because all the individual supply curves shift to the left, the market supply curve likewise shifts to the left. At any given price, firms supply less beer to the market. From the figure, we see that the higher price of hops leads to an increase in the price of beer and a reduction in the quantity of beer produced and sold.

Figure 7.7 A Shift in Market Supply

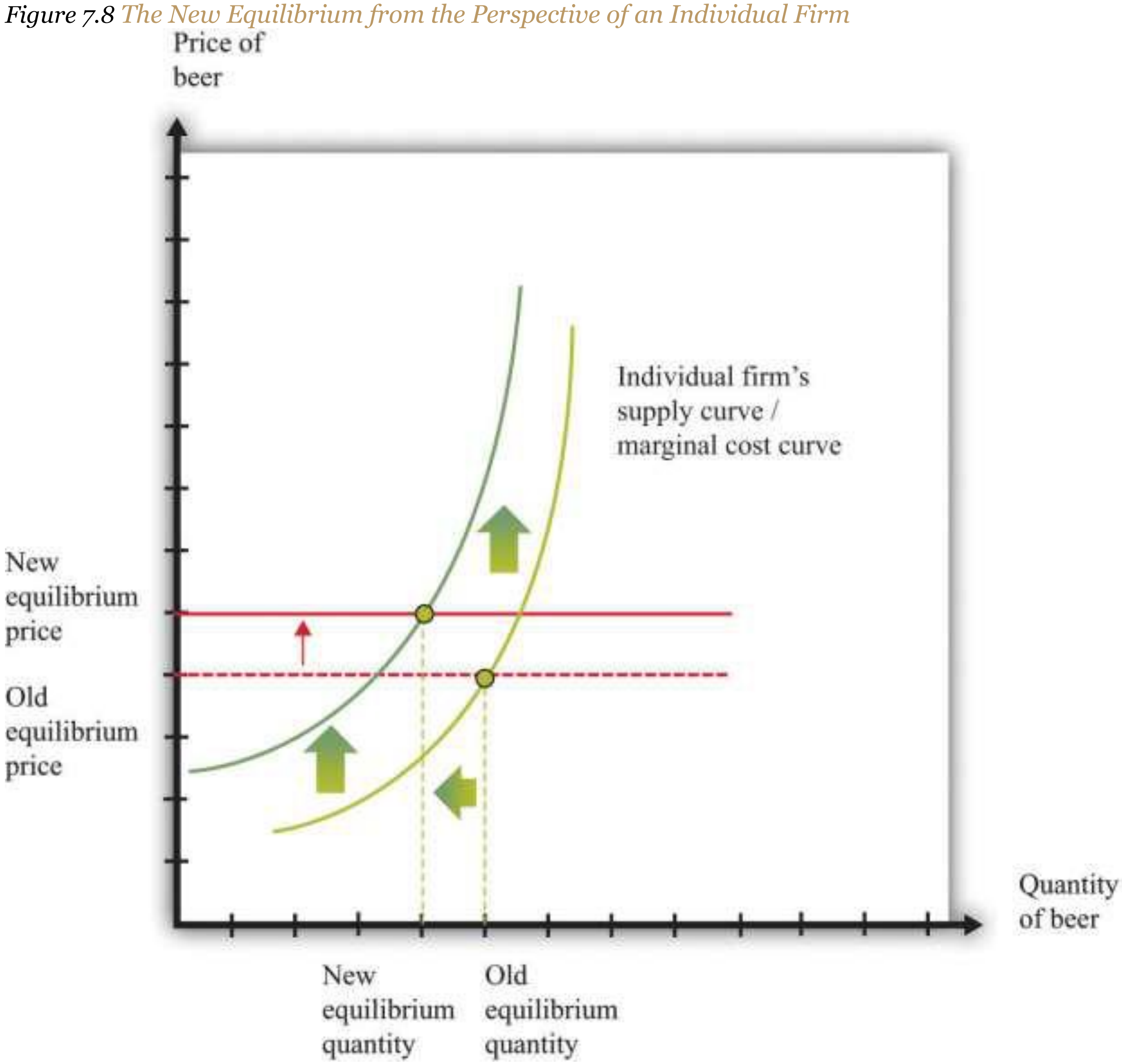


An increase in the price of hops causes all beer producers to produce less at any given price. This means that the market supply curve shifts to the left. The consequence is an increase in the equilibrium price and a decrease in the equilibrium quantity.

For the individual producer, what does this mean? The producer sees an increase in marginal cost. In the new equilibrium, the producer also obtains a higher price. However, the increase in price is not as big as the increase in marginal cost. Because the producer sets price equal to marginal cost, each individual brewer still produces less. We show this in Figure 7.8 "The New Equilibrium from the Perspective of an Individual Firm".

1. The price of hops, an input into beer production, has increased, which increases the marginal cost of producing beer.
2. At each given price, beer producers want to supply less beer: the firm supply curve shifts to the left.
3. Because all the individual supply curves shift to the left, the market supply curve also shifts to the left.
4. The beer market reaches a new equilibrium with a higher price and smaller quantity of beer produced and consumed.

Figure 7.8 The New Equilibrium from the Perspective of an Individual Firm



Following the increase in the price of hops, the equilibrium price of beer increases. An individual firm ends up with higher marginal cost but also receives a higher price for beer. Because the increase in price is smaller than the increase in marginal cost, beer production still decreases.

Comparative Statics

The approach that we used here is an illustration of a general technique used by economists to explain changes in prices and quantities and to make predictions about what will happen to market prices.

Toolkit: Section 17.16 "Comparative Statics"

Comparative statics is a technique that allows us to describe how market equilibrium prices and quantities depend on exogenous events. As such, much of economics consists of exercises in comparative statics. In a comparative statics exercise, you must do the following:

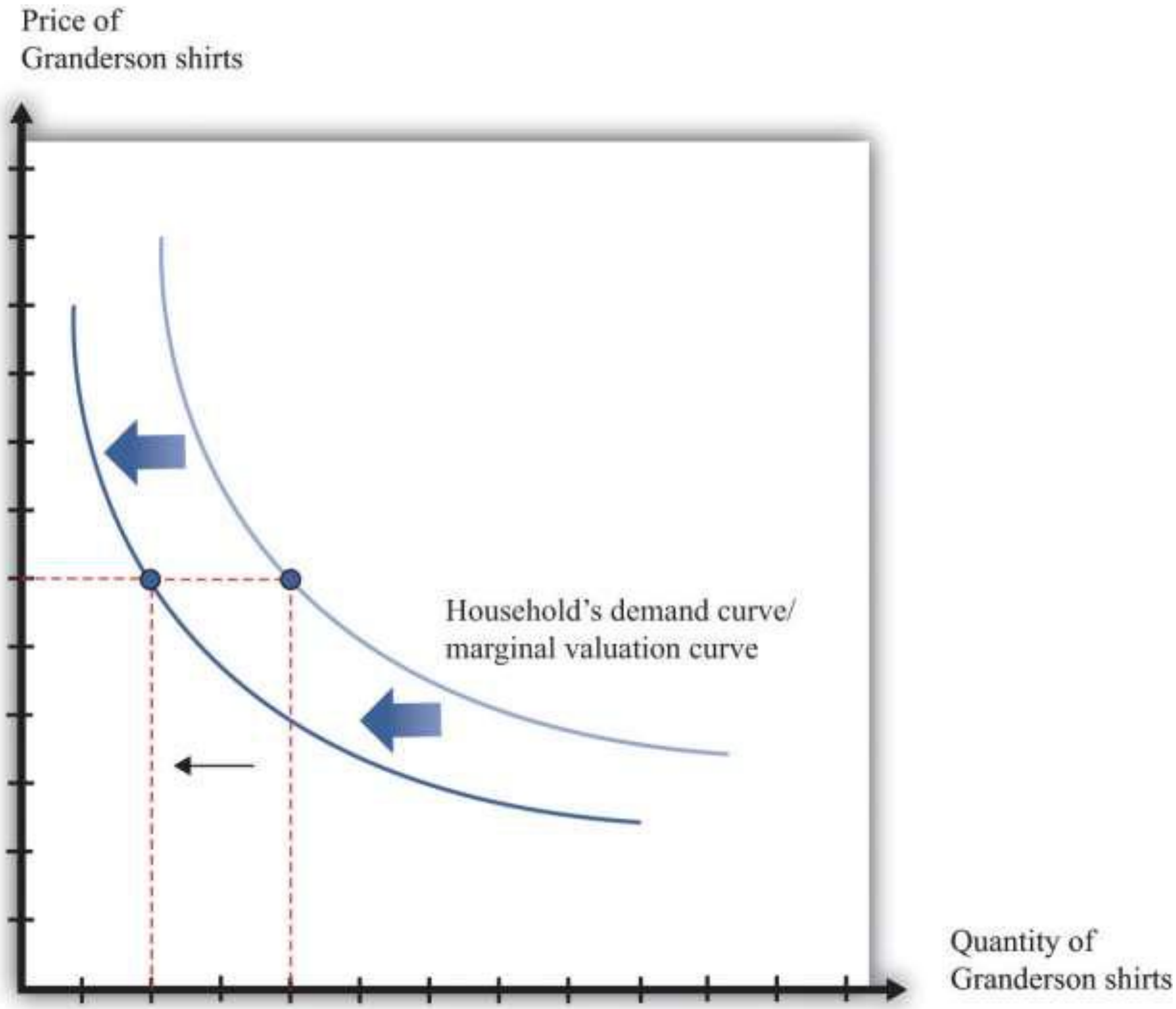
1. Begin at an equilibrium point where the quantity supplied equals the quantity demanded.
2. Based on the description of the event, determine whether the change in the exogenous factor shifts the market supply curve or the market demand curve.
3. Determine the direction of this shift.
4. After shifting the curve, find the new equilibrium point.
5. Compare the new and old equilibrium points to predict how the exogenous event affects the market.

The most difficult part of a comparative statics exercise is to determine, from the description of the economic problem, which curve to shift—supply or demand. Once you determine which curve is shifting, then it is only a matter of using the supply-and-demand framework to find the new equilibrium. The final step is to compare the new equilibrium point (the new crossing of supply and demand) with the original point. With this comparison, you can predict what will happen to equilibrium prices and quantities when something exogenous changes.

A Shift in Demand

Let us try this technique again. Recall the second story from the chapter introduction about Detroit Tiger merchandise. In that story, we learned that "most shops have marked down jerseys and T-shirts branded with ex-Tigers between 25 and 50 percent." Figure 7.9 "Shifts in Household Demand" shows the demand of a typical Detroit household's demand for Granderson shirts. Now that Granderson has left the Tigers for the New York Yankees, the household's marginal valuation for these shirts is lower. At any given price, a household wants to purchase fewer shirts, so the household's demand curve shifts to the left.

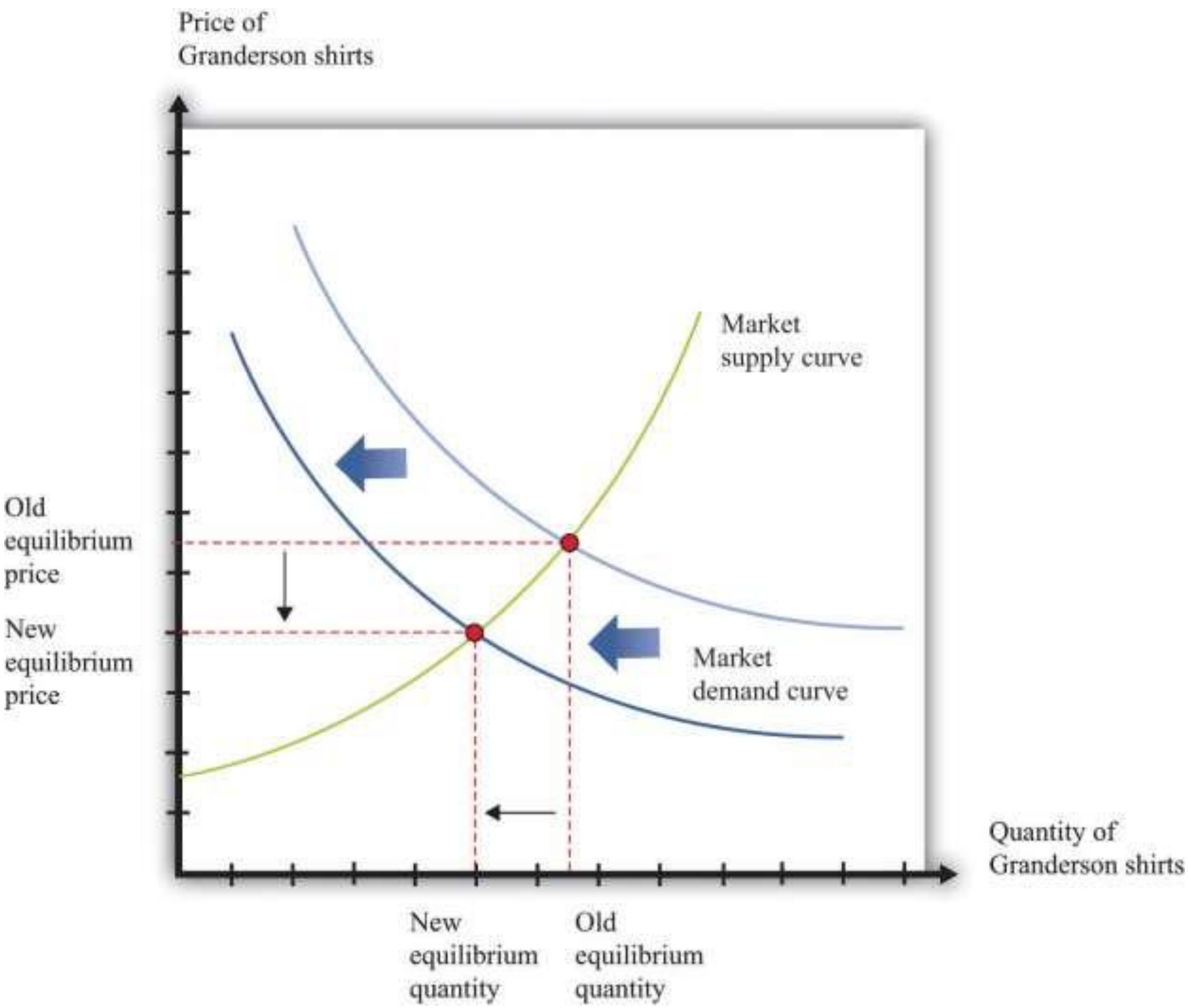
Figure 7.9 Shifts in Household Demand



A decrease in the marginal valuation of Granderson T-shirts leads a household to demand a smaller quantity at any given price. This means that a household's demand curve shifts to the left.

We would expect that this shift in demand would apply to most households that contain Detroit Tigers fans. If we now add all the demand curves together, we get the market demand curve. The market demand curve shifts to the left (Figure 7.10 "Shifts in Market Demand"). The end result is that we expect to see a decrease in the price of T-shirts—that is, the retailers put them in the discount bins—and also a decrease in demand.

Figure 7.10 Shifts in Market Demand



The decrease in demand causes both the equilibrium price and the equilibrium quantity of T-shirts to decrease.

Learning about the Slopes of the Supply and Demand Curves

Comparing our beer and T-shirt examples, we see that the quantity demanded decreased in both examples. In the first case, price increased; in the second case, price decreased. We can understand the difference by using the supply-and-demand framework. In the Detroit Tigers example, there is a decrease in the price of shirts and in the quantity sold. This might seem like a violation of the law of demand, which tells us that when price decreases, the quantity demanded increases. The explanation comes directly from Figure 7.10 "Shifts in Market Demand".

- Market demand is downward sloping and obeys the law of demand.
- Both equilibrium price and equilibrium quantity decrease after the departure of Granderson to the New York Yankees.

Curtis Granderson's move leads to a *shift in the demand curve* and a *movement along the supply curve*. The law of demand, by contrast, applies to the movement along a demand curve.

Shifts in a Curve versus Movements along a Curve

Understanding the distinction between moving along a curve (either supply or demand) and shifting a curve is the hardest part about learning to use the supply-and-demand framework. Journalists and others frequently are confused about this—and no wonder. It requires practice to learn how to use supply and demand properly.

Let's look at another example. An article in the British newspaper the *Guardian* reported about sales of beef when the news came out that eating beef might carry a risk of bovine spongiform encephalopathy (BSE), better known as mad cow disease. On November 1, 2000, the newspaper wrote, "Beef sales did drop after the link between BSE and deaths in humans was circumstantially established in 1996, but they have recovered as prices have fallen." [1]

The exogenous event here is the medical news about beef and mad cow disease. Presumably, this primarily affects the demand for beef: consumers decide to eat less beef and more of other products—such as chicken and pork. The demand curve for beef shifts to the left. As we saw in the T-shirt example, a leftward shift of the demand curve has two consequences: price decreases, and the quantity demanded and supplied also decreases. Thus the conclusion that the news should lead to a decrease in beef sales is perfectly consistent with our supply-and-demand analysis, as well as with common sense.

But what about the second part of the sentence? The article claims that beef sales "have recovered as prices have fallen." This is not consistent with our supply-and-demand analysis. The decrease in prices is intimately connected with the decrease in quantity: both were caused by the health news. They are two sides of the same coin, so it does not make sense to use the decrease in prices to explain a recovery in beef sales.

In fact, you should be able to convince yourself that an increase in beef sales together with a decrease in prices (as asserted by the article) would require a *rightward shift of the supply curve*. (Draw a diagram to make sure you understand this.) It seems unlikely that health concerns about beef led cattle farmers to *increase* their production of beef. It is hard to escape the conclusion that the journalist became confused about shifts in the demand curve and movements along the curve.

Estimating Demand and Supply Curves

Comparative statics allows us to make qualitative predictions about prices and quantities. Given an exogenous shock in a market, we can determine whether (1) the price is likely to increase or decrease and (2) the quantity bought and sold is likely to increase or decrease. Often, though, we would like to be able to do more. We would like to be able to make some predictions about the *magnitudes*of the changes.

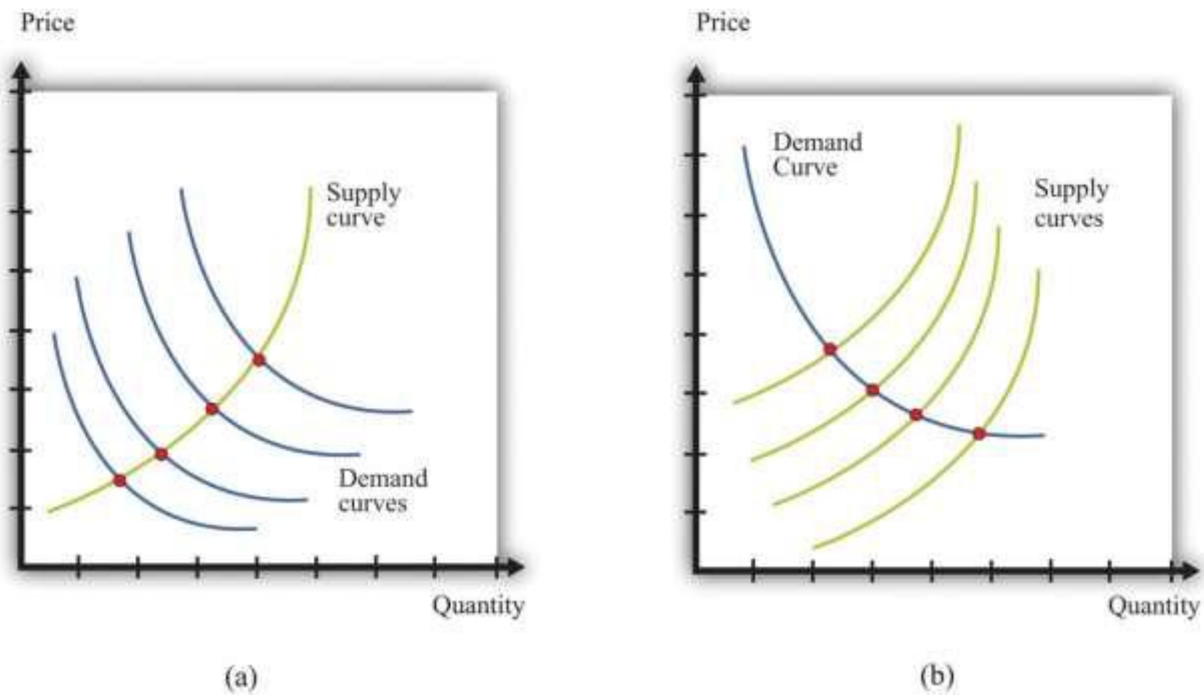
Figuring out what will happen to equilibrium prices and quantities requires economists to know the shapes of supply and demand. When the supply curve shifts, we need to know about the slope of the demand curve to predict the impact on price and quantity. When the demand curve shifts, we need to know about the slope of the supply curve to predict the impact on price and quantity. More precisely, we need measures of the **elasticity** of demand and of supply.

How do economists learn about these elasticities? The answer, perhaps surprisingly, is through the logic of comparative statics. For example, suppose the supply curve does not move, but the demand curve shifts around a lot. As the demand curve shifts, we observe different combinations of prices and quantities. Part (a) of Figure 7.11 "Finding the Elasticities of the Supply and Demand Curves" shows this in a supply-and-demand diagram. The different points that we observe are points on the supply curve. If the demand curve shifts but the supply curve does not, we eventually gather data on the supply curve. We can use these data to come up with estimates of the price elasticity of supply.

Toolkit: Section 17.2 "Elasticity"

The price elasticity of supply equals the percentage change in the quantity supplied divided by the percentage change in price.

Figure 7.11 Finding the Elasticities of the Supply and Demand Curves



Economists estimate the elasticities of supply and demand curves by looking for situations in which one curve is relatively stable while the other one is moving. What we actually observe are the equilibrium points. Movements in the demand curve (a) mean that the equilibrium points trace out the supply curve; movements in the supply curve (b) allow us to observe the demand curve. In most real-life cases, both curves move, and economists use sophisticated statistical techniques to tease apart shifts in supply from shifts in demand.

Part (b) of Figure 7.11 "Finding the Elasticities of the Supply and Demand Curves"shows the opposite case, where demand is stable and the supply curve is shifting. In this case, the data that we observe are different points on the demand curve. We can use this information to estimate theprice elasticity of demand, which is the percentage change in the quantity demanded divided by the percentage change in price. It is important to note that we are speaking here about the elasticity of the *market* demand curve, not the elasticity of the demand curve facing an individual firm.

This sounds straightforward in theory, but it is difficult in practice. Economic data are messy. Typically, both the demand curve and the supply curve are shifting simultaneously. If economists had access to controlled environments, perhaps like a biochemist does, we could "shift the demand curve" and see what happens in the laboratory. Occasionally, we get lucky. Sometimes we can isolate a particular event that we know is likely to shift only one of the curves. This is sometimes called a natural experiment. Most of the time, however, we are not so lucky. Economists and statisticians have come up with sophisticated statistical techniques to disentangle shifts in demand and supply in these circumstances. [2]

KEY TAKEAWAYS

- Changes in prices come from shifts in market supply, market demand, or both.
- Economists use comparative statics to predict changes in prices. This technique explains how changes in exogenous variables cause shifts in supply and/or demand curves, which lead to changes in prices.

CHECKING YOUR UNDERSTANDING

1. Suppose coffee crops in Brazil are destroyed by inclement weather. What happens to the supply curve for coffee? What happens to the price of coffee and the equilibrium quantity of coffee?
2. The discussion of the Detroit Tigers states that "it's too soon to tell which Tigers will prove popular at the checkout line." Suppose that Miguel Cabrera has an excellent season and breaks the home run record. What do you expect will happen to the price and quantity of T-shirts with his name on the back?
3. In our discussion of the demand for beef and mad cow disease, we said that an increase in quantity and a decrease in price require a rightward shift of the supply curve. Draw a diagram to illustrate this case.

[1] "First Beef—Now Lamb to the Slaughter?" Analysis, *Guardian*, November 1, 2000, accessed February 4, 2011, http://www.guardian.co.uk/uk/2000/nov/01/bse?INTCMP=SRCH.

[2] Chapter 6 "Where Do Prices Come From?" discusses how a firm can use a similar technique to learn about the demand curve that it faces. Chapter 10 "Raising the Wage Floor" discusses the difficulties of measuring the demand curve for labor.

7.3 Another Perspective on Changing Prices

LEARNING OBJECTIVES

1. Are price changes good for the economy?
2. How is information conveyed among households and firms in an economy?

Think back to our story of increasing beer prices. In Figure 7.6 "A Shift in the Supply Curve of an Individual Firm", we saw that an increase in the marginal cost of beer production led to an increase in the price and a decrease in the quantity supplied. In that explanation, we focused on what was happening to supply. But as the supply curve shifted, we moved along the demand curve to a new equilibrium. What was happening to the quantity demanded as the quantity supplied decreased? The answer is that as firms started decreasing their supply, the price in the market began to increase. Consumers of beer, confronted by these higher prices, bought less beer. Perhaps they switched to wine or spirits instead. The higher prices induced the quantity demanded to decrease in line with the decline in supply.

Something remarkable is happening in this story, however. Bad weather has affected the hops harvest, making beer more expensive to produce, relative to other goods and services. Because it is more expensive to make beer, it makes sense—from the point of view of society as a whole—to shift resources away from the production of beer and toward the production of other goods. And it makes sense—from the point of view of society as a whole—for people to consume less of the expensive-to-produce beer and more of other goods and services. If we imagine an all-knowing, all-powerful *central planner*, whose job is to allocate resources in the economy, we would expect this person to respond to the decrease in the hops harvest by ordering the production and consumption of less beer.

But this is exactly what happens in an economy, simply through the mechanism of supply and demand. The automatic adjustment of prices, resulting from shifts in supply and demand, brings about desirable shifts in production and consumption. Nobody orders producers to produce less or consumers to consume less. These outcomes result from the working of supply and demand.

Similarly, think about our T-shirt example. Consumers decide that they would like to consume fewer Granderson T-shirts. This change in their preferences shows up in the market as a shift in the demand curve, which causes the price of T-shirts to decrease. This decrease in the price encourages producers in the economy to adjust their behavior to fit the changed tastes of households. Firms stop producing Granderson shirts. Again, this is not because anyone has instructed them to do so. The changed tastes of households generate the price signal that induces firms to produce less.

So far, we have answered the question of the chapter by saying that prices change because of shifts in supply and/or demand. This answer is correct. But we could give a different answer from another perspective: prices change in order to provide signals to firms and households about what to produce and what to consume. In a market economy, households and firms decide what to consumer by considering the prices they face. Prices change in response to changes in costs and tastes, and these changes lead firms and households to adjust their decisions in line with the new economic reality.

It is fair to ask whether we should trust prices to play this role. Economics provides a very direct answer to this question: when markets are competitive, the price system delivers an efficient allocation of resources. In the following subsections, we develop the idea that markets deliver efficient outcomes by looking at a single market.

Buyer Surplus

Consider the market for chocolate bars, as shown in Figure 7.5 "Market Equilibrium". At the market clearing price, suppliers and demanders of chocolate bars trade the equilibrium quantity of chocolate bars. Imagine first that each household purchases no more than a single chocolate bar at the equilibrium price. For example, if 200 chocolate bars are sold, then 200 separate households bought a chocolate bar. Not all these households are alike, however: some like chocolate bars more than others. Most of them would have, in fact, been willing to pay more than the equilibrium price for the chocolate bar. Their **valuation** of a chocolate bar is greater than the price.

Any household that would have been willing to pay more than the equilibrium price gets a good deal. For example, suppose the equilibrium price is $5, but a household would have been willing to pay $7. Then that household receives a **buyer surplus** of $2. [1] This logic extends to the case where households consume more than one unit. The demand curve of a household indicates the maximum amount that a person would pay for each successive unit of a good. The demand curve shows the household's **marginal valuation** of a good. The individual household's demand curve slopes downward because the household is willing to pay less and less for each successive unit—the marginal unit—as the total quantity consumed increases.

In general, we know that a household purchases chocolate bars up to the point where

marginal valuation = price.

The household receives no surplus on the very last bar that it purchases because the marginal valuation of that bar equals price. But it receives surplus on all the other bars because its marginal valuation exceeds price for those bars. Diminishing marginal valuation means that the household obtains surplus from all the chocolate bars except the very last one.

Table 7.3 "Calculating Buyer Surplus for an Individual Household" gives an example of a household facing a price of $5. The first column is the quantity, the second is the price, the third is the marginal valuation (the extra value from the last chocolate), the fourth column measures the marginal surplus, and the last column is the total surplus.

Table 7.3 Calculating Buyer Surplus for an Individual Household

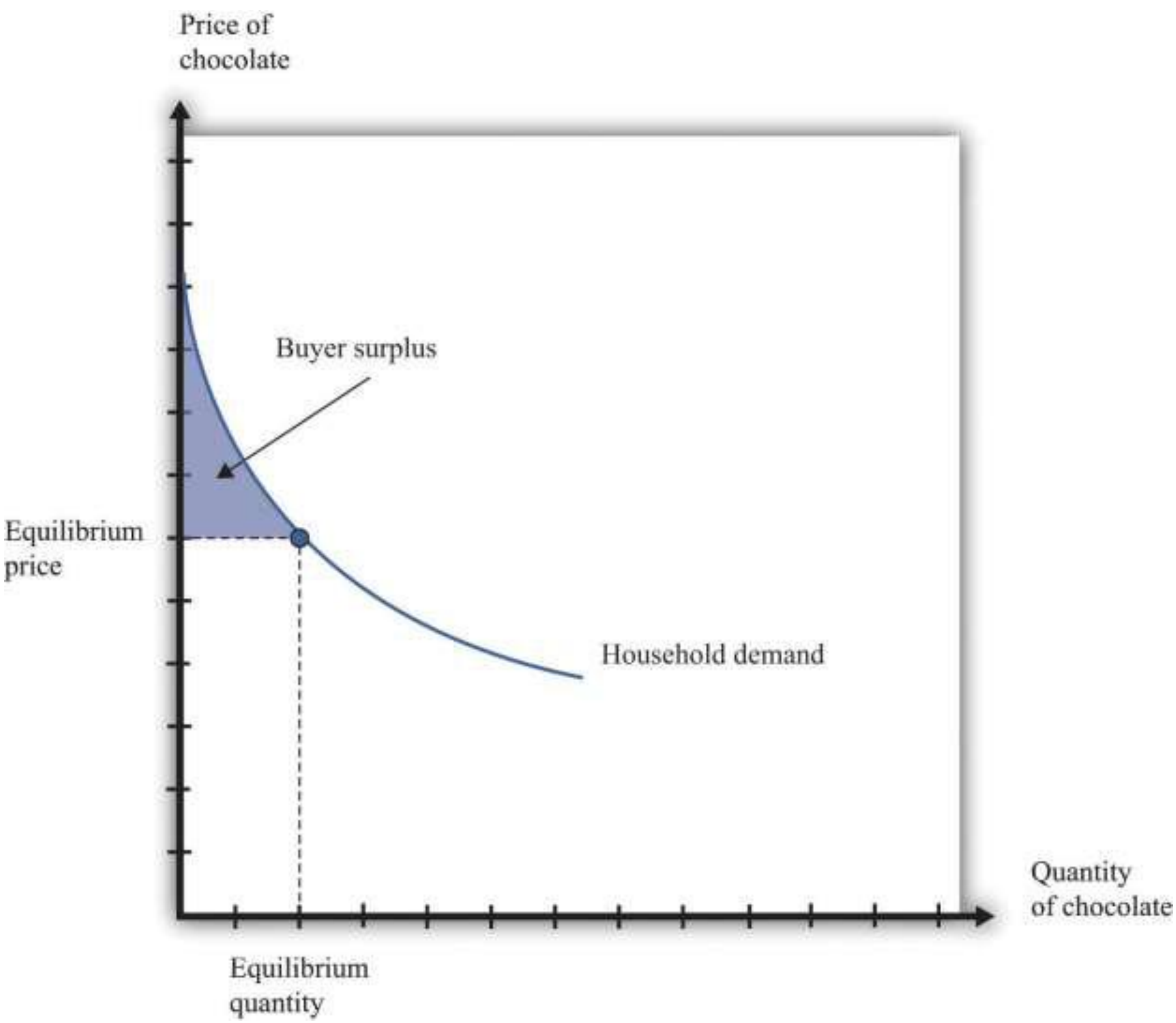
Quantity (Bars)	Price	Marginal Valuation	Surplus for Marginal Unit	Total Surplus
1	5	10	5	5
2	5	8	3	8
3	5	5	0	8
4	5	3	–2	6

The household is willing to buy three chocolate bars because the marginal value of the third bar is exactly equal to the price of $5. (In fact, the household would be equally happy buying either two or three bars. It makes no substantive difference to the discussion, but it is easier if we suppose that the household buys the last bar even though it is *indifferent* about making this purchase.) The household would not buy four bars because the marginal valuation of the last unit is less than the price, which means the surplus from a fourth chocolate bar would be negative.

The household obtains surplus from the first and second bars that it purchases. The household would have been willing to pay $10 for the first bar but only had to pay $5. It gets $5 of surplus from this first bar. The household would have been willing to pay $8 for the second bar but only had to pay $5. It gets $3 of surplus for this second bar. It gets no surplus from the third bar. So the total buyer surplus for this household is $5 + $3 = $8. Notice that by following the rule "buy until marginal valuation equals price," the household maximizes its total surplus from the purchase of chocolate bars.

More generally, the buyer surplus for this household is measured by the area under its demand curve (Figure 7.12 "Buyer Surplus for an Individual Household"). For each unit, the vertical difference between the price actually paid for each unit and the price the household would have been willing to pay measures the surplus earned for that unit. If we add the surplus over all units, we get the area between the demand curve and the price.

Figure 7.12 *Buyer Surplus for an Individual Household*



The buyer surplus is equal to the area between the demand curve and the price.

Seller Surplus

Sellers as well as buyers obtain surplus from trade. Suppose you won a used bicycle that you value at $20. If you can sell that bicycle for $30, you receive a**seller surplus** of $10—the difference between the price and your valuation of the good. It is worth your while to sell as long as the price is greater than your valuation. When a firm is producing a good for sale, the situation is analogous. If a firm can produce one more unit of a good at a marginal cost of $20, then the firm's valuation of the good is effectively equal to $20. If the firm can sell that unit for $30, it will receive a surplus of $10. The seller surplus earned by a firm for an individual unit is the difference between price and the marginal cost of producing that unit.

Given the price prevailing in a market, an individual firm in a competitive market will supply output such that the marginal cost of producing the last unit equals the price. The firm follows the rule: increase production up to the point where

price = marginal cost.

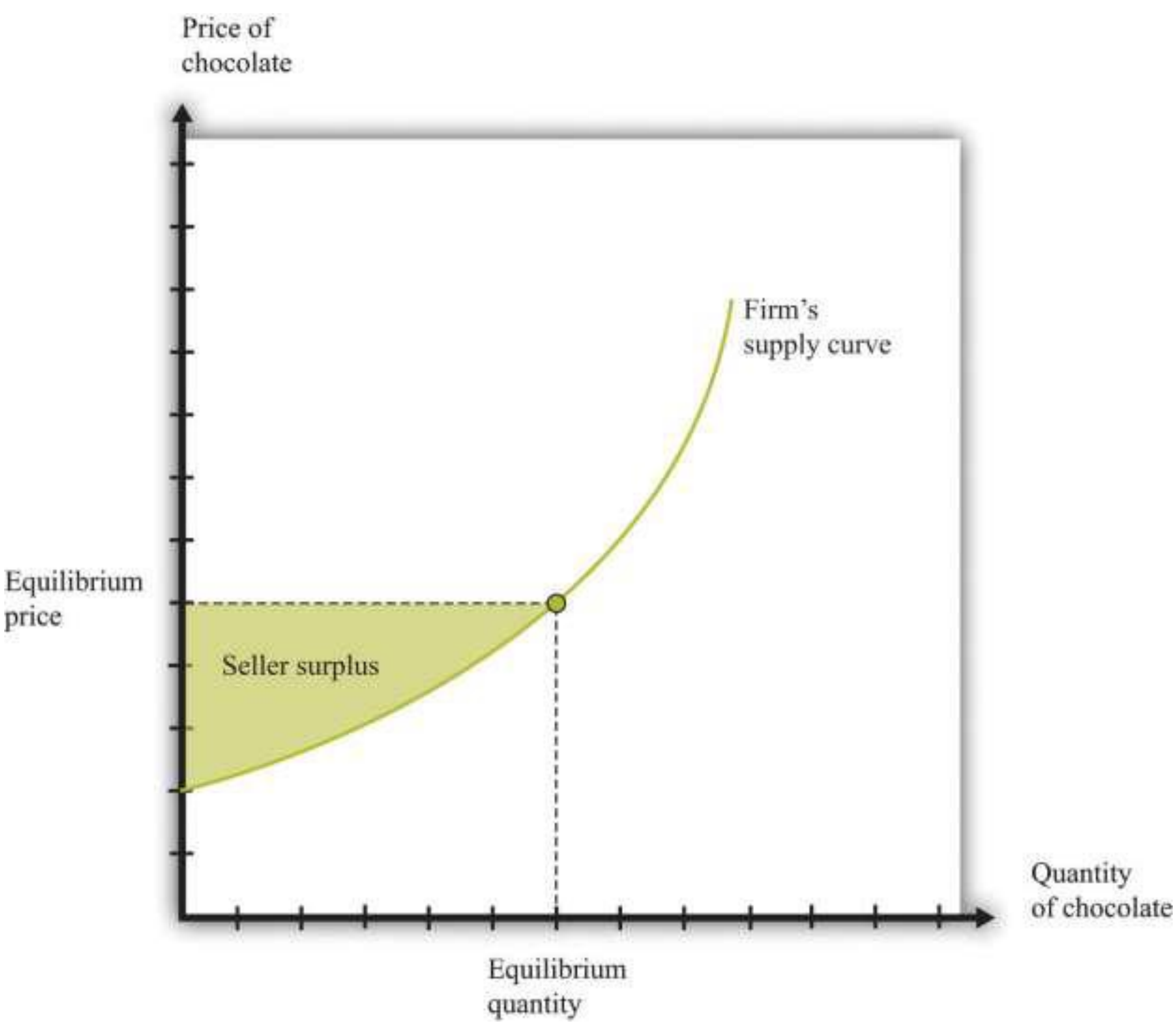
The example in Table 7.4 "Calculating Seller Surplus for an Individual Firm"gives the marginal cost of production for each unit and the surplus earned by a firm from producing that unit. If the firm produced only one unit, it would incur a marginal cost of $1, sell the unit for $5, and obtain a surplus of $4. The second unit costs $3 to produce, providing the firm with a surplus of $2. The third unit provides surplus of $1. The fourth unit costs $5 to produce, so the firm earns no surplus on this final unit. So the firm produces four units and obtains a total seller surplus of $7.

Table 7.4 Calculating Seller Surplus for an Individual Firm

Quantity	Price	Marginal Cost	Marginal Surplus	Total Surplus
1	5	1	4	4
2	5	3	2	6
3	5	4	1	7
4	5	5	0	7
5	5	6	–1	6

This difference between the price of a good and the marginal cost of producing the good is the basis of the seller surplus obtained by a firm. Exactly analogously to a household's buyer surplus, we measure the seller surplus by looking at the benefit a firm gets from selling each unit, and then we add them together. For each unit, the seller surplus is the difference between the price and the supply curve (remember that the supply curve and the marginal cost curve are the same thing). When we add the surplus for all units, we obtain the area above the supply curve and below the price (Figure 7.13 "Seller Surplus for an Individual Firm").

Figure 7.13 Seller Surplus for an Individual Firm



The seller surplus is the area between the equilibrium price and the firm's supply curve.

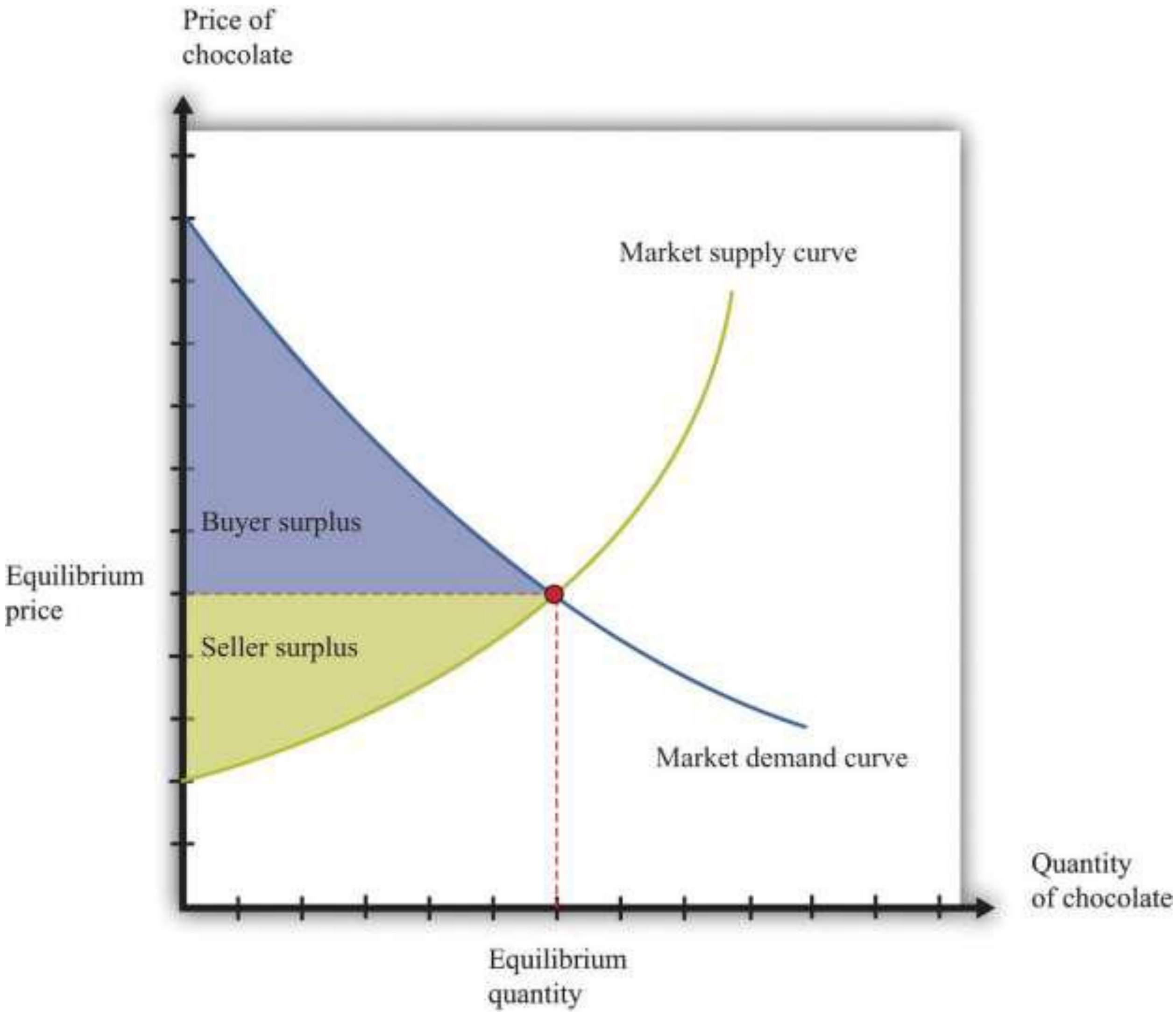
Toolkit: Section 17.1 "Individual Demand" and Section 17.10 "Buyer Surplus and Seller Surplus"

You can review the concepts of valuation, marginal valuation, buyer surplus, and seller surplus in the toolkit.

Buyer Surplus and Seller Surplus for the Entire Market

So far we have considered the buyer surplus and seller surplus for an individual household and an individual firm. Because the market demand and supply curves are obtained by adding together the individual demand and supply curves, the same result holds if we look at the entire market. We illustrate this inFigure 7.14 "Surplus in the Market Equilibrium", which shows the total surplus flowing to all households and firms in the market equilibrium. The area below the market demand curve and above the price level is the total buyer surplus. The area above the market supply curve and below the price is the total seller's (producer's) surplus. [2]

Figure 7.14 *Surplus in the Market Equilibrium*



The total surplus generated in a market is the sum of the buyer surplus and the seller surplus. It is therefore equal to the area below the demand curve and above the supply curve.

Markets and the Gains from Trade

The buyer surplus and the seller surplus tell us something remarkable about market outcome. If we add together the surplus for all buyers and sellers, we obtain the total surplus (gains from trade) in the market. In a competitive market, this is the maximum amount of surplus that it is *possible* to obtain—that is, exchange in a competitive market exhausts all the gains from trade.

There are two ways of seeing why this is true. First, we can ask what level of output would give us the largest total surplus. You might be able to see by looking at Figure 7.14 "Surplus in the Market

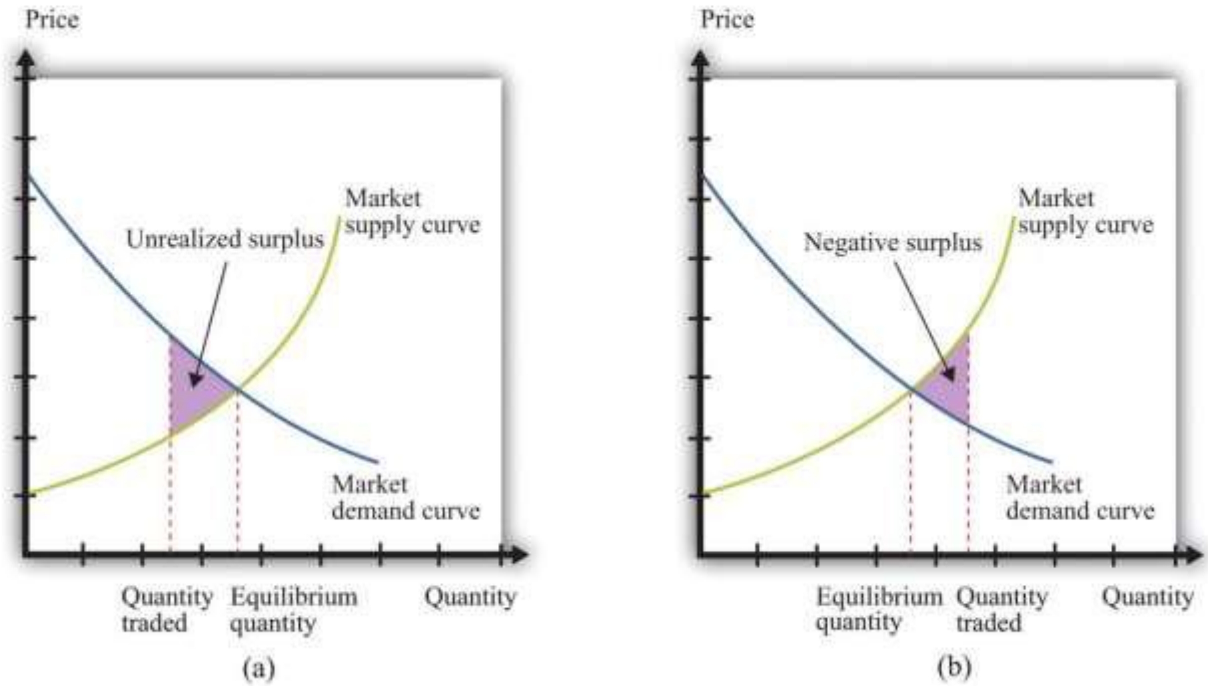
Equilibrium", where the equilibrium quantity yields the largest total surplus. Figure 7.15 "Surplus Away from the Market Equilibrium" explains why in more detail. If there are fewer trades, then some surplus goes unrealized: some transactions that would yield positive surplus do not take place. To put it another way, there are buyers whose marginal valuation exceeds the marginal cost of production but who are unable to purchase the good. By contrast, if there are more trades than the equilibrium quantity, then some trades generate a negative surplus. The marginal cost of producing output beyond the competitive level is *less* than the goods are worth to consumers.

Second, the following things are true at the market equilibrium:

- For each household, the marginal valuation for the last unit equals the price.
- For each firm, price equals the marginal cost for the last unit.

Combining these two pieces of information, we know that each household's marginal valuation of the last unit is the marginal cost of producing that unit. As quantity increases, marginal valuation decreases and marginal cost increases. Therefore, if more of the good were produced, the marginal cost of the extra units would be higher than the marginal valuation. By the same argument, if fewer units were produced, the reduction in the household's valuation would be higher than the reduction in cost. So producing one unit more or one unit less would not be beneficial to households and producers. Remember that it is the adjustment of prices that ensures that an economy trades at the point where supply and demand are equal. Price adjustment allows buyers and sellers to obtain all the gains from trade.

Figure 7.15 Surplus Away from the Market Equilibrium



If the quantity traded is less than the equilibrium quantity (a), then some gains from trade go unrealized. If the quantity traded is greater than the equilibrium quantity (b), then some trades generate negative surplus.

Why Markets?

We have been highlighting one of the principal messages in economics: markets are a mechanism to achieve the gains from trade. But are there other ways of achieving the same result? We previously introduced the fiction of an all-knowing, all-powerful central planner. Such a planner tells everyone what they should produce, takes all those goods, and distributes them throughout the economy. The planner tells everybody how much to work at each technology and decides exactly how to distribute all the goods and services that the economy produces.

Should we take this idea seriously or is it only a device to help us think about our theory? The answer is a bit of both. No economy has ever literally been run by a central planner. Historically, though, there have been many examples of so-called planned economies, where government bureaucracies played a major role in deciding what goods and services should be produced. For much of the 20th century, the economy of the Soviet Union operated under such a regime. China also used to be a largely planned economy. North Korea still operates as a largely planned economy.

Neither the Soviet Union nor China enjoyed much economic success under this system. The collapse of the Soviet Union's economy was a key reason why the country itself collapsed. China eventually changed its system of economic organization to one that gives more primacy to markets. Today, there are very few economies that operate under central planning and none that are significant in global economic terms. However, there are still several economies in which the government plays a significant role in the allocation of resources; so the analysis of the planner remains relevant.

Why were planned economies so unsuccessful? Books have been written on this topic, but there is one key insight. In order to make good decisions—decisions in the interest of individuals in the economy—the planner would need a lot of information. It is simply inconceivable that a planner could have sufficient knowledge about the abilities and skills of different individuals to make good decisions about where and how much they should work. Moreover, the planner also needs to know the tastes of everyone in the economy. Without that knowledge, the planner might instruct them to produce too many chocolate bars or not enough beer. If we think of an economy with millions of inhabitants, all with their own preferences and abilities, it is surely impossible that a planner could be sufficiently well informed to make decisions that are in the interest of all an economy's inhabitants.

KEY TAKEAWAYS

- The response of prices and quantities to exogenous events is key for the efficient allocation of resources in the economy.
- Information about the tastes of households and the costs of production for a firm is conveyed through the price system.
- Through price adjustments in competitive markets, all potential gains from trade are realized.

CHECKING YOUR UNDERSTANDING

1. If the demand for a good increases but the price is fixed, what gains from trade are lost?
2. Where in this section did we use the assumption of competitive markets?

[1] See Chapter 5 "eBay and craigslist" for more discussion.

[2] There is one slightly technical footnote we should add. In some circumstances, the seller surplus may not all go to the firm. Instead, it may be shared between the firm and its workers (or other suppliers of inputs to the firm). Specifically, this occurs when increases in the market supply are large enough to cause input prices to change.

7.4 Three Important Markets

LEARNING OBJECTIVES

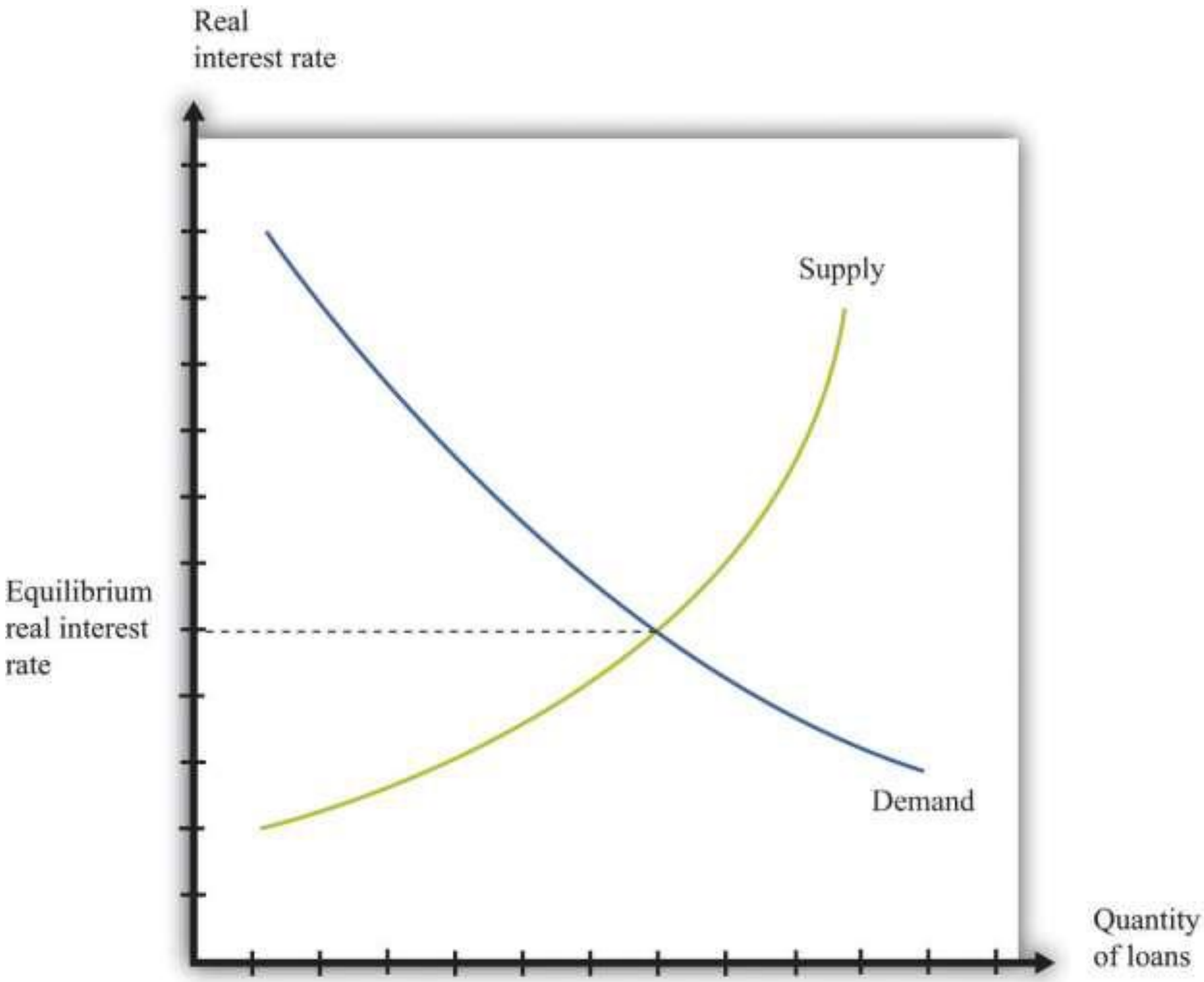
1. How do the tools of comparative statics extend to other markets, such as the market for credit or labor?
2. How do markets interact with one another?

In other chapters in this book, we use the supply-and-demand framework to look at how goods and services are traded. [1] Here we give a brief overview of these markets.

Credit Market

The credit market (or loan market)—we use the terms *loans* and *credit* interchangeably—is where suppliers and demanders of credit meet and trade (Figure 7.16 "Credit Market Equilibrium"). On the supply side are households and firms that, for various reasons, have chosen to save some of their current income. On the demand side are other households, firms, and (in some cases) the government. Households buy houses and cars, so they often need to borrow funds to finance those purchases. Firms seek credit to finance investment, such as the construction of a new production plant. Finally, governments borrow to finance some of their expenditures.

Figure 7.16 Credit Market Equilibrium



This diagram represents the loan or credit market.

The price of credit is the **real interest rate**, which is a measure of the value of the interest charged on a loan, adjusted for inflation. There are many different markets for credit because there are different kinds of loans in the economy. [2]Associated with these different credit markets are different interest rates. For simplicity, though, we often suppose that there is a single market for credit.

Toolkit: Section 17.6 "The Credit Market"

You can review the credit market and the real interest rate in the toolkit.

The demand for credit decreases as the real interest rate increases. When it becomes more expensive to borrow, households, firms, and even governments want fewer loans. The supply of credit by households increases with the real interest rate. When the return on savings increases, households and firms will typically save more and so supply more loans to the market. [3]

The news is filled with stories about interest rates increasing and decreasing. You can always use some version of Figure 7.16 "Credit Market Equilibrium" to understand why interest rates are changing. Ultimately, any change in the interest rate is due to a shift in either the supply of credit or the demand for credit. For example, if construction firms anticipate high future demand for housing, they will think that building new homes is a good use of investment funds. They will borrow to finance such construction. The increased demand for credit will shift the demand curve in Figure 7.16 "Credit Market Equilibrium"outward, and interest rates will increase. As another example, if individuals in other countries wish to increase their investment in US assets, this will shift the supply of credit outward, and interest rates will decrease.

Two of the most important players in the credit market are the government and the monetary authority. If the US federal government borrows more, this shifts the demand for credit outward and increases the interest rate. (The government is such a big player in this market that its actions affect the interest rate.) The monetary authority, meanwhile, buys and sells in credit markets to influence the real interest rate in the economy. [4]

Foreign Exchange Market

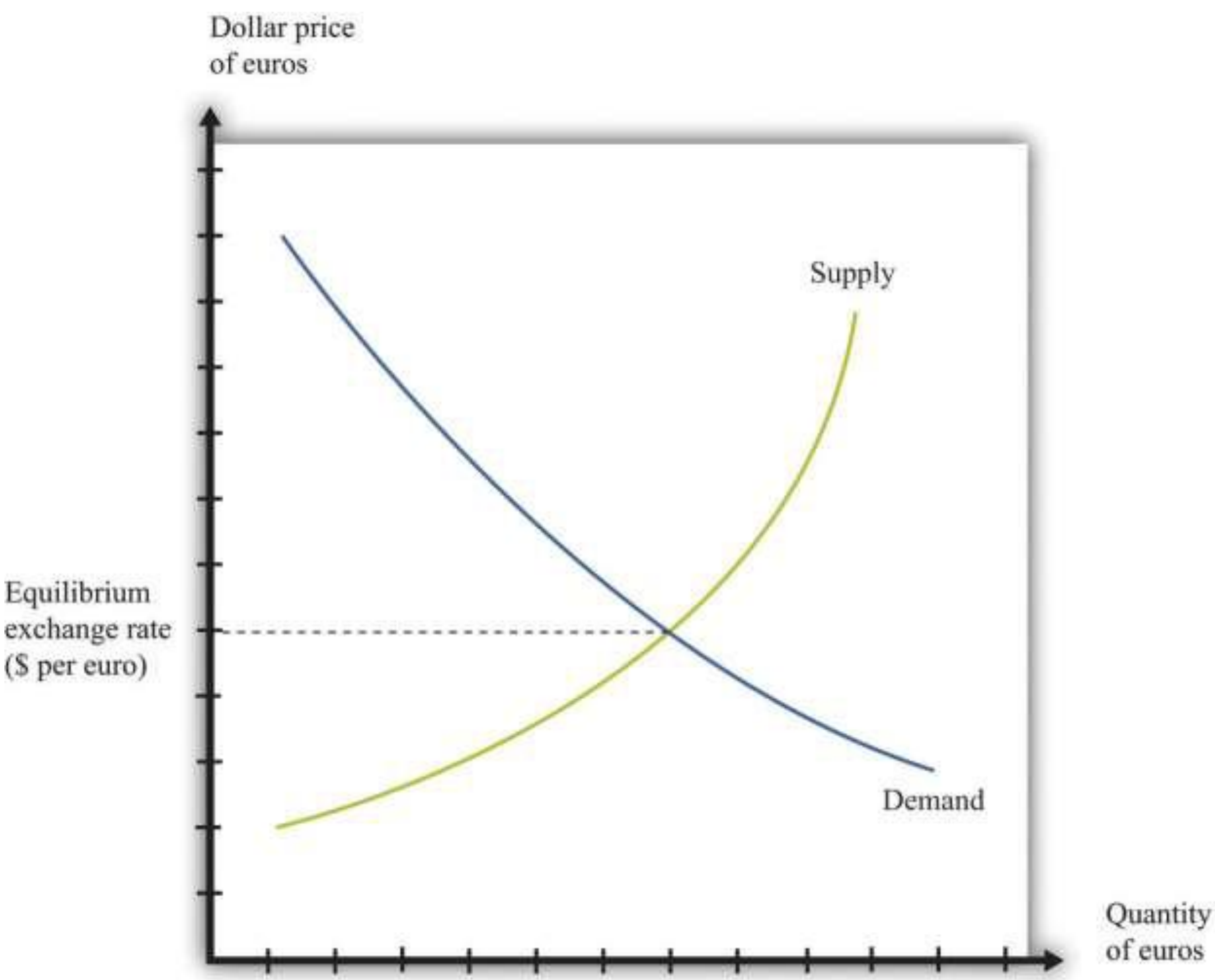
If you travel abroad, you need to acquire the currency used in that region of the world. If you take a trip to Finland, Russia, and China, for example, you will undoubtedly buy euros, rubles, and yuan along the way. To do so, you need to participate in the foreign exchange market, trading one currency for another. Foreign exchange markets operate like other markets in the economy. The price—which in this case is called the exchange rate—is determined by the interaction of supply and demand.

Toolkit: Section 17.20 "Foreign Exchange Market"

The foreign exchange market is the market where currencies are traded. The price in this market is the price of one currency in terms of another and is called the exchange rate.

Figure 7.17 "Foreign Exchange Market Equilibrium" shows an example of a foreign exchange market—the market in which US dollars are exchanged for euros. On the horizontal axis, we show the number of euros bought and sold on a particular day. On the vertical axis is the exchange rate: the price of euros in dollars. This market determines the dollar price of euros just like the gasoline market in the United States determines the dollar price of gasoline.

Figure 7.17 Foreign Exchange Market Equilibrium



This diagram represents the foreign exchange market in which euros are bought and sold with US dollars.

On the demand side, there are traders (households and firms) who want to buy European goods and services. To do so, they need to buy euros. This demand for euros expressed in dollars need not come from

US households and firms. Anyone holding dollars is free to purchase euros in this market. On the supply side, there are also traders (households and firms) who are holding euros and who wish to buy US goods and services. They need to sell euros to obtain US dollars.

There is another source of the demand for and the supply of different currencies. Households and, more importantly, firms often hold assets denominated in different currencies. You could, if you wish, hold some of your wealth in Israeli government bonds, in shares of a South African firm, or in Argentine real estate. But to do so, you would need to buy Israeli shekels, South African rand, or Argentine pesos. Likewise, many foreign investors hold US assets, such as shares in Dell Computer or debt issued by the US government. Thus the demand and the supply for currencies are influenced by the portfolio choices of households and firms. In practice, the vast majority of trades in foreign exchange markets are conducted by banks and other financial institutions that are adjusting their asset allocation.

In addition to households and firms, monetary authorities also participate in foreign exchange markets. For example, the US Federal Reserve Bank monitors the value of the dollar and may even intervene in the market, buying or selling dollars in an attempt to influence the exchange rate.

If you open a newspaper or browse the Internet, you can quickly find the current price of euros. This price changes all the time in response to changes in the currency's demand and supply. For example, if you read that the euro is getting stronger, this means that the euro is becoming more expensive: you must give up more dollars to buy a euro. This increase in the price of the euro could reflect either an outward shift in the demand for euros, say as US households demand more goods from Europe, or an inward shift of the euro supply curve if holders of euros are not as willing to sell them for dollars.

Labor Market

In the markets for goods and services, the supply side usually comes from firms. In some cases, buyers are other firms (businesspeople call these B2B transactions), whereas in other cases buyers are households (often called B2C transactions). For the most part, though, households are not on the supply side of these markets. In the **labor market**, by contrast, firms and households switch roles: firms demand labor and households supply it.

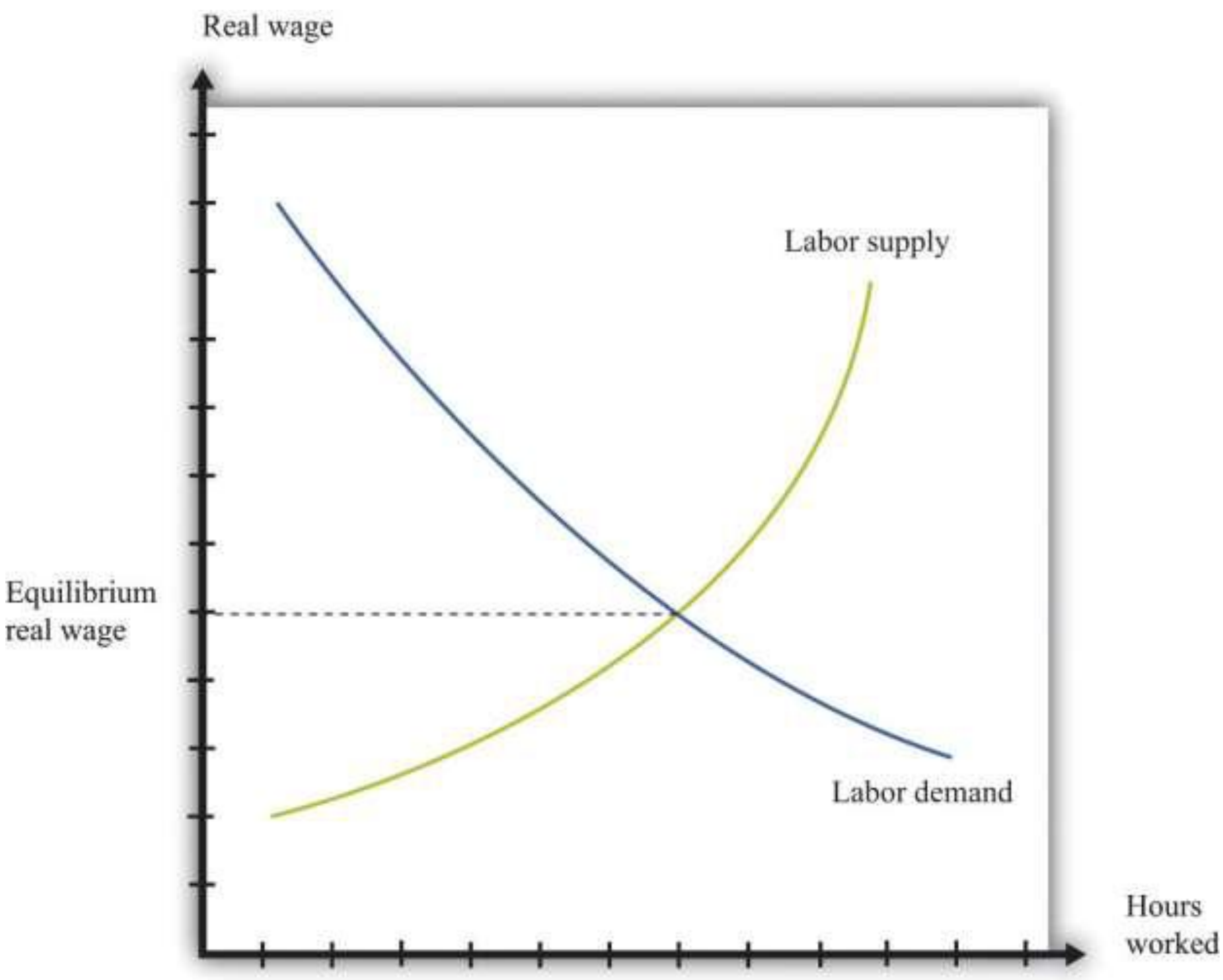
Supply and demand curves for labor are shown in Figure 7.18 "Labor Market Equilibrium". Here the price of labor is the **real wage**. The real wage measures how much in the way of goods and services an individual can buy in exchange for an hour's work. It equals the nominal wage (the wage in dollars) divided by the general price level.

Toolkit: Section 17.3 "The Labor Market"

You can review the labor market and the real wage in the toolkit.

The demand for labor comes from the fact that workers' time is an input into the production process. This demand curve obeys the law of demand: as the price of labor increases, the demand for labor decreases. The supply of labor comes from households that allocate their time between work and leisure activities. InFigure 7.18 "Labor Market Equilibrium", the supply of labor is upward sloping. As real wage increases, households supply more labor. There are two reasons for this: (1) higher wages induce people to work longer hours, and (2) higher wages induce more people to enter the labor force and look for a job. [5]

Figure 7.18 Labor Market Equilibrium



The equilibrium real wage is the price where supply equals demand in the labor market.

As with the other markets, we can use Figure 7.18 "Labor Market Equilibrium" to study comparative statics. For example, if an economy enters a boom, firms see more demand for their products, so they want to buy more labor to produce more product. This shifts the labor demand curve outward, with the result that real wages increase and employment is higher.

Multiple Markets

You have now seen equilibrium in a wide variety of markets: goods (chocolate), loans, foreign exchange, and labor. Actual economies contain hundreds of thousands of markets. Analyzing a single market would be enough if the markets in an economy were not connected, but markets are interrelated in many ways:[6]

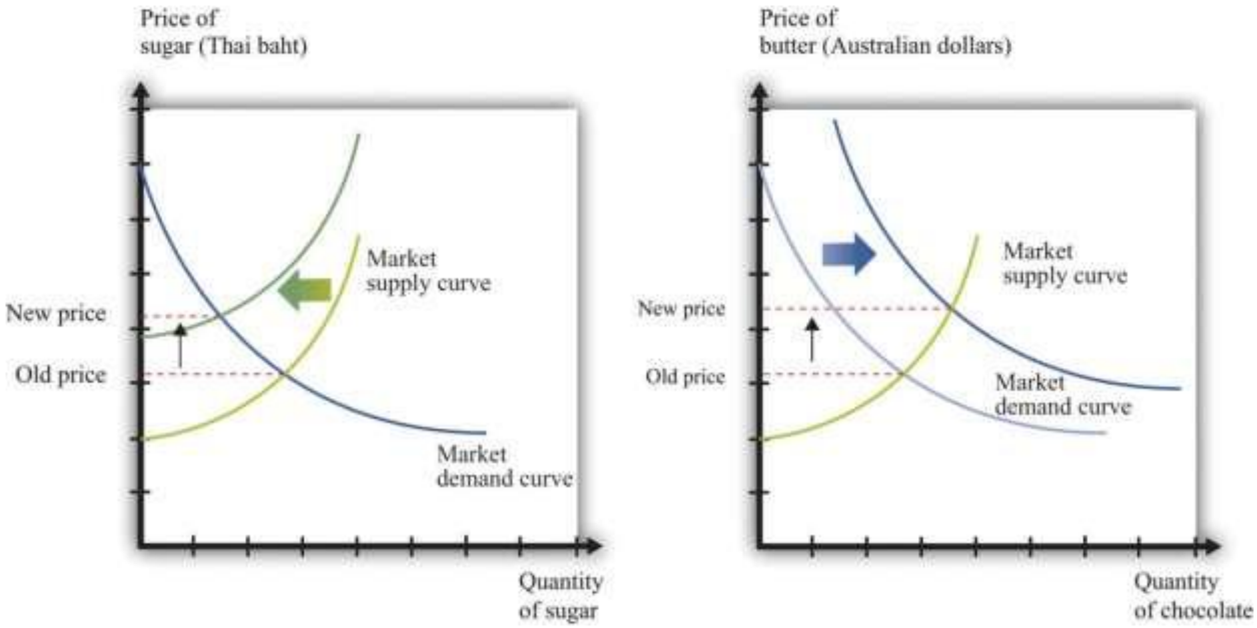
- Factors that shift the demand in one market may affect other markets as well. For example, an increase in energy costs will raise the marginal costs of firms in all sectors of an economy.
- The demand for one good depends on the prices of others. In the market for coffee, for example, the demand curve depends on the price of goods that are complements to coffee, such as milk, and the prices of goods that are substitutes for coffee, such as tea.
- The supply curve for most goods depends on the prices of inputs, such as labor. The real wage—the price of labor—is determined by the supply and demand for labor. Thus the outcome of the labor market influences the position of the supply curve in almost every other market.
- The income level of households affects the position of the demand curve for most goods and services. But the level of income comes, in part, from the labor market outcome because labor income is part of the income households have to spend.

The following newspaper story from the Singaporean newspaper the *Straits Times* nicely illustrates linkages across markets.

> *Singaporeans with a sweet tooth could soon find themselves paying more for their favourite treats, as bakers and confectioners buckle under soaring sugar prices.*
>
> *Since March last year, the price of white sugar has shot up by 70 per cent, according to the New York Board of Trade. As if that didn't make life difficult enough for bakers, butter and cheese prices have also risen, by 31 per cent and 17 per cent respectively.*
>
> *The increases have been caused by various factors: a steep drop in Thailand's sugarcane production due to drought, higher sea freight charges, increasing demand from China's consumers for dairy products and the strong Australian and New Zealand dollar.* [7]

Look at the last paragraph. First, we are told that a drought has caused a drop in sugarcane production in Thailand. Part (a) of Figure 7.19 "The Sugar Market in Thailand and the Butter Market in Australia" shows this market. We can see that a decrease in sugar production will increase the price of sugar. In this picture we are showing the market in Thailand, so the price is measured in Thai baht.

Figure 7.19 The Sugar Market in Thailand and the Butter Market in Australia



(a) The market for sugar in Thailand is affected by a drought, which has decreased the sugar supply, causing an increase in sugar prices measured in Thai baht. (b) In the Australian butter market, increased demand from China causes the demand curve to shift outward, increasing the price of butter measured in Australian dollars.

We are also told that there has been increased demand for dairy products coming from China. Australia and New Zealand are the major suppliers of dairy products in Southeast Asia. Part (b) of Figure 7.19 "The Sugar Market in Thailand and the Butter Market in Australia" shows the market for butter in Australia. Increased demand from China shifts the demand curve outward, leading to an increase in the price of butter. For this market, we measure the price in Australian dollars.

From the perspective of Singapore bakers, both of these price changes show up as increases in their marginal cost. Moreover, the article reveals that these changes are exacerbated by other factors. Shipping costs have increased, so it also costs more to obtain sugar from Thailand and butter from Australia. And the Australian dollar has appreciated relative to the Singapore dollar, making goods imported from Australia even more expensive.

These are examples of B2B transactions. In fact, there is likely a whole chain of such transactions between, say, the Australian dairy farmer and the Singaporean baker. Farmers sell milk to butter producers, butter producers sell to wholesalers, and wholesalers sell to Singaporean importers and bakeries.

This story also illustrates again the powerful way in which market prices provide information that helps us understand the efficient allocation of resources. Drought in Thailand has reduced the amount of sugar available in the world. Through the magic of a series of prices, one of the results is that people in Singapore are less likely to eat cake for dessert.

KEY TAKEAWAYS

- The supply-and-demand framework can be used to understand the markets for labor, credit, and foreign currency.
- Comparative statics can be used to study price and quantity changes in these markets.
- As markets interact with one another, sometimes comparative statics requires us to look at effects across markets.

CHECKING YOUR UNDERSTANDING

1. Figure 7.17 "Foreign Exchange Market Equilibrium" shows the market where euros are bought and sold using dollars. We could equivalently think of this as the market where dollars are bought and sold using euros. Draw the graph for this market. How are the supply and demand curves in the two markets related to each other?
2. Using supply and demand, explain how an increase in Chinese demand for Australian butter might be a factor that causes the Australian dollar to appreciate.
3. What might be the effect of the financial crisis in the United States in 2008–9 on the income of lawyers? How does your answer depend on the specialization of the lawyer?

[1] We focus on labor in Chapter 8 "Growing Jobs" and Chapter 10 "Raising the Wage Floor". Chapter 9 "Making and Losing Money on Wall Street" looks at both the loan/credit market and the foreign exchange market.

[2] Chapter 9 "Making and Losing Money on Wall Street" discusses these.

[3] The response of savings to changes in the real interest rate is discussed more fully inChapter 4 "Life Decisions".

[4] The actions of the Federal Reserve and other monetary authorities are studied in detail in macroeconomics courses.

[5] Chapter 10 "Raising the Wage Floor" explains more about nominal wages and real wages, and we study the individual demand for labor in Chapter 8 "Growing Jobs". The decisions underlying labor supply are explained more fully in Chapter 3 "Everyday Decisions".

[6] A topic in advanced studies of economics is the simultaneous equilibrium of all markets. Because all markets are linked, it is necessary to find prices for all goods and services and all inputs simultaneously such that supply equals demand in all markets. This is an abstract exercise and uses lots of mathematics. The bottom line is good news: we can usually expect an equilibrium for all markets.

[7] See http://straitstimes.asia1.com.sg. We also discuss this quote in Chapter 5 "eBay and craigslist".

7.5 Beyond Perfect Competition

LEARNING OBJECTIVE

1. How can I predict what will happen to prices when markets are not competitive?

Everything that we have discussed in this chapter applies, strictly speaking, only to perfectly competitive markets. Yet the conditions for perfect competition are quite stringent. For a market to be perfectly competitive, there must be a large number of sellers of an identical product. There also must be a large number of buyers. Each buyer and seller must be "small" relative to the market, meaning that they cannot influence market price.

There are certainly some markets that fit these criteria. Markets for commodities, such as wheat or gold, are one example. Markets for certain financial assets are another. Such examples notwithstanding, the vast majority of markets are *not* perfectly competitive. In most markets, firms possess some market power, meaning that the demand curve they face is not perfectly elastic.

You might think this greatly weakens the usefulness of the supply-and-demand framework. A firm with market power chooses a point on the demand curve that it faces. It sets a price as a markup over marginal cost and then produces enough to meet demand at that price. [1] A firm with market power does not take the price as given and then determine a quantity to supply. In fact—strictly speaking—there is no such thing as a supply curve when a firm has market power.

Economists understand this very well. Yet suppose you ask an economist to predict the likely effect of a worsening conflict in the Middle East on oil prices. The mental model she will use is almost certainly to imagine a supply curve for oil shifting to the left. Based on this model, she will predict higher prices and lower consumption. If you were to ask another economist to predict the effects of an economic recession on purchases of automobiles, he would imagine a demand curve shifting to the left and thus predict lower prices and lower output.

The first economist would use a supply-and-demand framework even though oil producers have market power. The second economist would use a supply-and-demand framework even though not all cars are identical. Although economists understand that many markets do not satisfy the strict conditions of perfect competition, they also know that the intuition from comparative statics carries over to more general market structures.

To see why, let us go back to our beer example again. We all know that not all beer is the same, and the beer companies spend a lot of money to convince us of this fact. Different beers have different tastes, and there are customers who are loyal to different beer brands. Breweries possess market power, meaning that we cannot—strictly speaking—draw a supply curve for individual beer producers or for the market as a whole.

Yet our comparative statics story, which supposes that the beer market is competitive, gives us an answer that makes sense. When the price of hops increases, this increases the marginal cost of production for all beer producers. Because they set prices based on a markup over marginal cost, the price of beer will increase, and less will be consumed. Output will be lower for all producers, and prices will be higher. Our comparative statics technique gives the right answer. Let us go through this more formally, first for a change in production costs and then for a change in demand.

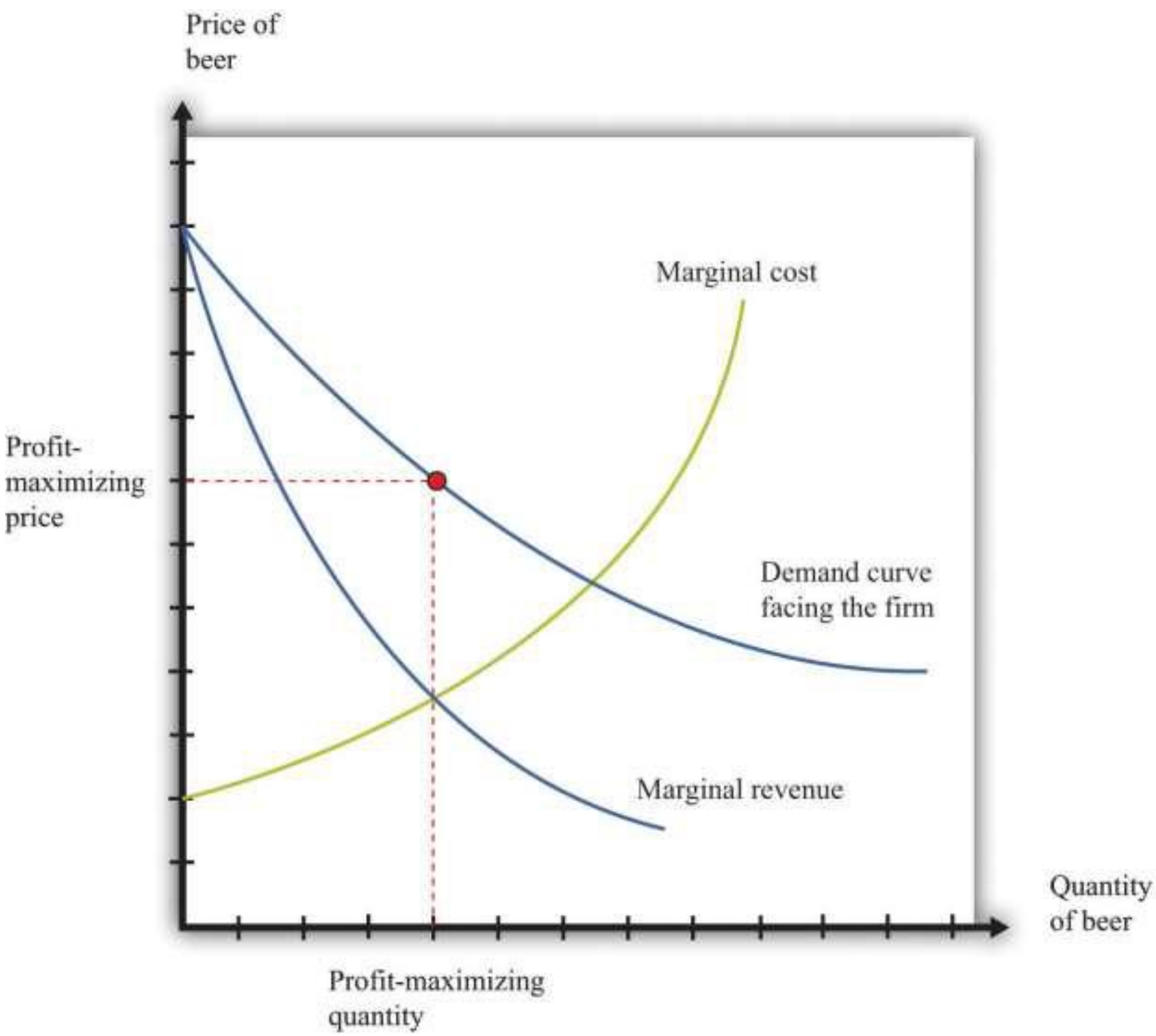
Shifts in the Marginal Cost of Production

Figure 7.20 "Finding the Profit-Maximizing Price and Quantity When a Firm Has Market Power" shows how a firm with market power sets its price. [2] To maximize its profits, a firm wants to produce the quantity where **marginal revenue** equals marginal cost. It sets the appropriate price as a **markup** over marginal cost.

Toolkit: Section 17.15 "Pricing with Market Power"

You can review the details of pricing with market power, including marginal revenue and markup, in the toolkit.

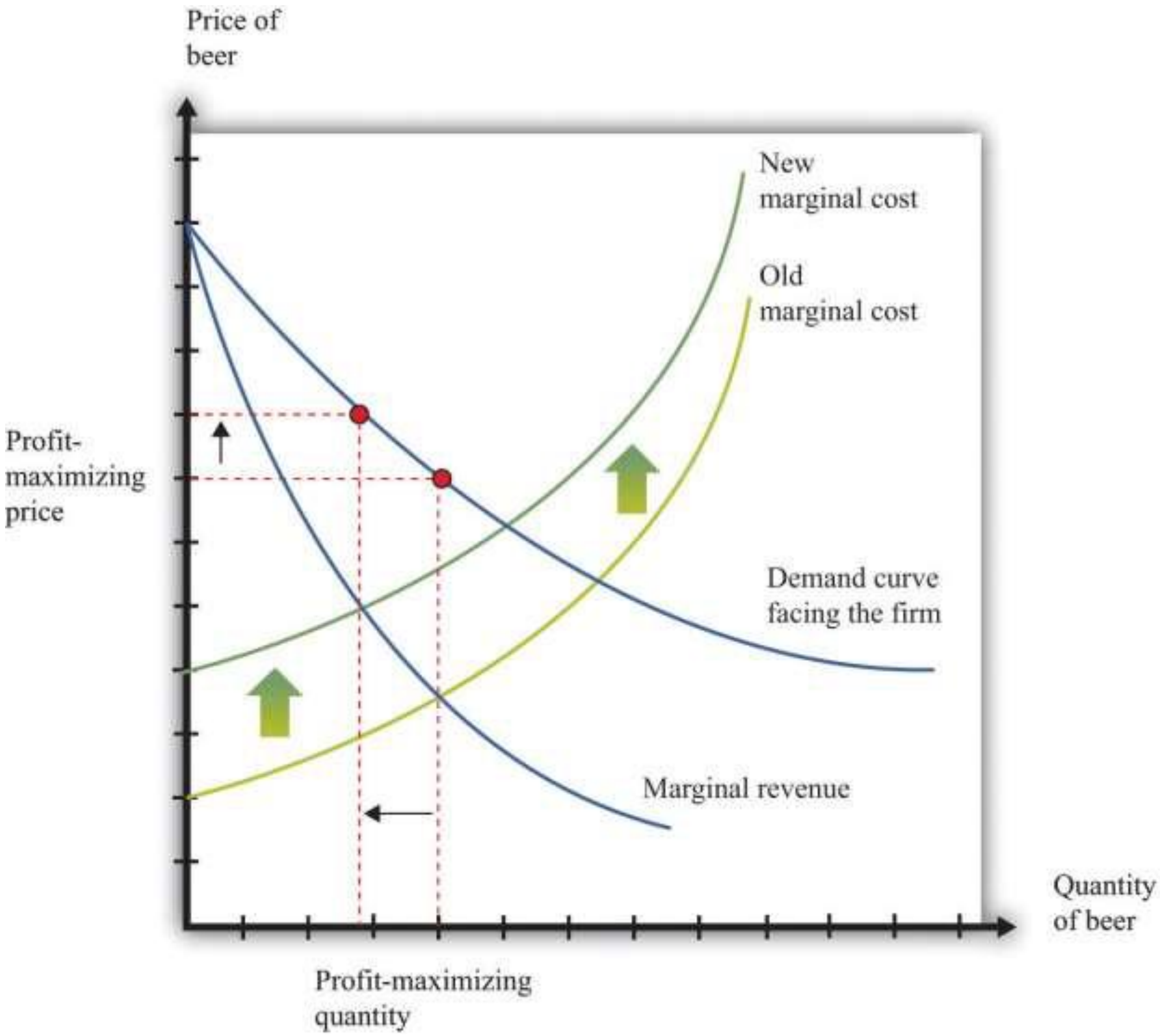
Figure 7.20 Finding the Profit-Maximizing Price and Quantity When a Firm Has Market Power



A firm with market power faces a downward-sloping demand curve and earns maximum profit at the point where marginal revenue equals marginal cost.

Now what happens if marginal cost increases? Think of a single beer producer and then imagine that the price of hops increases, so the marginal cost of producing an extra unit of output increases. This change in the marginal cost of production leads the brewer to decrease production (Figure 7.21 "An Increase in Marginal Cost"). Marginal cost decreases and marginal revenue increases until the two are again equal.

Figure 7.21 An Increase in Marginal Cost



In response to an increase in marginal cost, a firm now finds it optimal to set a higher price and produce a smaller quantity of beer.

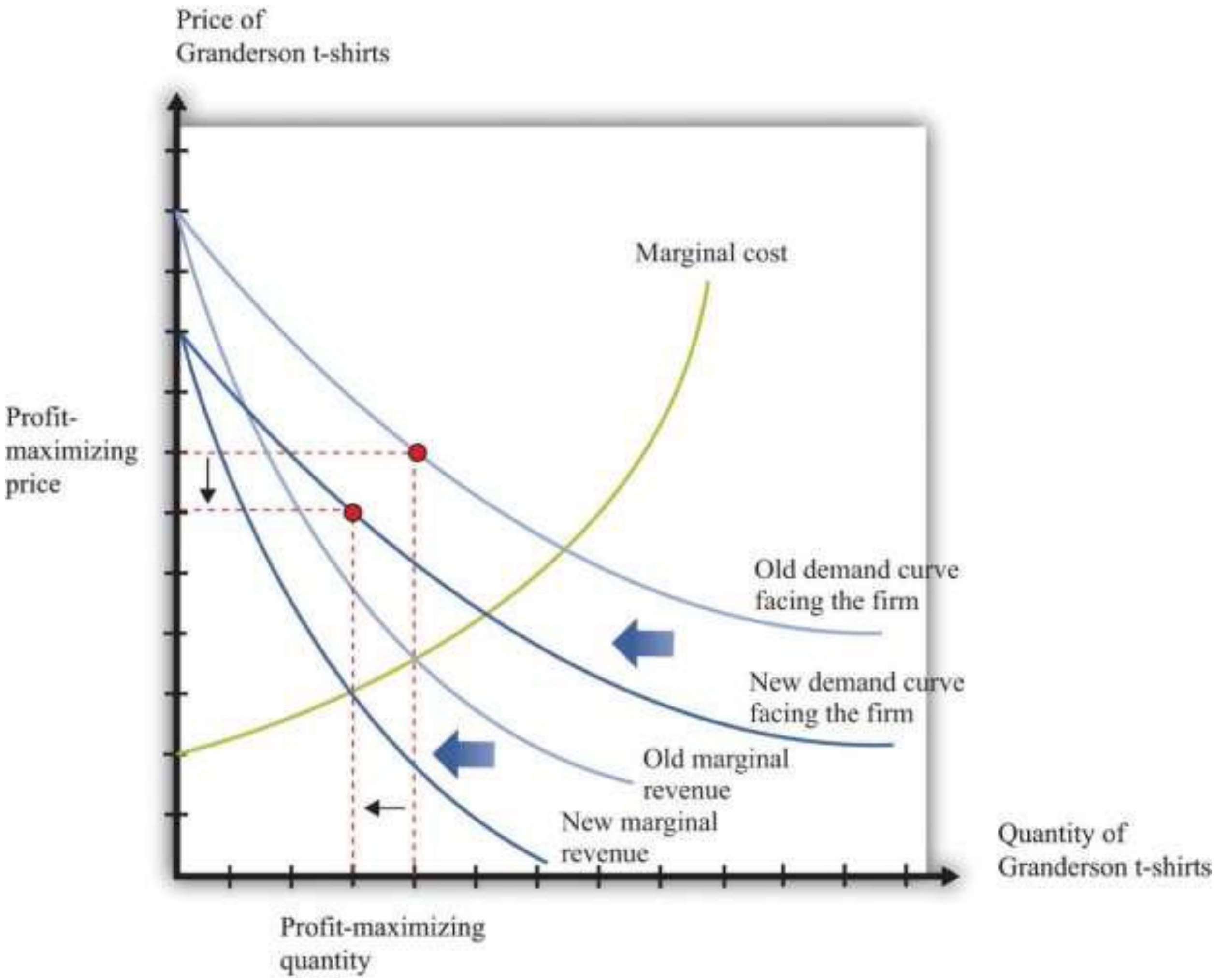
Although Figure 7.21 "An Increase in Marginal Cost" is drawn for a single seller, it captures the common experience and response of *all* sellers. The increase in the price of hops causes the marginal cost to increase for all brewers, and they all respond by producing less and increasing the price of beer. And *this is exactly what we predicted about prices and quantities when we considered an increase in marginal cost in a perfectly competitive market.*

Shifts in Demand

Now reconsider the T-shirt example. There may be only a small number of producers who are licensed to produce Detroit Tigers T-shirts. Although there are many different kinds of replica sporting shirts in the world—at least one for each major team in most of the sports you can imagine—these shirts are not all the same. So the market for replica T-shirts—and certainly the market for Granderson T-shirts—is not competitive. The producers of Granderson shirts for the Tigers choose quantities and set prices (see Figure 7.20 "Finding the Profit-Maximizing Price and Quantity When a Firm Has Market Power").

Figure 7.22 "Shifts in Demand and Marginal Revenue" shows the market as seen by one of these producers. Granderson's move from the Tigers shifts the demand curve inward for the shirts that they produce. This shift in the demand curve also shifts the marginal revenue curve inward. In response, the firm adjusts its output so that marginal revenue again equals marginal cost, choosing its price to match the point on the demand curve at the new quantity produced. Output and price both fall. Again, this is the same prediction that we obtained from the comparative statics of a competitive market.

Figure 7.22 Shifts in Demand and Marginal Revenue



A decrease in demand causes marginal revenue to shift to the left. Marginal cost and marginal revenue intersect at a lower level of output. This lower level of output means marginal cost is lower, so the firm will also decrease its price.

When Is Using Supply and Demand Misleading?

Even when markets are not perfectly competitive, the supply-and-demand framework is usually a good device for predicting what will happen to prices and quantities in a market following a shock. Even if firms have market power, an increase in marginal cost will typically lead to an increase in the price and a decrease in the quantity supplied, just as supply and demand predict. Similarly, an increase in demand will typically lead to an increase in the price and an increase in the quantity supplied, again as predicted by basic supply-and-demand reasoning.

Although the supply-and-demand framework can be used for most situations where markets are not perfectly competitive, we still need to know when it might mislead us.

- Though unlikely, it is possible that an increase in demand will lead to an increase in the price and a *decrease* in the quantity supplied. This is because it is possible, though unlikely, that an increase in demand will cause marginal revenue to decrease.
- It is possible that an increase in demand will lead to an increase in quantity but a *decrease* in the price. There are two ways this could happen, both stemming from the fact that a firm sets its price as a markup over marginal cost. First, even if demand is greater, the elasticity of demand could change so as to make the optimal markup. Second, some firms with market power may have decreasing marginal cost.

We also need to recognize that while we may be able to use supply-and-demand intuition for qualitative predictions, it is more difficult to make *quantitative* predictions when markets are not competitive. We cannot determine supply and demand elasticities as easily when firms have market power. The reason is that one firm's decisions depend on what other firms are doing. Consider, for example, an increase in marginal cost in the beer industry. We have said that each firm in the market will respond by increasing its price and decreasing its quantity (Figure 7.21 "An Increase in Marginal Cost"). For example, Miller in the United States will set higher prices in response to an increase in the price of hops. But when markets are not perfectly competitive, the story does not stop there. Firms also look at the prices set by their competitors. Miller's decisions on pricing depend also on the price chosen by Budweiser. If Budweiser sets a higher price as well, then Miller may want to increase its price still further, and so on. [3]

To sum up, the supply-and-demand framework can occasionally mislead when markets are not perfectly competitive. Yet most economists still begin with supply and demand when trying to explain a change in prices or quantities. Then they consider if there are reasons to expect either changes in the elasticity of demand or decreasing marginal cost. If neither of these seems likely, then the simple intuition of supply and demand will almost certainly give the right answer.

KEY TAKEAWAYS

- Even if markets are not competitive, the qualitative predictions from comparative statics in a competitive market remain.

- The prediction of the supply-and-demand framework could be misleading if a shift of the demand curve does not lead the marginal revenue curve to shift in the same direction.

CHECKING YOUR UNDERSTANDING

1. Suppose there is a decrease in marginal cost in some industry. What will happen to price and quantity if the industry is competitive? What will happen to price and quantity if firms set the price as a markup over marginal cost?

[1] We explain how firms set these prices in Chapter 6 "Where Do Prices Come From?".

[2] This figure is explained more fully in Chapter 6 "Where Do Prices Come From?".

[3] How exactly this plays out is a complicated problem, requiring some of the ideas that we introduce in Chapter 14 "Busting Up Monopolies" and Chapter 16 "Cars".

7.6 End-of-Chapter Material

In Conclusion

The supply-and-demand framework is the most powerful framework in the economist's toolkit. Armed with an understanding of this framework, you can make sense of much economic news, and you can make intelligent predictions about future changes in prices.

A true understanding of this framework is more than just an ability to shift curves around, however. It is an understanding of how markets and prices are one of the main ways in which the world is interlinked. Markets are, quite simply, at the heart of economic life. Markets are the means by which suppliers and demanders of goods and services can meet and exchange their wares. Because exchange creates value—it makes both buyers and sellers better off—markets are the means by which our economy can prosper. Markets are the means by which economic activity is coordinated in our economy, allowing us to specialize in what we do best and buy other goods and services.

Economists wax lyrical about these features of markets, but this should not blind us to the fact that markets can go wrong. There are many ways in which market outcomes may not be the most desirable or efficient. In other chapters, we look in considerable detail at all the ways that markets can fail us as well as help us.

EXERCISES

1. Fill in the missing values in Table 7.5 "Individual and Market Demand". What can you say about the missing price in the table?

TABLE 7.5 INDIVIDUAL AND MARKET DEMAND

Price of Chocolate Bars ($)	Household 1 Demand	Household 2 Demand	Market Demand
1	7		22
2		11	16
10	0.5	3	3.5
	0.75	4	4.75

2. If the income levels of all households increase, what happens to the individual demand curves? What happens to market demand?

3. Suppose the price of coffee increases. Household 1 always eats chocolate bars while drinking coffee. What will happen to household 1's demand for chocolate bars when the price of coffee increases? Household 2 either has coffee or a chocolate bar for dessert. What happens to household 2's demand for chocolate bars when the price of coffee increases? What happens to the market demand for chocolate bars when the price of coffee increases?
4. In Figure 7.4 "Market Supply", list the factors that would imply that firm 1 produces fewer chocolate bars than firm 2 when the price is $5. The figure is drawn so that firm 1 produces less than firm 2 at all prices. Does this have to be the case? Could the firms' supply curves cross?
5. (Advanced) Draw a version of Figure 7.22 "Shifts in Demand and Marginal Revenue" if there is an outward shift in demand but no shift in marginal revenue. What would happen to the market price?
6. Consider the operation of a café. Describe the types of trades in terms of whether they are B2C or B2B. In what ways do you think that B2B trades are different from B2C trades?
7. Economists often say that individual decisions are "made at the margin." How do you see that in the determination of market supply and market demand?
8. If there are fewer sellers in a market, what will happen to total output? What will happen to the output of each seller?
9. Explain why an increase in the mortgage rate, which reduces the demand for new houses, can teach researchers about the elasticity of the supply curve.
10. (Advanced) Using the credit market, show how governmental borrowing increases interest rates. Could governmental borrowing also lead to an outward shift in the supply of credit as households save more to pay off the future debt? How would you show this in a supply-and-demand diagram?
11. If the US Federal Reserve Bank takes actions to lower interest rates in the United States relative to other countries, what will happen to the euro price of the dollar? Explain.
12. Draw a figure showing an outward shift in a demand curve along with a reduction in marginal revenue. Explain what is going on in the diagram and how a monopolist would respond to the situation.

Spreadsheet Exercise

1. Using a spreadsheet, construct a version of Table 7.2 "Market Equilibrium: An Example" assuming that market demand = 50 − 5 × price. Fill in all the prices from 1 to 100. What is the equilibrium price and the

equilibrium quantity in the market? How would you explain the difference between this equilibrium and the one displayed in Table 7.2 "Market Equilibrium: An Example"?

Economics Detective

1. Find a newspaper article that describes a price change for a good or service. Why did the price change? What happened to the quantities produced and sold?

7.7 Appendix: Algebraic Presentation of Supply and Demand

The supply-and-demand framework can be analyzed with algebra. We start with supply and demand and then talk about market equilibrium. This presentation uses some notation rather than only words:

- p is the price of a chocolate bar.
- I measures the income of individuals in a market.
- q_d is the quantity demanded.
- q_s is the quantity supplied.
- A measures the technology of chocolate bar production.

Market Demand

With this notation, we represent the demand curve as follows:

Equation 7.1

$$q_d = \alpha_d - \beta_d p + \gamma_d I.$$

α_d, β_d, and γ_d are constants that characterize the effects of prices and income on the quantity demanded. With the restriction that $\beta_d > 0$, the demand curve is downward sloping because an increase in p implies a reduction in the quantity demanded. It is natural to assume $\gamma_d > 0$, so an increase in income leads to an increase in the quantity demanded. This is represented as a shift in the demand curve.

Market Supply

With this notation, we represent the supply curve as follows:

Equation 7.2

$$q_s = \alpha_s - \beta_s p + \gamma_s A.$$

α_s, β_s, and γ_s are constants that characterize the effects of prices and income on the quantity demanded. With the restriction that $\beta_s > 0$, the supply curve is upward sloping because an increase in p leads to an increase in the quantity supplied by all firms. It is natural to assume $\gamma_s > 0$, so an increase in the productivity of the current technology leads to an increase in the quantity produced at a given price. This is represented as a shift in the supply curve.

Market Equilibrium

The market is in equilibrium if there is a price and quantity combination, denoted (p^*, q^*) such that at the price p^*, the quantity demanded, and the quantity supplied equal q^*. Equilibrium is the simultaneous solution of supply and demand and can be found using the substitution method outlined in the toolkit.

Using $q_s = q_d = q^*$, we can substitute Equation 7.2 into Equation 7.1 yielding:

Equation 7.3

$\alpha_d - \beta_d p^* + \gamma_d I = \alpha_s - \beta_s p^* + \gamma_s A.$

This is a single equation in a single unknown, p^*. Solving the equation for p^* implies

Equation 7.4

$$p^* = \frac{\alpha_d - \alpha_s + \gamma_d I - \gamma_s A}{\beta_d + \beta_s}.$$

The denominator is positive because we have assumed that both β_d and β_s are positive. The numerator is positive as long as the vertical intercept of the demand curve is greater than the vertical intercept of the supply curve: $(\alpha_d + \gamma_s I) > (\alpha_s + \gamma_d A)$. This condition, combined with the restrictions on the slopes of supply and demand, is sufficient to guarantee that an equilibrium price exists in the market.

Using this calculation of p^* in, say, the supply curve, we find

Equation 7.5

$$q^* = \alpha_s + \beta_s \frac{\alpha_d - \alpha_s + \gamma_d I - \gamma_s A}{\beta_d + \beta_s} + \gamma_s A.$$

Grouping the terms into a constant, $\gamma_d I$ and $\gamma_s A$, this becomes

Equation 7.6

$$q^* = \alpha_s + \beta_s\left(\frac{\alpha_d - \alpha_s}{\beta_d + \beta_s}\right) + \gamma_d I\left(\frac{1}{\beta_d + \beta_s}\right) + \gamma_s A\left(1 - \frac{1}{\beta_d + \beta_s}\right).$$

Looking at Equation 7.4 and Equation 7.6, these expressions determine the equilibrium price and the equilibrium quantity depending on the two (exogenous) factors that impact supply and demand: income level I and state of technology A. Though income influences only the position of the demand curve, variations in income influence both the equilibrium price and the equilibrium quantity. The same is true for variations in technology that shift only the supply curve.

Chapter 8
Growing Jobs

Changing Jobs

Figure 8.1 "Walmart Fact Sheet" is a fact sheet from Walmart. The fact sheet tells you—if you didn't know already—that Walmart is everywhere. It has over 7,000 retail units in the world, with over 3,000 outside the United States. Walmart employs about 1.36 million people in the United States alone, which is about 1 percent of the total number of workers employed in the United States. It seems that Walmart means jobs.

Figure 8.2 Walmart: Growing Jobs?

Source: This image is taken fromhttp://en.wikipedia.org/wiki/Image:Walmartbizarro.png. The image is copyright Dan Piraro 2006.

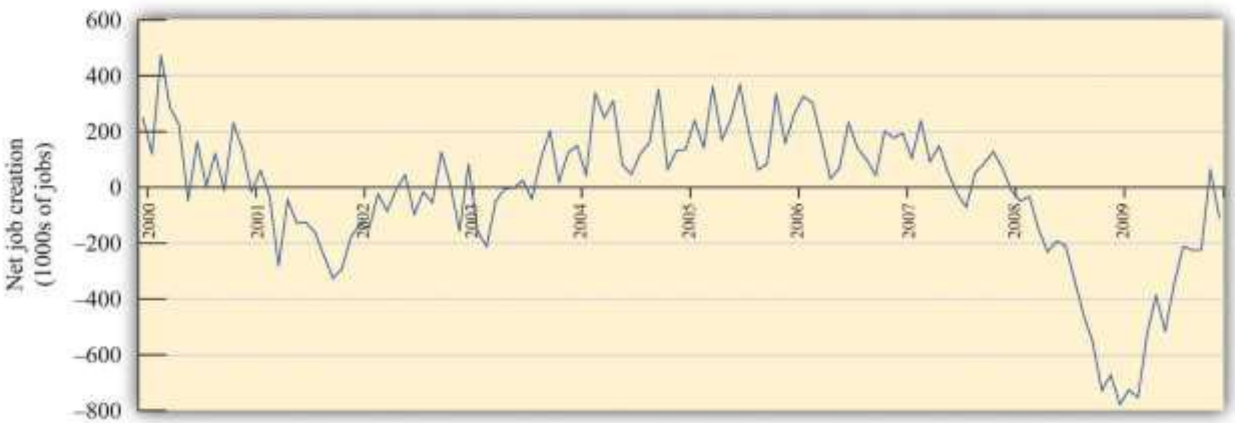
Despite the fact that Walmart provides so many jobs, the announcement of a new Walmart store is often greeted with trepidation or outright opposition. There are websites and even a film (http://www.walmartmovie.com) dedicated to criticism of Walmart. It is true that the arrival of Walmart in a town will mean the creation of new jobs, including checkout clerks, shelf packers, and many other positions. Yet the arrival of Walmart will also mean that its competitors will lose jobs. The overall effect on jobs is unclear.

The arrival of a Walmart in a town has implications beyond the effects on jobs. Consumers are likely to face a different menu of goods. Walmart will bring some goods that were previously unavailable, while at the same time other goods sold by unsuccessful specialty stores will disappear. The shopping experience will change for consumers because shopping in a Walmart is not like shopping in a series of small stores. Consumers will face different and—for the most part—lower prices: Walmart is able to obtain goods more cheaply from suppliers and also has a very efficient distribution system that decreases costs. As fewer consumers visit the smaller shops, other nearby businesses, such as local restaurants, may suffer a decline in demand. The patterns of life in the town will change in numerous yet subtle ways.

This scenario has played out in many countries with many different stores. In France, there are hypermarkets such as Carrefour that likewise have had major effects on local businesses and sometimes encounter opposition. In England, the same is true of the Asda supermarket chain. In this chapter, we look at the economics that lies behind a firm's decision to enter new markets (that is, open new stores) and exit from markets (that is, close down and go out of business). Although we begin here with Walmart, we tell a story that is about much more. Over the course of every year, jobs are created when existing firms expand and new firms enter. At the same time, jobs are lost when firms contract their workforces or close down completely.

A job is created when either an existing firm or a new firm hires workers. Jobs are destroyed when a firm fires some of its workers, some workers quit, or a firm exits a market. Figure 8.3 "US Net Job Creation from 2000 to 2009" provides data on net job creation in the United States from 2000 to 2009. By net job creation, we mean the number of new jobs created minus the number of jobs destroyed. For example, if some firms expand their employment by 200 workers and other firms reduce their workforce by 150 workers, we say that 200 jobs are created, 150 are destroyed, and the net job creation is 50.

Figure 8.3 US Net Job Creation from 2000 to 2009



Job creation and destruction take place all the time. Within an industry, and sometimes even within a firm, we see job creation and destruction occurring simultaneously. The Bureau of Labor Statistics now regularly produces quarterly job creation and destruction rates for the US economy. [1] For the US private sector over the period 1990–2010, the average quarterly job creation and destruction rates were 7.5 percent and 7.3 percent, respectively. To put it another way, if you looked at 1,000 typical private sector jobs right now, then 75 of them didn't exist last quarter, and 73 won't exist next quarter. This implies that about 15 percent of jobs are either destroyed or created in a given quarter. Steven Davis, John Haltiwanger, and Scott Schuh, who were perhaps the first economists to study these processes in detail, call this *job reallocation* because it reflects the reshuffling of jobs across production locations. Sometimes a car produced at one particular automobile factory is selling well, so new jobs are created there. In the same quarter, another automobile factory may be shut down because the models produced there are not selling. The picture you should take away from these numbers is one of a very fluid labor market. One of our goals in this chapter is to understand how this labor market works.

One way that jobs can be created is by expanding an existing plant or firm. Another way that jobs are created is by the entry of a new plant or firm. The situation is analogous for job destruction. Some jobs are lost when existing plants contract (such as a plant eliminating a shift). Others are lost through the exit of a plant or a firm. The Davis, Haltiwanger, and Schuh data suggest that, each quarter, 11.5 percent of the jobs destroyed come from plants closing. An additional 20 percent of jobs are destroyed when plants undertake large workforce adjustments of more than 50 percent. On the entry side, about 8.4 percent of jobs created in a quarter are due to start-ups. [2] Of course, if these start-ups are successful, then they create more jobs in later years. Ultimately, we want to answer the following question:

What are the economic forces driving job creation and destruction?

Road Map

The first part of this chapter examines labor demand by firms. We begin by looking at the decision of how many hours to hire. Then we turn to the entry and exit decisions of firms. We find the decision rules that govern these decisions.

After that, we consider a different way of looking at the labor market by examining the process of search and bargaining. We look at how workers who supply labor and firms that demand labor actually interact. Finally, we examine the effects of various government policies on labor market outcomes.

[1] See Bureau of Labor Statistics, Business Employment Dynamics, "Table 3. Private sector gross job gains and losses, as a percent of employment, seasonally adjusted," accessed March 1, 2011, http://www.bls.gov/web/cewbd/table3_1.txt.

[2]

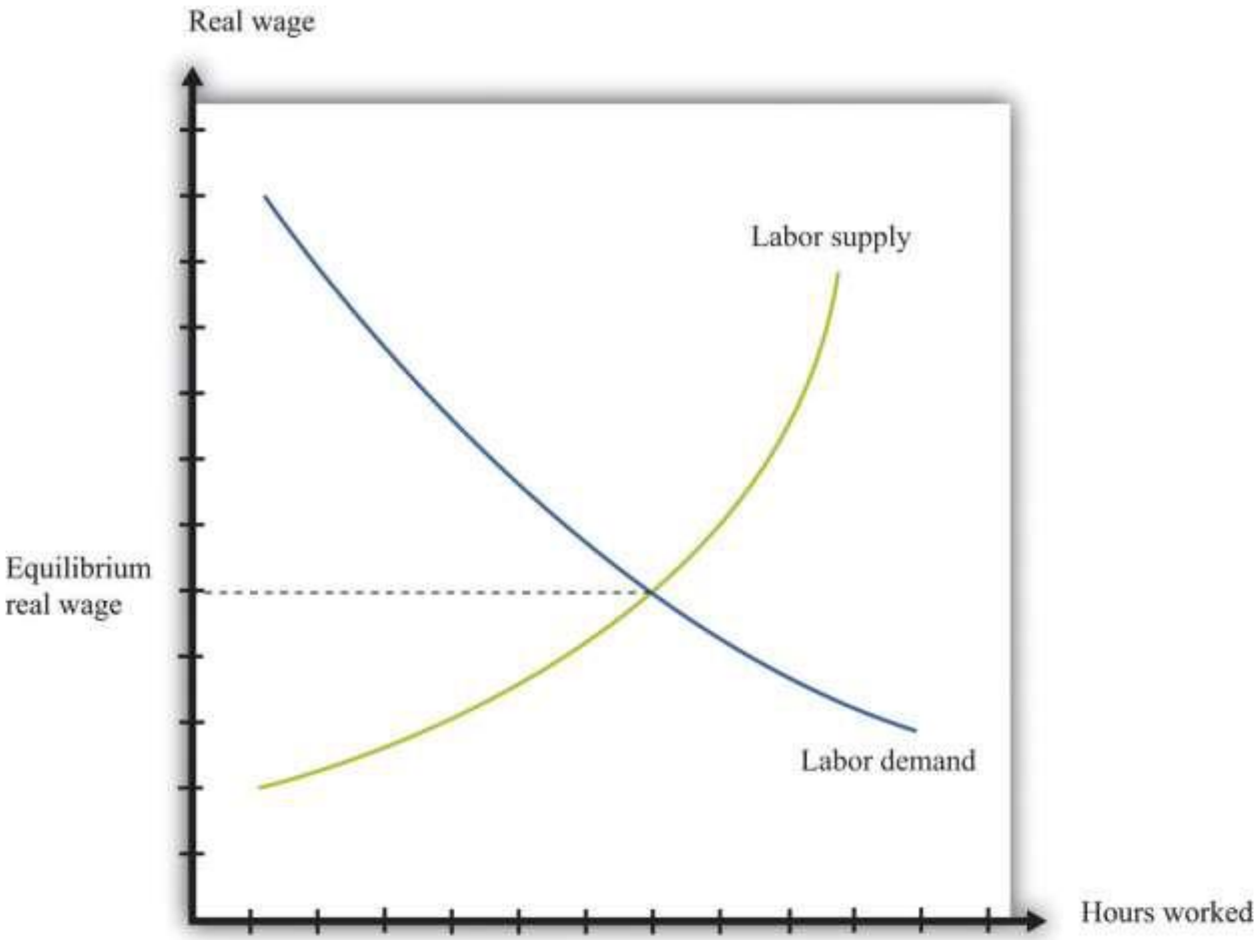
8.1 How Do Firms Decide How Many Hours of Labor to Hire?

LEARNING OBJECTIVES

1. What is a production function?
2. How does a firm decide how much to produce?
3. What factors determine a firm's labor demand?

When economists are asked to explain the creation and destruction of jobs in an economy, they will typically begin with a diagram of supply and demand in the **labor market**. In the labor market, the **real wage** (on the vertical axis) and the total number of hours worked (on the horizontal axis) are determined by the interaction of labor supply and labor demand. As shown in Figure 8.4 "Labor Market Equilibrium", equilibrium in the labor market occurs at the wage and employment level such that the number of hours supplied and demanded is equal.

Figure 8.4 Labor Market Equilibrium



The equilibrium real wage in the labor market is the price where supply equals demand.

Toolkit: Section 17.3 "The Labor Market"

See the toolkit for more discussion on the labor market.

The upward-sloping supply curve tells us that households will want to supply more labor time as wages increase. [1]

- As the wage increases, some people find it worthwhile to enter the labor force and look for a job.
- As the wage increases, some of those with jobs will find it worthwhile to work more hours.

  Labor demand slopes downward for two analogous reasons:

- As the wage increases, some firms find that they are no longer profitable and close down.
- As the wage increases, those firms that are in business choose to hire fewer hours of labor.

Thus an increase in wages will induce job destruction, and a decrease in wages will induce job creation.

The Decisions of a Firm

Firms hire labor to help them produce output. The amount of labor that a firm needs depends on the amount of output that it wants to produce. At the same time, its decision about how much to produce depends on its costs of production, which include the cost of labor. Our task here is to combine these ideas. The decision about how much labor to hire is only one of a large number of choices made by a firm's managers. Of these, the most fundamental decisions are the following:

- Should a firm be in business at all?
- How much should a firm produce, and what price should its managers set?
- How should a firm produce its desired level of output?

A firm's managers should actually answer these questions in the reverse order:

1. Managers should first determine the best way to produce output.
2. Then managers need to make a price/output decision. A firm is fundamentally constrained by the desires of the market. If managers choose the price of output, they must accept whatever sales are demanded by consumers at that price. If they choose the level of output, they can only charge the price that the market will bear for that quantity. In other words, a firm's managers must choose a point on the **demand curve facing a firm**.
3. Then, and only then, can a firm's managers decide if it is worth being in business at all. An existing firm can stay in the market or exit, thus destroying jobs. Potential new firms can enter the market and create jobs.

We follow this logic in our discussion.

Toolkit: Section 17.15 "Pricing with Market Power"

You can review the demand curve facing a firm and the details of pricing with market power in the toolkit.

The Production Function

A firm possesses a means of turning inputs into outputs. For example, Starbucks produces, among other things, grande vanilla low-fat decaf lattes. This drink is an example of a Starbucks output. The list of inputs that Starbucks needs to produce this product is much too long to write out in full but includes the following:

- Water, coffee beans, milk, vanilla syrup, paper cups, lids, and electricity
- An espresso machine
- The time of the barista—the person making the drink
- Starbucks' "blueprints" (that is, the instructions for how to make a grande vanilla low-fat decaf latte)

This list doesn't include any of the "back office" aspects of Starbucks' operations, such as accounting, payroll, or the logistics of sourcing coffee beans and delivering them to individual stores. Other firms, of course, would have a very different list of inputs. So if we want to talk in abstract terms about the production of a firm, we need a description of production that could apply not only to Starbucks but also to General Motors, IKEA, your local computer repair store, and a manufacturer of paper clips. Therefore, we group inputs into broad categories called factors of production.

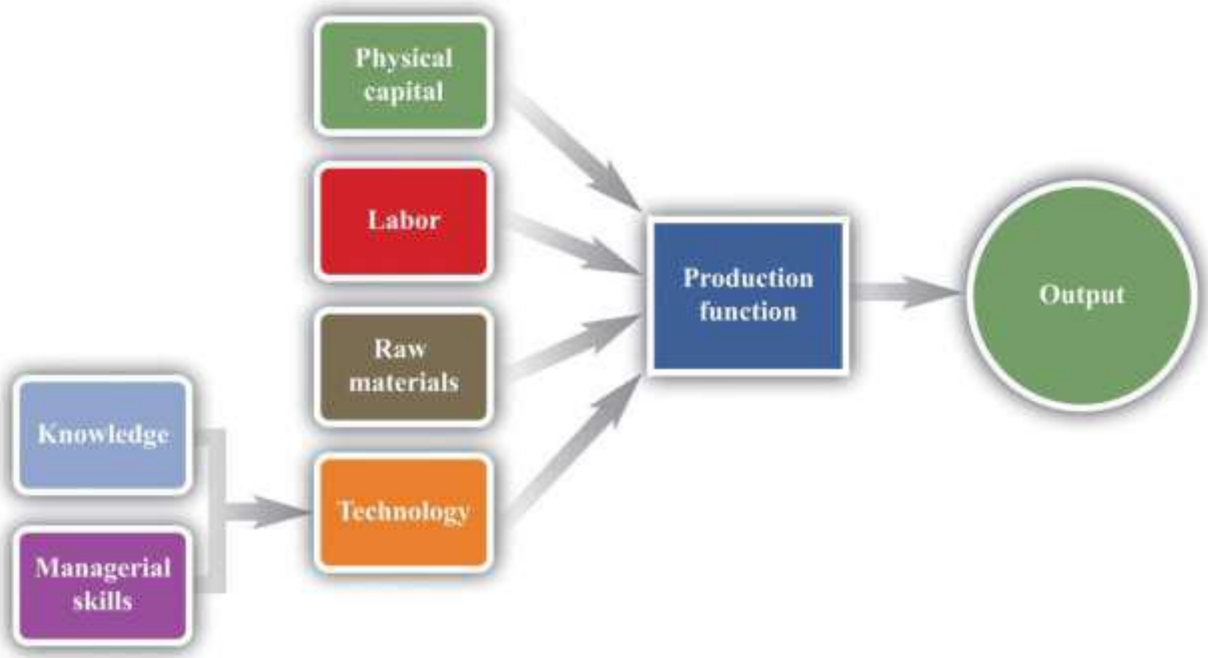
Toolkit: Section 17.17 "Production Function"

Economists group the inputs of any firm into a small number of general categories: raw materials, capital, and labor. We call these inputs a firm's factors of production.

You can think of raw materials as the things that are transformed in the production process. In our Starbucks example, these include milk, coffee beans, and electricity. Labor refers to the time of the employees who work at a firm, so the time put in by a Starbucks' barista counts as labor. Capital refers to goods that are used to help with production but are not used up in the process. The espresso machine is one of Starbucks' capital goods; others are the tables and chairs in the café.

Starbucks' technology—which we also think of as a factor of production—is the knowledge that allows it to take all these inputs and turn them into an output—the final product that people actually want to buy. It is this knowledge that ultimately lies behind Starbucks' existence as a firm. Included in the technology are the managerial skills that allow Starbucks to operate effectively.

Figure 8.5 The Technology of a Firm



The production function combines a firm's physical capital stock, labor, raw materials, and technology to produce output. Technology is the knowledge (the blueprints) that a firm possesses, together with managerial skills.

We represent the production process of a firm schematically in Figure 8.5 "The Technology of a Firm". Our description is quite general and can apply to nearly any kind of firm—for example, a lawyer's office, Walmart, a university, and a child's lemonade stand. Most people find it easiest to visualize a production function in terms of physical manufacturing, such as a production line for automobiles. Think of a firm's capital as factory buildings and machinery; its labor as the workers on the production line; and its raw materials as the steel, plastic, and glass that it purchases.

A Production Function That Uses Only Labor

We summarize the technological possibilities of a firm using aproduction function, which is a description of how much output a firm can produce as it varies its inputs. Even though a typical firm's production function contains many different inputs, we can understand most of the key features of the production function using an example where labor is the only factor of production. Although there are few goods or services that literally require no inputs other than labor, there are many services that are highly labor intensive, such as babysitting, housecleaning, and personal training at a gym.

To be concrete, think about housecleaning and suppose it has the following production function:

output = productivity × hours of labor input,

where we think of productivity as just some number. If output measures clean houses, and if it takes 5 hours of labor to produce one clean house, then productivity is 0.2, and the production function is

output = 0.2 × hours of labor input.

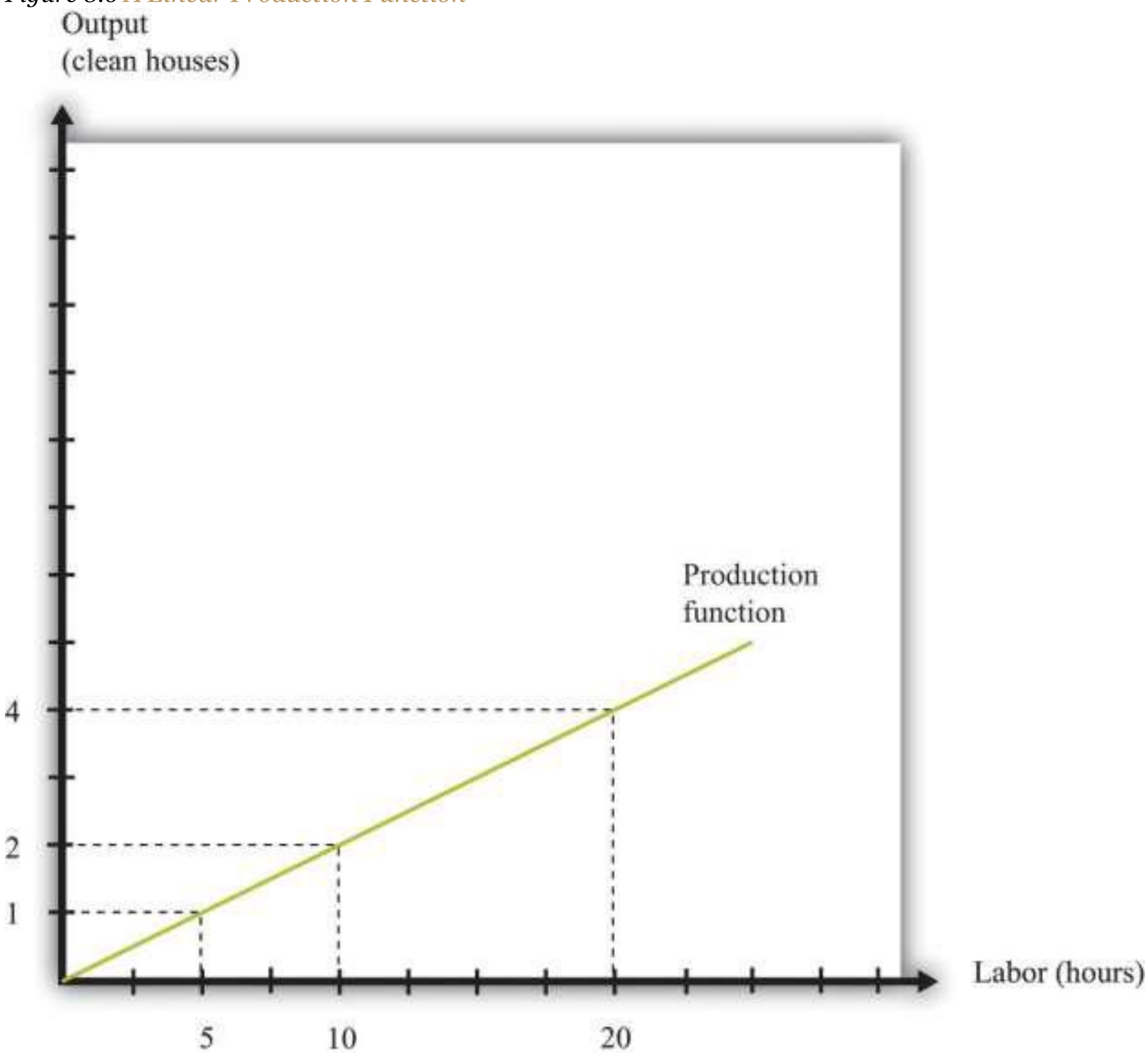
The production function tells us the level of output of a firm for given levels of labor input. Labor input is the total hours of labor time used by a firm. At this point, we are not distinguishing between hours worked per person and the number of people working, so a firm with 8 employees each working 20 hours per week has the same weekly labor input as a firm with 4 employees each working 40 hours per week. Table 8.1 "Production Function for Housecleaning"lists the amount of output that a housecleaning firm can obtain from various levels of input. We call this a *linear* production function because its graph is a straight line, as shown in Figure 8.6 "A Linear Production Function".

Table 8.1 Production Function for Housecleaning

Labor Input (Hours)	Output (Clean Houses)
0	0
1	0.2
2	0.4
3	0.6
4	0.8
5	1
10	2

Labor Input (Hours)	Output (Clean Houses)
15	3
20	4
25	5
30	6
35	7
40	8

Figure 8.6 *A Linear Production Function*



A production function shows the maximum amount of output produced, given a level of labor input.

The marginal product of labor is the amount of extra output produced from one extra hour of labor input and is defined as

$$\text{marginal product of labor} = \frac{\text{change in output}}{\text{change in labor input}}.$$

When the production function is linear, the marginal product of labor is constant. It is equal to the number we labeled *productivity* in our original production function.

In most cases, the marginal product of labor is not constant. To understand why, imagine you are managing a Starbucks outlet. You already have the machines to produce espresso, and you have lots of coffee beans on hand. You also have 500 square feet of space for making coffee and charging customers. But you still need labor. If you have no barista to operate the espresso machine, then you will have no output. If you hire one worker, you will be able to serve coffee to people. Adding the first worker will increase output considerably. However, that person must not only make the coffee but also clear the tables and handle the cash register. Adding a second worker to help with the register and clear tables will increase output even more. Now suppose you keep increasing the number of workers in the 500 square feet of space. After the third or fourth worker, they will start to bump into each other, and the barista will start to be very annoyed and unproductive. In other words, because one of your inputs—the amount of available space—is fixed, each additional worker contributes less and less to output. We call this the diminishing marginal product of labor.

Table 8.2 "Production Function for Coffee with a Diminishing Marginal Product of Labor" is an example of a production function with a diminishing marginal product of labor. In creating this table, the labor input is changed while holding all other inputs (the size of the café, the number of espresso machines, etc.) fixed.

Table 8.2 Production Function for Coffee with a Diminishing Marginal Product of Labor

Labor Input (Hours)	**Output (Cups of Coffee)**	**Marginal Product of Labor**
0	0	
1	10	10
2	14.1	4.1
3	17.3	3.2

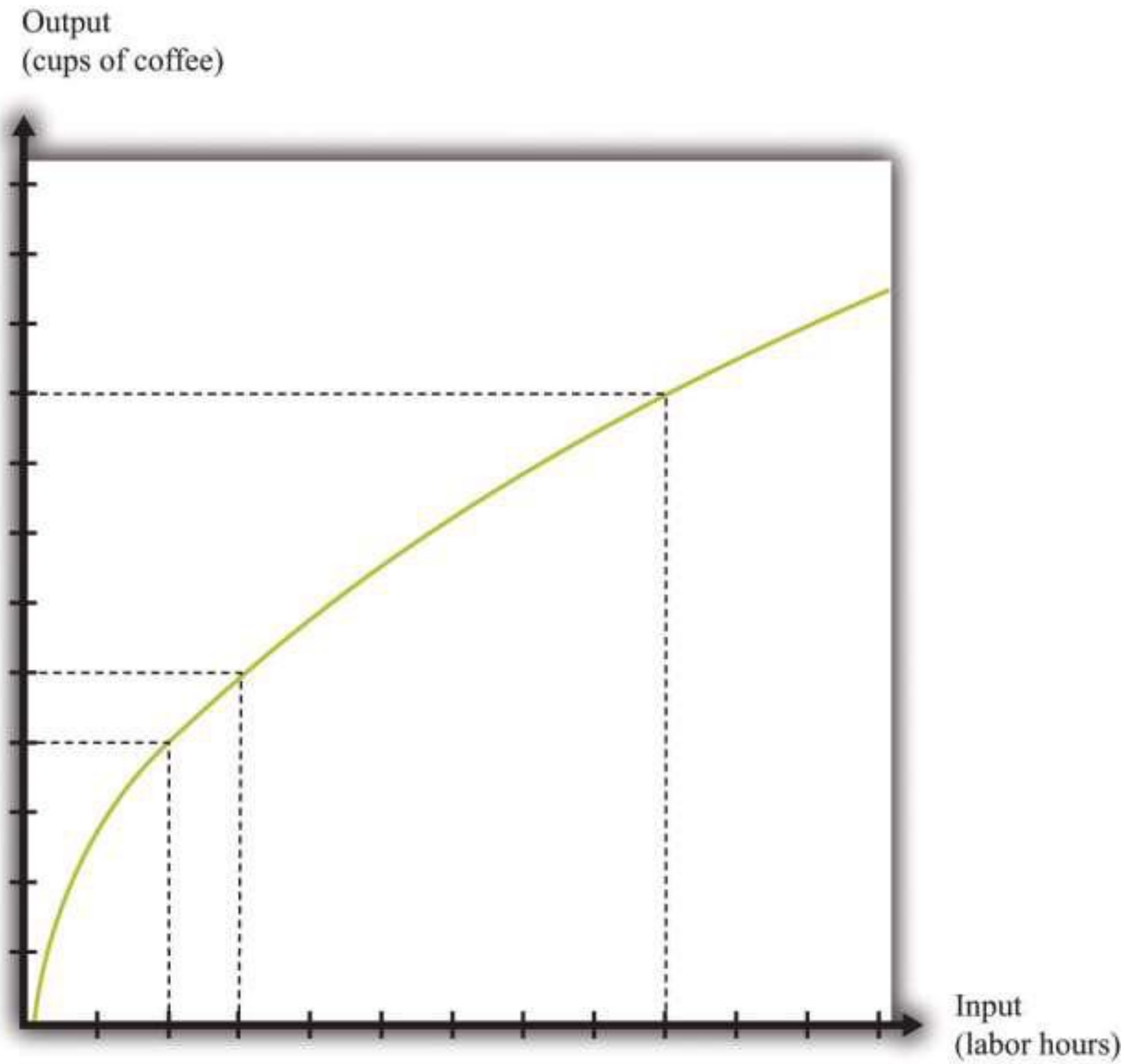
Labor Input (Hours)	Output (Cups of Coffee)	Marginal Product of Labor
4	20	2.7
5	22.4	2.4
…		
10	31.6	1.6
15	38.7	1.3
20	44.7	1.1
25	50.0	1.0
30	54.8	0.9
35	59.2	0.9
40	63.2	0.8

The marginal product of labor is shown in the third column. For the first few entries, you can calculate it directly from the table because you can easily determine how much output changes from one row to the next. For example, the marginal product of the third hour of labor is 17.3 − 14.1 = 3.2. Finding the marginal product of, say, the 40th unit of labor from the table is trickier because the table doesn't tell us how much we can produce with 39 hours of labor. Looking back at the formula for the marginal product of labor, however, we can calculate it:

$$\text{marginal product of labor} = \frac{\text{change in output}}{\text{change in labor input}} = \frac{63.2 - 59.2}{40 - 35} = \frac{4}{5} = 0.8.$$

We illustrate this production function in Figure 8.7 "A Production Function with a Diminishing Marginal Product of Labor". Notice that while the slope of the production function is always positive, the slope decreases as the labor input increases.

Figure 8.7 A Production Function with a Diminishing Marginal Product of Labor



This production function exhibits diminishing marginal product of labor: as more labor is added to a firm, output increases at a decreasing rate.

Toolkit: Section 17.17 "Production Function"

The production function is a description of how much output a firm can produce as it varies its inputs. Typically, we suppose that the production function exhibits the following:

- Positive marginal product of labor
- Diminishing marginal product of labor

The first property means that adding more labor into production means more output—that is, the slope of the production function is positive. The second property explains how the marginal product of labor varies as labor input increases. Though the marginal product of labor is always positive, it will generally decrease as more labor is added to a production process. That is why the second property is called *diminishing* marginal product of labor. (It is technically possible that the marginal product of labor could even become negative. But because a firm would never pay for workers when they decrease output, we never expect to see a firm operating with a negative marginal product of labor.)

The Cost Function

Now that we have a way of describing a firm's ability to produce goods, we are well on our way to understanding how a firm produces output. This then allows us to understand how much it will cost a firm to produce different levels of output. Our next goal is to describe these costs. The total cost of producing some specified level of output represents the cost of acquiring all the inputs needed.

To see how this works, let us determine the costs for our earlier housecleaning example. Recall that the production function is

output = 0.2 × number of hours of labor input.

Suppose that housecleaners can be hired at $10 per hour:

cost of one clean house = 5 hours × $10 per hour = $50.

The cost of two clean houses is $100, the cost of three clean houses is $150, and so on.

More generally, suppose we take the linear production function and divide both sides by the level of productivity. We get

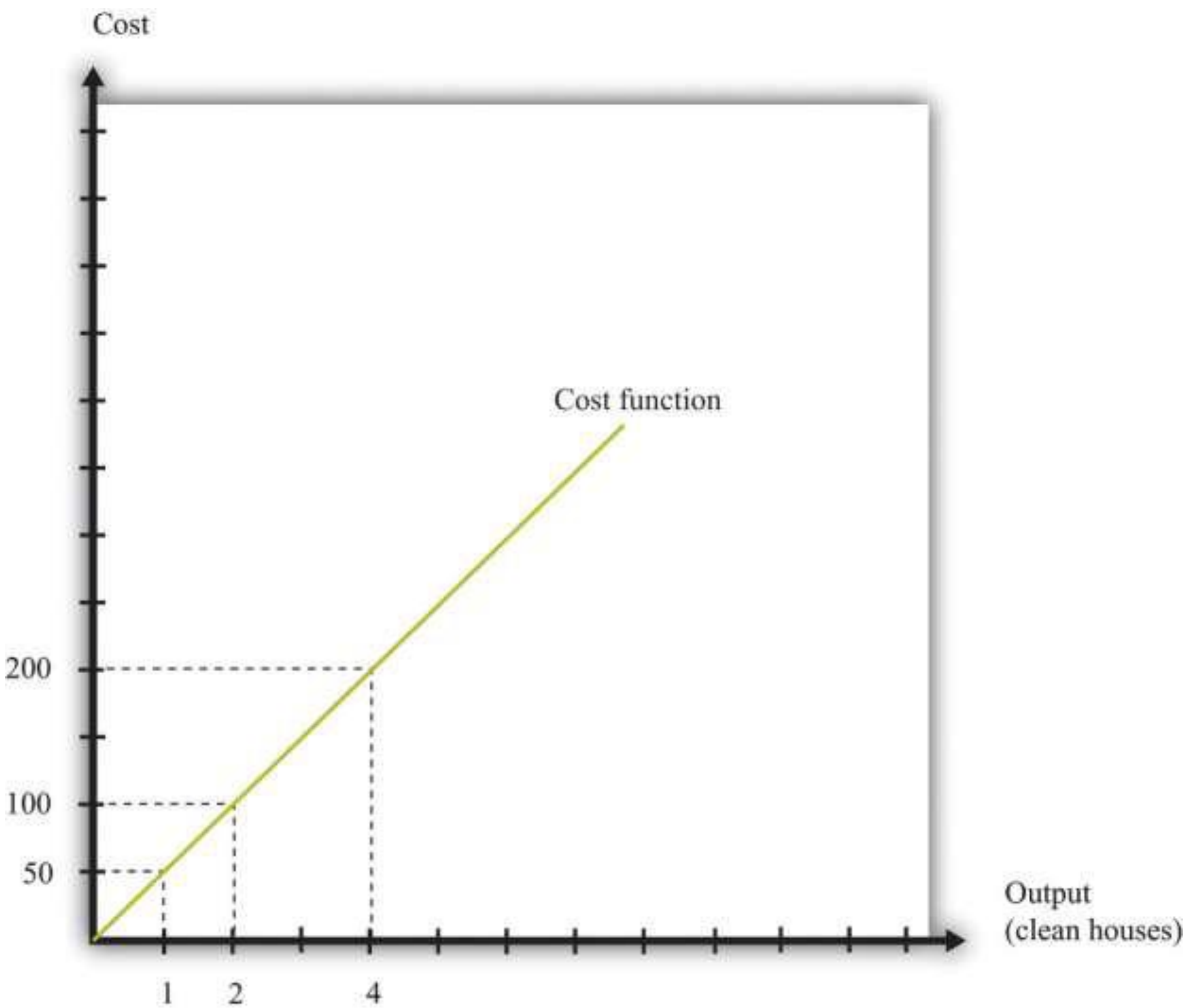
$$\text{hours of labor input} = \frac{\text{output}}{\text{productivity}}.$$

The cost of a single hour of labor is given by the wage. Thus we can write

$$\text{cost of production} = \text{wage} \times \text{hours of labor input} = \frac{\text{wage}}{\text{productivity}} \times \text{output}.$$

This is the cost function of a firm, which is illustrated in Figure 8.8 "The Cost Function". Pay careful attention to the axes in Figure 8.6 "A Linear Production Function" and Figure 8.8 "The Cost Function". Figure 8.6 "A Linear Production Function" has hours of labor on the horizontal axis and output on the vertical axis. Figure 8.8 "The Cost Function" has output on the horizontal axis and costs (= labor hours × wage) on the vertical axis.

Figure 8.8 The Cost Function



The cost function shows the cost of producing different levels of output.

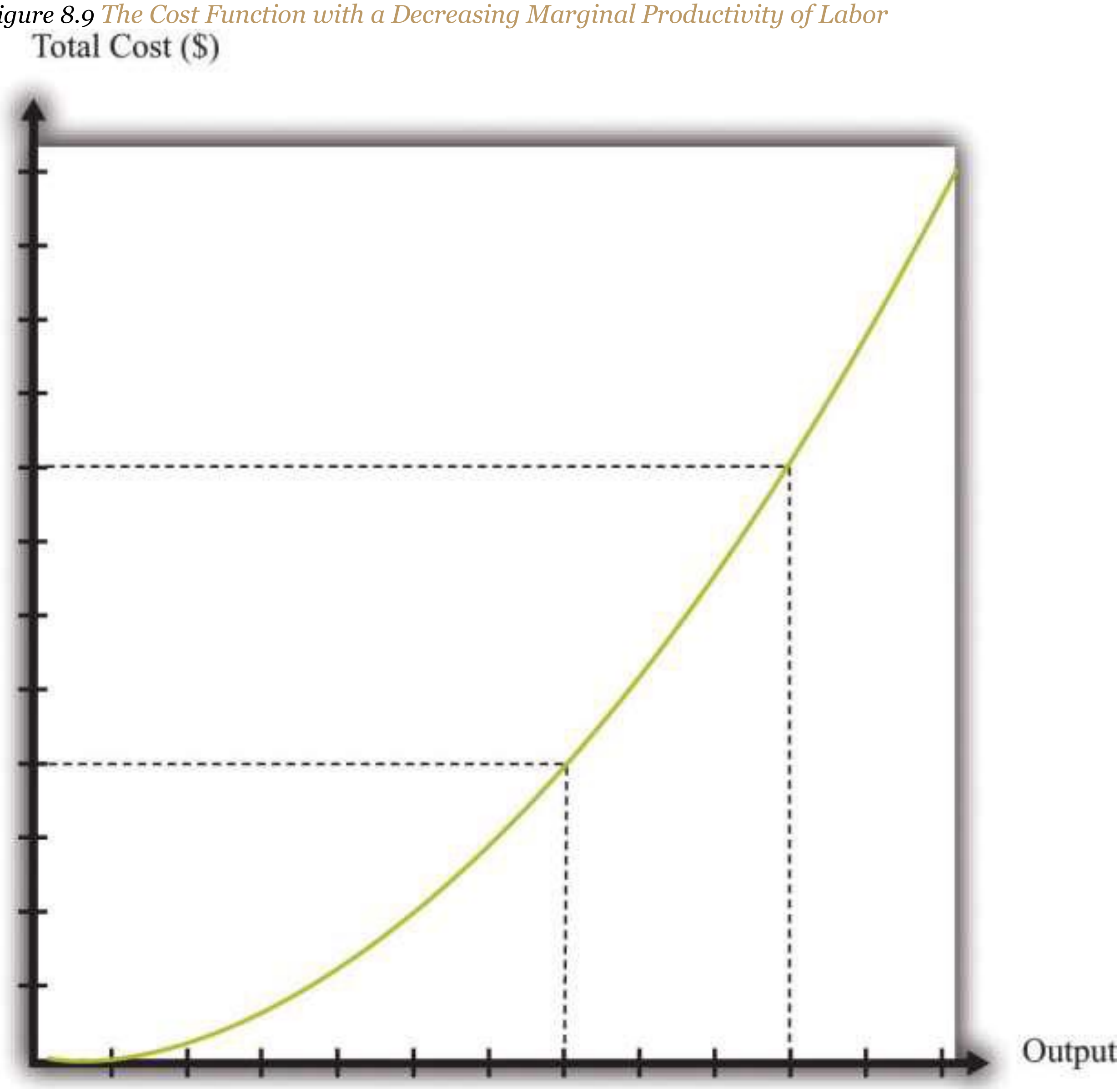
Marginal Cost

The cost function in Figure 8.8 "The Cost Function" is linear. Because the production function has a constant marginal product of labor, the cost function displays constant **marginal cost**. What about the case in which the production function has a diminishing marginal product? Then additional labor provides less and less output. Turning this around, it follows that producing each additional unit of output requires more and more labor. We show this in Figure 8.9 "The Cost Function with a Decreasing Marginal Productivity of Labor". In this figure, the marginal cost is increasing, so the cost function gets steeper as we produce more output.

Toolkit: Section 17.14 "Costs of Production"

You can review the definition of marginal cost in the toolkit.

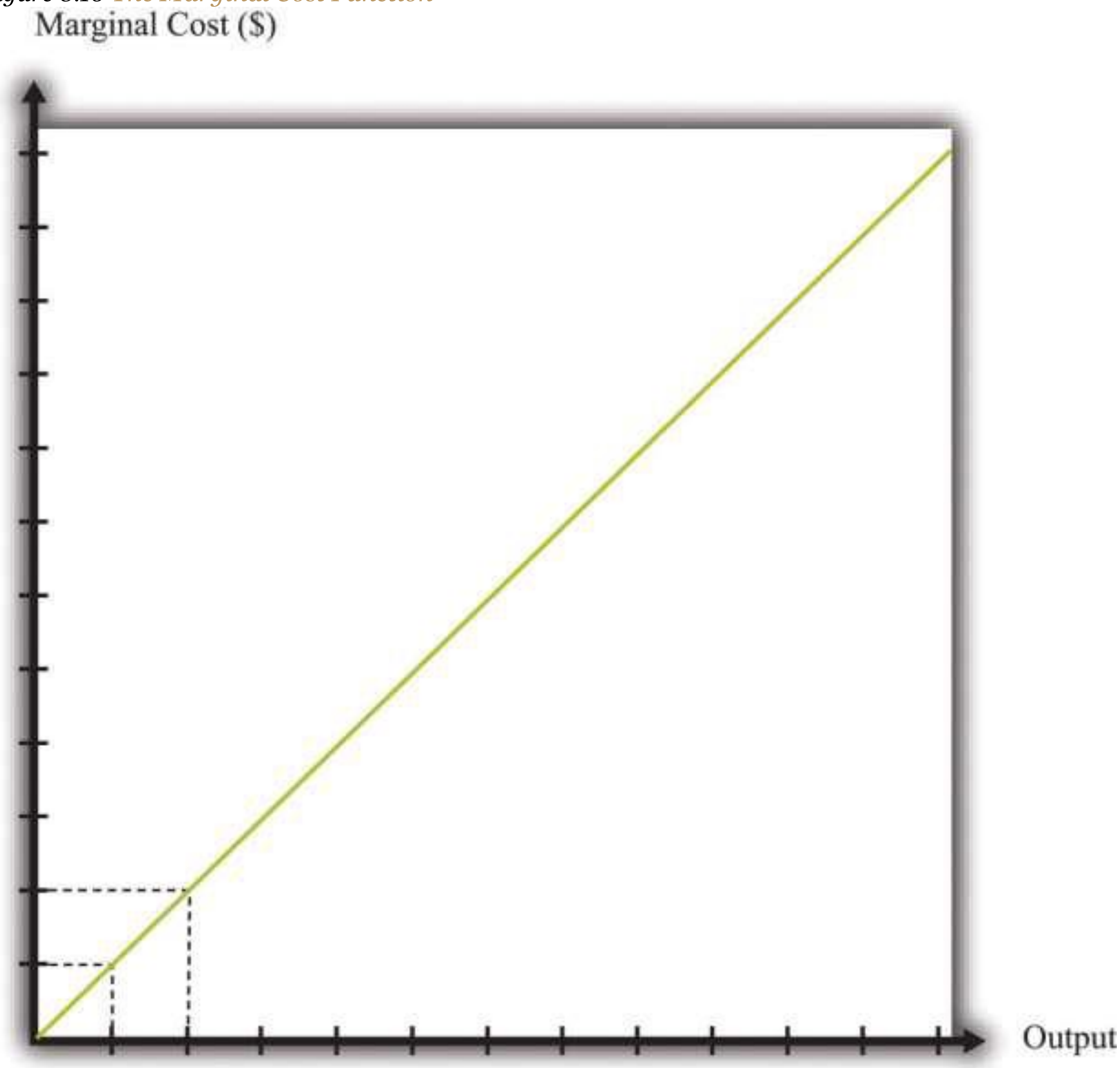
Figure 8.9 The Cost Function with a Decreasing Marginal Productivity of Labor



If a firm's technology exhibits a diminishing marginal product of labor, the cost function will increase at an increasing rate.

We show the marginal cost curve in Figure 8.10 "The Marginal Cost Function". In this example, marginal cost is a straight line, but this need not be the case in general.

Figure 8.10 *The Marginal Cost Function*



If a firm's technology exhibits a diminishing marginal product of labor, then the marginal cost will increase as output increases.

Marginal cost depends on the following:

- The *cost* of inputs into the production process
- The *productivity* of the inputs into the production process

If the costs of inputs increase, then the marginal cost is higher is well. If the productivity of the inputs into the production function is higher, then the marginal cost is lower. In fact, marginal cost can be written as

$$\text{marginal cost} = \frac{\text{wage}}{\text{marginal product of labor}}.$$

We can see from this equation that when the marginal product of labor decreases, the marginal cost of production increases. We see also that an increase in the cost of inputs—in this case an increase in wages—leads to an increase in the marginal cost of production.

The Choice of Inputs

There is one issue that we are ignoring here. Firms typically have many different ways in which they can produce the same quantity of output. A firm might have a choice between two production processes: (1) a process that is simple and cheap to operate but wasteful of raw materials and (2) a process with recycling that uses fewer materials but is more complicated and costly to run. As another example, if a construction company needs to dig a ditch, it could employ 20 people and equip each with a shovel, or it could hire a single individual and a backhoe. Economists say that the first process is *labor intensive* because it requires a lot of labor relative to capital; they call the second process *capital intensive* because it requires a relatively large amount of capital.

In medium-sized or large firms, there is usually a specific functional area, called*operations*, that decides how to produce output. Operational decisions are governed in large part by technical or engineering considerations: what are the ways in which it is physically possible to transform inputs into the desired output? Operational decisions also have an economic component. Given that there may be many different ways to get the same final amount of output, which is the most cost effective? In economics, not surprisingly, we focus on the second of these questions and leave the first to engineers and other technical experts.

The basic principle is intuitive: operations managers tend to choose methods of production that economize on relatively expensive inputs. For example, much garment manufacture takes place in countries like China or Vietnam, where wages are low (that is, labor is relatively cheap). As a result, the production processes tend to be highly labor intensive, using a lot of workers relative to the amount of machinery. By contrast, garment manufacture in richer countries (where labor is much more expensive) tends to use methods of production that require fewer people and more machines.

We simply presume that the operations function of a firm is doing a good job and has succeeded in finding the cheapest way of producing the firm's output, taking into account both the technical aspects of production and the costs of different inputs. When we talk about the cost function of a firm, therefore, we are assuming that it gives us the *lowest cost* for producing each given level of output.

The Price/Output Decision

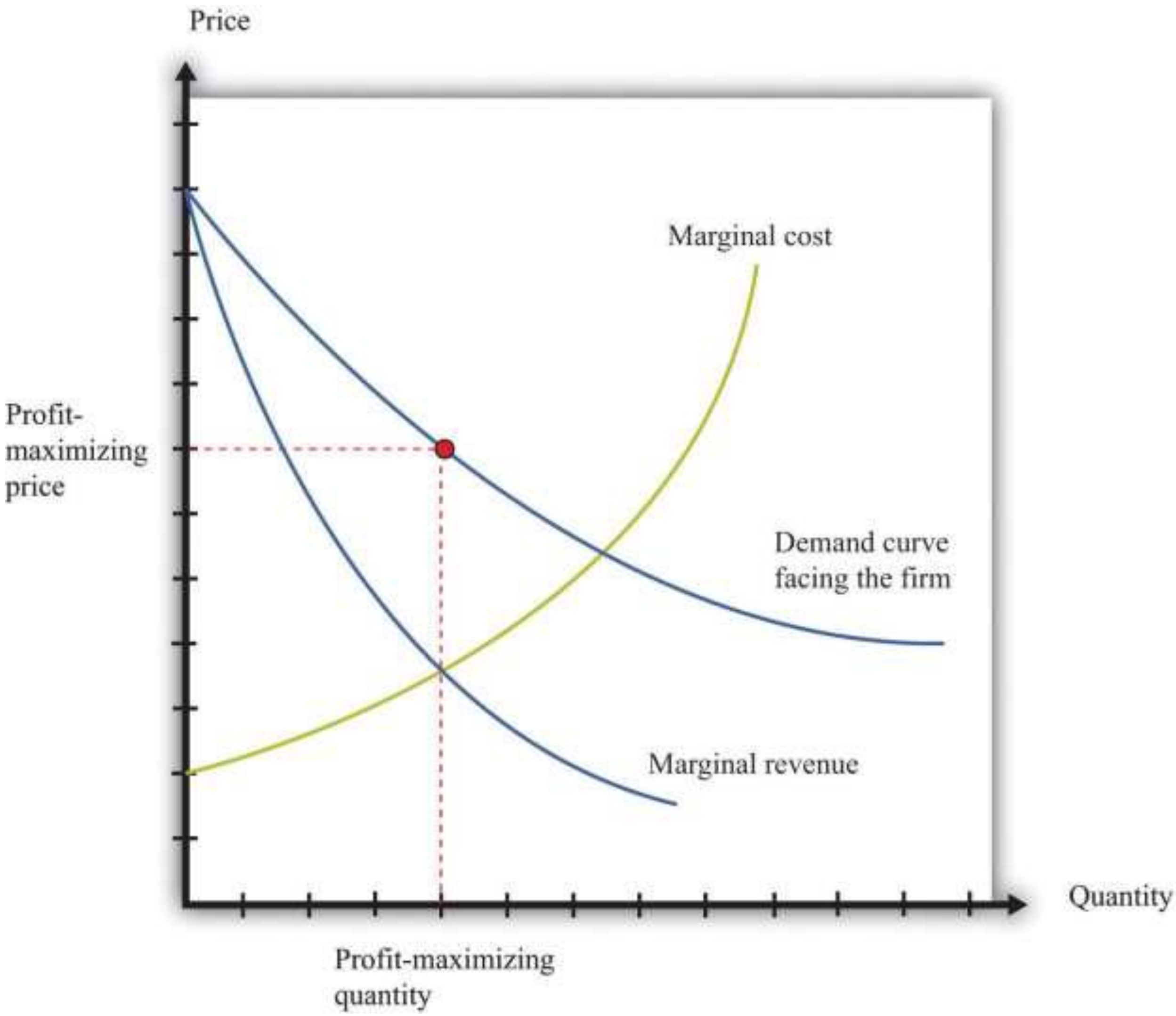
We have now completed our discussion of the first decision that managers must make: how to produce output. The production function tells us what a firm needs in terms of inputs—in this case, labor—to produce a given level of output. The more output a firm wants to produce, the more labor it will hire and the more jobs it will create. The cost function tells us the cost of producing different levels of output, and the marginal cost function tells us the cost of producing additional output.

Marginal cost is the critical ingredient in the next decision made by managers, which is selecting a point on the demand curve. We can think of managers as either choosing the price and then selling the quantity demanded at that price or choosing the level of output and selling it at the price that the market will bear. In either case, they are picking the point on the demand curve where [2]

marginal revenue = marginal cost.

We show this decision graphically in Figure 8.11 "Output and Price Decisions of a Profit-Maximizing Firm".

Figure 8.11 Output and Price Decisions of a Profit-Maximizing Firm



A profit-maximizing firm produces a quantity such that marginal revenue equals marginal cost, and the price is determined by the demand curve.

In our discussion of costs to this point, we have not specified whether we were talking about the nominal wage (that is, measured in dollars) or the real wage (that is, adjusted for inflation). The most important thing is being consistent. If we use the nominal wage when calculating our cost functions, then we end up with nominal costs. If we use the real wage, then we end up with real costs. And when we equate marginal revenue and marginal cost, we must be sure that we measure in nominal terms or real terms (not a mixture).

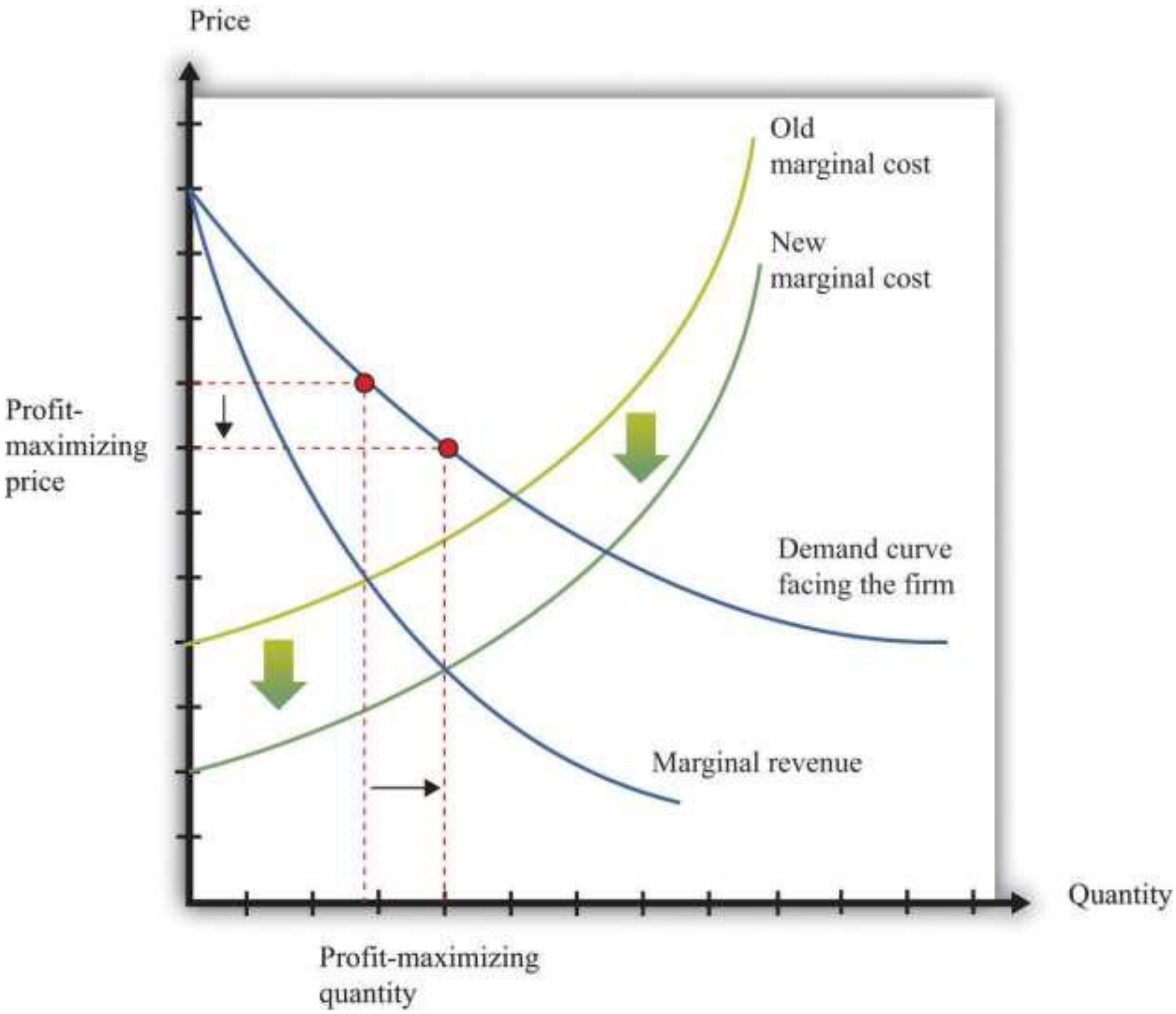
The distinction becomes important only when the general price level changes, so it is not central to our discussion here. When the price level is constant, we can always just suppose that it is equal to 1, in which case the nominal wage and the real wage are equal. Still, when we draw diagrams of the labor market, we typically put the real wage on the axis, so from here on we will explicitly suppose that we are measuring everything in real terms.

We can now explain labor demand by a firm. There are two steps:

1. As in Figure 8.11 "Output and Price Decisions of a Profit-Maximizing Firm", a firm produces a level of output such that marginal revenue equals marginal cost.
2. Using Figure 8.6 "A Linear Production Function", a firm determines the amount of labor it needs to produce the output chosen in step 1.

We already know that the marginal cost of production depends on the real wage: decreases in the real wage lead to decreases in real marginal costs. Figure 8.12 "The Effect of a Change in Marginal Cost on a Firm's Choice of Output and Employment" shows how a decrease in the real wage affects the output and price decision of a firm. As the real wage decreases, the marginal cost of an additional unit of output decreases, so a firm will choose to produce more output. The price will decrease because the firm must lower the price to sell the additional output.

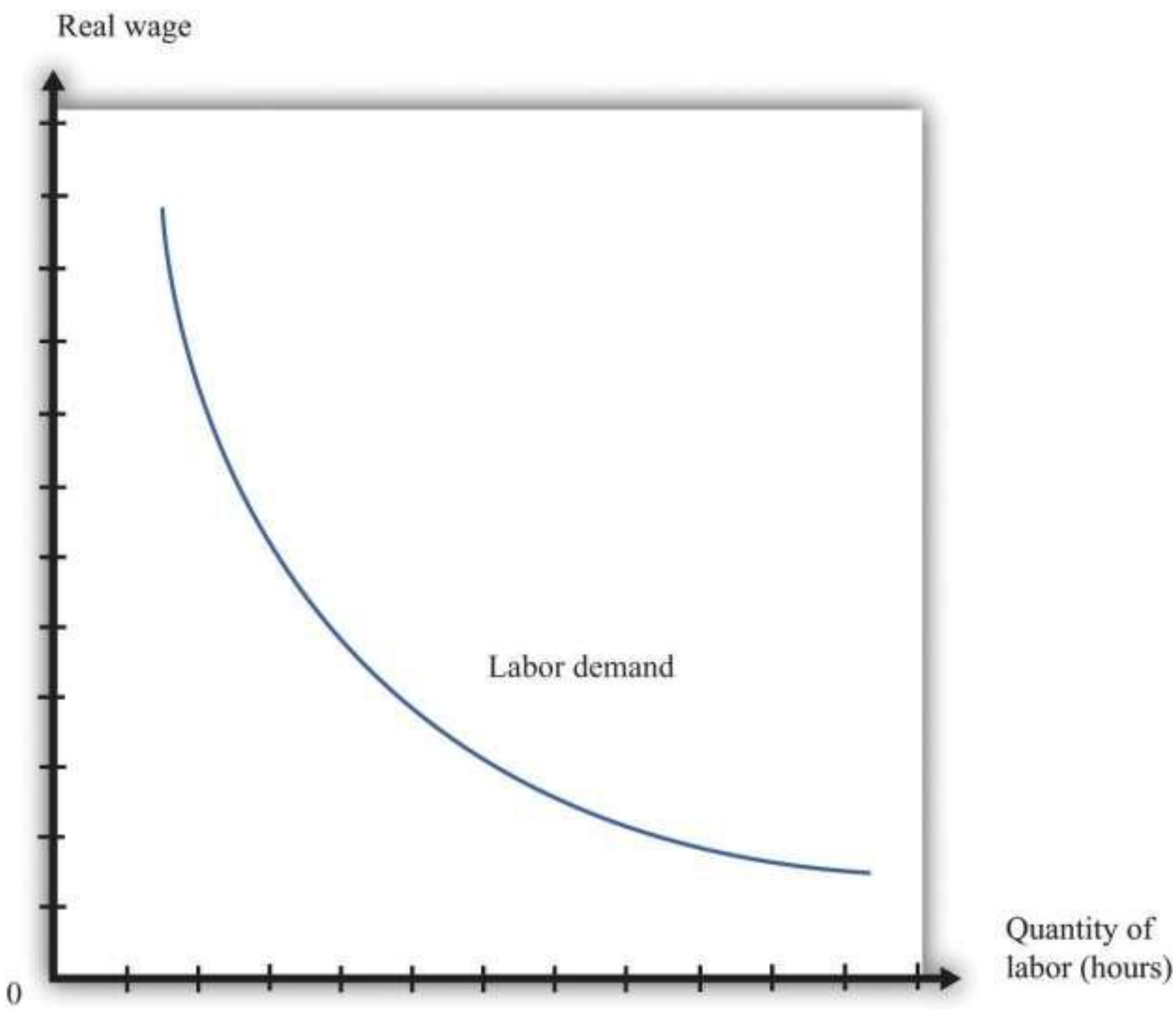
Figure 8.12 The Effect of a Change in Marginal Cost on a Firm's Choice of Output and Employment



When the real wage decreases, marginal cost decreases, so the firm reduces price, increases output, and creates jobs.

Because a firm wants to produce more output, it will demand more hours of labor. In other words, a decrease in wages leads to an increase in the quantity of labor demanded. The resulting inverse relationship between the real wage and the amount of labor demanded is shown in Figure 8.13 "The Demand for Labor".

Figure 8.13 The Demand for Labor



As the real wage increases, the demand for labor input decreases.

The labor demand curve for a single firm is downward sloping. This is true for every firm in the labor market. The market demand curve for labor is obtained by adding together the demand curves of individual firms. So the market demand for labor is downward sloping as well.

Changes in Employment

We can now connect our understanding of the labor market with the data on net job creation that we showed in Figure 8.3 "US Net Job Creation from 2000 to 2009". Based on what we have learned, there are three main reasons why jobs might be created or destroyed: (1) changes in the real wage, (2) changes in productivity, and (3) changes in demand.

Changes in the Real Wage

Changes in the cost of labor are one reason firms create or destroy jobs. Decreases in the real wage lead firms to produce more output and hire more workers, thus creating jobs. Increases in the real wage cause firms to produce less output and lay off workers. Going deeper, we can ask *why* the real wage might change. The answer comes from looking back at Figure 8.4 "Labor Market Equilibrium". The real wage changes, causing a change in the quantity of labor demanded, if the labor supply curve shifts.

Population growth is one source of a shift in labor supply. As the number of workers in the economy grows, then total labor supply will shift. At a given wage, there will be more workers and hence the labor supply curve will shift to the right. Other things equal, this causes a decrease in the real wage.

Changes in labor market participation also shift the labor supply curve. A leading example of this is the increased participation of women in the labor market. In the United States, the fraction of women in the labor force rose from about 20 percent in 1950 to about 70 percent in 2000. Participation might also change because of workers' expectations about the *future* state of the labor market. If you are worried you won't have a job next year, you might want to work this year.

Changes in Productivity

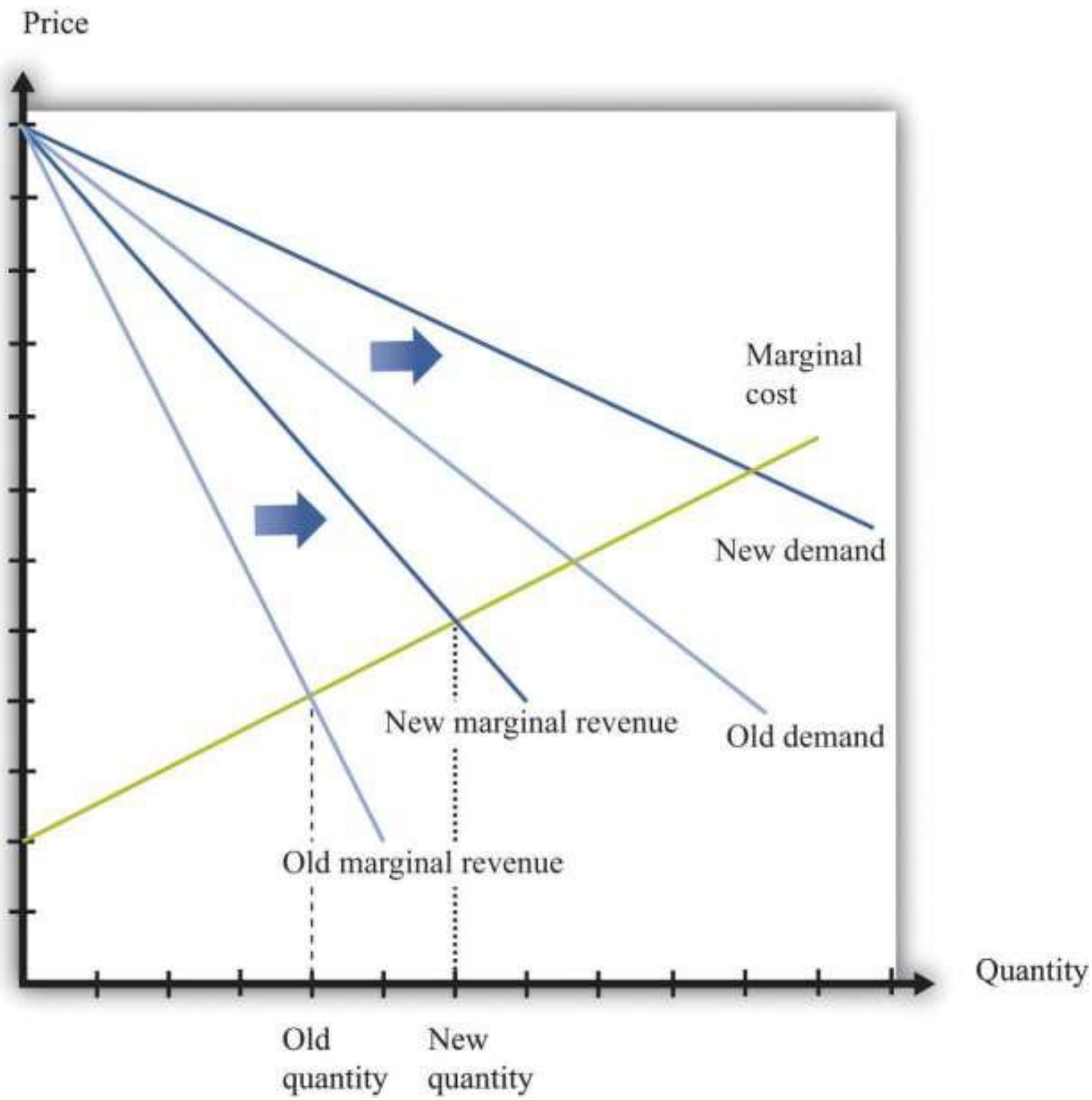
Changes in productivity—more precisely, in the marginal product of labor—work exactly like changes in the real wage. Remember that marginal cost depends on both real wages and productivity. If productivity increases, perhaps because a firm has upgraded its capital equipment, then marginal cost decreases. Firms will produce more output and hire more labor. The opposite is true if productivity decreases: in this case, firms will produce less and destroy jobs.

Over long periods of time, productivity in an economy increases. This increase in productivity is driven largely by technological advances: firms get better at producing goods and services and so are able to produce them more cheaply. As workers' productivity increases, firms demand more labor at any given wage.

Changes in Demand

When a firm's product becomes more desirable in the market, the demand curve that it faces shifts outward. This shift in demand typically leads to an outward shift in marginal revenue, inducing a firm to produce more output and demand more labor. We show this in Figure 8.14 "The Effect of a Change in Demand on a Firm's Choice of Output and Employment": an outward shift of a firm's demand curve typically leads to an outward shift in labor demand.

Figure 8.14 The Effect of a Change in Demand on a Firm's Choice of Output and Employment



An increase in demand typically leads to an increase in marginal revenue, which in turn induces firms to produce more output and create jobs.

KEY TAKEAWAYS

- A firm produces a quantity such that the marginal cost of producing an extra unit of output equals the marginal revenue from selling that extra unit of output.
- The demand for labor depends on the level of productivity, the demand for a firm's product, and the cost of labor compared to the cost of other inputs in the production process.

CHECKING YOUR UNDERSTANDING

1. Suppose the production function exhibits increasing marginal product of labor. What would it look like? What would the cost function look like in this case?
2. Marginal cost is defined as

$$\text{marginal cost} = \frac{\text{change in cost}}{\text{change in output}}.$$

We also know that

$$\text{marginal product of labor} = \frac{\text{change in output}}{\text{change in labor input}},$$

and

change in cost = wage × change in labor input.

Show how you can use these three equations to derive the condition in the text that

$$\text{marginal cost} = \frac{\text{wage}}{\text{marginal product of labor}}.$$

3. Using Figure 8.14 "The Effect of a Change in Demand on a Firm's Choice of Output and Employment", what would happen to a firm's decision on prices, the quantity of output, and labor demand if the demand curve and marginal revenue curves shifted inward?

[1] Labor supply is discussed in Chapter 3 "Everyday Decisions".

[2] This decision of the firm is also covered in detail in Chapter 6 "Where Do Prices Come From?".

8.2 Entry and Exit

LEARNING OBJECTIVES

1. What is the difference between a fixed cost and a variable cost?
2. What factors determine if a firm should remain in business?
3. What is a sunk cost?

So far we have been thinking about a firm simply changing the number of labor hours that it wants to plug into its technology. Such job creation and destruction takes place at individual firms all the time. Some firms see an increase in productivity and hire more workers. Other firms see reduced demand for their output and destroy existing jobs. The joint creation and destruction of jobs underlies the net job creation we displayed earlier in Figure 8.3 "US Net Job Creation from 2000 to 2009".

Yet the expansion and contraction of employment in existing plants is only one source of job creation and destruction. During an economic downturn, such as the severe recession that began in 2008, some firms closed factories, and other firms went completely out of business. For example, US automobile manufacturers, such as General Motors, responded to the decreased demand for cars by closing some of their existing manufacturing plants. This led to job destruction at these plants. At other times, when an economy is expanding, new firms enter into business and existing ones open new plants. Thus a complete picture of the job creation and destruction process requires us to understand the economics of entry and exit.

Only when the firm's managers know how much the firm is going to produce, the price at which it will sell it, and the cost of producing that output can they figure out profits and decide whether it is sensible to be in business at all. This logic applies to both managers of firms that are already in business and entrepreneurs who are considering starting a business. Firms also apply this logic to parts of their operations—for example, a firm may want to decide whether to shut down an existing plant or open a new one.

If a firm that is already in business discovers that its profits are too small to justify its other costs, then it should exit the market, shutting down its operations completely. If an entrepreneur is contemplating starting a new firm and calculates that the profits it will earn justify the costs of setting up operations, then we say that a firm enters the market.

In the previous section, we explained how job creation and destruction take place as firms expand and contract. When Walmart comes to town, however, much more is going on. The opening of a new Walmart means that some new jobs are created. Against that, existing stores may be forced to close down completely. Now that we have looked at a firm's price and output decisions, we are able to analyze entry and exit decisions.

Exit

Businesses do not stay around forever. At some point, they exit the market, destroying jobs in the process. Restaurants that were a big hit only a few years ago can quickly lose their luster and disappear from the scene. The same is true for many retail outlets. Manufacturing plants also close, taking jobs with them. Imagine, for example, that you own a small clothes retailing store. Then Walmart comes to your town. Now your customers have another place to buy their clothes, and you must decide whether to stay in business. You need to decide which is more profitable: staying in business or closing your business down and selling off any assets you possess.

To understand the factors influencing firm exit, we begin with a key distinction between different kinds of costs.

Toolkit: Section 17.14 "Costs of Production"

Costs that are the same at all levels of production are called fixed costs. Costs that vary with the level of production are called variable costs. As an accounting identity, the total costs of a firm are divided up as follows:

total costs = fixed costs + variable costs.

Table 8.3 "Monthly Costs of Production" shows an example of fixed costs, variable costs, and total costs for your store. (To keep life simple, we treat all the different kinds of clothing you sell as if they were the same. Let us call them blue jeans.) The numbers in Table 8.3 "Monthly Costs of Production" are based on the following equation for costs:

total costs = 14,000 + 10 × quantity.

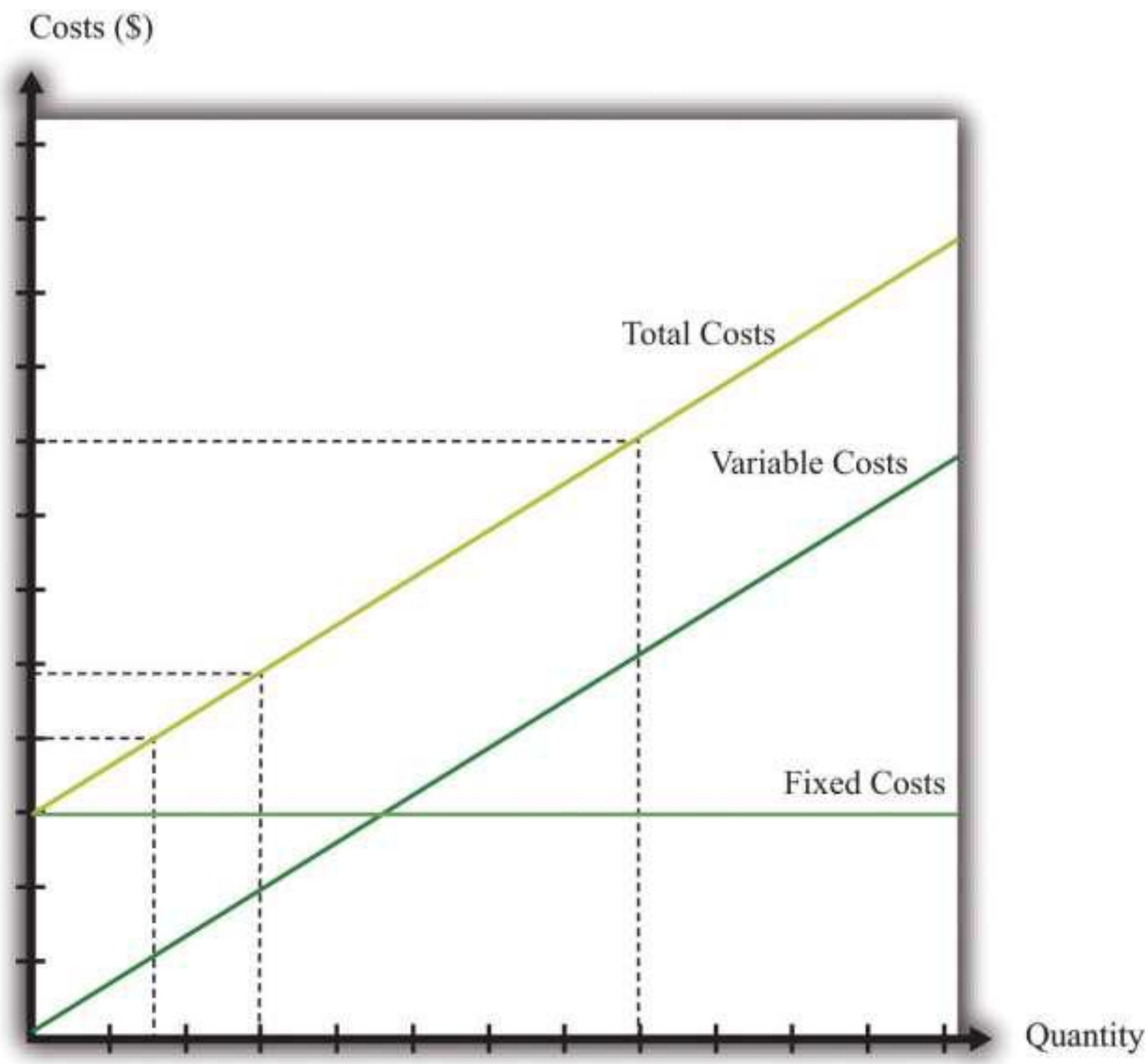
Suppose your firm has fixed costs every month of $14,000. By definition, these fixed costs do not change as your level of output changes. Think of these as your *overhead* costs—for example, the cost of renting your retail space, utility bills, the wage of your sales clerk, and so on.

Table 8.3 Monthly Costs of Production

Quantity	Fixed Costs ($)	Variable Costs ($)	Total Costs ($)
0	14,000	0	14,000
1	14,000	10	14,010
2	14,000	20	14,020
…	…	…	…
200	14,000	2,000	16,000
400	14,000	4,000	18,000
600	14,000	6,000	20,000
800	14,000	8,000	22,000
1,000	14,000	10,000	24,000
1,200	14,000	12,000	26,000
1,400	14,000	14,000	28,000
1,600	14,000	16,000	30,000
1,800	14,000	18,000	32,000
2,000	14,000	20,000	34,000
2,200	14,000	22,000	36,000
2,400	14,000	24,000	38,000

By contrast, variable costs increase as the level of output increases. In this example, if output increases by one, variable costs increase by $10. You can think of this as the cost of purchasing your blue jeans from the wholesaler. For you, the cost of "producing"—that is, making available for sale—one more unit of output is $10. Figure 8.15 "Total Costs, Fixed Costs, and Variable Costs" graphs the data from this table. Notice that even if your firm produces no output at all, it still incurs fixed costs.

Figure 8.15 Total Costs, Fixed Costs, and Variable Costs



Fixed costs are the same at all levels of output. Variable costs increase as the quantity of output increases. Total costs equal fixed costs plus variable costs.

The Exit Decision

We are now ready to study the decision to continue in business or exit. You need to compare your revenues, defined as price times quantity, with the cost of running your business. Profit is the difference between revenues and costs, so

profit = total revenues − total costs = total revenues − variable costs − fixed costs.

Table 8.4 Demand and Profit before Walmart Comes to Town

Quantity	Price ($)	Total Revenues ($)	Variable Costs ($)	Fixed Costs ($)	Profits ($)
0	30	0	0	14,000	–14,000
200	29	5,800	2,000	14,000	–10,200
400	28	11,200	4,000	14,000	–6,800
600	27	16,200	6,000	14,000	–3,800
800	26	20,800	8,000	14,000	–1,200
1,000	25	25,000	10,000	14,000	1,000
1,200	24	28,800	12,000	14,000	2,800
1,400	23	32,200	14,000	14,000	4,200
1,600	22	35,200	16,000	14,000	5,200
1,800	21	37,800	18,000	14,000	5,800
2,000	20	40,000	20,000	14,000	6,000
2,200	19	41,800	22,000	14,000	5,800
2,400	18	42,550	24,000	14,000	5,200

The demand for your blue jeans is shown in the first two columns of Table 8.4 "Demand and Profit before Walmart Comes to Town". Looking at this table, your profit is at its highest when you set a price at $20 and sell 2,000 pairs of jeans. In this case, you earn $6,000 per month. Your revenues are enough to cover your variable costs and your fixed operating costs.

After Walmart comes to town, the demand for your jeans shifts inward because shoppers start going to Walmart instead. Now your demand is as shown in Table 8.5 "Demand and Profit after Walmart Comes to Town".

Table 8.5 Demand and Profit after Walmart Comes to Town

Quantity	Price ($)	Revenues ($)	Variable Costs ($)	Fixed Costs ($)	Profit ($)
0	26	0	0	14,000	–14,000
200	25	5,000	2,000	14,000	–11,000
400	24	9,600	4,000	14,000	–8,400
600	23	13,800	6,000	14,000	–6,200
800	22	17,600	8,000	14,000	–4,400
1,000	21	21,000	10,000	14,000	–3,000

Quantity	Price ($)	Revenues ($)	Variable Costs ($)	Fixed Costs ($)	Profit ($)
1,200	20	24,000	12,000	14,000	–2,000
1,400	19	26,600	14,000	14,000	–1,400
1,600	18	28,800	16,000	14,000	–1,200
1,800	17	30,600	18,000	14,000	–1,400
2,000	16	32,000	20,000	14,000	–2,000
2,200	15	33,000	22,000	14,000	–3,000
2,400	14	33,600	24,000	14,000	–4,400

In response to this decrease in demand, you should drop your price. Your profits are now maximized at $18. Unfortunately, at this price, you don't earn enough to cover your fixed costs. Your profits are −$1,200 a month. Should you remain in business? The answer to this question depends on when you ask.

Suppose you ask this question just *after* you have paid your monthly fixed operating costs. During the course of a month, you should stay in business because you are earning enough revenues to cover your variable costs. As soon as it is time to pay your monthly fixed cost again, though, you should choose to exit and close down your store. In this case, you would engage in job destruction by firing your sales clerk.

In this simple example, it is easy to see exactly when and why you should exit. A more general rule for when to exit is as follows. You should exit if

discounted present value of expected future profits < value of recoverable assets.

To make sense of this rule, we need to look at each part of it in turn.

- **Discounted present value.** Our previous example considered only a single month. In fact, you must think about the entire future. This means you must use the tool of **discounted present value** to add up profits earned in different months. [1]
- **Expected future profits.** Even though your price has decreased this month, it might not stay low forever. Perhaps your customers will decide, after they have tried Walmart, that they prefer your store after all. It would then be foolish to close down your store immediately just because you fail to cover fixed costs in one month. This means you must make a decision in the presence of uncertainty: you don't know for sure if your customers will come back, and if they do, you don't know for sure that they will not go away again. As a simple example, suppose you think there is a

75 percent chance that the shift in your demand curve is permanent and a 25 percent chance that it will go back to its original position. Looking ahead and using the tool of **expected value**, you would calculate your expected profit as follows:

$$\text{expected profit}=14\times\$6{,}000+34\times(-\$1{,}200)=\$1{,}500-\$1{,}200=\$300.$$

In this case, you still expect to make a small profit every month. Provided you were not too **risk-averse**, you would keep your store in business. Of course, after some months had gone by, you would probably have better information about whether your customers are truly coming back or not.

- **Value of recoverable assets.** If you are thinking of closing down your store, then you also need to look at your existing assets in the store. You may be able to sell off some of these assets. For example, you could perhaps sell your cash register or computers. We say that such assets are (partially) recoverable assets because you can get back a portion of what you originally paid for these assets.

Toolkit: Section 17.5 "Discounted Present Value", and Section 17.7 "Expected Value"

You can review the meaning and calculation of discounted present value, expected value, and risk-aversion in the toolkit.

Defining Fixed Costs

Our definition of fixed costs seems very straightforward. Unfortunately, it is not always easy to decide in practice whether a cost is fixed or variable. There are two main reasons for this:

1. **Time horizon.** Business planning must be carried out over multiple time horizons. You must decide what to do from one week to the next, from one month to the next, and from one quarter to the next. Costs that are fixed over short time horizons may be variable over longer time horizons. For example, suppose your contract with your employee says you must give her six weeks' notice prior to letting her go. Then her wages are a fixed cost when you are planning for the next six weeks but a variable cost over a longer horizon. Similarly, you may have to lease your store space yearly, in which case that cost is fixed until your lease next comes up for renewal. The bottom line is that whether you think of a cost as fixed or variable depends on your time horizon.

2. **“Lumpiness.”** Some inputs are easier to vary than others. You can freely decide how many pairs of jeans to buy from your wholesaler, so your purchase of jeans to sell is a variable cost. Other inputs are “lumpy”; they are fixed over some ranges of output but variable over others. This means that some costs are fixed over some ranges of output but variable over others. For example, you might be able to sell up to, say, 10,000 pairs of jeans a month in your current store space. However, if you wanted to expand beyond that, you would no longer have enough room to store your inventory and provide an acceptable shopping space for your customers. Because the size of your store is not something you can vary smoothly, this is a lumpy input.

In fact, if we take a very long time horizon and very large ranges of output, there are few costs that are truly fixed.

Entry

We use very similar reasoning to think about a firm’s decision to enter a market. When Walmart’s senior management team contemplates opening a new store, they compare the costs of entry against the (discounted present value of the) profit they expect to earn once they enter.

What are some of Walmart’s costs when it wants to open a new store?

- Searching for a suitable location
- Going through the necessary legal processes in the particular location
- Purchasing the land
- Dealing with public opposition (through lobbying, advertising, sending representatives to town council meetings, etc.)
- Designing the store
- Building the store
- Adjusting their supply chain logistics so as to be able to supply the store
- Hiring people to work there

You can probably think of many more. We call these Walmart’s entry costs.

Toolkit: Section 17.14 "Costs of Production"

Entry costs are the one-time fixed costs that a firm incurs when establishing a business. The toolkit has more discussion of other kinds of costs.

Such costs of establishing a business can be very substantial. Notice, by the way, that entry costs are for the most part truly fixed costs. Walmart must incur these costs before it can let a single customer inside; these costs are fixed no matter how long the time horizon; these costs are largely independent of Walmart's scale of operation.

The senior management team must also predict how much profit they expect the store to make. These forecasts are based on the idea that, once the store is opened, the store will set its prices and manage its operations to maximize its profits. Because the team will be uncertain about profits, they will need to use expected value calculations. They will also be counting on a profit flow for years if not decades and will need to use discounted present value calculations. Thus the appropriate decision rule for a firm is to enter if [2] discounted present value of expected future profits > entry costs.

A firm is more likely to enter if

- the costs of entry are low,
- variable costs are low,
- the revenues from operating are high, and
- demand for its product is very inelastic.

Sunk Costs and Recoverable Costs

Firms that enter markets know that it is possible that they will exit again in the future. Because their profit flow is uncertain, they recognize that there may come a point where they will judge it better to close down their operations. If they close down their operations, they may be able to sell off some of their existing assets. Therefore, when deciding to enter, managers also take into account the extent to which their assets are recoverable. If they are likely to be able to reclaim most of the value of their assets, then entry is more likely to be profitable even if demand turns out to be lower than expected.

Specifically, we can divide entry costs into sunk costs and recoverable costs.

Toolkit: Section 17.14 "Costs of Production"

A sunk cost is a cost that, once incurred, cannot be recovered. A recoverable cost is a cost that, once incurred, can be recovered.

Looking back at our list of Walmart's entry costs, we can see that many of these costs are sunk costs. All the planning and legal fees are completely tied to this store; if they end up not building the store, they cannot get any of the monies back. The building is a sunk cost as well. Other costs are at least partly recoverable. If they decide not to build the store, they can resell the land. If they have equipped the building with shelving and cash registers and then decide not to open the store, they can resell these assets or move them to another Walmart store instead.

Economic reasoning gives clear instructions about sunk costs: they should be irrelevant for any future decision. Whether it was a good or bad idea to build a store, any decisions made going forward should take into account only the future profitability of the store. For example, suppose that Walmart's entry costs were $100 million, of which $30 million were recoverable. Suppose also that Walmart's managers estimated the discounted present value of expected profits at $120 million and therefore decided to build the store. However, once it was built, they discover that they have badly overestimated demand. The managers revise their estimate of future profits by half to $60 million. They now regret having built the store; it was a bad investment. *But they should still keep the store open because it is earning more than they could obtain by closing the store and selling its assets.*

Even though the economic principle is clear, people frequently include sunk costs in their calculations instead of ignoring them. This is such a pervasive problem that it is given the name the sunk cost fallacy.

Toolkit: Section 17.14 "Costs of Production"

The sunk cost fallacy is the mistake of including sunk costs in future-looking decisions, even though they should properly be ignored.

KEY TAKEAWAYS

- A fixed cost is paid regardless of the level of output produced; a variable cost depends on the level of output produced.

- A firm should exit when the discounted present value of its future profits decreases below the value it can receive from selling its assets.
- A sunk cost is a cost that cannot be recovered, such as the cost of entry. This cost should have no effect on the decision to exit.

CHECKING YOUR UNDERSTANDING

1. Go back to our discussion of the data in Table 8.5 "Demand and Profit after Walmart Comes to Town". Explain why you should reduce your price after Walmart arrives in town.
2. Give an example of a fixed cost associated with taking this economics class. Is that cost sunk? How much of it can you recover if you stop taking the class?
3. Suppose the interest rate increases. Explain how that will lead more firms to exit the market. (Hint: think about discounted present value.)

[1] For examples of discounted present value in action, look at Chapter 4 "Life Decisions"and Chapter 9 "Making and Losing Money on Wall Street".

[2] In Chapter 14 "Busting Up Monopolies", you will find an example of very similar economic reasoning. There we present a parallel rule for the situation where a firm is deciding whether or not to engage in innovation.

8.3 Search

LEARNING OBJECTIVES

1. What is the process of matching workers and vacancies?
2. What is the optimal strategy to follow when looking for a job?
3. How are wages determined in labor markets?

As we have seen, job creation and destruction occur because of the entry and exit of firms. Jobs are created when firms enter into an industry and destroyed when firms exit. Job creation and destruction also arise as a result of the hiring and firing decisions of existing firms. We have used the labor market as a device to help us understand these hiring and firing decisions.

If you have ever looked for a job, though, then you know there is more to the labor market than supply and demand. Several aspects of the way labor is traded do not fit neatly into this framework. Workers and firms devote time and money to finding one another: search is an important element of the job market. And wages are often determined by some type of bargaining process, perhaps between a single worker and a firm or between a firm and a union that represents many workers.

Internet Job Search

Internet job searches are now an established part of the way labor markets operate. If you are a worker looking for a job, you can go to a site like Monster.com (http://www.monster.com) or CareerBuilder.com (http://www.CareerBuilder.com) to search for vacancies posted by firms. [1]When you search on one of these sites, you are asked to provide information about the type of job you are looking for by providing the following:

- Keywords (the type of work you want to do)
- Categories (a description of the occupation)
- Location of the job
- Career level

In addition, you provide information about yourself, such as your work experience and education level. The search engine then provides a list of vacancies posted by firms matching these characteristics.

If there are potential matches for you, you are provided with a list of potential employers. Each will typically provide some information about the job. Sometimes this will include a salary range. These postings often include a description of the type of worker the firm is searching for, using phrases such as "team player," "responsible," "leadership skills," or "people skills." The next step is then up to you. Along with the job postings comes information about how to contact the firm. You can indicate your interest to the firm, and you may be called in for an interview. If that goes well, a job offer will follow. At this point, negotiation over compensation comes into play.

Eventually, you must decide whether or not to accept the job. What should you do? If you knew for sure that this was your dream job of a lifetime, the decision would be obvious: accept the job. But life is never that easy. In reality, you face considerable uncertainty over any job you are offered:

- What will the job really be like?
- What other options are there?

The first type of uncertainty has multiple dimensions. No matter how many brochures you read about a job, how many other workers you talk to, or how much time you spend watching someone at work, you still will not know everything about a job until you actually go to work. Even then, there are elements of a job that you will not know about until you have worked for many years. An example is promotion. When you consider a job, you will probably hear about opportunities for advancement if you stay with the firm. But whether or not you will be promoted is something that will be resolved in the future and is part of the uncertainty you face when you think about accepting a job.

The second type of uncertainty concerns the alternatives to the job you are considering. If you had a list of all possible jobs available to you, then you could consult that list and pick the best job. But, of course, there is no such list. Instead of being presented with a list, you have to search for a job. If you turn down the job you are offered today, you will not know for sure what job will be available to you tomorrow. Uncertainty over how to respond to a job offer is very important for some workers but less so for others. The difference is determined by how easy it is to change jobs. We can illustrate the point with two extreme examples.

The first case is a job that offers lifetime employment. If you accept this job, it is yours forever. You will never be fired and—let us suppose—you can never quit either. Given this situation and faced with a job offer, what would you do? Presumably the first thought that comes to your mind is "be careful." You would not accept this job unless you were very sure it was a good match for you. If you are not sure, you should reject the job offer and search more.

The second case is a job that offers very short-term employment, on a week-by-week basis. If you accept this job, you are employed for the week; then you can choose to remain in the job (if it is still available) or leave to search for another one. Also, suppose that during your work time you can still keep an eye out for other jobs. It might be that you can check a computer that displays job ads, look at classified ads in the newspaper, or pass by a few shops advertising job openings during lunch. If you are offered this second type of job, there is no need to be very selective. Your employment is very temporary, and it is easy to change jobs.

The first kind of job is more descriptive of professional positions available to highly skilled workers, where employers are very selective about the type of person they hire. For these types of individuals, searching and changing jobs can be very lengthy and expensive. If they accept a job, it had better be right for them. The second kind of job is one you might be more familiar with as a student—a short-term job such as waiting tables, working as a secretarial temp, or selling in a retail outlet.

Search and Bargaining

The existence of Monster.com and similar search engines makes clear that the trade of labor services is quite different from the trade of, say, US government bonds. The return on a bond is the same regardless of who owns it. But the match between a firm and a worker is special. No two jobs are the same, and no two workers are the same. Also, if you want to buy a US government bond, you can simply call a broker to buy one for you. But if you run a restaurant and want to hire a worker with some very special skills in a particular location, there is no obvious person to call or place to go.

There are three stages of search and bargaining:

1. The *meeting* of workers looking for jobs with firms looking to fill vacancies

2. The *matching* of workers with certain characteristics with jobs requiring certain characteristics
3. The *determination of wages* through a bargaining process

These three elements correspond to the stages you might encounter when you look for a job. First, there is time spent looking for job opportunities. This might involve a recruiting program or search on the Internet. Once you have found a job opening, there is normally a second stage: an interview process. You will typically be interested in the characteristics of the job (such as wages, hours, benefits, promotion possibilities, and job security), and the firm will be interested in your characteristics (such as skills, experience, and trustworthiness). If both you and the firm think that the match is a good one, then the process moves to a third and final stage of bargaining to determine the compensation you will receive.

Searching with a Reservation Wage

We suppose that the bargaining process between a worker and a firm is very simple. The firm makes a take-it-or-leave-it offer. In other words, the firm gives the terms of its offer, including the compensation package and the working conditions, and the worker can then either accept or reject this offer. We also suppose that the offer can be summed up in terms of wages.

To see how this works, imagine that there are two firms each offering jobs at $10 per hour. One firm provides very flexible working hours, while the other requires you to work from 10 p.m. to 6 a.m. and sometimes on the weekend. The first job is evidently more desirable than the second one. If you would be willing to pay $4 an hour for the flexibility of the first job, then it is *as if* the second firm was offering a job at $6 per hour.

Once a worker has a wage offer in hand, should that person accept or reject it? The answer comes from balancing the benefits of having a job (and therefore a wage) right now versus waiting for another job to come along in the hope that it will pay a higher wage. To see how a worker would make this choice, here is a simple numerical example.

- There are only two possible wage offers: some firms offer $500 per week, and some offer $1,000 per week.

- It takes a week to search for a job. If a worker turns down the offer he gets this week, he will not get another offer until next week.
- If a worker accepts a job, he cannot then search for another job.
- The government offers unemployment insurance of $300 per week.

Suppose a worker gets a job offer of $1,000 per week. Then the decision is easy: accept that job. The more difficult case is when the worker gets an offer of $500 per week. By accepting this job, the worker gets $200 more than with unemployment insurance. But accepting the job also has an opportunity cost. It means that the worker loses out on the chance of getting the higher paying job next week. So what should the worker do?

If you think about this problem, you will probably realize that the answer depends on how likely the worker is to get the better job by waiting. If most of the available jobs are the ones that pay $1,000 per week, then it is likely to be worth waiting. On the other hand, if most of the jobs pay only $500 a week, then the worker might have to wait a long time for the better job, so it is likely better to accept the one that pays $500 a week. More generally, in a world where there are lots of different jobs paying lots of different wages, the best thing for the worker to do is to pick something we call the reservation wage. Workers can follow this rule:

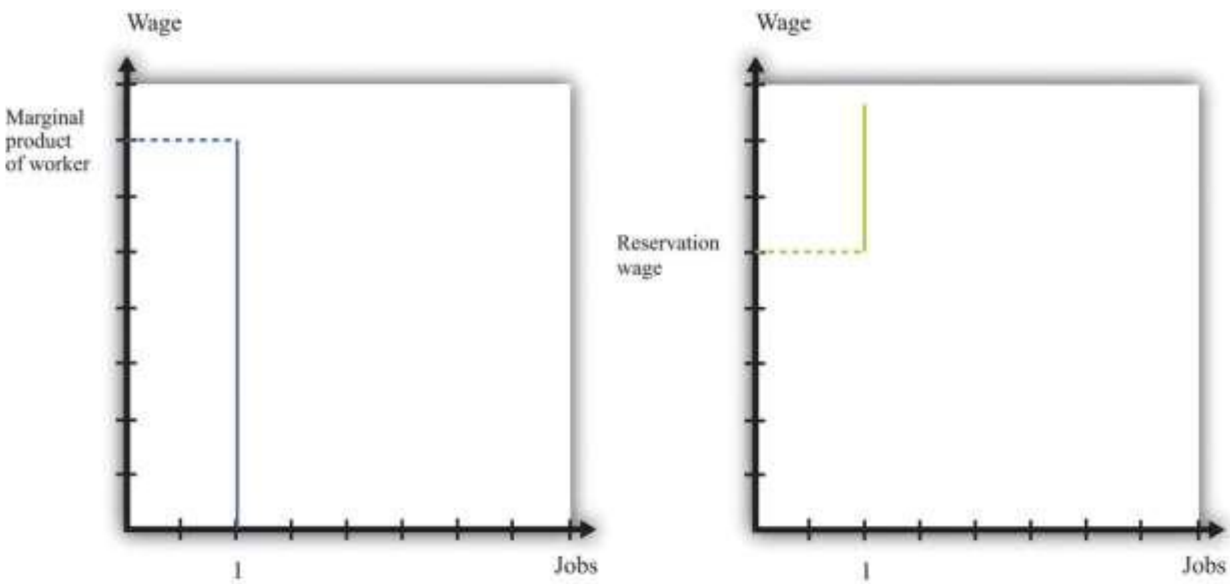
- Accept a job if it offers a wage above the reservation wage.
- Reject a job if it offers a wage below the reservation wage.

Bargaining

If a worker and a firm meet and determine that the match is good, then they proceed to determine wages. There are two ways in which this might happen. [2]One possibility is the one we just discussed. Firms post vacant jobs and at the same time advertise the wage. If a worker qualifies for the job, then that worker will accept the job if the wage exceeds the reservation wage.

There can also be bargaining between a worker and a firm. A firm will make a profit based on the difference between the marginal product of hiring the extra worker and wages paid to the worker. So a firm will choose to hire the worker as long as the wage is below the marginal product of labor. This is a firm's valuation of the job. A worker will be willing to take a job as long as the wage exceeds the reservation wage. This is a worker's valuation of the job. In Figure 8.16 "The Valuation of a Job" we show both.

Figure 8.16 The Valuation of a Job



A firm follows the decision rule: "Offer the job if a worker's marginal product exceeds the wage." A worker follows the decision rule: "Accept the job if the wage exceeds the reservation wage."

As long as the marginal product of labor for a worker is higher than a worker's reservation wage, there is something to gain by employing a worker. The match is potentially a good one. But how will these gains be split? The answer depends on the relative bargaining power of a worker and a firm, which in turn depends on the information that they each possess.

As an example, suppose that a firm knows a worker's reservation wage and can make a take-it-or-leave-it offer. It would then offer a real wage slightly above the worker's reservation wage. The worker would accept the job, and all the surplus from the employment relationship would flow to the firm. At the other extreme, suppose that a worker knows his marginal product at the firm, so the worker can make a take-it-or-leave-it offer. Then the worker will offer to work at a real wage slightly below his marginal product. The firm would accept the offer, and all the surplus of the employment relationship would flow to the worker.

Another source of bargaining power for a worker is the other options available to him. If a worker comes into a negotiation with a good job offer from another firm in hand, then this will increase his reservation wage.

Workers can also enhance their bargaining power by negotiating together. Firms typically have many workers. Sometimes these workers group together, form a union, and bargain jointly with a firm. When workers organize in this way, they generally have more bargaining power because they can threaten to strike and shut down the firm. In this way, unionized workers get more of the surplus from their jobs.

In reality, workers don't know their exact marginal product and firms don't know the exact reservation wage of their workers. Not surprisingly, bargaining in such situations is more complicated to analyze. Sometimes, gains from trade are not realized. Suppose, for example, that a worker's reservation wage is below his marginal product. But a firm, thinking that a worker's reservation wage is really low, makes the worker a very low wage offer. If this wage offer is below the worker's reservation wage, then the worker will decline the offer and search again—despite the fact that there were gains to trade. Unfortunately, private information prevented the firm and the worker from realizing these gains.

Posting Vacancies

The final element in the search process is the vacancies posted by firms. You can see these vacancies on Monster.com, in the newspaper, and in magazines. Vacancies are costly to post, and it is expensive to evaluate workers. They are the analogy on a firm's side to the costly search on a worker's side.

You can think of a firm's decision of posting vacancies as being very similar to the labor demand for a firm. Firms want to expand output and thus post more vacancies whenever the marginal revenue of selling an extra unit of output increases relative to marginal cost. This could happen, for example, if the demand curve faced by a firm shifts outward. To expand output, the firm needs to hire more workers. It does so by posting vacancies, interviewing workers, and eventually bargaining with the best qualified ones to fill the open positions.

We noted earlier that labor demand also depends on wages: as the real wage decreases, a firm's real marginal cost decreases, so it will want to hire more workers and expand output. When we think of search and bargaining, say through Monster.com, there is no "market wage" that a firm takes as given. Instead, the wage comes from a bargain between a worker and the firm. But the wage that is eventually agreed on will depend on the outside options of workers and firms. As the prevailing wage in the market decreases, a firm will be able to hire workers at a lower wage and will choose to post more vacancies and expand its workforce.

KEY TAKEAWAYS

- The search process brings together workers and vacancies of firms. This process lies behind the supply and demand curves for labor.
- For many searches, it is best to follow a reservation wage strategy: accept a job if and only if the wage exceeds the reservation wage.
- Wages are determined through a bargaining process. Sometimes this is through a take-it-or-leave-it offer of the firm. Often there is bargaining between a firm and its workers (or their union) to share the surplus of the employment relationship.

CHECKING YOUR UNDERSTANDING

1. If a worker becomes very impatient, what will happen to his reservation wage?
2. When Walmart comes to town, what will happen to the vacancies posted by competing stores? What will happen to the reservation wages of salespeople?

[1] The Department of Labor sponsors a website (http://www.careeronestop.org) filled with information, including compensation levels, for different occupations. Help with job search is available here as well.

[2] If you read Chapter 5 "eBay and craigslist", you will see some close parallels between the mechanisms we discuss here and the ways in which a buyer and a seller may agree on a price.

8.4 Government Policies

LEARNING OBJECTIVES

1. What government policies impact job creation and destruction?
2. What are the effects of trade on job creation and destruction?

Governments are very interested in job creation. A political leader whose economy loses a large number of jobs without creating new ones is unlikely to be reelected. On a local level, state and local governments compete fiercely to have firms locate in their region by offering lucrative tax reductions. This is seen as a way to create jobs in the local economy. We now examine some of these policy interventions and trace out their implications, focusing on three policies: restrictions on closing plants, policies that promote small businesses, and trade policies.

Plant-Closing Restrictions

In the United States, if you want to close a factory, you do not have to have approval of the government, but an act of Congress—the Worker Adjustment and Retraining Notification Act [1]—requires you to announce your intentions ahead of time. According to the US Department of Labor, "The Worker Adjustment and Retraining Notification Act (WARN) protects workers, their families, and communities by requiring employers to provide notification 60 calendar days in advance of plant closings and mass layoffs." [2]

This law was passed in 1988 during a time of higher than average unemployment in the United States. Similar restrictions apply in some European countries, such as Spain and France. You cannot simply close an unproductive plant; employees must be given advance notice, and government approval may be required. Such restrictions on plant closings are intended to reduce job destruction. After all, if you make something more expensive to do, then less of it will be done. If it becomes more expensive to close plants, then fewer jobs will be destroyed by the exit of plants.

Economists point out, however, that the incentives of such policies are complicated and go beyond the effects on job destruction. To see why, think about our earlier discussion of entry and exit. When a firm decides to enter into an industry, it compares the profit flow from operating to the entry cost. When a firm thinks about the profits it will earn if it enters, it recognizes that if the demand for its product disappears, it can exit and thus avoid periods of negative profits. But if you take this option to exit a market away from a firm, then the value of entering an industry will decrease. Fewer firms will enter, and fewer jobs will be created. Thus laws that make it costly to close plants will also reduce job creation. The effect on *net* job creation is unclear.

Small Business Promotion

Started in the 1950s, the Small Business Administration (SBA;http://www.sba.gov) is a US government agency whose goal is to protect and promote small businesses. Small firms obtain preferential treatment in terms of taxes, regulation, and other policies. Part of the argument in favor of promoting and protecting small businesses is the view that job creation is centered on these firms. According to the SBA website, small businesses [3]

- represent 99.7 percent of all employer firms and
- have generated 60–80 percent of net new jobs annually over the last decade.

So it appears that small businesses are critical to an economy.

We must remember that these are indeed small firms, however. Suppose there were 5 firms in an economy. Four of them have 2 workers, and the fifth has 92 workers. The typical firm then has 2 workers: 80 percent of the firms in this economy have 2 or fewer workers. From the perspective of workers, though, things are rather different. Ninety-two percent of the workers are employed by the single large firm. If you ask workers how many employees are in their firm, the typical worker will say 92. *Most firms have few workers, but most workers are employed by the large firm*.

This is not far from the reality of the US economy, where much economic activity (employment and output) is centered on relatively few firms. A recent study of about 5.4 million businesses found that 182,000 of them operate multiple units. Dividing 182,000 by 5.4 million, we learn that these larger firms are less than 4 percent of the total number of firms. But they account for about 61 percent of the revenue of the business sector of the economy. So most firms are relatively small, but those that are large are huge compared to the rest. [4]

Davis, Haltiwanger, and Schuh point out that "large firms and plants dominate the creation and destruction of jobs in the manufacturing sector." [5] Larger firms and plants both destroy and create more jobs. For example, at a job destruction rate of 10 percent a year, a small firm of 50 workers will destroy 5 jobs, while a large plant with 1,000 workers will destroy 100 jobs. So even if the job creation and destruction *rates* are higher for small firms, this does not necessarily mean that these small entities create and destroy more *jobs* than large firms do.

This is not to say that small firms are unimportant. Most of the large firms in the economy started small. Likewise, all the older, more successful firms were once young firms. So any impact the SBA has on either small or young firms will influence these firms as they age and grow. However, the rationale for the SBA is not completely clear. Normally, government interventions are based on the idea of either correcting some problem with the operation of free markets or redistributing resources. It is not clear whether the SBA fulfills either of these roles.

Trade Policy

Job creation and destruction are also affected by things that happen outside US borders. The removal of trade barriers allows countries to benefit more fully from the gains from trade. But in the process, some jobs are destroyed, while others are created.

Job destruction frequently takes center stage during debates on trade policy. In the early 1990s, for example, the United States was contemplating a reduction in trade barriers with its neighbors—Canada and Mexico—through negotiation of the North American Free Trade Agreement (NAFTA). Ross Perot, a third-party candidate for the US presidency in 1992 and 1996, was extremely critical of NAFTA. His focus was on job destruction, and he was famous for forecasting "a giant sucking sound" as employment opportunities moved from the United States to Mexico in response to NAFTA.

The loss of some jobs from a reduction in trade barriers is part of the adjustment one would expect. For countries to reap the gains from trade brought about by the removal of trade barriers, production patterns across countries must change. That process leads to job destruction and creation. Firms that used to produce certain goods in one country exit, as firms in other countries start to produce those goods instead. Workers at the exiting firms will certainly lose their jobs, but other jobs are created in the economy at the same time.

NAFTA was implemented in January 1994. More than 15 years later, it is still difficult to say exactly what the effects were and will be of NAFTA. Economics is not a laboratory science. It is not possible to subject the economies of Canada, Mexico, and the United States to this reduction in trade barriers, holding everything else the same. Instead, we have to look at data from before and after 1994 to try to infer the effects of NAFTA. But of course many other economic factors have also changed over this period. In parts of the United States where manufacturing jobs have been lost over the last 15 years, there is a tendency to hold NAFTA responsible. In fact, there is little evidence that NAFTA led to net job destruction.

What *has* happened over the last decade is that the US manufacturing sector has been exposed to increased competition from other countries, most notably China. It is this trade that has had a bigger impact on US manufacturing. At the same time, this has meant that NAFTA has been less of a success story for the *Mexican* economy than was predicted and hoped, as US consumers have purchased very cheap goods from China rather than Mexico.

KEY TAKEAWAYS

- In the United States, firms are able to close plants if they choose to do so. This is not the case in all countries.
- The government promotes small businesses, viewing them as a source of job creation.
- The reduction of trade barriers creates new jobs and destroys others.

CHECKING YOUR UNDERSTANDING

1. If plants, once opened, were never allowed to be closed, what would this do to the incentives of a firm to open a plant?
2. In your college classes, what is the analogue of the statement that *most firms have few workers, but most workers are employed by the large firm*? Is it that most classes are small, but most students are in large classes?

[1] US Department of Labor, "Other Workplace Standards: Notices for Plant Closings and Mass Layoffs," elaws Employment Law Guide, accessed March 14, 2011,http://www.dol.gov/compliance/guide/layoffs.htm.

[2] US Department of Labor, *The Worker Adjustment and Retraining Notification Act (WARN)*, accessed January 22, 2011, http://www.dol.gov/compliance/laws/comp-warn.htm.

[3] See http://www.sba.gov/sites/default/files/files/leg_priorities112th.pdf.

[4] Steven J. Davis et al., "Measuring the Dynamics of Young and Small Business: Integrating the Employer and Non-employer Universes" (NBER Working Paper 13226, 2007), accessed January 30, 2011, http://www.nber.org/papers/w13226.

[5] See Steven J. Davis, John C. Haltiwanger, and Scott Schuh, *Job Creation and Destruction*(Boston, MA: MIT Press, 1998), chap. 7, sect. 4.

8.5 End-of-Chapter Material

In Conclusion

During election season in the United States, the adverse effects of trade on jobs are often talked about extensively. In the 2008 Michigan presidential primary election, for example, candidates offered different ways in which they claimed they would help the automobile industry and bring more jobs to the Michigan economy. In the South Carolina primary in the same year, job losses in the textile industry received lots of attention. One large textile manufacturer in the region, Swift, had been cutting jobs steadily and was closing up, apparently planning to move production to South America. Individuals who lost their jobs due to this closing reported that they experienced a period of unemployment as they searched for a new job. Some found new jobs, either in an automobile assembly plant or working on optic fibers. Others moved from working in manufacturing to services. This is job creation and destruction in action. It happens all the time, all across the world.

Viewed abstractly, job creation and destruction are healthy processes for an economy. Through the process of job creation and destruction, workers are induced to move from less productive to more productive jobs. Such movement enhances the overall productivity of an economy. From the perspective of individual workers, however, the process looks very different. Job destruction means that people lose their jobs, which are a source of income and perhaps also of pride and dignity. They may have to spend some period of time unemployed, and they may lose important benefits, such as health insurance. They may have to relocate in search of jobs that are being created elsewhere; such relocation can be difficult and costly.

In sum, although the productivity of an economy as a whole may increase, this need not translate into improvements for workers who lose their jobs. Some find higher-paying jobs, but others, particularly those with few skills, see their wages decrease. One of the big challenges faced by governments and policymakers is to encourage the efficient reallocations of workers while minimizing the individual hardships that workers confront.

Key Links

- US Census Bureau website on employment dynamics:http://lehd.did.census.gov/led/index.html

- Bureau of Labor Statistics data on the labor force characteristics:http://www.bls.gov/cps
- Facts about working for Walmart:http://walmartstores.com/pressroom/factsheets
- Information about NAFTA:http://www.export.gov/fta/NAFTA/index.asp?dName=NAFTA
- Department of Labor on WARN:http://www.dol.gov/compliance/laws/comp-warn.htm

EXERCISES

1. A statistic called "unit labor cost," which is the cost per unit of output of the labor input, is often calculated. How is this different from marginal cost?
2. Would you think that a firm's managers would have a different viewpoint than its workers on whether or not a plant should be shut down?
3. All else being the same, which firms would you expect to be more capital intensive—those with labor contracts that pay high wages or those with easy access to funds on capital markets?
4. How is the labor demand curve of a firm influenced by the cost of other inputs, such as energy?
5. If a firm operates with high fixed costs, should it set a higher price for its output to be able to cover those fixed costs?
6. Would a firm ever remain in business even though it is earning negative profits in the current year? How does this decision depend on the interest rate?
7. Besides labor, what other markets can you think of where search is important?
8. All else being the same, who will have a higher reservation wage—someone who can receive unemployment insurance for 13 weeks or someone who can receive unemployment insurance for 26 weeks?
9. What do you think is the role of "friends" in helping you find your first job? What about subsequent jobs?
10. In economic downturns, what happens to the ratio of unemployed workers to vacancies?
11. Why do people quit jobs? Do you think that the number of job quitters is higher or lower during economic downturns??
12. In many European countries, it is very difficult for a firm to close one of its plants. What might be the effects of an increase in the cost of closing a plant on job creation, destruction, and reallocation?
13. Suppose that the establishment of a Walmart in a nearby town led to the creation of 100 jobs in that store and the destruction of 150 retail jobs in the town. Is the net loss of 50 jobs enough reason to

oppose the opening of the Walmart? What other benefits might a Walmart bring? What other costs might it impose?

14. What might be the effects of a reduction in child-care costs on an unemployed worker's reservation wage?

Spreadsheet Exercise

1. Revisit Table 8.3 "Monthly Costs of Production" through Table 8.5 "Demand and Profit after Walmart Comes to Town". Suppose your fixed operating costs were $12,000 instead of $14,000. Redo the tables with this change. Should you stay in business after Walmart arrives? Explain.

Economics Detective

1. The Bureau of Labor Statistics produces data on labor turnover called JOLTS (http://www.bls.gov/jlt). Using that information, create a table and a plot of data to illustrate what has happened to job openings and the quit rate since January 2007. How would you explain these findings?

Chapter 9
Making and Losing Money on Wall Street
Financial Roller Coasters

In 2006 and 2007, the financial district in Shanghai, China, was in a frenzy.Figure 9.1 "Shanghai Stock Exchange Index" shows the value of stocks in that market since its inception in 2000. Starting in early 2006, the value of stocks traded on this market exploded. The market rose by 130 percent in 2006; by May 2007, it was up over 50 percent for that year. [1] The market peaked in late 2007 and is currently at about 50 percent of that value. A lot of money was made by those who invested in the Shanghai market. And unfortunately a lot of money was lost.

Figure 9.1 Shanghai Stock Exchange Index

This figure shows the closing prices on the Shanghai stock exchange between January 2000 and April 2010. Stock prices rose rapidly from 2006, peaking in October 2007, but decreased substantially over the next year. In late 2009, stocks rebounded again.

These gains attracted many investors. Funds from abroad poured into Shanghai. The savings accounts of Chinese households were another source of investment funds. From a *People's Daily* Internet article posted on May 13, 2007, we learn the following: "More than 70 billion yuan (9.1 billion U.S. dollars) was transferred from savings accounts in Shanghai to stock trading accounts in the first four months of this year, the Shanghai branch of the People's Bank of China estimated on Saturday. In April alone, [savings deposits denominated in Chinese currency] with Chinese banking institutions decreased by 8.5 billion yuan (1.1 billion U.S. dollars)." [2]

During May 2007, stories circulated about households spending many hours carefully evaluating individual stocks and market returns. At the same time, it appeared that many relatively uninformed individuals were simply betting on the market, gambling on a quick return.

We said that some investors made money in the Shanghai market. Does that mean there is a lot of money *to be made* by investing in that market? These phrases sound similar but mean very different things. It is one thing to look back at a market and say you could have made money investing in that market. It is quite another to forecast that you will be able to make a high return in a market in the future. Investors who were attracted to the market in late 2007 had a very different experience: they lost a lot of money. Those who came into market in late 2008 were again able to profit as the market value rose over the following year.

In this chapter, we study the markets for different kinds of assets. Assets include stocks—such as are traded in Shanghai, on Wall Street, and in other financial centers around the globe—but, as we will see, there are many other kinds of assets as well. Information on assets is easy to obtain. If you open almost any newspaper, the business section contains an enormous amount of detailed information on stocks sold in a variety of markets. That same section will contain information on bonds, which are another type of frequently traded asset. Part of our interest in this chapter is defining these assets more precisely. The terms *stocks* and *bonds* are used commonly, but we want to understand exactly what these assets are and how they are traded.

As we wrote this chapter, we had no idea whether we, too, should be putting our personal savings in the Shanghai stock exchange or in some other market around the globe. In the middle of 2007, it looked as if the surge in the Shanghai market was over. Market participants were concerned that the time of high gains had ended. Yet by November 2007, market values had again started to escalate. And then, as we said, the market peaked in late 2007 and decreased rapidly for the next year. This is part of the story of asset markets. They are extremely volatile and unpredictable. When you see these high returns in Shanghai and other markets, you might wonder:

"Can I get rich by trading stocks and bonds?"

Road Map

This chapter begins with a walk down a fictionalized Wall Street, where we describe various kinds of assets. We focus mainly on financial markets, although we will look at other assets as well. Financial markets are familiar to many of us from the financial pages of newspapers or reports on the evening news. Such markets provide a link between borrowers and lenders (Figure 9.2). Many of us are borrowers from banks, perhaps because we have a student loan, a car loan, or a mortgage for a house. Much of what we borrow from banks comes from deposits placed in banks by other households. Firms also borrow in the financial markets. They issue stock and sell bonds in financial markets to finance their investment in new factories and machines.

Figure 9.2

Financial markets link borrowers and lenders.

We then turn to a discussion of the *pricing* of assets. We begin by thinking about an unusual asset: a fruit tree. A fruit tree gives us a certain amount of fruit each year, and the value of the tree depends on the value of the fruit it produces. We explain how to calculate the *value* of a fruit tree that lives for several years and yields an uncertain crop, and we show how exactly the same principles apply to the valuation of stocks, bonds, houses, and other assets. Finally, we explain why—if financial markets are functioning well—the *price* of an asset will equal its value.

Finally, we ask whether it is easy to make money by trading assets. We explain that the gains and losses from trading assets are based on two factors: (1) luck and (2) the skill of investors who quickly recognize profit opportunities before others notice these opportunities. If financial markets are functioning well, then it is very difficult for the casual investor to make money consistently by trading financial assets. And even if—as many believe—financial markets do not function perfectly, this still does not mean that there is easy money to be made.

[1] See http://finance.yahoo.com/q/hp?s=000001.SS&=00&=1&=2006&=11&=31&=2007&=d&=66&=330.

[2] "Chinese Pour Savings Deposits into Stock Market," *People's Daily*, May 13, 2007, accessed March 14, 2011,http://english.peopledaily.com.cn/200705/13/eng20070513_374113.html.

9.1 A Walk Down Wall Street

LEARNING OBJECTIVES

1. What are the different types of assets traded in financial markets?
2. What can you earn by owning an asset?
3. What risks do you face?

Wall Street in New York City is the financial capital of the United States. There are other key financial centers around the globe: Shanghai, London, Paris, Hong Kong, and many other cities. These financial centers are places where traders come together to buy and sell assets. Beyond these physical locations, opportunities for trading assets abound on the Internet as well.

We begin the chapter by describing and explaining some of the most commonly traded assets. Ownership of an asset gives you the right to some future benefit or a stream of benefits. Very often, these benefits come in the form of monetary payments; for example, ownership of a stock gives you the right to a share of a firm's profits. Sometimes, these benefits come in the form of a flow of services: ownership of a house gives you the right to enjoy the benefits of living in it.

Stocks

One of the first doors you find on Wall Street is called the *stock exchange*. The stock exchange is a place where—as the name suggests—stocks are bought and sold. A stock (or share) is an asset that comes in the form of (partial) ownership of a firm. The owners of a firm's stock are called the shareholders of that firm because the stock gives them the right to a share of the firm's profits. More precisely, shareholders receive payments whenever the board of directors of the firm decides to pay out some of the firm's profits in the form of dividends.

Some firms—for example, a small family firm like a corner grocery store—are privately owned. This means that the shares of the firm are not available for others to purchase. Other firms are publicly traded, which means that anyone is free to buy or sell their stocks. In many cases, particularly for large firms such as Microsoft Corporation or Nike, stocks are bought and sold on a minute-by-minute basis. You can find information on the prices of publicly traded stocks in newspapers or on the Internet.

Stock Market Indices

Most often, however, we hear not about individual stock prices but about baskets of stocks. The most famous basket of stocks is called the *Dow Jones Industrial Average* (DJIA). Each night of the week, news reports on the radio and television and newspaper stories tell whether the value of the DJIA increased or decreased that day. The DJIA is more than a century old—it started in 1896—and is a bundle of 30 stocks representing some of the most significant firms in the US economy. Its value reflects the prices of these stocks. Very occasionally, one firm will be dropped from the index and replaced with another, reflecting changes in the economy. [1]

Figure 9.3 "The DJIA: October 1928 to July 2007" shows the Dow Jones Industrial Average from 1928 to 2011. Over that period, the index rose from about 300 to about 12,500, which is an average growth rate of about 4.5 percent per year. You can see that this growth was not smooth, however. There was a big decrease at the very beginning, known as the stock market crash of 1929. There was another very significant drop in October 1987. Even though the 1929 crash looks smaller than the 1987 decrease, the 1929 crash was much more severe. In 1929, the stock market lost about half its value and took many years to recover. In 1987, the market lost only about 25 percent of its value and recovered quite quickly.

One striking feature of Figure 9.3 "The DJIA: October 1928 to July 2007" is the very rapid growth in the DJIA in the 1990s and the subsequent decrease around the turn of the millennium. The 1990s saw the so-called Internet boom, when there was a lot of excitement about new companies taking advantage of new technologies. Some of these companies, such as Amazon, went on to be successful, but most others failed. As investors came to recognize that most of these new companies would not make money, the market fell in value. There was another rise in the market during the 2000s, followed by a substantial fall during the global financial crisis that began around 2008. Very recently, the market has recovered again.

If these ups and downs in the DJIA were predictable, it would be easy to make money on Wall Street. Suppose you knew the DJIA would increase 10 percent next month. You would buy the stocks in the average now, hold them for a month, and sell them for an easy 10 percent profit. If you knew the DJIA would decrease next month, you could still make money. If you currently owned DJIA stocks, you could sell them and then buy them back after the price decreased. Even if you don't own these stocks right now, there is still a way of selling first and buying later. You can sell (at today's high price) a promise to deliver the stocks in a month's time. Then you buy the stocks after the price has decreased. This is called a *forward sale*. If this sounds as if it is too easy a way to make money, that's because it is. The ups and downs in the DJIA are not perfectly predictable, so there are no easy profit opportunities of the kind we just described. We have more to say about this later in the chapter.

Although the DJIA is the most closely watched stock market index, many others are also commonly reported. The *Standard and Poor's 500* (S&P 500) is another important index. As the name suggests, it includes 500 firms, so it is more representative than the DJIA. If you want to understand what is happening to stock prices in general, you are better off looking at the S&P 500 than at the DJIA. The *Nasdaq* is another index, consisting of the stocks traded in an exchange that specializes in technology-based firms.

We mentioned earlier that the DJIA has increased by almost 5 percent per year on average since 1928. On the face of it, this seems like a fairly respectable level of growth. Yet we must be careful. The DJIA and other indices are averages of stock prices, which are measured in dollar terms. To understand what has happened to the stock market in real terms, we need to adjust for inflation. Between 1928 and 2007, the price level rose by 2.7 percent per year on average. The average growth in the DJIA, adjusted for inflation, was thus 4.8 percent − 2.7 percent = 2.1 percent.

The Price of a Stock

As a shareholder, there are two ways in which you can earn income from your stock. First, as we have explained, firms sometimes choose to pay out some of their income in the form of dividends. If you own some shares and the company declares it will pay a dividend, either you will receive a check in the mail or the company will automatically reinvest your dividend and give you extra shares. But there is no guarantee that a company will pay a dividend in any given year.

The second way you can earn income is through capital gains. Suppose you own a stock whose price has gone up. If that happens, you can—if you want—sell your stock and make a profit on the difference between the price you paid for the stock and the higher price you sold it for. Capital gains are the income you obtain from the increase in the price of an asset. (If the asset decreases in value, you instead incur a *capital loss*.)

To see how this works, suppose you buy, for $100, a single share of a company whose stock is trading on an exchange. In exchange for $100, you now have a piece of paper indicating that you own a share of a firm. After a year has gone by, imagine that the firm declares it will pay out dividends of $6.00 per share. Also, at the end of the year, suppose the price of the stock has increased to $105.00. You decide to sell at that price. So with your $100.00, you received $111.00 at the end of the year for an annual return of 11 percent:

$$\$106.00+\$5.00\$100.00=0.11=11\%.$$

(We have used the term *return* a few times. We will give a more precise definition of this term later. At present, you just need to know that it is the amount you obtain, in percentage terms, from holding an asset for a year.)

Suppose that a firm makes some profits but chooses not to pay out a dividend. What does it do with those funds? They are called *retained earnings* and are normally used to finance business operations. For example, a firm may take some of its profits to build a new factory or buy new machines. If a firm is being managed well, then those expenditures should allow a firm to make higher profits in the future and thus be able to pay out more dividends at a later date. Presuming once again that the firm is well managed, retained earnings should translate into extra dividends that will be paid in the future.

Furthermore, if people expect that a firm will pay higher dividends in the future, then they should be willing to pay more for shares in that firm today. This increase in demand for a firm's shares will cause the share price to increase. So if a firm earns profits but does not pay a dividend, you should expect to get some capital gain instead. We come back to this idea later in the chapter and explain more carefully the connection between a firm's dividend payments and the price of its stock.

The Riskiness of Stocks

Figure 9.3 "The DJIA: October 1928 to July 2007" reminds us that stock prices decrease as well as increase. If you choose to buy a stock, it is always possible its price will fall, in which case you suffer a capital loss rather than obtain a capital gain. The riskiness of stocks comes from the fact that the underlying fortunes of a firm are uncertain. Some firms are successful and earn high profits, which means that they are able to pay out large dividends—either now or in the future. Other firms are unsuccessful through either bad luck or bad management, and do not pay dividends. Particularly unsuccessful firms go bankrupt; shares in such a firm become close to worthless. When you buy a share in a firm, you have the chance to make money, but you might lose money as well.

Bonds

Wall Street is also home to many famous financial institutions, such as Morgan Stanley, Merrill Lynch, and many others. These firms act as the financial intermediaries that link borrowers and lenders. If desired, you could use one of these firms to help you buy and sell shares on the stock exchange. You can also go to one of these firms to buy and sell bonds. A bond is a promise to make cash payments (the coupon) to a bondholder at predetermined dates (such as every year) until the maturity date. At the maturity date, a final payment is made to a bondholder. Firms and governments that are raising funds issue bonds. A firm may wish to buy some new machinery or build a new plant, so it needs to borrow to finance this investment. Or a government might issue bonds to finance the construction of a road or a school.

The easiest way to think of a bond is that it is the asset associated with a loan. Here is a simple example. Suppose you loan a friend $100 for a year at a 6 percent interest rate. This means that the friend has agreed to pay you $106 a year from now. Another way to think of this agreement is that you have *bought*, for a price of $100, an asset that entitles you to $106 in a year's time. More generally (as the definition makes clear), a bond may entitle you to an entire schedule of repayments.

The Riskiness of Bonds

Bonds, like stocks, are risky.

- The coupon payments of a bond are almost always specified in dollar terms. This means that the real value of these payments depends on the **inflation rate** in an economy. Higher inflation means that the value of a bond has less worth in real terms.

- Bonds, like stocks, are also risky because of the possibility of bankruptcy. If a firm borrows money but then goes bankrupt, bondholders may end up not being repaid. The extent of this risk depends on who issues the bond. Government bonds usually carry a low risk of bankruptcy. It is unlikely that a government will default on its debt obligations, although it is not impossible: Iceland, Ireland, Greece, and Portugal, for example, have recently been at risk of default. In the case of bonds issued by firms, the riskiness obviously depends on the firm. An Internet start-up firm operated from your neighbor's garage is more likely to default on its loans than a company like the Microsoft Corporation. There are companies that evaluate the riskiness of firms; the ratings provided by these companies have a tremendous impact on the cost that firms incur when they borrow.

Inflation does not have the same effect on stocks as it does on bonds. If prices increase, then the fixed nominal payments of a bond unambiguously become less valuable. But if prices increase, firms will typically set higher nominal prices for their products, earn higher nominal profits, and pay higher nominal dividends. So inflation does not, in and of itself, make stocks less valuable.

Toolkit: Section 17.8 "Correcting for Inflation"

You can review the meaning and calculation of the inflation rate in the toolkit.

One way to see the differences in the riskiness of bonds is to look at the cost of issuing bonds for different groups of borrowers. Generally, the rate at which the US federal government can borrow is much lower than the rate at which corporations borrow. As the riskiness of corporations increases, so does the return they must offer to compensate investors for this risk.

Real Estate and Cars

As you continue to walk down the street, you are somewhat surprised to see a real estate office and a car dealership on Wall Street. (But this is a fictionalized Wall Street, so why not?) Real estate is another kind of asset. Suppose, for example, that you purchase a home and then rent it out. The rental payments you receive are analogous to the dividends from a stock or the coupon payments on a bond: they are a flow of money you receive from ownership of the asset.

Real estate, like other assets, is risky. The rent you can obtain may increase or decrease, and the price of the home can also change over time. The fact that housing is a significant—and risky—financial asset became apparent in the global financial crisis that began in 2007. There were many aspects of that crisis, but an early trigger of the crisis was the fact that housing prices decreased in the United States and around the world.

If you buy a home and live in it yourself, then you still receive a flow of services from your asset. You don't receive money directly, but you receive money indirectly because you don't have to pay rent to live elsewhere. You can think about measuring the value of the flow of services as rent you are paying to yourself.

Our fictional Wall Street also has a car dealership—not only because all the financial traders need somewhere convenient to buy their BMWs but also because cars, like houses, are an asset. They yield a flow of services, and their value is linked to that service flow.

The Foreign Exchange Market

Further down the street, you see a small store listing a large number of different three-letter symbols: BOB, JPY, CND, EUR, NZD, SEK, RUB, SOS, ADF, and many others. Stepping inside to inquire, you learn that that, in this store, they buy and sell foreign currencies. (These three-letter symbols are the currency codes established by the International Organization for Standardization (http://www.iso.org/iso/home.htm). Most of the time, the first two letters refer to the country, and the third letter is the initial letter of the currency unit. Thus, in international dealings, the US dollar is referenced by the symbol *USD*.)

Foreign currencies are another asset—a simple one to understand. The return on foreign currency depends on how the exchange rate changes over the course of a year. The (nominal) **exchange rate** is the price of one currency in terms of another. For example, if it costs US$2 to purchase €1, then the exchange rate for these two currencies is 2. An exchange rate can be looked at in two directions. If the dollar-price of a euro is 2, then the euro price of a dollar is 0.5: with €0.5, you can buy US$1.

Suppose that the exchange rate this year is US$2 to the euro, and suppose you have US$100. You buy €50 and wait a year. Now suppose that next year the exchange rate is US$2.15 to the euro. With your €50, you can purchase US$107.50 (because US$(50 × 2.15) = US$107.50). Your return on this asset is 7.5 percent. Holding euros was a good investment because the dollar became less valuable relative to the euro. Of course, the dollar might increase in value instead. Holding foreign currency is risky, just like holding all the other assets we have considered. [2]

The **foreign exchange market** brings together suppliers and demanders of different currencies in the world. In these markets, one currency is bought using another. The law of demand holds: as the price of a foreign currency increases, the quantity demanded of that currency decreases. Likewise, as the price of a foreign currency increases, the quantity supplied of that currency increases. Exchange rates are determined just like other prices, by the interaction of supply and demand. At the equilibrium exchange rate, the quantity of the currency supplied equals the quantity demanded. Shifts in the supply or demand for a currency lead to changes in the exchange rate.

Toolkit: Section 17.20 "Foreign Exchange Market"

You can review the foreign exchange market and the exchange rate in the toolkit.

Foreign Assets

Having recently read about the large returns on the Shanghai stock exchange and having seen that you can buy Chinese currency (the yuan, which has the international code CNY), you might wonder whether you can buy shares on the Shanghai stock exchange. In general, you are not restricted to buying assets in your home country. After all, there are companies and governments around the world who need to finance projects of various forms. Financial markets span the globe, so the bonds issued by these companies and governments can be purchased almost anywhere. You can buy shares in Australian firms, Japanese government bonds, or real estate in Italy. [3] Indeed, television, newspapers, and the Internet report on the behavior of both US stock markets and those worldwide, such as the FTSE 100 on the London stock exchange, the Hang Seng index on the Hong Kong stock exchange, the Nikkei 225 index on the Tokyo stock exchange, and many others.

You could buy foreign assets from one of the big financial firms that you visited earlier. It will be happy to buy foreign stocks or bonds on your behalf. Of course, if you choose to buy stocks or bonds associated with foreign companies or governments, you face all the risks associated with buying domestic stocks and bonds. The dividends are uncertain, there might be inflation in the foreign country, the price of the asset might change, and so on. In addition, you face exchange rate risk. If you purchase a bond issued in Mexico, you don't know what exchange rate you will face in the future for converting pesos to your home currency.

You may feel hesitant about investing in other countries. You are not alone in this. Economists have detected something they call *home bias*. All else being equal, investors are more likely to buy assets issued by corporations and governments in their own country rather than abroad.

A Casino

Toward the end of your walk, you are particularly surprised to see a casino. Stepping inside, you see a casino floor, such as you might find in Las Vegas, Monaco, or Macau near Hong Kong. You are confronted with a vast array of betting opportunities.

The first one you come across is a roulette wheel. The rules are simple enough. You place your chip on a number. After the wheel is spun, you win if—and only if—you guessed the number that is called. There is no skill—only luck. Nearby are the blackjack tables where a version of 21 is played. In contrast to roulette, blackjack requires some skill. As a gambler in blackjack, you have to make choices about taking cards or not. The objective is to get cards whose sum is as high as possible without going over 21. If you do go over 21, you lose. If the dealer goes over 21 and you don't, you win. If neither of you goes over 21, then the winner is the one with the highest total. There is skill involved in deciding whether or not to take a card. There is also a lot of luck involved through the draw of the cards.

You always thought of stocks and bonds as serious business. Yet, as you watch the players on the casino floor, you come to realize that it might not be so peculiar to see a casino on Wall Street. Perhaps there are some similarities between risking money at a gambling table and investing in stocks, bonds, or other assets. As this chapter progresses, you will see that there are some similarities between trading in financial assets and gambling in a casino. But you will learn that there are important differences as well.

KEY TAKEAWAYS

- Many different types of assets, such as stocks, bonds, real estate, and foreign currency, are traded in financial markets.
- Your earnings from owning an asset depend on the type of asset. If you own a stock, then you are paid dividends and also receive a capital gain or incur a capital loss from selling the asset. If you own real estate, then you have a flow of rental payments from the property and also receive a capital gain or incur a capital loss from selling the asset.
- Risks also depend on the type of asset. If you own a bond issued by a company, then you bear the risk of that company going bankrupt and being unable to pay off its debt.

CHECKING YOUR UNDERSTANDING

1. If you live in a house rather than rent it, do you still get some benefits from ownership? How would these benefits compare with the income you could receive if you rented out the house?
2. What assets are subject to the risk of bankruptcy?

[1] You can learn more about the DJIA if you go to NYSE Euronext, “Dow Jones Industrial Average,” accessed March 14, 2011,http://www.nyse.com/marketinfo/indexes/dji.shtml.

[2] The currency market is also discussed in Chapter 7 "Why Do Prices Change?".

[3] Some countries have restrictions on asset purchases by noncitizens—for example, it is not always possible for foreigners to buy real estate. But such restrictions notwithstanding, the menu of assets from which you can choose is immense.

9.2 The Value of an Asset

LEARNING OBJECTIVES

1. What factors determine the value of an asset?
2. How do you use discounted present value to calculate the value of an asset?
3. How is risk taken into account when valuing an asset?

Our basic explanation of assets reveals that there are two ways in which you can earn money from holding an asset: (1) You may receive some kind of payment that we call a *flow benefit*—a dividend payment from a stock, a coupon payment from a bond, a rental check from an apartment, and so on. (2) The price of the asset may increase, in which case you get a capital gain. You might guess that the price of an asset should be linked in some way to the payments you get from the asset, and you would be right. In this section, we explain how to determine the price of an asset. To do so, we use two tools: **discounted present value**and **expected value**. [1]

Toolkit: Section 17.5 "Discounted Present Value" and Section 17.7 "Expected Value"

You can review the meaning and calculation of discounted present value and expected value in the toolkit.

The Value of an Orange Tree

Imagine that you own a very simple asset: an orange tree. The orange tree pays a "dividend" in the form of fruit that you can sell. What is the value to you of owning such a tree? You can think of this value as representing the most you would be willing to pay for the orange tree—that is, your valuation of the tree. As we proceed, we will link this value to the price of the orange tree.

We begin by supposing your orange tree is *very* simple indeed. Next year, it will yield a crop of precisely one orange. That orange can be sold next year for $1. Then the tree will die. We suppose that you know all these things with certainty.

The value to you of the orange tree today depends on the value of having $1 next year. A dollar next year is not worth the same as a dollar this year. If you have a dollar this year, you can put it in the bank and earn interest on it. The technique of discounted present value tells us that you must divide next year's dollar by the **nominal interest factor** to find its value today:

$$\text{value of tree} = \text{discounted present value of \$1 next year} = \frac{\$1}{1 + \text{nominal interest rate}} = \frac{\$1}{\text{nominal interest factor}}.$$

Here and for the rest of this chapter we use the nominal interest factor rather than the **nominal interest rate** to make the equations easier to read. The interest factor is 1 plus the interest rate, so whenever the interest rate is positive, the interest factor is greater than 1. We use the *nominal* interest factor because the flow benefit we are discounting has not been corrected for inflation. If this flow were already corrected for inflation, then we would instead discount by the **real interest factor**.

Toolkit: Section 17.6 "The Credit Market"

You can review nominal and real interest rates and nominal and real interest factors in the toolkit.

To see why this formula makes sense, begin with the special case of a nominal interest rate that is zero. Then using this formula, the discounted present value of a dollar next year is exactly $1. You would be willing to pay at most $1 today for the right to receive $1 next year. Similarly, if you put $1 in a bank paying zero interest today, you would have exactly $1 in the bank tomorrow. When the nominal interest rate is zero, $1 today and $1 next year are equally valuable. As another example, suppose the nominal interest rate is 10 percent. Using the formula, the discounted present value is $\frac{\$1.00}{1.1}$ = $0.909. If you put $0.909 in a bank account paying a 10 percent annual rate of interest (an interest factor of 1.1), then you would have $1 in the bank at the end of the year.

A Tree That Lives for Many Years

Our orange tree was a very special tree in many ways. Now we make our tree more closely resemble real assets in the economy. Suppose first that the tree lives for several years, yielding its flow benefit of fruit for many years to come. Finding the value of the tree now seems much harder, but there are some tricks that help us determine the answer. Orange trees—like stocks, bonds, and other assets—can be bought and sold. So suppose that next year, you harvest the crop of one orange, sell it, and then also sell the tree. Using this strategy, the value of the tree is as follows:

$$\text{value of tree this year} = \frac{\text{value of crop next year} + \text{price of tree next year}}{\text{nominal interest factor}}.$$

The first term is the same as before: it is the discounted present value to you of the crop next year ($1.00 in our example). The second term is the price that you can sell the tree for next year. After all, if the tree lives for 10 years, then next year it will still have 9 crops remaining and will still be a valuable asset.

This expression tells us something very important. The value of an asset depends on

- the value of the flow benefit (here, the crop of oranges) that you obtain while owning the asset,
- the price of the asset in the market when you sell it.

The insight that the value of the tree equals the value of the crop plus next year's price greatly simplifies the analysis. If you know the price next year, then you know the value of the tree to you this year. Of course, we do not yet know how the price next year is determined; we come back to that question later.

We can now give a more precise definition of the return on an asset: it is the amount you obtain, in percentage terms, from holding the asset for a year. The return has two components: a flow of money (such as a dividend in the case of a stock) and the price of the asset. In the case of the orange tree, the return is calculated as

1 + nominal return = value of crop next year + price of tree next yearvalue of tree this year.

Because we know that

value of tree this year = value of crop next year + price of tree next yearnominal interest factor,

it follows that

1 + nominal return = nominal interest factor = 1 + nominal interest rate.

In this simple case, the return on the asset is equal to the nominal interest rate. If we wanted the real return, we would use the real interest factor (1 + the real interest rate) instead.

A Tree with a Random Crop

So far we have assumed that you know the orange crop with certainty. This is a good starting point but is not realistic if we want to use our story to understand the value of actual assets. We do not know for sure the future dividends that will be paid by a company whose stock we might own. Nor do we know the future price of a stock or a bond.

Looking back at the tree that lives for one year only, imagine you do not know how many oranges it will yield. Start by assuming that you can buy a tree that lasts for one period and whose crop is not known with certainty. The value of the tree depends on the following.

- The expected value of the crop. You must list all the possible outcomes and the probability of each outcome. For example, Table 9.1 "Expected Crop from an Orange Tree" shows the case of a tree where there are three possible outcomes: 0, 1, or 2 oranges. The probability of 0 oranges is 10 percent—that is, 1 in 10 times on average, the tree yields no fruit. The probability of 1 orange is 50 percent: half the time, on average, the tree yields 1 fruit. And the probability of 2 oranges is 40 percent. The expected crop is obtained by adding together the numbers in the final column: 1.3 oranges.
- A risk premium is an addition to the return on an asset that is demanded by investors to compensate for the riskiness of the asset. This adjustment reflects the riskiness of the crop and how **risk-averse** the owner of the tree is. If the owner is **risk-neutral**, there is no need for a risk premium. Obviously enough, if the crop is known with certainty, there is also no need for a risk premium.

Toolkit: Section 17.7 "Expected Value"

You can review the concepts of risk aversion and risk-neutrality in the toolkit.

Table 9.1 Expected Crop from an Orange Tree

Outcome (Number of Oranges)	Probability	Probability × Outcome
0	0.1	0
1	0.5	0.5
2	0.4	0.8

The easiest way to see how the risk premium works is to recognize that someone who is risk-averse will demand a higher return to hold a risky asset. Earlier, we said that the return on an asset without risk equals the nominal interest rate. In the case of a risky asset, however,

1 + nominal return = expected value of crop next yearvalue of tree = nominal interest factor + risk premium.

From this we can see that there is a relationship between risk and return. If the crop is not risky, then the risk premium is zero, so the return equals the nominal interest rate. As the crop becomes riskier, the risk premium increases, causing an increase in the return per dollar invested.

We can see how the risk premium affects the value of the tree by rearranging the equation:

$$\text{value of tree} = \frac{\text{expected value of crop next year}}{\text{nominal interest factor} + \text{risk premium}}.$$

For a given expected crop, the higher is the risk premium, the lower is the value of the tree.

The Value of an Asset in General

We have been talking about orange trees because they nicely illustrate the key features of more complex assets. We can combine the insights from our analysis of the orange tree to obtain a fundamental equation that we can use to value all kinds of assets:

$$\text{value of asset this year} = \frac{\text{flow benefit from asset} + \text{price of asset next year}}{\text{nominal interest factor} + \text{risk premium}}.$$

We apply this equation throughout the remainder of the chapter. To keep things simple, however, we will suppose most of the time that there is no risk premium—that is, we will discount using the nominal interest factor alone, except when we explicitly want to talk about the riskiness of different assets. We can now use this formula to value assets that are more familiar, such as bonds, stocks, cars, and houses.

The Value of a Bond

Suppose that you want to value a bond that lasts only one year. You will receive a payment from the borrower next year and then—because the bond has reached its maturity date—there will be no further payments. Naturally enough, the bond is worthless once it matures, so its price next year will be zero. This bond is like the first orange tree we considered: it delivers a crop next year and then dies. Hence we can value the bond using the formula

$$\text{value of bond this year} = \frac{\text{coupon payment next year}}{\text{nominal interest factor}}.$$

For example, if the coupon on the bond called for a payment of $100 next year and the nominal interest rate was zero, then the value of the bond today would be $100. But if the nominal interest rate was 10 percent, then the value of the bond today would be $\frac{\$100}{1.1}$ = $90.91.

If the bond has several years until maturity,

value of bond this year = coupon payment next year + price of bond next yearnominal interest factor.

This expression for the value of a bond is very powerful. It shows that a bond is more valuable this year if

- the coupon payment next year is higher,
- the bond will sell for a higher price next year, or
- interest rates are lower.

We explained earlier that bonds are subject to inflation risk. There are two ways of seeing this in our example. Imagine that inflation increases by 10 percentage points.

This inflation means that the coupon payment next year will be worth less in real terms—that is, in terms of the amount of goods and services that it will buy. Also, from the **Fisher equation**, we know that increases in the inflation rate translate into changes in the nominal interest rate. If inflation increases by 10 percentage points and the real rate of interest is unchanged, then the nominal rate increases by 10 percentage points. So the discounted present value of the bond decreases. Inflation risk might cause a bondholder to include a risk premium when valuing the bond.

Toolkit: Section 17.8 "Correcting for Inflation"

You can review the Fisher equation in the toolkit.

The Value of a Stock

Now let us use our general equation to evaluate the dividend flow from stock ownership. Imagine you are holding a share of a stock this year. You can hold it for a year, receive the dividend payment if there is one, and then sell the stock. For now we treat both the dividend and the price next year as if they are known for sure. What is the value of a share under that plan?

value of share this year = dividend payment next year + price of share next yearnominal interest factor.

This equation is similar to the one we used for the fruit tree and the bond. The flow benefit in this case is the dividend paid on the stock. Because the dividend is received next year, we have to discount it back to the current year using the nominal interest factor. The other part of the value of the share comes from the fact that it can be sold next year. Again, that share price must be discounted to put it in today’s terms. If the share does not pay a dividend next year, then its value is even simpler: the value of the share this year equals its price next year discounted by the nominal interest factor.

The return to owning the share comes in two forms: the dividend and the gain from selling the share next year. To calculate the return per dollar invested, we divide the dividend and future price by the value of a share this year:

return per dollar = dividend next year + price of share next yearvalue of share this year.

Table 9.2 "Discounted Present Value of Dividends in Dollars" shows an example where we calculate the value of a stock using two different interest rates: 5 percent and 10 percent.

Table 9.2 Discounted Present Value of Dividends in Dollars

Dividend	Price Next Year	Discounted Present Value (5%)	Discounted Present Value (10%)
1	2	2.86	2.73
1	4	4.76	4.55
2	4	5.71	5.45

The Value of a House

There are other familiar assets that can also be valued in the same way. A house is an asset that delivers a benefit each year in the form of providing shelter. The value of a house is the flow of services that it provides over the coming year plus the price it could be sold for next year. Of course, instead of living in your house and enjoying the service flow, you could rent it out instead. Then

value of house this year = value of rental payments over the next year + price of house next yearnominal interest factor.

For a house and similar assets, the value today reflects

- the flow of services of the asset over a year,
- the resale value next year, and
- the interest rate that is used to discount the future flows.

This completely parallels what we have already found for both bonds and stocks.

KEY TAKEAWAYS

- The value of an asset is the most you would pay to own that asset. The value today is the discounted value of the sum of the dividend (or service flow) plus the future price of the asset.
- Because the return of owning an asset comes in the future, you use discounted present value to calculate the current value of the asset. If the dividend and future price are not corrected for inflation, then you discount using the nominal interest rate. If the dividend and future price have already been corrected for inflation, then you discount using the real interest rate.
- The value of an asset is reduced by a risk premium that takes into account the riskiness of the asset and your risk aversion.

CHECKING YOUR UNDERSTANDING

1. Explain why an increase in the price of an asset in the future will increase its value today. Is this a violation of the law of demand?
2. In Section 9.2.4 "The Value of a House", we talked about houses. Can you think of other assets that could be valued using a similar formula?
3. Revise Table 9.1 "Expected Crop from an Orange Tree" so that the probability of getting 0 oranges is 0 and the probability of getting 3 oranges is 0.1. What is the expected crop from this tree? Is it more or less valuable to you?

[1] These tools are discussed at length in Chapter 4 "Life Decisions".

9.3 Asset Markets and Asset Prices

LEARNING OBJECTIVES

1. What is arbitrage?
2. How are asset prices determined?

So far we have focused on the value of an asset to an individual: "What is the value to you of the asset (fruit tree, bond, stock, car, house, etc.) you are holding?" Now we want to go a step further and see what the market price is for the asset. We already know that the two are connected. For example, when we valued a bond, we wrote

value of bond this year = coupon payment next year + price of bond next yearinterest factor.

Part of the value of a bond to you is the price you can sell it for on the market next year. Now we explain that the *current* price of a bond is closely connected to its current value.

Arbitrage

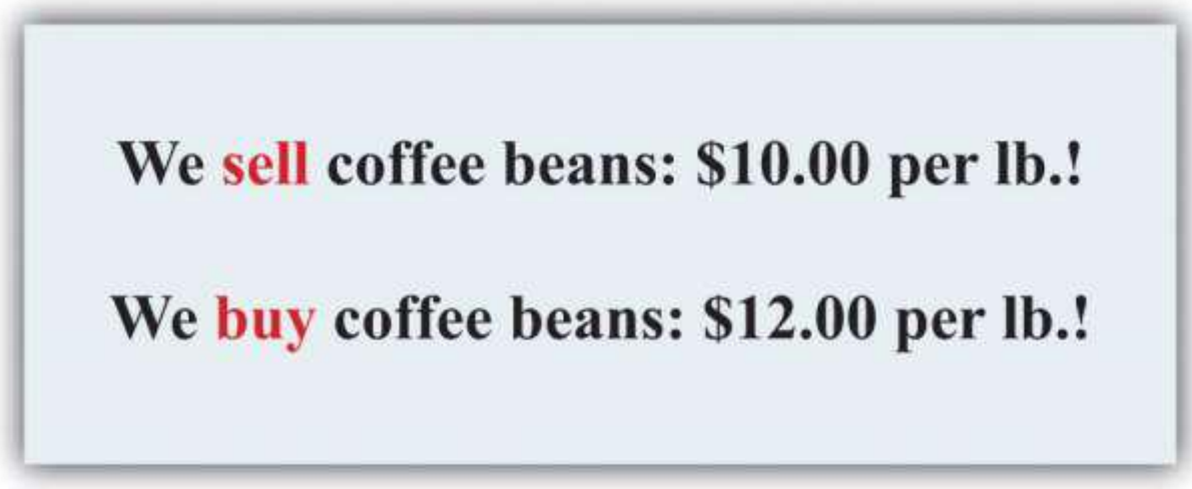
Assets are traded in markets around the world. Typically, there are a large number of (potential) buyers and sellers for any given asset: thousands of people might be willing to buy Microsoft Corporation stock or sell government bonds if they felt the price was right. Also, assets are homogeneous: one US government 10-year bond is the same as another. This means that asset markets are a good example of **competitive markets**, which means that we can look at asset markets using **supply and demand**.

Toolkit: Section 17.9 "Supply and Demand"

You can review supply and demand and competitive markets in the toolkit.

To derive the supply and demand curves for assets, we use the idea of arbitrage. This is the act of buying and then reselling an asset to take advantage of profit opportunities. The idea of arbitrage is to "buy low" and "sell high." Arbitrage is usually carried out across two markets to profit from any difference in prices. The strict definition of arbitrage refers to buying and selling where there is no risk, meaning that profits can be made with certainty. A weaker meaning of arbitrage allows risk to be associated with the process.

Figure 9.4 Arbitrage at a Coffee Shop



Imagine you passed a coffee shop and saw the sign shown in Figure 9.4 "Arbitrage at a Coffee Shop". This would make an economist salivate, not because of the prospect of good coffee but because it presents an opportunity for arbitrage. Facing an offer like this, you could immediately go and buy a pound of coffee beans for $10. Then you could turn around and sell the coffee at $12 per pound. You would have made $2 easy profit. Forget about drinking the coffee: just buy and sell, buy and sell, pound after pound—and become a billionaire. This is an example of arbitrage.

Sadly, you will never see a coffee shop making you an offer like this. We are confident of this because any coffee shop that made such an offer would very quickly go out of business. After all, if you can make a profit by buying at a low price and selling at a high price, then whoever is on the other side of these transactions is making a loss.

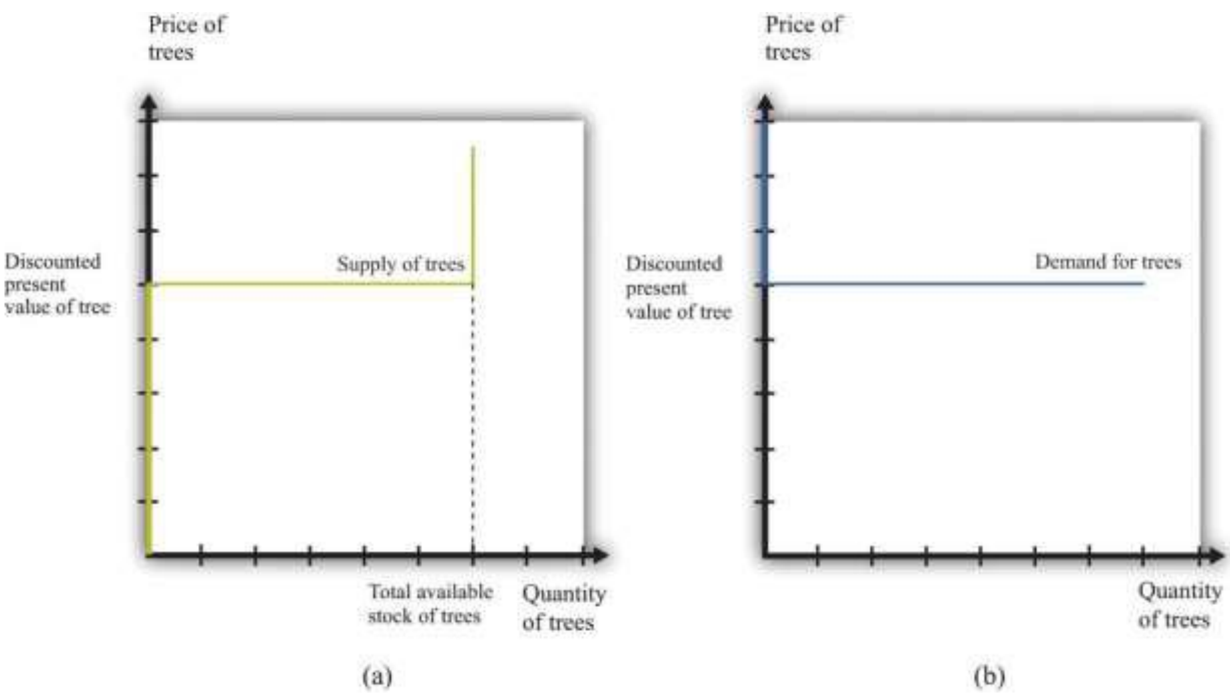
Arbitrage in the Supply-and-Demand Framework

We can think about arbitrage using the supply-and-demand framework. There are two markets: in one the coffee shop sells coffee, while in the other the coffee shop buys coffee. The demand of potential buyers would be extremely large, and the supply of coffee (from people selling it back) would likewise be very large. With the prices for buying and selling coffee as stated in the sign, demand could never equal supply in these two markets. An arbitrage possibility like this is not consistent with market equilibrium.

Using similar logic, we can argue that *the price of an asset will equal its value*. To see why, we begin again with an orange tree that will yield an orange worth $1 next year. Owners of this asset value it at value of tree = $1nominal interest factor.

They will be willing to sell the asset at this price but not if the price is any lower. They would definitely want to sell if the price were higher. But buyers can perform exactly the same calculation. They would be willing to buy the asset at this price but not if the price were any higher. They would definitely want to buy if the price were lower. Figure 9.5 "Asset Demand and Supply" shows the supply of and demand for trees in this case.

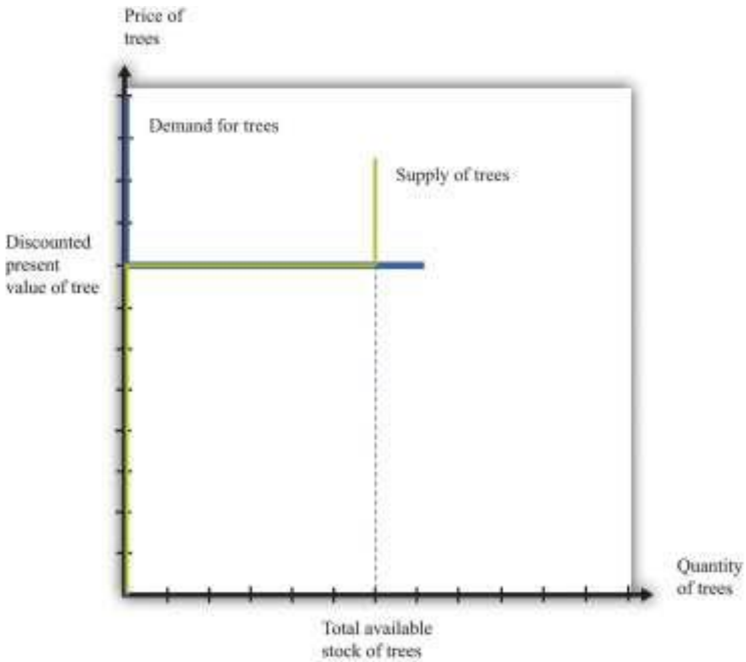
Figure 9.5 Asset Demand and Supply



(a) Owners of trees are willing to sell at a price equal to the discounted present value of the tree, and the supply curve is flat (perfectly elastic) up to the total available stock of trees. (b) Potential buyers of trees are willing to buy at a price equal to the discounted present value of the tree, and the demand curve is flat (perfectly elastic).

The supply side is shown in part (a) of Figure 9.5 "Asset Demand and Supply". There is a given stock of trees available. For prices below the asset value, no one wants to sell the asset. At prices above the value, everyone wants to sell the asset. So the supply curve is horizontal at a price equal to the asset value, all the way up to the point where every tree is on the market. At that point, the supply curve becomes vertical. The demand function is in part (b) of Figure 9.5 "Asset Demand and Supply". At a price above the discounted present value of the tree, the quantity demanded is zero: no one will pay more than the discounted present value for the asset. If the price equals the value, the demand is flat (horizontal). At a price below the value, the asset looks like a great deal because there are arbitrage opportunities. So demand is very large.

Figure 9.6 Asset Market Equilibrium



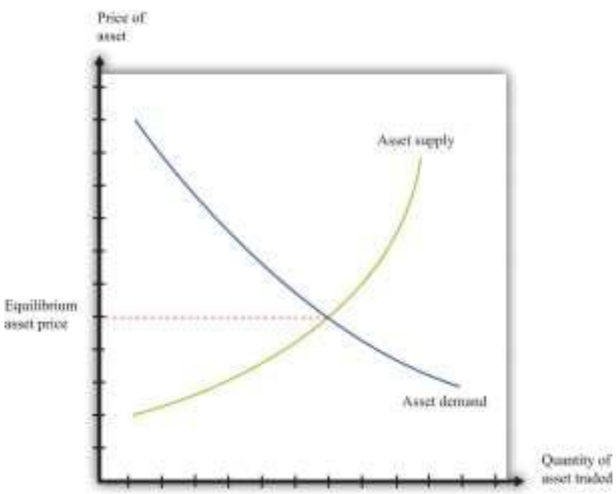
Because buyers and sellers place the same value on the tree, the demand and supply curves lie on top of each other at this value, so the price will equal the discounted present value of the tree.

We put these curves together in Figure 9.6 "Asset Market Equilibrium". Both supply and demand are horizontal at a price equal to the discounted present value of the asset. Thus at this price, and at this price only, supply equals demand. We obtain a powerful prediction: *assets will be priced at their discounted present value*. If we see the prices of assets (such as stocks or bonds) increase or decrease, this model of the asset market tells us to attribute these variations to changes in the discounted present value of dividends.

The supply and demand curves in these figures look rather untraditional. We are used to seeing upward sloping supply curves and downward sloping demand curves. But in the market for the tree, everyone values the asset in exactly the same way. That valuation is given by the discounted present value of the dividend stream. As a result, Figure 9.6 "Asset Market Equilibrium" does not tell us how much people will trade or if they will trade at all. When the price equals the discounted present value, buyers are indifferent about buying, and sellers are indifferent about selling. Everyone is happy to trade, but no one particularly wants to trade. In reality, though, the market for an asset may look much more like a "normal" supply-and-demand diagram, as in Figure 9.7 "Asset Market Equilibrium: A More Familiar View", with an upward-sloping supply curve and downward-sloping demand curve. The reason is that different individuals may have differing views about the discounted present values of the asset, because [1]

- different buyers and sellers have different information that causes them to make different forecasts of future dividends, or
- different buyers and sellers differ in their attitudes toward risk.

Figure 9.7 Asset Market Equilibrium: A More Familiar View



If potential buyers and sellers of an asset differ in their beliefs about the dividend from that asset or differ in terms of their degree of risk aversion, then we obtain demand and supply curves of the familiar form.

For example, suppose some buyers are optimistic about future dividends from a stock, while others are pessimistic. Optimistic buyers will calculate a high discounted present value and have a high willingness to pay. Pessimistic buyers will calculate a lower discounted present value and be willing to pay less for the asset. Such differences in views can hold for sellers as well. Alternatively, suppose some buyers and sellers are more risk-averse than others. The less risk-averse the buyer, the higher the price he is willing to pay because he uses a lower risk premium when calculating his discounted present value. The less risk-averse the seller, the higher the price she is willing to accept.

There is one last, more subtle point. We have been imprecise—intentionally—about what exactly it means for an asset to be risky. Buyers and sellers care not only about assets in isolation but also about how those assets fit into their entire portfolio—that is, the entire collection of assets that they own. An asset that seems very risky to one person may appear less risky to another because he holds other assets that balance out the risks. The riskiness of an individual asset depends on the **diversification** of the portfolio as a whole.

In Figure 9.6 "Asset Market Equilibrium", all traders in the market valued the asset in exactly the same way, so arbitrage guaranteed that the price equals the discounted value of the flow benefit. In Figure 9.7 "Asset Market Equilibrium: A More Familiar View", there is no immediate guarantee that this will still be true. Even with differences in valuation, however, we expect that the price of an asset is still likely to be (at least approximately) equal to its true discounted present value. In particular, if some traders in the market do not care about risk and are accurately informed about the flow benefit, arbitrage will still keep the market price close to the discounted present value of the stock.

We have not yet explained how supply and demand actually come together in financial markets—that is, who actually makes the market? If you study the workings of a market such as the New York Stock Exchange, you will learn that it works through specialized traders. If you want to buy a stock, you typically contact a stockbroker who communicates your demand to his firm on Wall Street. Another broker then takes that order onto the floor of the stock exchange and looks for a seller. If a seller is found, then a deal can be made. Otherwise, the broker can place your order with another specialist who essentially "makes the market" by buying and selling securities at posted prices. So in the end, the market has some elements of posted prices (take-it-or-leave-it offers) and some elements of a double-oral auction. [2]

The Price and Value of a Long-Lived Asset

Previously, we explained how to value an asset assuming you hold it for one year, receive the flow benefit (the fruit, the dividend, or the coupon payment), and then sell it at the current market price. We said that (assuming no risk premium)

$$\text{value of an asset this year} = \text{flow benefit from asset} + \frac{\text{price of asset next year}}{\text{nominal interest factor}}.$$

We have also discovered that, in general,

value of asset = price of asset.

We combine those two pieces of information to complete our study of the valuation of assets. Imagine an orange tree that lives for two years and yields a crop valued at \$1 each year. We already know that we can value the tree as follows:

$$\text{value of tree this year} = \$1 + \frac{\text{price of tree next year}}{\text{nominal interest factor}}.$$

But now we know that the *price* of the tree next year will equal the *value* of the tree next year:

$$\text{value of tree this year} = \$1 + \frac{\text{value of tree next year}}{\text{nominal interest factor}}.$$

Next year, we can apply exactly the same formula:

$$\text{value of tree next year} = \$1 + \frac{\text{value of tree the year after next}}{\text{nominal interest factor}} = \frac{\$1}{\text{nominal interest factor}}.$$

Why is this true? The year after next, this particular tree will be worthless because it will be dead. So the value of the tree today is

$$\text{value of tree this year} = \$1 + \frac{\frac{\$1}{\text{nominal interest factor}}}{\text{nominal interest factor}}.$$

This is a more complicated formula. It tells us that the value of the tree today is the discounted present value of the flow benefit tomorrow plus the discounted present value of the value of the tree tomorrow—*which is itself the discounted present value of the flow benefit the year after*. In other words, the value of the tree today is the discounted present value of the flow benefits over the entire lifetime of the tree. What is more, we could use exactly the same logic if the tree were to yield a crop for 3 years, 10 years, or 100 years.

There is one last step. If we again use the idea that the price of the tree should equal its value, then we can conclude the following:

price of tree this year = discounted present value of the flow of benefits from the tree.

This logic applies to all assets, not only trees, so we can now apply it to bonds and stocks:

price of bond today = discounted present value of the flow of payments from the bond

and

price of share today = discounted present value of the flow of dividend payments.

A final note on uncertainty. We have been assuming that dividends are known with certainty. If they are not, then we need to modify the valuation of the stock by (1) replacing "dividend" with "expected dividend" and (2) adding a risk premium to the interest rate. As discussed previously, the adjustment for risk will reflect both the riskiness of the stock and the aversion to risk of investors. Riskier stocks generally have a lower value and a higher expected return.

KEY TAKEAWAYS

- Arbitrage entails the buying and selling of assets to make a profit. In equilibrium, there are no profits to be made through arbitrage.
- The price of an asset is (approximately) equal to the discounted present value of the flow of benefits (dividends, service flow, etc.) from the asset.

CHECKING YOUR UNDERSTANDING

1. Suppose an orange tree lives for two years, with a crop of five oranges in the first year and three in the second year. The price is $2 in the first year and $5 in the second year. If the interest rate is 20 percent, what is the price of the orange tree?
2. Explain why an increase in interest rates will reduce the price of a bond.

[1] Chapter 5 "eBay and craigslist" discusses in some detail the reasons why people trade. We explain that important motives for trade are that people have different tastes and skills. To these we can add the two motives just mentioned here.

[2] Both of these are discussed in more detail in Chapter 5 "eBay and craigslist".

9.4 Efficient Markets

LEARNING OBJECTIVES

1. Is it possible to make large profits in asset markets?
2. Is it easy to make large profits in asset markets?
3. What are some factors that cause asset prices to increase and decrease?
4. Why do asset prices respond to new events but not forecasted ones?
5. Are markets efficient?

The title of this chapter speaks of making and losing money on Wall Street. We have gone into considerable detail about what determines the price of assets, but we have not yet discussed how easy or hard it is to make money by buying and selling these assets.

Can Easy Profits Be Made on Wall Street?

Our fictional Wall Street contained places where you could buy many different kinds of assets, such as real estate and automobiles as well as stocks and bonds. But it also contained a building that wasn't selling assets at all: the casino.

Is Wall Street Like a Roulette Wheel?

Is buying and selling shares like gambling on a roulette wheel, where gains and losses are purely a matter of luck? To answer this question, think more about the uncertainty associated with buying stocks and bonds. Suppose we are buying a stock that will pay dividends over four years, as in Table 9.3 "Discounted Present Value of Dividends in Dollars", and suppose that the interest rate is 5 percent. From Table 9.3 "Discounted Present Value of Dividends in Dollars", we know that the discounted present value of the stock is $609.61. We then expect this will also be the price of the stock.

Table 9.3 Discounted Present Value of Dividends in Dollars

Year	Dividend ($)	Discounted Present Value ($, Interest Rate = 5%)
1	100	95.2381
2	90	81.63265
3	120	103.6605

Year	Dividend ($)	Discounted Present Value ($, Interest Rate = 5%)
4	400	329.081
Discounted present value (all years)		609.61

Can you make money buying and selling this stock? It seems unlikely. If the price of a stock is equal to the present discounted value of the flow of dividends, then you get what you pay for. If you sell the stock, then instead of an asset that would have paid you the equivalent of $609.61, you receive $609.61 in cash. If you buy the stock, the reverse is true. Either way, you are no richer or poorer after the transaction; you are just holding your wealth in a different form.

Economists often use the metaphor of $100 bills lying on the ground to describe a situation where easy money can be made. Our example of the arbitrage opportunity in the coffee shop, where you could buy beans at $10 a pound and resell them at $12 a pound, is an example: getting rich in that coffee shop is as easy as picking up money on the floor. But if the value of a stock is the discounted present value of the dividends that it will pay, then there is no easy money to be made. There are no $100 bills lying around. You should not anticipate spectacular earnings from owning assets. You can earn a reasonable rate of return, equal to the nominal interest rate, but no more. A market with the characteristic that you cannot expect to earn abnormal profits is called an efficient market. In such a market, the price of an asset accurately reflects the best forecast of the value of that asset. The value of an asset is the discounted present value of the flow benefits.

Yet this may strike you as odd. There are many people who certainly make it rich by buying and selling stocks, bonds, and other assets. Some of these individuals—like Warren Buffett or George Soros—are household names. Does the fact that some people get very rich on Wall Street mean that markets are not efficient?

There are a couple of possible answers to this question. The first is that even when markets are efficient, some people may get rich. Go back to the stock we were considering earlier but with one small change. Imagine that the numbers in Table 9.3 "Discounted Present Value of Dividends in Dollars" are now *expected* dividends. They tell you what this stock can be expected to pay out on average, but they are not guaranteed. For example, it might be the case that in year four, the firm will pay a dividend of $800 with 50 percent probability but will also pay $0 with 50 percent probability. The expected dividend is

$400. If people care only about the expected value of the dividends (if they are risk-neutral), the price of the stock still equals the (expected) discounted present value.

Now, there are still no $100 bills on the ground. You cannot expect to make unusual profits by buying this stock or others like it. However, some people will get lucky and thus get rich. If you bought this stock, and it ended up paying the high dividend, you would end up with a return nicely above the market interest rate. If it failed to pay the dividend, however, you would get a lower return than the market interest rate. As we looked around the economy, we would see some lucky investors earning high returns and other unlucky investors earning low returns. In this world, buying and selling assets in the stock market is really not that different from betting on a roulette wheel. Buying an asset is like placing your chip on a certain number. If the number comes up, you get rich. If it does not, you lose.

When you go to a casino, you should not *expect* to win at roulette. But this does not mean that you can never earn spectacular amounts of money. It happens frequently: casinos thrive on advertising these success stories. The same goes for buying an asset, such as a stock. Suppose you buy a share in a pharmaceutical company. The price of the share when you purchase it might indeed equal the expected discounted present value of dividends. Yet the following week the company could have a major discovery that will allow it to be much more profitable in the future. Expected dividends will increase, and so will the stock price. This is certainly good news for you. But it is also no different from getting lucky on the roulette wheel.

Spectacular successes tend to be more visible than losses. In the 1990s, some people earned large amounts (sometimes spectacularly large amounts) from certain successful Internet companies. But many other people lost money on Internet companies that were ultimately unsuccessful, and you are less likely to hear about them.

Is Wall Street Like a Blackjack Table?

More than luck is required when investing on Wall Street, however. Just as there are skilled players of blackjack in the casino, there are people who are skilled in assessing the prospects of different firms in an economy.

If the price of a share equals the discounted present value of the dividends that a company will pay, then the total value of all the shares in a company should equal the discounted present value of the total profits the firm will pay out—both now and in the future. The total value of all the outstanding shares in a firm is called its market capitalization.

The price of a share increases whenever a firm's market capitalization increases, and a firm's market capitalization should increase whenever there is reason to think that the firm has become more profitable—either now or in the future. If markets are efficient, therefore, we expect share prices to respond to *new information* about a firm. Traders in large financial firms make their money in part by gathering new information about firms and then acting very quickly on that new information. News about a firm—good or bad—is likely to be incorporated quickly into a firm's market price, but a trader who can move fast can make money from these movements.

When economists use the metaphor of $100 bills lying on the ground, they are pointing out that if opportunities for easy profits arise, they will disappear very quickly. It is not impossible that someone will drop a $100 bill. But it is highly unlikely that that bill will lie unretrieved for more than a few minutes. If there is easy money to make in the market by buying a stock, professional traders will jump on the opportunity quickly. This has an important implication for the rest of us. If you read in the newspaper today that Merck Pharmaceuticals has just announced a new pharmaceutical compound that is highly effective in treating lung cancer, there is no point in calling your broker and instructing him to buy Merck stock on the basis of this news. Somebody—a very smart trader with her ear to the ground and lots of knowledge about the pharmaceutical industry—might well have made money buying Merck stock at the first hint of this news. But by the time there is an announcement in the paper, the increase in Merck's expected profits has long been factored into the price.

Are Markets Efficient?

The theory of efficient financial markets is very powerful because it gives us a key to understanding the prices of assets. Go back to our equation for the value of an asset:

value of asset this year = flow benefit from asset + price of asset next yearnominal interest factor.

In the case of a share, the flow benefit is the value of the dividend. If the price next year is the discounted present value of further dividends from that point in time onward, then this equation is another way of saying that the value of the share today equals the discounted present value of dividends.

But what if the price that everyone expects an asset to sell for next year is—for some reason—much higher than the discounted present value of the flow benefit from the asset starting next year? As an example, consider a house. You buy a house in part to enjoy living in it—to enjoy "housing services." You also buy a house as a possible source of capital gains if the price of the house increases. Imagine you live somewhere where everyone seems to think housing prices will increase a lot over the next few years. Then you should be willing to pay more for a house. After all, if you anticipate a large capital gain in five years from selling the house, you can pay more for it now and still expect a good return. So if everyone expects the price of houses to increase in the future, then the current demand for houses increases, so the current price increases. The price increase today reflects the expectation of higher prices in the future.

Now fast forward to next year. We can say the same thing applied to next year: "If everyone expects the price of houses to increase in the future, then the current demand for houses increases, so the current price increases." This can go on from year to year: housing prices are high because everyone expects higher prices in the future. Higher prices have an element of self-fulfilling prophecy: prices increase because everyone thinks prices will continually increase.

This is sometimes called a housing *bubble*. In a bubble, the increase in the price of housing does *not* reflect an increase in the value people place in housing services. In the language of economics, the price is not changing because of any change in the *fundamentals*. Furthermore, it is possible that prices will not actually keep increasing. If everyone suddenly becomes more pessimistic about the future of housing prices, then the capital gains that everyone anticipated are gone, and housing prices collapse. In this case, the bubble bursts, and prices fall—sometimes very rapidly.

Many economists think that something like this happened in the early stages of the global financial crisis that started in around 2007. The price of housing in many markets had been increasing substantially, and people expected this to continue. At some point, people stopped being so confident that house prices would keep increasing—and, sure enough, the price of houses then decreased rapidly.

The same idea applies to stocks. If everyone believes the value of a particular stock will be higher in the future, then the price will be bid up today. If these beliefs persist, they can sustain a bubble in the stock. If everyone believes that stocks will generally increase in price, then this can lead to a bubble in the entire stock market.

Data on the prices of stocks can perhaps help us see if the efficient market view is accurate, or if we instead see lots of bubbles. This sounds like an easy exercise but is actually very hard. It is difficult to calculate the discounted present value of expected dividends because it requires forecasts of the future. The usual interpretation of this evidence is that the efficient market hypothesis is not capable of explaining all the variations in asset prices, particularly over short periods of time.

An extreme example illustrates this point well. Here is a quotation from an article that appeared in the *Wall Street Journal* after the collapse in the US stock market in 1987: "Calmly appraised, the intrinsic value of American industry didn't fall 23 percent in a day." [1] "The intrinsic value of American industry" refers to the total market capitalization of all the firms quoted on the stock exchange. It is hard to explain such short-run variations in asset prices from the perspective of discounted present value of dividends.

Economist Robert Shiller claims that stock markets can exhibit "irrational exuberance." He argues that asset prices move around too much to be explained by theories that rely on the discounted present value of dividends to the price of assets. Instead, the fluctuation of prices, at least over short periods of time, might also be influenced by expectations and bubbles. More generally, Shiller is one of many financial economists who believe that economic theory needs to be supplemented with some ideas from social psychology to do a better job of explaining the performance of financial assets.
Such behavioral finance has identified several *anomalies*—that is, occasions on which asset prices apparently diverge from the values predicted by efficient market theory.

Despite the insights of behavioral finance, most economists take the view that, at the very least, efficient market theory is the best starting point for thinking about asset prices. Efficient market theory provides us with two key insights: (1) the price of a stock should reflect expectations about future profits, which means that (2) the price of a stock should change when—and only when—there is new information that changes those expectations. Many economists nonetheless think this approach is incomplete and that behavioral finance can also help us understand financial markets. A word of caution: even if markets are not always efficient, this still does not mean that there are easy ways to make money on Wall Street.

Changes in Asset Quantities and Asset Prices

Each weekday in the United States, around 5 p.m. Eastern Standard Time, there are reports on the performance of the markets that day. At other times of the day, you can learn about other markets around the world. Newscasters report on the volume of trade (the number of shares exchanged) and some index of the price of stocks, such as the DJIA. In economic terms, these are reports about the price and quantity in a market. Therefore we can use the supply-and-demand framework, and more specifically the tool of **comparative statics**, to consider what makes asset prices increase and decrease.

Toolkit: Section 17.16 "Comparative Statics"

You can review how to carry out a comparative statics exercise in the toolkit.

If we take the efficient-market view that the price of an asset equals its discounted present value, then any change in the price of an asset must be due to some change in its expected discounted present value. If, for example, the price of General Motors stock increases, this should reflect some information about the prospects and hence future dividends of this company. Moreover, any information that makes the price increase or decrease must be *new*. If it were old information that everyone in the market already knew, then it would have already been factored into the stock price. Hundred-dollar bills do not stay on the ground for long. So what are some of the big events that can change asset prices?

Product Development News

Part of the profitability of a firm comes from its innovative activities in developing and marketing new products. Open a computer magazine, for example, and you will see hundreds of advertisements for a wide range of new products. How does news about a new product affect the price of a firm's stock?

Consider the following story:

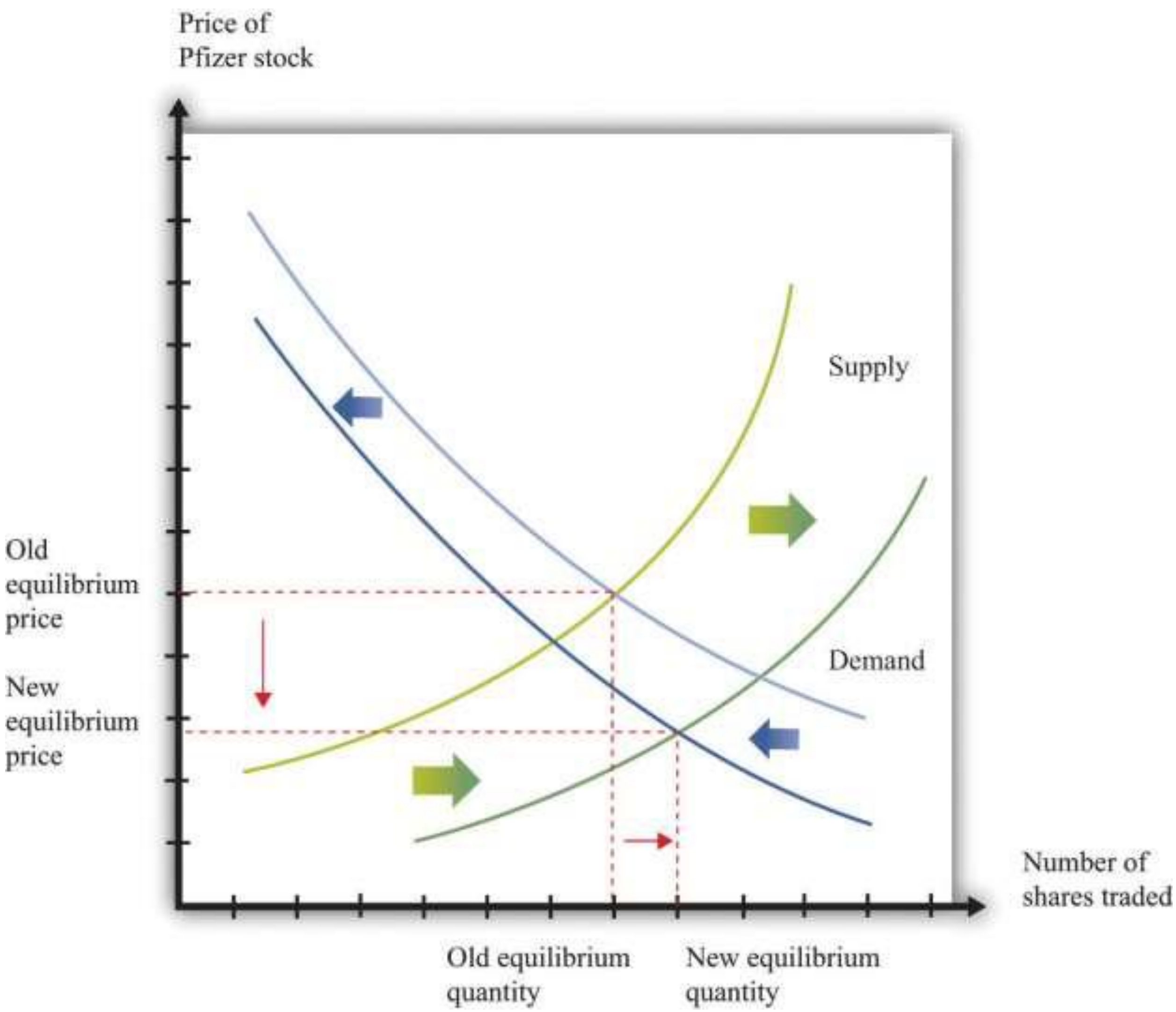
> *Pfizer stock tumbled Monday after the world's biggest drug maker abruptly pulled the plug on its most important experimental medicine—a drug meant to treat heart disease that instead caused an increase in deaths and heart problems in people taking it in a clinical trial.*
>
> *Shares of Pfizer sank about 11 percent in afternoon trading as investors worried what the New York-based company would do to replace the product, torcetrapib, in its pipeline.*
>
> *Pfizer CEO Jeffrey Kindler unveiled his company's pipeline at an analyst meeting last week, before the bad news on torcetrapib.*
>
> *Trading was heavy with more than 235 million shares changing hands by mid-afternoon—nearly seven times the stock's average daily volume.*
>
> *On Saturday, Pfizer and the Food and Drug Administration announced that the drug company would halt a clinical trial of torcetrapib due to an increased rate of death and heart problems in patients who took it.*
>
> *Just two days earlier, Pfizer's new CEO Jeffrey Kindler had told hundreds of investors and analysts at a research meeting that the drug maker could seek approval for the medicine as early as next year if clinical data supported it.* [2]

Pfizer's announcement is the **exogenous variable** in this comparative statics exercise. The story tells us that the announcement led to a decrease in price and a large amount of trading. Let us try to make sense of this. First the announcement clearly contained new information that the markets had not anticipated. Indeed, the article tells us that, a few days previously, there had been positive information about this product.

Because market participants did not previously know the results of the clinical trials, and because the announcement is bad news, analysts and market professionals immediately revise downward their estimates of the future profitability of Pfizer. They now expect that dividends in future years will be lower than they had previously thought. This reduction in the discounted present value of dividends causes the stock price to decrease.

The way this works in the supply-and-demand framework is shown in Figure 9.8 "Bad News about a Firm's Product Reduces the Value of Its Stock.". The bad news influences both the demand and supply curves. Prospective buyers of the stock are not willing to pay as much for it, given the bad news. So at a given stock price, the quantity demanded decreases. In Figure 9.8 "Bad News about a Firm's Product Reduces the Value of Its Stock.", the demand curve shifts to the left. Owners of the stock are now less interested in holding their shares. So at a given price, the quantity supplied increases. This means that the supply curve shifts to the right.

Figure 9.8 *Bad News about a Firm's Product Reduces the Value of Its Stock.*



Bad news about a clinical trial causes many current holders of Pfizer stock to want to sell their shares and makes people less likely to want to buy Pfizer stock.

The bad news has an unambiguous effect on the price of the stock: it decreases. The effect of the announcement on the quantity traded is not clear. It depends on the steepness of the curves and the relative shifts in supply and demand. In this particular example, we know that there was an unusually large amount of trading, so the equilibrium quantity increased.

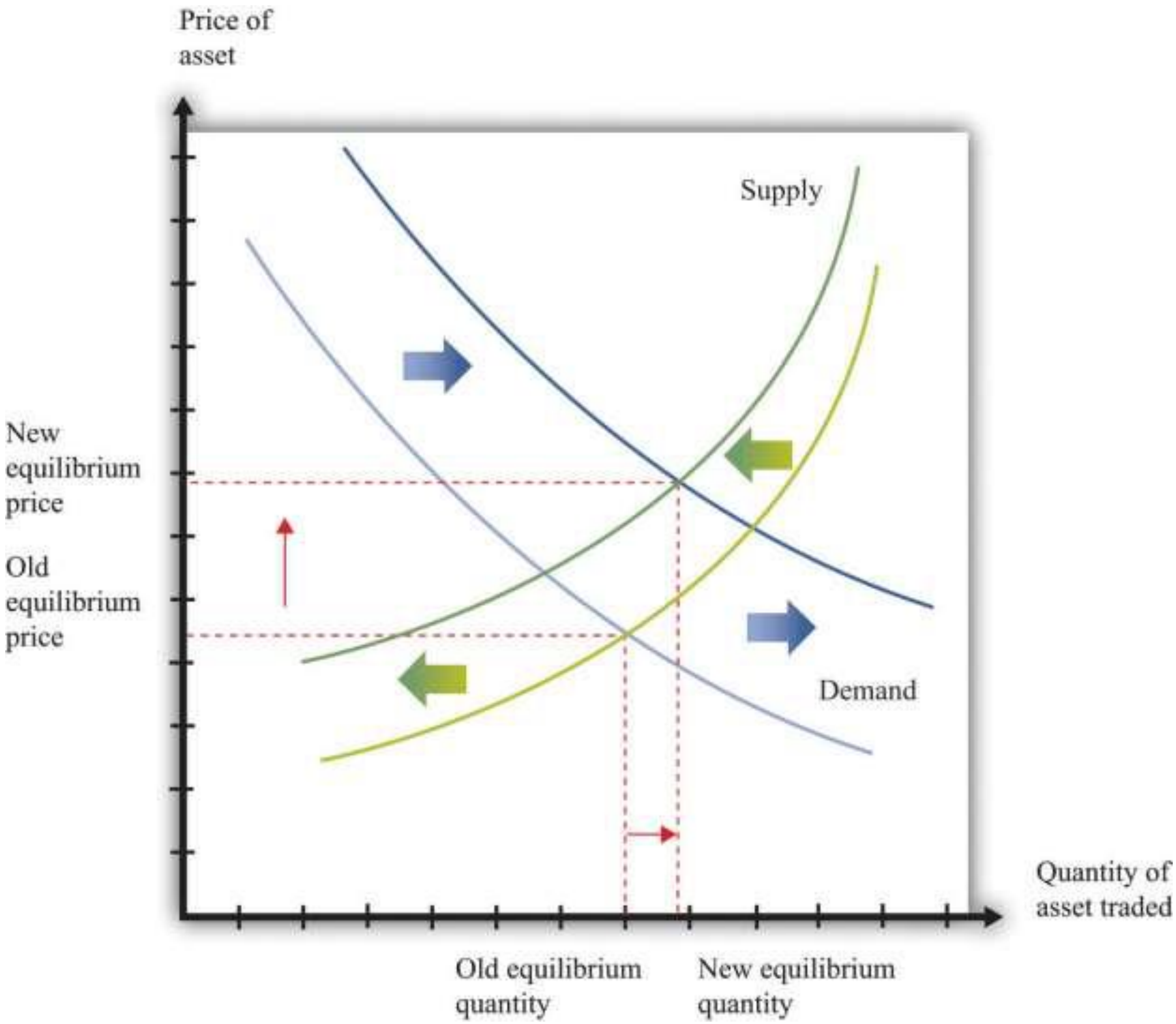
Had the results of the clinical trials already been leaked to the market, there would have been no new information from the announcement. The test results would already have been incorporated in the asset price. The supply and demand curves would not have moved at all.

Monetary Policy

In most modern economies, interest rates are set (or more precisely heavily influenced) by a central bank through its conduct of monetary policy. This process and its effects on an economy are studied in detail in macroeconomics courses. Here, we look at the effect of monetary policy on asset prices.

Because monetary policy influences interest rates, the link to asset prices is immediate. We know that the price of an asset equals its discounted present value, and that interest rates are used for the discounting process. A change in interest rates therefore directly affects asset prices. Using the example in Table 9.3 "Discounted Present Value of Dividends in Dollars", if a monetary authority were to increase the interest rate from 5 percent to 10 percent, then the asset price would decrease by about $81.

Figure 9.9 Equilibrium Asset Prices Respond to a Decrease in Interest Rates



A reduction in interest rates shifts the demand curve for assets (stocks, bonds, and other assets) to the right and shifts the supply curve to the left.

Figure 9.9 "Equilibrium Asset Prices Respond to a Decrease in Interest Rates"shows the impact of monetary policy on the supply-and-demand framework. As the interest rate decreases, the discounted present value of an asset increases. So at a given price, the demand for an asset increases. This is shown as an outward shift in demand. Also, at a given price, the supply of the asset decreases. As the interest rate decreases, more holders of the asset want to hold onto it, so the quantity supplied to the market is lower at each price. A rightward shift of the demand curve, combined with a leftward shift in the supply curve, tell us that the price of an asset increases when the interest rate decreases. The effect of the interest rate change on the quantity traded is ambiguous: it depends on the relative slopes of the curves and how much the curves shift.

Multiple Asset Markets

On the walk down Wall Street, we noticed that there were many markets, all trading simultaneously. Yet we have looked at the market for each particular asset in isolation. This is a fine tactic for understanding how to do comparative statics. But the real world is more complex: a single event can impact multiple markets.

Take, for example, the bad news on the test results for Pfizer. We saw that this news forced the firm's stock price to decrease. But Pfizer may have also borrowed in years past by issuing bonds. What will happen to them when the bad news hits the bond market? The price of a bond is the discounted present value of the interest payments on a bond over its lifetime, taking into account the possibility of bankruptcy. So the bad news from the pharmaceutical company ought to depress the price of its bonds. This happens largely because the chance of bankruptcy—while surely small—increases with the bad news.

We observed earlier that the global financial crisis had its origins, at least in part, in the real estate market. The effects quickly spread beyond the housing market. Many people had borrowed to buy houses, and these loans—known as mortgages—were financial assets held by banks. When housing prices decreased, the value of these mortgages decreased as well. Moreover, these mortgages were often combined in various ways to make new assets. When housing prices decreased, the price of these related assets decreased as well. And because financial institutions sometimes had difficulty working out the value of their assets, there was a risk that they would go bankrupt. This in turn meant that anyone who had lent to such an institution now had an asset that was less valuable. Thus a collapse in the price of one asset—houses—led to a decrease in the price of all sorts of other assets in the economy as well.

KEY TAKEAWAYS

- If markets are efficient, then, on average, there are no excessive profits to be made in asset markets. Some people will be lucky and do better than average, while others will be unlucky and do worse than average.

- Efficient markets provide a benchmark for asset valuation, though asset prices may sometimes deviate from these values.
- We use comparative statics to study the effects of changes in asset supply and demand on prices. Shifts in asset demand may come from new information about a new product or a new technique established by a firm. Monetary policy may also influence the demand for an asset.
- Asset markets respond to events, like the surprise announcement of a new product. Asset markets do not respond to changes today that were announced in the past, such as a change in interest rates by a central bank that was announced (or forecasted) weeks earlier.

CHECKING YOUR UNDERSTANDING

1. Explain how the expected return from playing slot machines is negative even though lots of people have great stories about winning money playing slot machines. Can you tell a similar story about stock markets?
2. If you see housing prices in a city decrease by 50 percent over six months, how would you explain this using the efficient markets viewpoint?
3. Suppose there is good news about a company's future product. What happens to the value of its stock?
4. Using the supply-and-demand framework, show how an increase in interest rates can increase the quantity of an asset sold. Now show how a decrease in interest rates can increase the quantity of an asset sold. What are the key differences between the two figures you just created?

[1] Roger Lowenstein, "After the Fall: Some Lessons Are Not So Obvious," *Wall Street Journal*, August 25, 1997.

[2] Aaron Smith, "Heart Drug Pulled, Pfizer Tumbles," *CNNMoney.com*, December 4, 2006, accessed January 29, 2011,http://money.cnn.com/2006/12/04/news/companies/pfizer_stock/index.htm.

9.5 End-of-Chapter Material

In Conclusion

The performance of the stock market is one of the most closely watched of all economic statistics. This chapter provided some clues as to why people care so much about the value of stocks and other assets.

One reason is that people save by purchasing stocks and other assets. Thus savers want to know what determines the value of assets in the economy. Having read this chapter, you should now understand that the value of any asset is closely linked to the flow of benefits that the asset provides. Indeed, if markets are efficient, then the value of any asset should equal the discounted present value of the flow of benefits.

There are two other reasons why we pay so much attention to the stock market. (1) If the value of a stock reflects the discounted present value of expected dividends, then the market capitalization of a firm represents the best guess as to the value of that firm—which depends ultimately on the profits that it will generate in the future. In that case, a stock market index represents our best guess of the overall value of all firms. It truly is a measure of an economy as a whole. (2) The stock market plays a key role in allocating an economy's saving to those firms that can make the most profitable use of those funds.

Key Links

- NYSE: http://www.nyse.com/about/education/1095056911801.html
- S&P 500: http://www.standardandpoors.com/home/en/us
- Robert Shiller
  - Home page: http://www.econ.yale.edu/~shiller
  - Irrational exuberance: http://www.irrationalexuberance.com

EXERCISES

1. List the factors you think would make stock prices increase and decrease in the Shanghai stock exchange.
2. An October 2007 article in the *Economist* magazine discusses land prices and office rents. According to the article, rents have recently increased, and land prices have been increasing in the past few years as well. Why would land prices and rents move together?

3. Following from Question 2, what do you think has happened to the price of office buildings as rental rates have increased?
4. Suppose an orange tree yields a crop of one orange after the first year and then two oranges in the second year. As before, let the price of an orange be $1 in both years. What is the value to you of buying the tree today and then selling it next year, after you have harvested the first orange? (Hint: first find the value of the tree tomorrow and then use that as the price for selling the tree.)
5. Suppose an orange tree lives for two years, with a crop of five oranges in the first year and three in the second year. The price is $2 in the first year and $5 in the second year. If the interest rate is 20 percent, what is the price of the orange tree?
6. (Advanced) The table titled Discounted Present Value Exercise provides information about the crop from an orange tree as well as the interest rate for a tree that lives four years. Assume that the price of oranges is $1 in the first year and then increases at 10 percent per year. What is the discounted present value of this tree?
7. Suppose prospective buyers of houses become very optimistic about the future prices of houses. Existing owners, on the other hand, become very pessimistic about the future value of houses. What happens to the price of houses today?
8. Suppose housing markets are efficient. If you see rapidly increasing prices in a market, do you think that rental rates are increasing as well?
9. Explain how contractionary monetary policy can reduce housing prices.
10. (Advanced) In the first row of Table 9.2 "Discounted Present Value of Dividends in Dollars", we considered a stock that pays a dividend of $1 this year and that will have a price of $2 next year. Suppose the inflation rate from this year to next year is 5 percent. There are two ways that you can correct for this inflation.
    a. You can leave next year’s price in nominal terms and deflate by the nominal interest rate, as we did in the table.
    b. You can adjust next year’s price and put it in terms of today’s dollars, so next year’s price is a “real price.” Then you can discount using the real interest rate, which you can get from the Fisher equation.

Show that you get the same answer for the discounted present value using the second method as using the first method. (Note: when the interest rate is 10 percent, you should get exactly the same answer; when the interest rate is 5 percent, there will be a very small difference because the Fisher equation is an approximation.)

Can you think of an exogenous event that would cause the demand curve but not the supply curve for an asset to shift?

(Advanced) Explain why the DJIA and other stock market indices are more useful after they have been adjusted for inflation.

Economics Detective

1. Find data from a stock exchange in another country. Create a version ofFigure 9.3 "The DJIA: October 1928 to July 2007" for that stock exchange.
2. What is the current annual return on US government bonds? What is the current annual return on government bonds issued by Argentina? How would you explain the differences in returns?
3. Find recent data on the yields on the debt of Ireland, Spain, and Portugal. What happened to these yields, relative to the yield on German debt, in both October 2010 and November 2010? How might you explain the patterns you find? (Hint: think about our discussion of the riskiness of bonds.)
4. The chapter opened with a discussion of the stock market in Shanghai. Suppose you wanted to buy shares of a company trading on that exchange. How would you go about doing that?
5. Look at data on housing prices in your area. Do they fluctuate as much as stock prices?

Spreadsheet Exercise

1. Suppose an orange tree lives for three years, with a crop of 5 oranges in the first year, 3 in the second year, and 10 in the third year. The price is $2 in the first year and $5 in the second and third years. If the interest rate is 20 percent the first year and then 10 percent the next two years, what is the price of the orange tree?

TABLE 9.4 DISCOUNTED PRESENT VALUE EXERCISE

Year	Number of Oranges	Price of Orange	Revenue	Interest Rate
1	5	1.00		0.05
2	6			0.10

Year	Number of Oranges	Price of Orange	Revenue	Interest Rate
3	4			0.075
4	10			0.20

9.6 Appendix: A General Formulation of Discounted Present Value

This section presents a more general way of thinking about discounted present value. The economic idea is the same as the one we encountered when discussing the pricing of orange trees. Here the idea is to isolate the central ideas of discounted present value. We then use this more general formulation to talk about the pricing of stocks in an asset market.

We begin by defining the *t*-period real interest factor between the present date and some future date *t* years from now. The *t*-period real interest factor is simply the amount by which you must discount when calculating a discounted present value of a flow benefit (already adjusted for inflation) that will be received *t*years from now.

Suppose we have an asset that will provide real dividend payments every year for *t* years. Suppose that D_t is the real dividend in period *t*, and R_t is the real interest factor from the current period to period *t*. Then the price of the asset is given by

price = dividend in year 1 1-period real interest factor +dividend in year 22-period real interest factor + ... +dividend in year T T-period real interest factor

or

$$q=D_1R_1+D_2R_2...+D_TR_T.$$

All we did was to divide the dividends (*D*) due in period *t* by the interest factor R_t and then add them together.

If interest rates are constant over time, then the interest factors are easy to determine. Suppose that the annual real interest rate for one year is *r*. Then $R_1 = (1 + r)$ because this is the factor we would use to discount from next year to the present. What about discounting dividends two periods from now? To discountD_2 to period 1, we would divide by $(1 + r)$. To discount that back again to the current period we would again divide by $(1 + r)$. So to discount D_2 to the present we divide D_2 by $(1 + r) \times (1 + r) = (1 + r)^2$. That is, $R_2 = (1 + r)^2$. In general, $R_t = (1 + r)^t$ when interest rates are constant.

If real interest rates are not constant over time, the calculation of R_t is more tedious. If $R_1 = (1 + r_1)$, then $R_2 = (1 + r_1) \times (1 + r_2)$, where r_2 is the real interest rate between period 1 and period 2. In the calculation of R_2, you can think of $(1 + r_2)$ as discounting the flow from period 2 to period 1 and then $(1 + r_1)$ as discounting the flow from period 1 to period 0.

Chapter 10
Raising the Wage Floor
Working at Minimum Wage

Even in rich economies like those of the United States or Western Europe, there are numerous jobs where the level of pay is very low. Perhaps you have experienced this yourself—for example, waiting tables, bagging groceries, or working at a fast-food restaurant. Strikingly, many of these jobs pay exactly the same hourly wage. In 2010 in the United States, for example, the wage for jobs in fast-food restaurants was often $7.25. If you worked for Burger King in Georgia or Arby's in Iowa, you were likely to receive exactly the same wage. In Washington State, you would have earned more—$8.67 an hour—but you would have again found that many different employers were offering exactly the same wage. Had you looked for a job in a fast-food restaurant in 1995 in the United States, you would probably have been offered $4.25. The story is similar in many other countries. In New Zealand, the wage at fast-food restaurants in 2010 was typically NZ$12.75 (about $9.50); in France it was €9.00 (about $12.50).

The fact that different US employers from Wisconsin to Pennsylvania offer the same hourly wage is not a coincidence. It is the result of legislation by the federal government that sets a lower limit on the wage that firms can pay. Such regulations are called *minimum wage laws*, and they are found in many different countries. Figure 10.1 "US Department of Labor Poster" is a poster you might have seen where you have worked. This is from the US Department of Labor and outlines your rights as an employee. This chapter is about the origins and consequences of the government intervention summarized in this poster.

Governments enact such laws because they want to ensure that those who work earn a "living wage." Were you to work in the United States at the current federal minimum wage for 40 hours a week, 50 weeks a year (a total of 2,000 hours), you would earn $14,500. This is slightly above the current poverty level for an individual (which is $11,369) but is well below the average income in the United States. Without minimum wage legislation, the wage earner in a family could have a full-time job, work hard every day, and still not be able to keep the family out of poverty.

Minimum wage laws have been in existence in some parts of the world for a long time. The Industrial Conciliation and Arbitration Act of New Zealand set a minimum wage more than a century ago, in 1894. The first minimum wage in the United States was established in Massachusetts in 1912. Working conditions at that time were terrible in comparison to those in modern developed economies. Women, men, and children worked long hours in very dangerous working conditions for extremely low pay. It was quite natural, confronted with these sweatshops, to feel that the government could do more to actively protect the rights of workers and secure a fair standard of living for them. Those of us fortunate to live in rich economies are now largely spared from such working conditions, but in much of the world, people continue to work in unsafe and unhealthy conditions for very low pay.

The US federal government first established a minimum wage in 1938, as part of the Fair Labor Standards Act. Not all workers were covered by this act, however. The US Constitution charges the federal government with the duty of regulating interstate trade, so the act originally covered only those workers who were involved with trade that crossed state lines. Over time, however, amendments to the legislation have increased its coverage, and it now applies to all workers. [1]

Prior to the Fair Labor Standards Act, many states instituted their own minimum wage laws, and minimum wages still differ from state to state. For example, Oregon's current minimum wage is $8.50, about 17 percent higher than the federal minimum. These state-by-state differences complicate the life of economic historians who wish to study minimum wages. But for the economic analyst, these differences are extremely valuable because they are like an experiment: we can compare the experiences of different states with different laws.

In the United States, the minimum wage was raised in 2007, 2008, and 2009. Prior to that, the minimum wage had been constant for a decade; it had last been raised in 1997 following an act of Congress passed in 1995. President Clinton's 1995 message to the Congress, accompanying his minimum wage proposal, laid out arguments for the minimum wage increase. The following quote is from the Congressional Record:

To the Congress of the United States:

I am pleased to transmit for your immediate consideration and enactment the 'Working Wage Increase Act of 1995.' This draft bill would amend the Fair Labor Standards Act to increase the minimum wage in two 45 cents steps—from the current rate of $4.25 an hour to $4.70 an hour on July 4, 1995, and to $5.15 an hour after July 3, 1996.

…

To reform the Nation's welfare system, we should make work pay, and this legislation would help achieve that result. It would offer a raise to families that are working hard, but struggling to make ends meet. Most individuals earning the minimum wage are adults, and the average worker affected by this proposal brings home half of the family's earnings. Numerous empirical studies indicate that an increase in the minimum wage of the magnitude proposed would not have a significant impact on employment. The legislation would ensure that those who work hard and play by the rules can live with the dignity they have earned.

I urge the Congress to take prompt and favorable action on this legislation. [2]

President Clinton's words were forceful, and legislators in many different countries have been convinced by arguments such as these. Yet despite their widespread existence, minimum wage laws are highly contentious. Some commentators and analysts think that minimum wage laws are badly misguided and do much more harm than good.

The superficial appeal of minimum wage legislation is clear: it allows us to know that people who hold down jobs will at least earn a basic living wage—hard work will be rewarded by a minimum standard of living. Sometimes this is expressed differently: those who work hard should receive a "fair" wage for their efforts. (Fair is in quotation marks because not everyone agrees on what is fair and what is unfair, so it is hard to define exactly what the word means here.) [3] In other words, minimum wage legislation has a redistributive goal: the aim is to put more of society's resources in the hands of the working poor.

Economics, however, teaches us that many policy actions can have unintended consequences. To assess whether it is a good idea for the government to intervene in this manner, we need to develop a framework for understanding the effects of minimum wage laws. In particular, we want to answer the following questions:

1. *What are the consequences of a statutory minimum wage?*
2. *Why is there so much disagreement about whether the minimum wage is a good idea?*

We will not tell you whether or not the minimum wage is a good thing. By the end of this chapter, you should be in a position to make your own informed opinion about this controversial public policy question.

Road Map

Because the minimum wage says that firms are not allowed to pay below a certain price for the labor they hire, it is natural that our analysis focuses on the labor market. Of course, there is no single labor market—rather, we might think about there being many different markets for different types of skilled individuals. Lawyers, plumbers, engineers, web designers, and airline pilots earn much more than the minimum wage. In this chapter, though, we are focusing on people who earn relatively low wages, which means that we should look at the unskilled labor market. Sellers of labor in this market are not bringing any specialized skills; buyers of labor are not looking for any particular qualifications. The unskilled labor market is largely a market for time.

The minimum wage is set in terms of money—dollars in the United States, euros in France, and so on. Over time, increases in prices can erode the value of the minimum wage. We therefore begin this chapter by explaining how to adjust for the effects of higher prices. We then turn to the unskilled labor market. We look at what happens when we impose a minimum wage in that market, and then we look at what happens when the minimum wage changes.

Who is affected by minimum wage changes? Recognizing that people move in and out of jobs, we go beyond a supply-and-demand framework and consider these dynamic changes in the labor market. When we take into account these movements, we obtain a more sophisticated answer to this question. Then, once we are done with theory, we turn to evidence. We look at what different studies have found about the effects of changes in the real wage, and we assess how well these studies match up with our theory.

[1] See the following discussion for more details: US Department of Labor, Wage and Hour Division (WHD), "Fact Sheet #14: Coverage Under the Fair Labor Standards Act (FLSA)," revised July 2009, accessed March 14, 2011,http://www.dol.gov/whd/regs/compliance/whdfs14.htm.

[2] *Congressional Record*, February 13, 1995 (House), page H1677-H1678, accessed March 3, 2011, http://frwebgate2.access.gpo.gov/cgi-bin/TEXTgate.cgi?WAISdocID=C6M4yE/0/1/0&WAISaction=retrieve.

[3] Chapter 12 "Superstars" contains more discussion.

10.1 Nominal Wages and Real Wages

LEARNING OBJECTIVES

1. What is the difference between the real minimum wage and the nominal minimum wage?
2. What determines the equilibrium real wage and the level of employment?

When the federal minimum wage was first introduced in the United States in 1938, it was set at $0.25 per hour. Since then, Congress has raised the minimum wage several times. Figure 10.2 "Nominal Federal Minimum Wage in the United States" shows the minimum wage since 1938. You can see that the wage increases in steps whenever Congress enacts an increase in the wage.

The repeated increases in the minimum wage are not primarily due to the increased generosity of the US Congress. As you probably know, prices and wages have also tended to increase over time—a process we call inflation. The price level in 2007 was, on average, 14.8 times higher than in 1938, so $0.25 per hour is equivalent in modern terms to $3.70 ($0.25 × 14.8 = $3.70). Most of the increase in the minimum wage has simply been about keeping up with inflation. That said, the current minimum wage is $7.25, so the federal minimum wage has increased faster than the rate of inflation since its inception.

We call the wage in dollars the *nominal wage*. It is not the most useful measure of the amount that workers are receiving. Inflation means that a given nominal wage becomes worth less and less over time in terms of the goods and services that it buys. Between 1938 and 1957, for example, the general price level in the United States doubled. Had there been no change in the minimum wage, the $0.25 per hour minimum wage would have been worth only half as much in 1957 as it was when it was established.

From Nominal to Real Wages

The nominal wage is the wage measured in money (dollars in the United States). The real wage is the nominal wage in an economy adjusted for changes in purchasing power. It is defined as the nominal wage divided by the general price level:

$$\text{real wage} = \text{nominal wage}\text{price level}.$$

Workers care about the real wage, not the nominal wage, because the real wage captures the trade-off between leisure time and goods and services. Firms care about the real wage, not the nominal wage, because it measures the true cost of hiring labor. Figure 10.3 "Real Minimum Wage in the United States" shows the*real minimum wage*—that is, the minimum wage adjusted for inflation. The real minimum wage is defined as

$$\text{real minimum wage} = \frac{\text{nominal minimum wage}}{\text{price level}}.$$

Toolkit: Section 17.8 "Correcting for Inflation"

The conversion from nominal wages to real wages is an example of the more general idea of correcting for inflation. If you have data expressed in nominal terms (for example, in dollars) and want to covert them to real terms, you should follow the following four steps:

1. Select your deflator. In most cases, the Consumer Price Index (CPI) is the best deflator to use. You can find data on the CPI (for the United States) at the Bureau of Labor Statistics website (http://www.bls.gov).
2. Select your base year. Find the value of the index in that base year.
3. For all years (including the base year), divide the value of the index in that year by the value in the base year. Notice that the value for the base year is 1.
4. For each year, divide the value in the nominal data series by the number you calculated in step 3. This gives you the value in "base year dollars."

The real minimum wage increases in jumps whenever the nominal wage is increased, but it declines over time as it is eroded by inflation. The erosion of the minimum wage by inflation was recognized by President Clinton in the address that we quoted in our introduction. In that request to Congress he also said:

> *The first increment of the proposal simply restores the minimum wage to its real value following the change enacted in 1989.*
>
> *If the Congress does not act now, the minimum wage will fall to its lowest real level in 40 years. That would dishonor one of the great promises of American life—that everyone who works hard can earn a living wage. More than 11 million workers would benefit under this proposal, and a full-time, year-round worker at the minimum wage would get a $1,800 raise—the equivalent of 7 months of groceries for the average family.*

When President Clinton referred to “7 months of groceries,” he was converting the increase in the minimum wage into real terms, just as our technique for converting to inflation does. Instead of using the bundle of goods that goes into the CPI, however, he was using a bundle of goods representing groceries for the average family.

Nominal and Real Wages in the Labor Market

The challenge when analyzing the minimum wage is that it is set in nominal terms, but workers and firms care about the real minimum wage. To help us understand the difference, we begin with a specific numerical example of the**labor market**. Suppose we have the following labor supply-and-demand equations, where labor supply and labor demand are measured in hours:

labor supply = 10,000 × real wage

and

labor demand = 72,000 – 8,000 × real wage.

Think of this example as referring to the weekly demand for and supply of unskilled labor in a small city. It is reasonable to think of this as a **competitive market**, in which market participants will typically agree on a price at or close to the point where supply equals demand. In the supply-and-demand framework, the intersection of the supply and demand curves tells us the equilibrium price in the market and the equilibrium quantity traded. In the labor market, the place where supply and demand meet tells us the equilibrium wage and the equilibrium number of hours worked.

Toolkit: Section 17.3 "The Labor Market", and Section 17.9 "Supply and Demand"

You can find more detail about the underpinnings of labor market supply and demand and the workings of the competitive market in the toolkit.

First we solve for the equilibrium in this market. In equilibrium, the quantity of labor supplied equals the quantity of labor demanded, so

10,000 × real wage = 72,000 – 8,000 × real wage.

Add (8,000 × real wage) to each side:

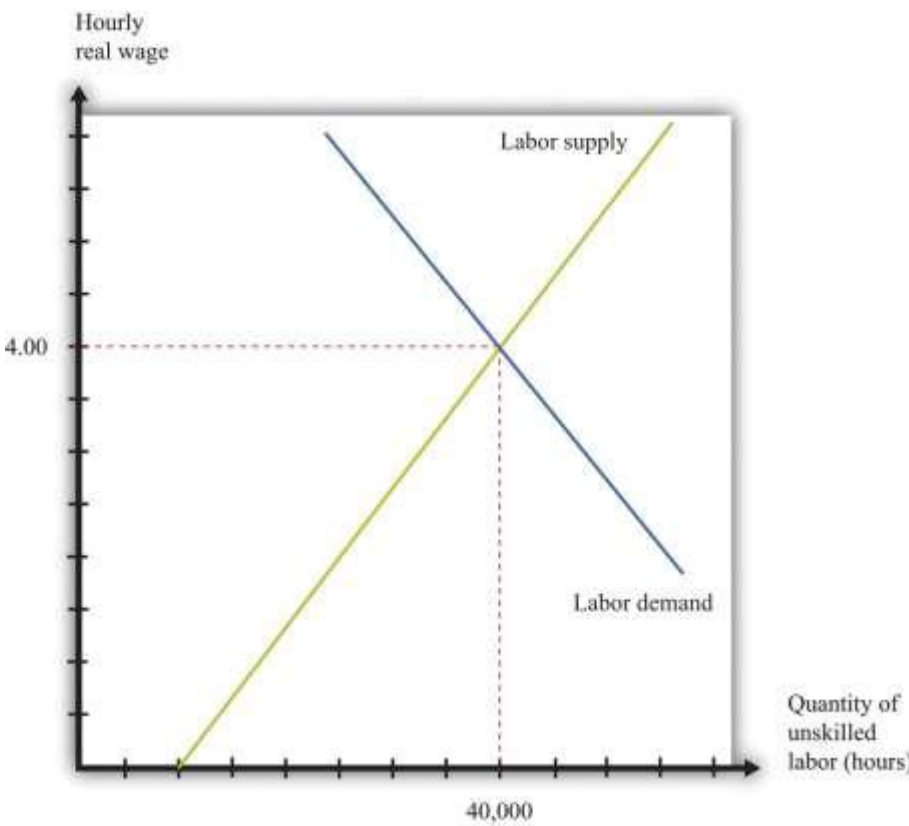
18,000 × real wage = 72,000.

Then divide both sides by 18,000 to obtain

real wage = 4.

If we plug this value of the real wage back into either the supply or the demand equation, we find that the equilibrium quantity of hours worked is 40,000 hours. For example, we might have 1,000 workers, each of whom works a 40-hour week. The equilibrium is illustrated in Figure 10.4 "Labor Market Equilibrium".

Figure 10.4 Labor Market Equilibrium



The market for unskilled labor is in equilibrium at an hourly wage of $4 and a total of 40,000 hours. In this diagram, we assume that the price level is 1, so the real wage equals the nominal wage.

Suppose that this example pertains to the base year. From our discussion of correcting for inflation, we know that in the base year we set the price level equal to 1. When the price level is 1, the real wage equals the nominal wage. In the initial year, therefore,

$$\text{labor supply} = 10{,}000 \times \frac{\text{nominal wage}}{1} = 10{,}000 \times \text{nominal wage}$$

and

$$\text{labor demand} = 72{,}000 - 8{,}000 \times \frac{\text{nominal wage}}{1} = 72{,}000 - 8{,}000 \times \text{nominal wage}.$$

In the base year, the nominal wage is $4 per hour.

Now imagine we have 10 percent inflation, which means that the price of all goods and services in the economy increases by 10 percent over the course of a year. If a household paid $100 a week for groceries last year, it must pay $110 this year; if a household used to pay $500 a month in rent, it must now pay $550; and so on. Turning this around, a dollar is worth less than it used to be; you need $1.10 to purchase what you could have bought for $1 this year. The price level has increased from 1 to 1.1.

To see what this means in terms of the labor market diagram, think about the situation at a given nominal wage, such as $2.20 per hour. Last year, when the price level was 1, households were willing to supply 22,000 hours (= 10,000 × 2.2). But $2.20 now is worth the equivalent of only $2, so households are willing to supply only 20,000 (10,000 × $\frac{2.2}{1.1}$ = 10,000 × 2) hours of labor instead. The same idea applies at every wage; households will supply only the amount of labor that they would previously have supplied when the wage was 10 percent higher.

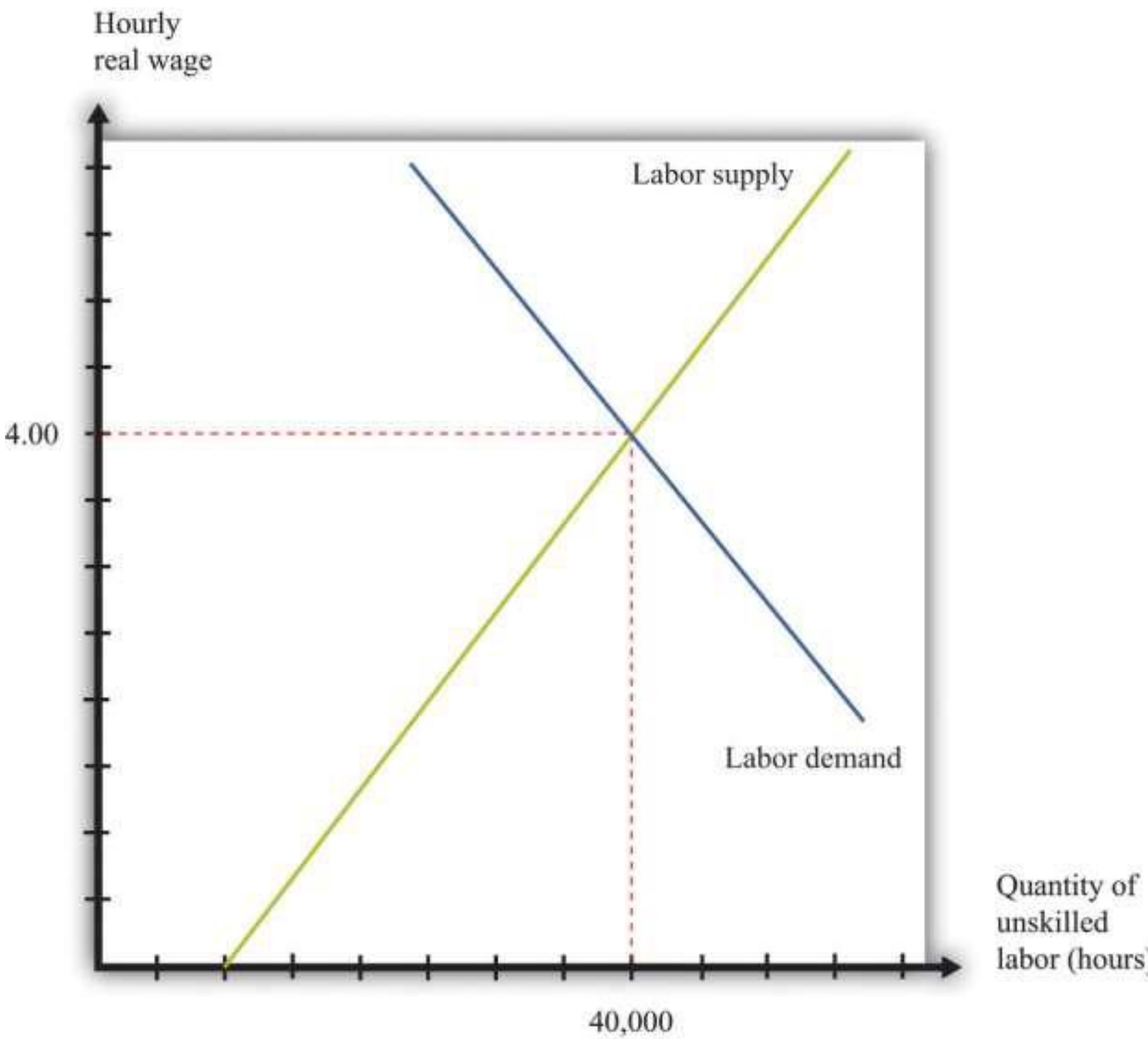
A similar logic applies to the demand for labor. The increase in the price level means that firms get 10 percent more dollars for the goods that they sell. As a consequence, the labor performed by workers generates more dollars than it used to. If it was worth paying $7 for an hour of work before, it is now worth paying $7.70 for that same hour of work.

In terms of real wages, however, nothing has changed. The equilibrium real wage is still $4, as it was before. But because

$$\text{real wage} = \frac{\text{nominal wage}}{\text{price level}},$$

the nominal wage must increase by 10 percent to match the increase in the price level. The equilibrium in the labor market is shown in Figure 10.5 "Labor Market Equilibrium after 10 Percent Inflation". It is no coincidence that this diagram looks exactly the same as Figure 10.4 "Labor Market Equilibrium"; that is the point. An increase in the price level is matched by an increase in the nominal wage, and nothing changes in terms of the real wage or the real equilibrium quantity of labor.

Figure 10.5 Labor Market Equilibrium after 10 Percent Inflation



If there is 10 percent inflation, the price level increases from 1 to 1.1, the real wage is unchanged, and the nominal wage increases by 10 percent.

KEY TAKEAWAYS

- The nominal minimum wage is set by governments. The real minimum wage is the real value of the nominal minimum wage. It is determined by dividing the nominal minimum wage by the price level.
- The levels of the real wage and employment are determined by labor market equilibrium.

CHECKING YOUR UNDERSTANDING

1. Looking at Figure 10.3 "Real Minimum Wage in the United States", explain why the real minimum wage increases very quickly but never decreases very quickly.
2. Why do labor demand and supply depend on the real and not the nominal wage?

10.2 The Effects of a Minimum Wage

LEARNING OBJECTIVES

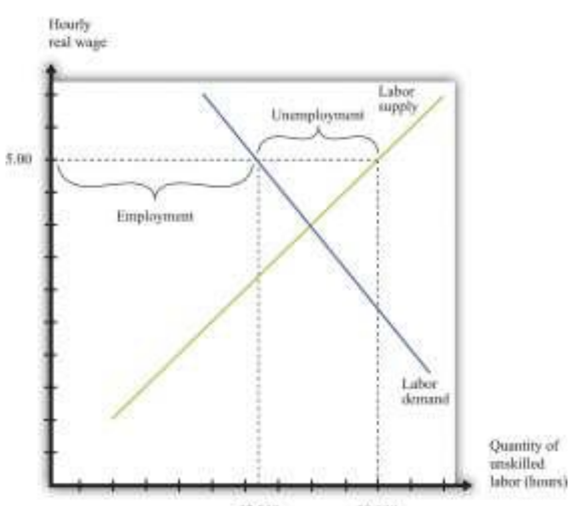
1. What happens when a government imposes a minimum wage?
2. If there is inflation under a minimum wage system, what happens to the level of employment?
3. What are the efficiency costs of a minimum wage?

Adam Smith, the 18th-century economist who founded modern economics, had a vivid metaphor for the idea that supply would equal demand in a competitive market: he referred to an "invisible hand" guiding markets to equilibrium. Joan Robinson, a famous economist at Cambridge University in the first half of the 20th century, wrote that "the hidden hand will always do its work but it may work by strangulation." [1] What she meant by this was there is no guarantee that the *equilibrium* wage in the market would in fact be a *living* wage.

The Imposition of a Minimum Wage

When the government imposes a minimum wage, firms are not permitted to pay less than the amount that the government mandates. Suppose we are again in the base year, so the price level is 1. Imagine that the market equilibrium wage is $4 per hour, but the government now passes legislation stating that all firms must pay at least $5 per hour. At this wage, supply does not equal demand. Figure 10.6 "Labor Market with a Minimum Wage" illustrates what happens.

Figure 10.6 Labor Market with a Minimum Wage



With a minimum wage of $5, the supply of labor is 50,000 hours, but firms demand only 32,000 hours of labor, so the labor market is not in equilibrium.

Markets are based on voluntary trades. In Figure 10.6 "Labor Market with a Minimum Wage", we see that sellers (the workers who supply labor) would like to sell 50,000 hours of labor to the market at the set minimum wage—that is, 250 more people would like to have a 40-hour-a-week job when the wage increases from $4 to $5. But firms wish to purchase only 32,000 hours of labor—firms want to hire 200 *fewer* workers (8,000 fewer hours). In a market with voluntary trade, no one can force firms to hire workers. As a result, the equilibrium quantity of labor traded in the market will be determined by how much the firms wish to buy, not how much workers want to sell.

We can now answer our first motivating question of the chapter: what is the consequence of imposing a minimum wage? Two things happen when the government imposes a minimum wage:

1. The amount of labor hired in the market decreases. In our example, the number of unskilled workers employed decreases from 1,000 to 800. Thus while those who have jobs earn a higher wage, there are now some individuals who no longer have jobs. Employment has decreased.
2. At the government-imposed wage, there are more people who want to work than are able to find jobs. Thus the minimum wage has created unemployment. Because 1,250 people would like jobs at a wage of $5 but only 800 jobs are available, 450 people are unemployed; they would like a job at the prevailing wage, but they are unable to find one.

The number of unemployed workers is 450, even though employment decreased by only 200 workers. The difference comes from the fact that the higher wage also means that more people want to work than before. In this case, the higher wage means 250 more people would like a job.

We have assumed in this discussion that everyone works for 40 hours, in which case the number of people employed must decrease by 200. Another possibility is that everyone who wants a job is able to get one, but the number of hours worked by each individual decreases. Because there are 32,000 hours of work demanded and 1,250 people who want jobs, each worker would work 25.6 hours a week. In this situation, we say that there is *underemployment* rather than unemployment. Yet another possibility is that, after the introduction of the minimum wage, the number of people employed stays the same as before (1,000), but those individuals are allowed to work only 32 hours per week. In this case, we have both underemployment (of the previously employed) and unemployment (of the extra workers who want a job at the higher wage). In real-life situations, there may be both unemployment and underemployment.

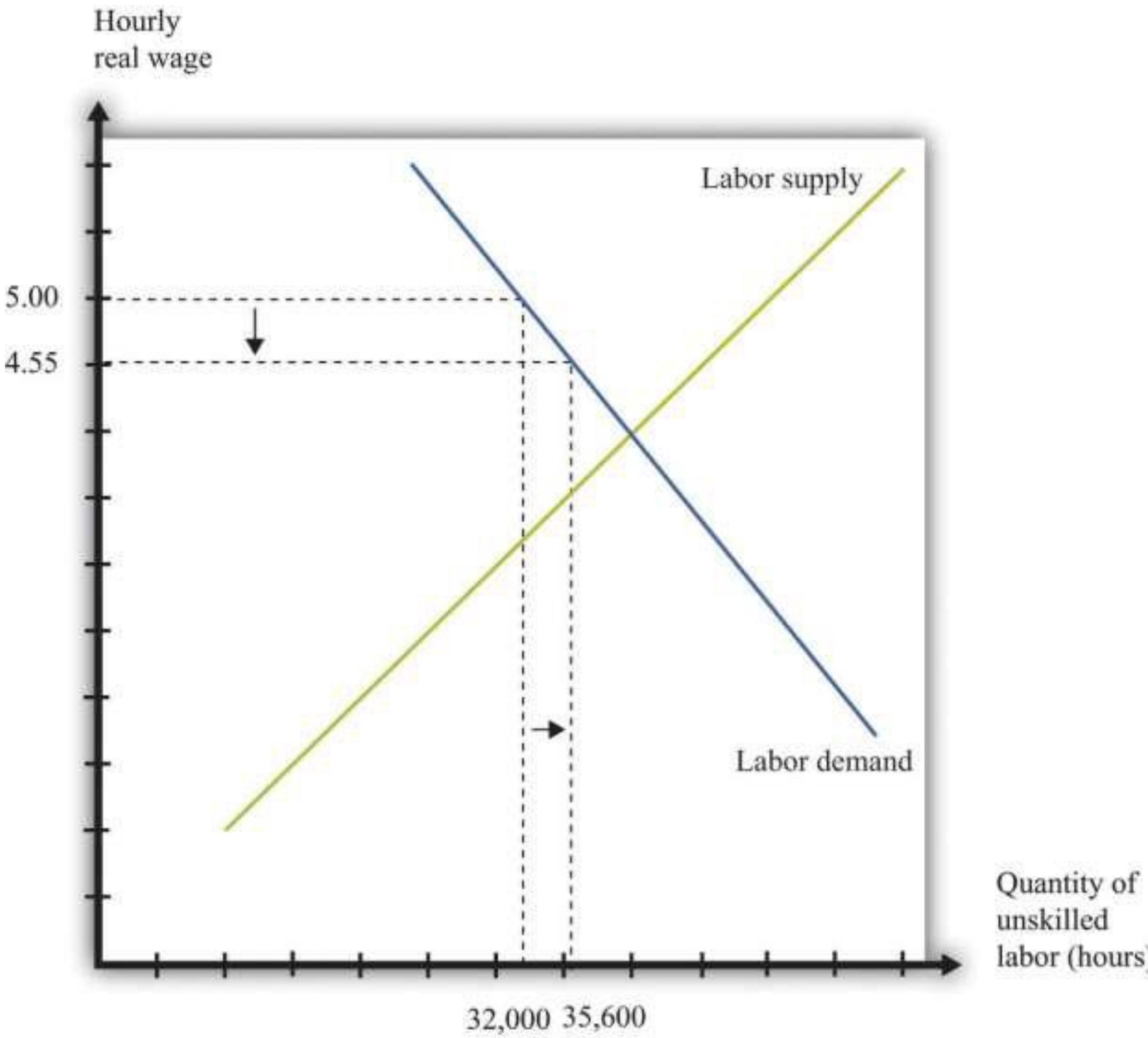
Inflation and the Minimum Wage

Although inflation made no difference to our basic analysis of the labor market, it does change our analysis of the minimum wage. Minimum wages are fixed in nominal terms and do not automatically change when there is inflation. So if the minimum wage is set at $5 and the price level increases from 1 to 1.1, the real minimum wage declines. Looking back at the definition of the real minimum wage, we find that

$$\text{real minimum wage} = \frac{\text{nominal minimum wage}}{\text{price level}} = \frac{5}{1.1} = 4.55.$$

The effect of a reduction in the real minimum wage is shown in Figure 10.7 "A Reduction in the Real Minimum Wage". At the lower real wage, firms are willing to hire more workers. Employment increases from 32,000 hours to 35,600 hours: 90 more people can find jobs.

Figure 10.7 A Reduction in the Real Minimum Wage

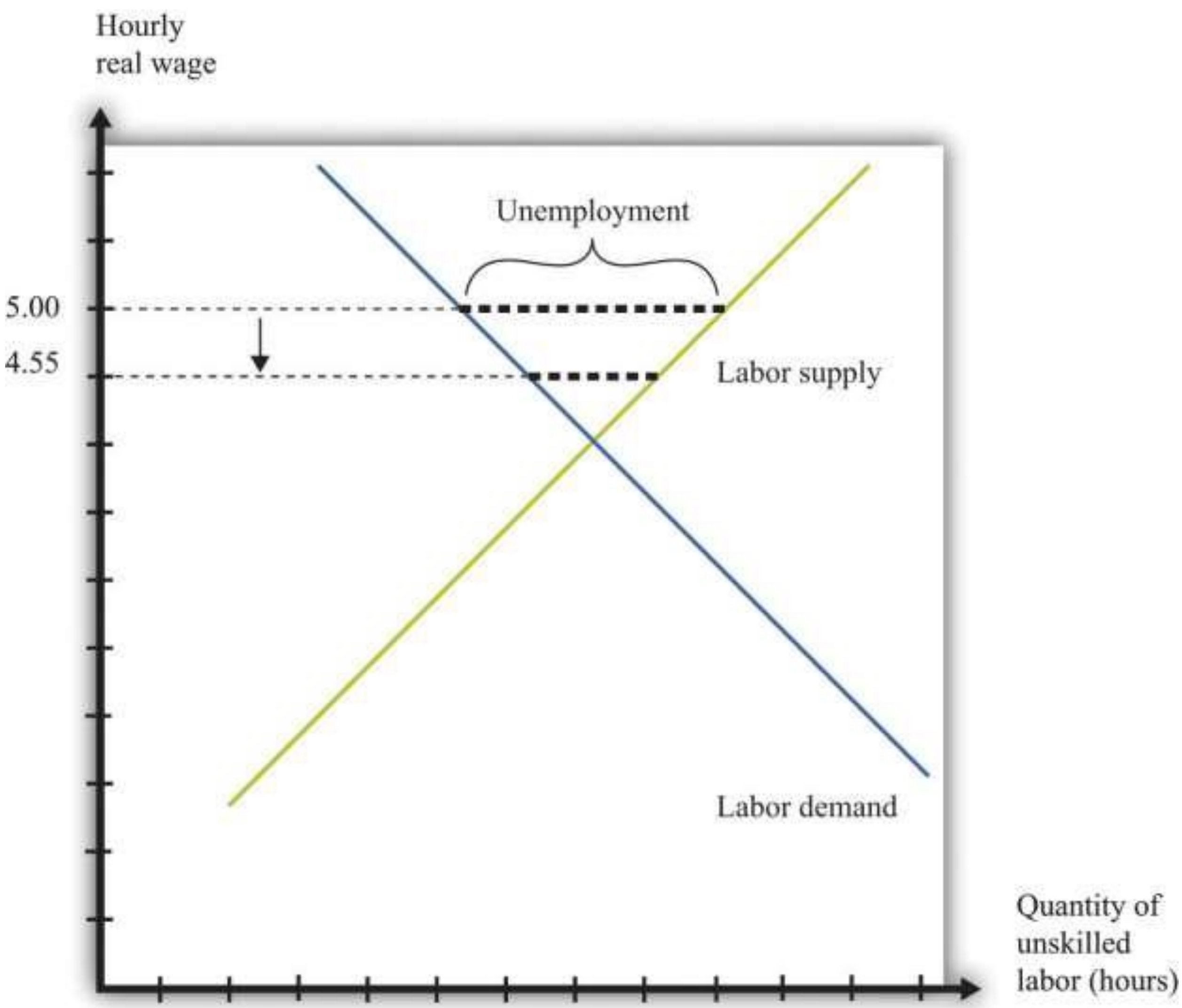


A 10-percent increase in the price level leads to a reduction in the real minimum wage to $4.55 and an increase in employment from 32,000 to 35,600.

In Figure 10.7 "A Reduction in the Real Minimum Wage", the real minimum wage of $4.55 is still higher than the equilibrium wage of $4.00. To put the same point another way, the equilibrium nominal wage has increased to $4.40, but this is still below the nominal minimum wage of $5.00. However, if the price level were to increase by 25 percent or more from the base year, the minimum wage would become completely irrelevant. The minimum wage would be below the market wage. Economists say that the minimum wage would no longer be "binding" in this case.

It is exactly this process of increasing prices that lies behind Figure 10.3 "Real Minimum Wage in the United States". As the price level increases, the minimum wage becomes worth less in real terms (and has less of an effect on employment). Eventually, Congress acts to increase the minimum wage to bring it back in line with inflation—although, as Figure 10.3 "Real Minimum Wage in the United States" shows, Congress has allowed the real minimum wage to decline substantially from its high point in the late 1950s. The reduction in the real minimum wage also leads to a reduction in unemployment, as shown in Figure 10.8 "Effects on Unemployment of a Reduction in the Real Minimum Wage".

Figure 10.8 Effects on Unemployment of a Reduction in the Real Minimum Wage



The reduction in the real minimum wage to $4.55 leads to a decrease in unemployment.

Efficiency Implications of a Minimum Wage

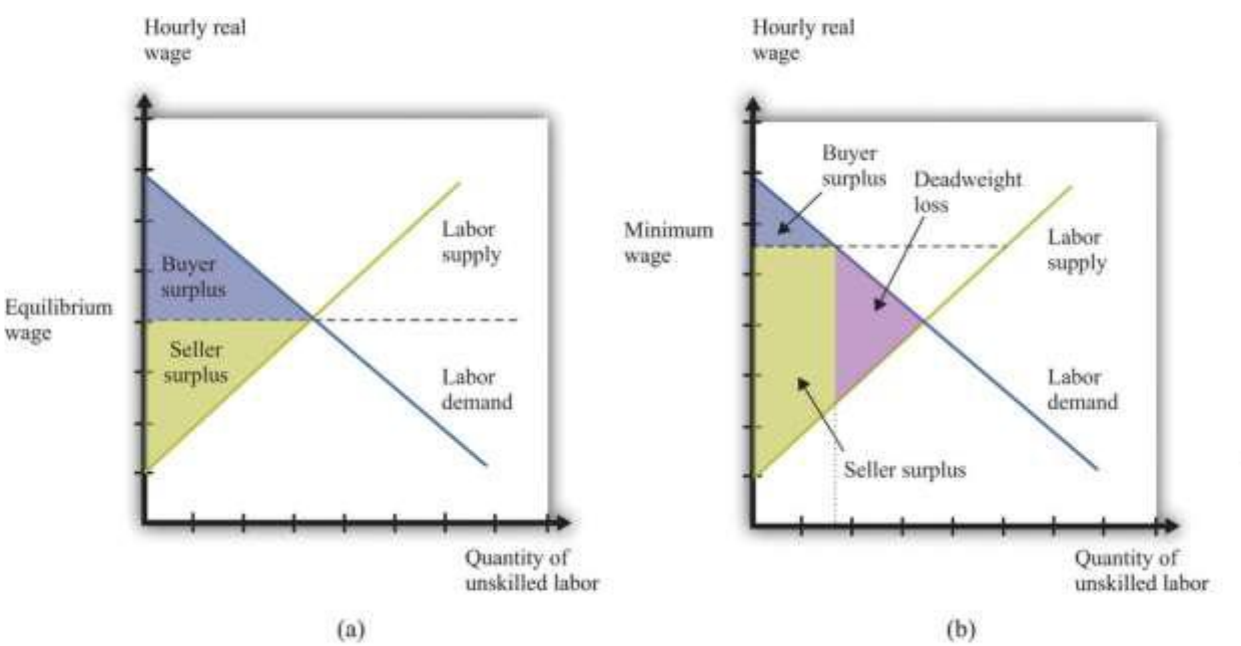
Markets are a mechanism that allow individuals to take advantage of gains from trade. Whenever a buyer has a higher valuation than a seller for a good or service, they can both benefit from carrying out a trade. This is how economies create value—by finding opportunities for mutually beneficial trades.

The minimum wage interferes with this process in the unskilled labor market. It reduces employment, which is the same as saying that fewer transactions take place. Because each voluntary transaction by definition generates a surplus, anything that reduces the number of transactions causes a loss of surplus. In economists' terminology, the minimum wage leads to a departure fromefficiency. We can represent that inefficiency graphically. Figure 10.9 "Deadweight Loss from Minimum Wage" shows the effect of the minimum wage, using the ideas of **buyer surplus** and **seller surplus**.

Toolkit: Section 17.10 "Buyer Surplus and Seller Surplus" and Section 17.11 "Efficiency and Deadweight Loss"

You can review the different kinds of surplus, as well as the concepts of efficiency and deadweight loss, in the toolkit.

Figure 10.9 Deadweight Loss from Minimum Wage



With no minimum wage (a), all the possible gains from trade in the market are realized, but with a minimum wage (b), some gains from trade are lost because there are fewer transactions.

In part (a) of Figure 10.9 "Deadweight Loss from Minimum Wage", we see the market without the minimum wage. In the labor market, it is the firm who is the buyer. The total buyer surplus is the profit that firms obtain by hiring these workers; it is the difference between the cost of hiring these workers and the revenues that they generate. Graphically, it is the area below the labor demand curve and above the market wage. The total sellersellers' surplus#8217;s surplus is the benefit that accrues to workers from selling labor time. Sellers of labor (workers) receive surplus equal to the area below the market wage and above the supply curve.

In part (b) of Figure 10.9 "Deadweight Loss from Minimum Wage", we show the effect of the minimum wage. As we already know, the higher wage leads to a reduction in employment. Fewer transactions occur, so the total surplus in the market is reduced. Economists call the lost surplus the deadweight loss from the minimum wage policy.

The most obvious cost of the minimum wage is this loss of surplus. But there may be other hidden costs as well. Whenever people are prevented from carrying out mutually beneficial trades, they have an incentive to try to get around these restrictions. Firms and workers may try to "cheat," conducting hidden transactions below the minimum wage. For example, a firm might pay a worker for fewer hours than he or she actually worked. Alternatively, a firm might reduce other benefits of the job. Such cheating not only subverts the minimum wage law but also uses up resources because the firm and the worker must devote effort to devising ways around the law and ensuring that they do not get caught.

The loss in surplus could also be greater than is shown in Figure 10.9 "Deadweight Loss from Minimum Wage". The figure is drawn under the presumption that the trades taking place in the labor market are the ones that generate the most surplus. But suppose that the minimum wage is $5.00. It is possible that someone who would be willing to work for, say, $2.00 an hour loses her job, whereas someone willing to work for, say, $4.50 is employed.

Remember that if there were no restrictions in the labor market, the wage would adjust so that anyone wanting to work could find a job. This means that both the person willing to work for $2.00 and the person willing to work for $4.50 could both find a job as long as the wage is above $4.50. If the equilibrium wage were $4.00, then the person willing to work for only $4.50 would not be employed. In either case, this is the efficient outcome, consistent with obtaining all the gains from trade.

KEY TAKEAWAYS

- When the government imposes a minimum wage, the real wage is determined by the minimum wage divided by the price level, not by the interaction between labor supply and demand.
- If there is inflation and a fixed nominal minimum wage, then the level of employment will increase and the real minimum wage will decrease.
- The minimum wage creates deadweight loss because some trades of labor services do not take place.

CHECKING YOUR UNDERSTANDING

1. Draw a diagram for a labor market where the minimum wage is not binding.
2. What happens to the real minimum wage and the level of employment if there is *deflation*—that is, if the price level decreases?

[1] Joan Robinson, “The Pure Theory of International Trade,” in *Collected Economic Papers*(Oxford: Basil Blackwell, 1966), 189.

10.3 Minimum Wage Changes

LEARNING OBJECTIVES

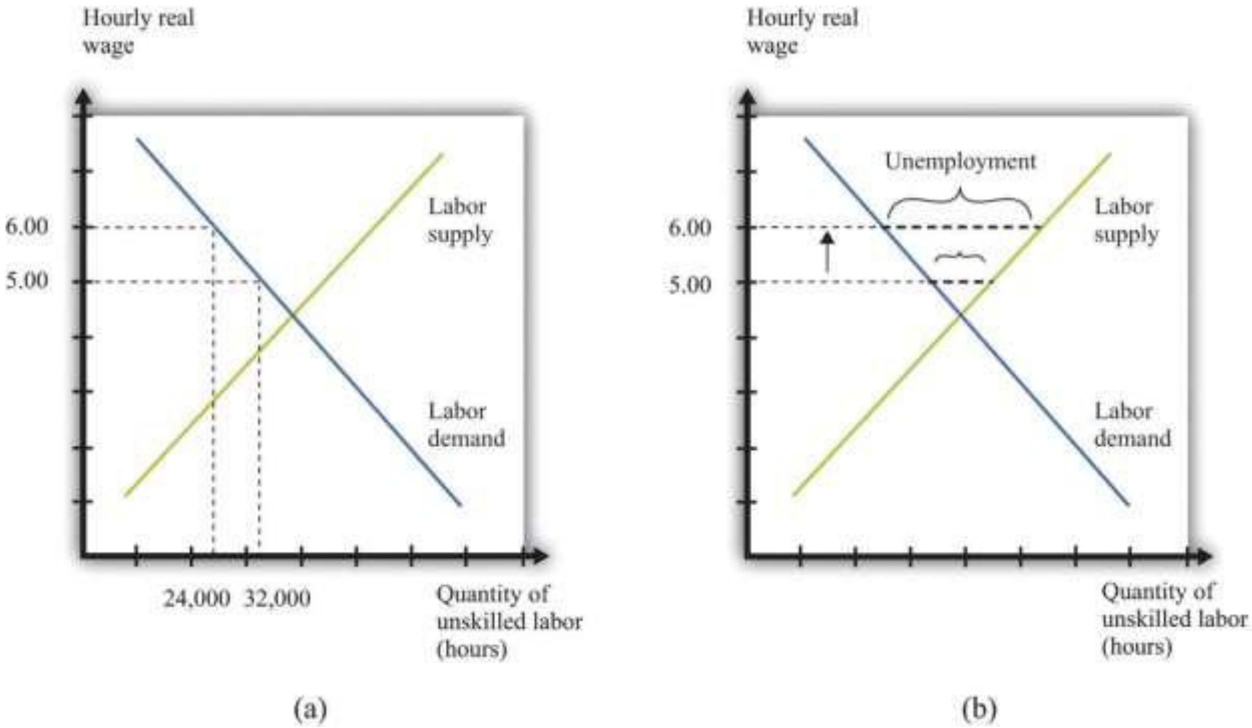
1. What happens to the levels of employment and unemployment if the real minimum wage increases?
2. What determines the size of the change in employment when the real minimum wage increases?
3. What determines the size of the change in unemployment when the real minimum wage increases?

Suppose that the government is considering an increase in the minimum wage. What should we expect to happen? How will firms and workers respond? One might be tempted simply to ask firms what they would do in the face of an increase in the minimum wage. Unfortunately, this is likely to be both infeasible (or at least prohibitively expensive) and inaccurate. It would be an immense amount of work to interview all the firms in an economy. What is more, there is no guarantee that managers of firms would give accurate answers if they were asked hypothetical questions about a change in the minimum wage. Instead, government statisticians use statistical sampling techniques to interview a random sample of firms in an economy, and they ask them about their actual behavior—they ask questions such as the following: "How many workers do you employ at present?" and "How much do you pay them?" The data from such surveys are useful but do not directly help us determine the effects of a change in the minimum wage. For this we need more theory.

The Effect of a Minimum Wage Increase on Employment and Unemployment

Figure 10.10 "Effects of Increasing the Real Minimum Wage" amends our view of the labor market to show an increase in the minimum wage from $5 to $6. (We suppose that the price level is constant, so an increase in the nominal minimum wage implies an increase in the real minimum wage.) The increase in the minimum wage leads to a reduction in the level of employment: employment decreases from 32,000 to 24,000. Labor is now more expensive to firms, so they will want to use fewer hours. At the same time, the higher minimum wage means that more people would like jobs. The increase in the amount of labor that people would like to supply, and the decrease in the amount of labor that firms demand, both serve to increase unemployment.

Figure 10.10 Effects of Increasing the Real Minimum Wage



An increase in the value of the hourly real minimum wage from $5 to $6 leads to a decrease in employment from 32,000 hours to 24,000 hours (a) and an increase in unemployment (b).

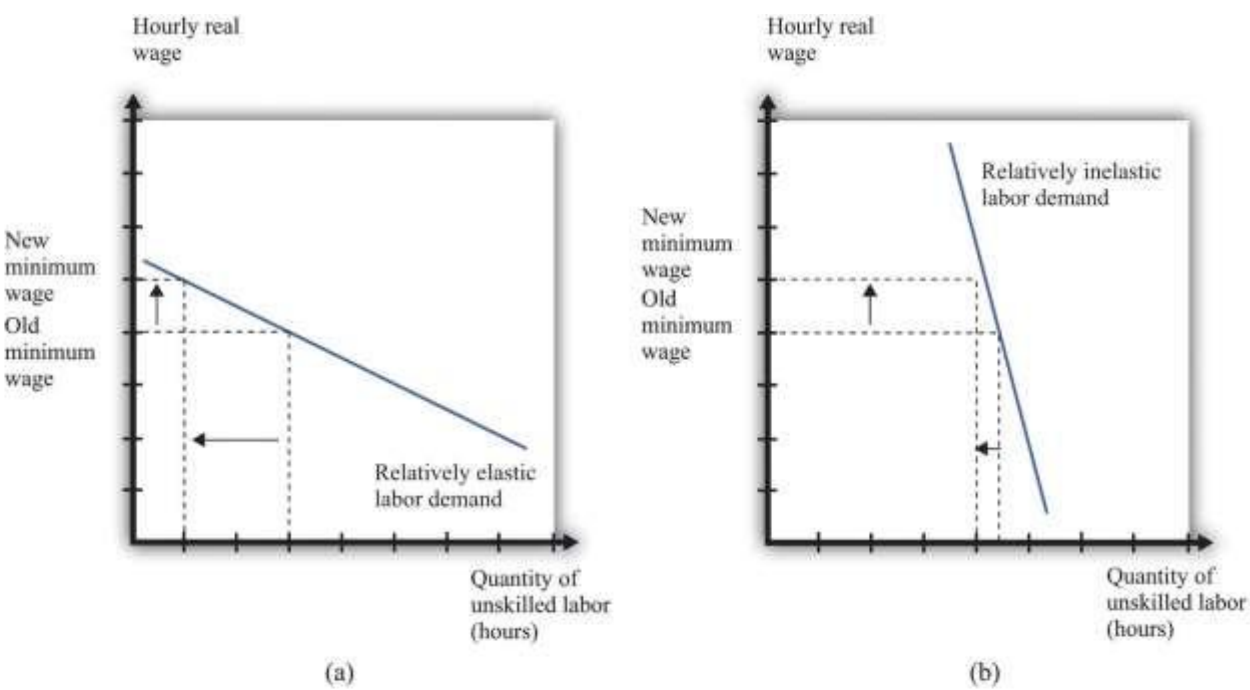
Our model generates a *qualitative* prediction: an increase in the minimum wage will decrease employment and increase unemployment. At the same time, the wage increase will ensure that those with jobs will earn a higher wage. So we can see that there may be both advantages and disadvantages of increasing the minimum wage. To go further, we have to know how big an effect such a change would have on employment and unemployment—that is, we need the *quantitative* effects of a higher minimum wage.

To understand the quantitative effects, we want to know when to expect big or small changes in employment or unemployment—which depends on the wage elasticity of labor demand and the labor supply. Remembering that the wage is simply the price in the labor market, the wage elasticity of demand is an example of the price elasticity of demand in a market:

wage elasticity of labor demand = percentage change in labor demandpercentage change in real wage.

From Figure 10.10 "Effects of Increasing the Real Minimum Wage", we can see that the wage elasticity of labor demand tells us everything we need to know about the effects of a change in the wage on employment. If the demand curve is relatively elastic, then a change in the minimum wage will lead to a relatively large change in employment. If the demand curve is relatively inelastic, then a change in the minimum wage will lead to a relatively small change in employment. This is intuitive because the elasticity of labor demand tells us how sensitive firms' hiring decisions are to changes in the wage. An elastic demand for labor means that firms will respond to a small change in the wage by laying off a large number of workers, so the employment effect will be large. The elasticity of labor supply is not relevant if we are concerned only with employment effects. This is illustrated in Figure 10.11 "The Employment Effect of a Change in the Minimum Wage" and summarized in Table 10.1 "Employment Effects of a Change in the Real Minimum Wage".

Figure 10.11 The Employment Effect of a Change in the Minimum Wage



If labor demand is relatively elastic (a), a change in the minimum wage has a big effect on employment, while if labor demand is relatively inelastic (b), the same change in the minimum wage has a much smaller effect on employment.

Table 10.1 Employment Effects of a Change in the Real Minimum Wage

	Effect on Employment
Elastic demand	Large change
Inelastic demand	Small change

If we are interested in the effect on unemployment, however, we must look at both demand and supply. A worker is counted as unemployed if he or she is looking for a job but does not currently have a job. The labor supply curve tells us how many workers are willing to work at a given wage; those who are not employed are looking for a job. To understand the effects of the minimum wage on unemployment, we need to look at the mismatch between supply and demand at the minimum wage, so we must look at the supply of labor as well as the demand for labor. The price elasticity of supply measures the responsiveness of the quantity supplied to a change in the price: in the case of the labor market, we obtain the wage elasticity of labor supply:

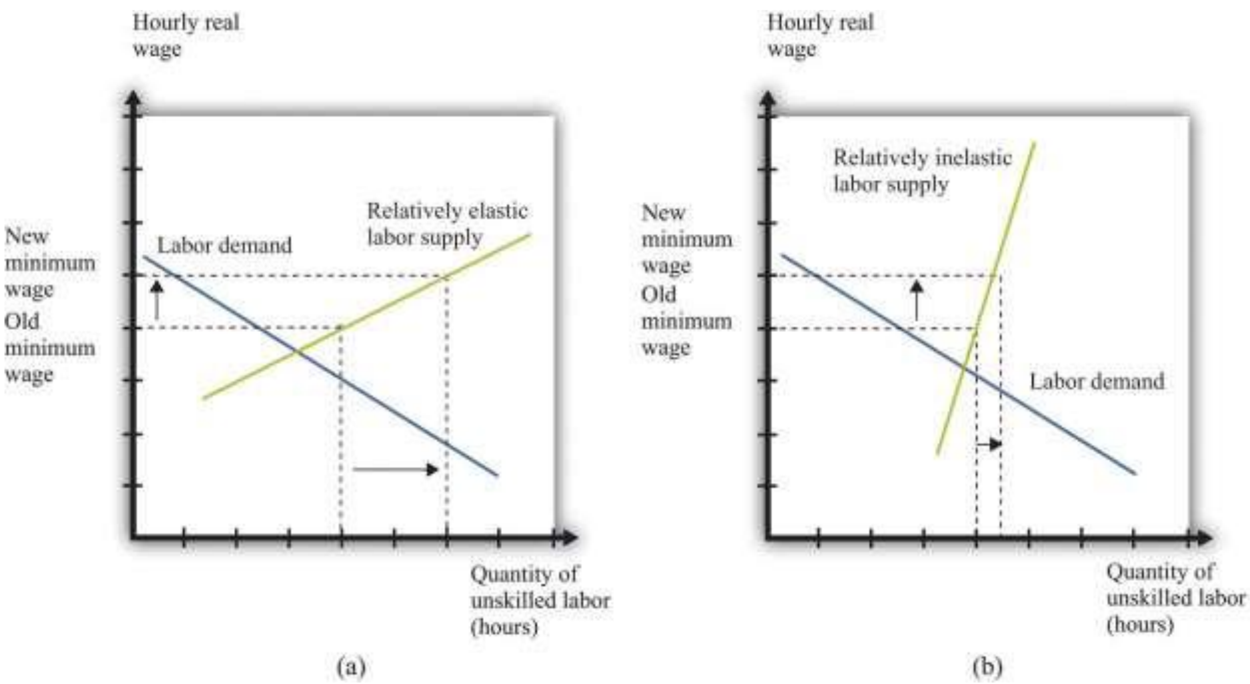
wage elasticity of labor supply = percentage change in labor supplypercentage change in real wage.

Toolkit: Section 17.2 "Elasticity"

You can review the general definition and calculation of elasticities in the toolkit.

The more elastic the labor supply curve, the bigger the change in labor supply for a given change in the real wage (Figure 10.12 "The Unemployment Effect of a Change in the Minimum Wage"). A bigger change in labor supply means a bigger change in unemployment. Combining this with Table 10.1 "Employment Effects of a Change in the Real Minimum Wage", we get the results summarized in Table 10.2 "Unemployment Effects of a Change in the Real Minimum Wage". If demand and supply are both inelastic, the change in the minimum wage has little effect on unemployment. The higher wage does not make much difference to firms' hiring decisions (inelastic demand), and it does not induce many additional workers to look for a job (inelastic supply). The overall effect on unemployment is small. By contrast, if both curves are elastic, then an increase in the wage will lead to a big decrease in the number of jobs available and a big increase in the number of job seekers. If we can find good estimates of the elasticities of labor demand and supply, we will be able to make good predictions about the likely effect of an increase in the minimum wage.

Figure 10.12 The Unemployment Effect of a Change in the Minimum Wage



If the labor supply is relatively elastic (a), a change in the minimum wage has a big effect on unemployment, while if the labor supply is relatively inelastic (b), the same change in the minimum wage has a much smaller effect on unemployment.

Table 10.2 Unemployment Effects of a Change in the Real Minimum Wage

	Effects on Unemployment	
	Elastic Supply	**Inelastic Supply**
Elastic demand	Very large change	Large change
Inelastic demand	Large change	Small change

KEY TAKEAWAYS

- All else being the same, an increase in the real minimum wage will reduce employment and increase unemployment.
- The size of the change in employment when the minimum wage increases is determined by the elasticity of the labor demand curve.
- The size of the change in unemployment when the minimum wage increases is determined by the elasticities of the labor demand and supply curves.

CHECKING YOUR UNDERSTANDING

1. Why doesn't the elasticity of labor supply matter for the effects of changes in the real minimum wage on employment?
2. If prices increase, what will happen to the level of unemployment when there is a binding minimum wage?

10.4 The Minimum Wage and the Distribution of Income

LEARNING OBJECTIVES

1. Which parts of the economy are affected by the minimum wage?
2. When the minimum wage increases, who gains and who loses?
3. What is an equity-efficiency trade-off?

We said earlier that governments impose minimum wages because they care about ensuring that the working poor earn a fair wage. Another way of saying this is that the minimum wage is an intervention by the government that is meant to change the distribution of society's resources. If unskilled workers are going to earn more, then this means they are obtaining more of the total resources available in an economy. And if they are getting more, then somebody else must be getting less. We would like to have some way of thinking about the effects of the minimum wage on the distribution of income.

To talk about distribution, we need to divide society into groups and then examine how much each group gets. One group is obviously those who receive the minimum wage—the working poor. Another group we need to consider is the unemployed. And then there is everybody else: all of those individuals who are sufficiently skilled to have jobs that pay more than the minimum wage. This is a large group, encompassing electricians and CEOs, but—for our present purposes—it makes sense to group them all together. Let's call them the "relatively rich," by which we mean that they are richer than unemployed or employed unskilled workers. So we have three groups: the unemployed, the working poor, and the relatively rich. How are these groups affected by an increase in the minimum wage?

Winners and Losers from the Minimum Wage

We know that the working poor are made better off by an increase in the minimum wage—after all, the whole point of the increase is to ensure that these individuals earn more. We can see this gain in Figure 10.9 "Deadweight Loss from Minimum Wage": it is the extra surplus that sellers obtain at the expense of buyers. Remember, though, that some of the working poor will lose their jobs as a result of the higher minimum wage. So our first conclusion is that those members of the working poor who keep their jobs are unambiguously made better off.

The increase in the minimum wage means that there are more people who are unemployed. Those who are *already* unemployed are not directly affected by the increase in the minimum wage. These unemployed individuals may be indirectly affected, however, because it becomes harder for them to find jobs.

Are there any effects on the relatively rich? The answer is yes. The increase in the minimum wage means that firms will earn lower profits. We can see this because buyer surplus is reduced in Figure 10.9 "Deadweight Loss from Minimum Wage". Although firms are just legal entities, they are owned by individuals. When a firm earns lower profits, the shareholders of that firm receive lower income. The working poor and the unemployed are not, for the most part, individuals with portfolios of stocks; the shareholders of firms are the relatively rich. Thus the relatively rich are made worse off by the increase in the minimum wage. Our broad conclusion is therefore that the working poor benefit from an increase in the minimum wage, but everybody else in society is made worse off.

There is another concern when we think about the distribution of income. One consequence of the minimum wage is that jobs become a scarce resource: more people want jobs than there are jobs available. We must consider how this scarce resource is allocated. Do workers line up outside factory gates? In this case, the time that they spend waiting in line is an additional cost of the minimum wage. Does some individual control who gets hired? Then there is the potential for corruption, whereby jobs are sold, meaning that the gains from the minimum wage flow not to workers but to this individual instead.

Does the Minimum Wage Benefit Unskilled Labor as a Whole?

We have concluded so far that the minimum wage benefits the working poor but at the cost of creating unemployment: some people who used to have jobs will lose them as a result of the minimum wage. Because of the flows between unemployment and employment in a dynamic labor market, it does not really make sense to think of the unemployed and the employed as different people. If we instead think about unskilled labor as a whole—a group that includes both those with jobs and those unable to find jobs—what can we conclude about the effects of an increase in the minimum wage? There are a few different ways of looking at this question.

The Wage Bill

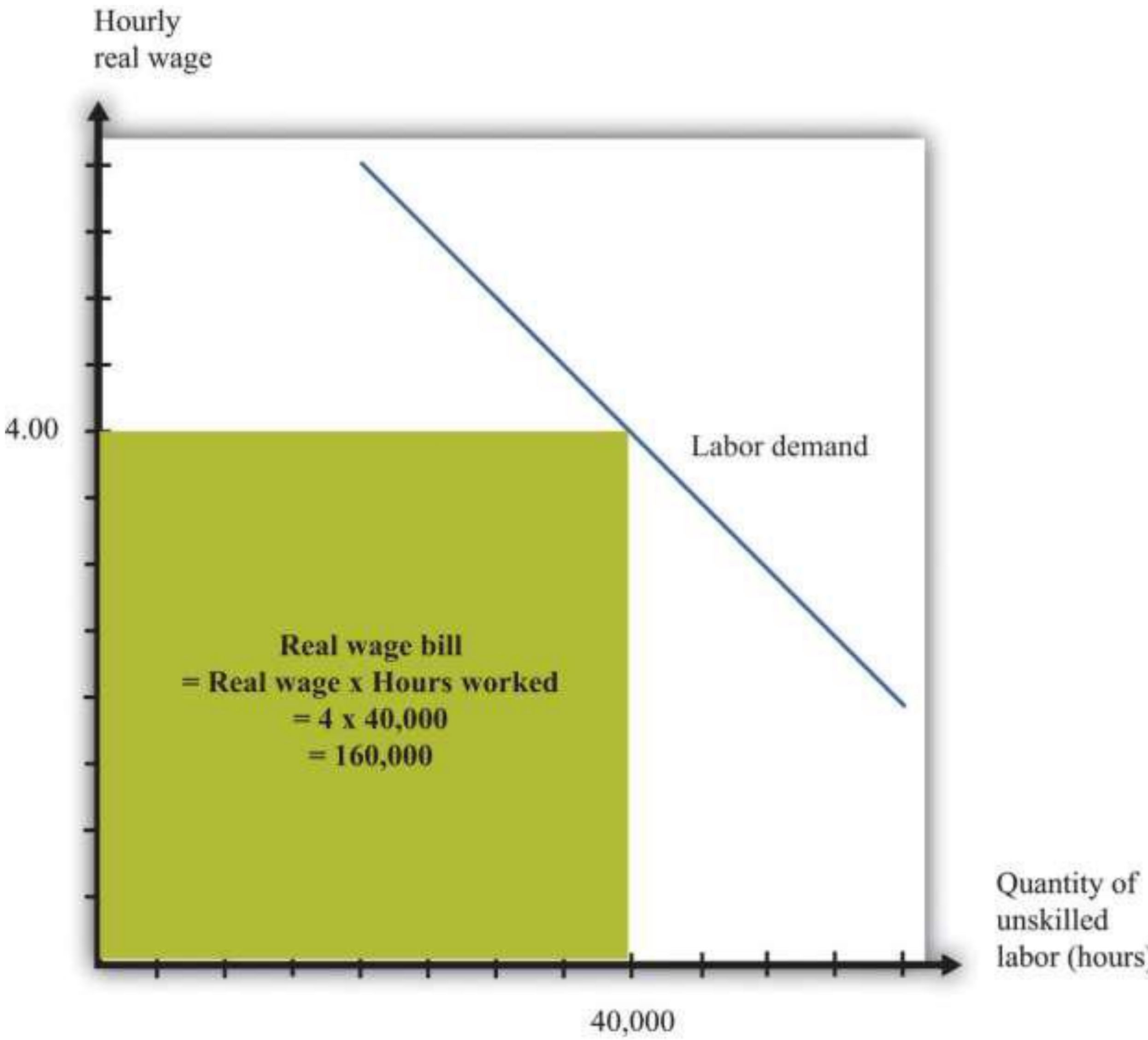
First, we can look at total wages paid in the labor market, sometime called the*wage bill*. By looking at the wage bill, we can find out if the additional wages earned by the working poor exceed the wages lost by those who find themselves unemployed.

Total wages are equal to the total hours worked multiplied by the hourly wage:

total wages = real minimum wage × hours worked.

Because we are measuring the wage in real terms, total wages are likewise measured in real terms. Figure 10.13 "The Wage Bill" provides a graphical interpretation of the wage bill: total wages paid are given by the shaded rectangle.

Figure 10.13 *The Wage Bill*



The wage bill is equal to the rectangle under the demand curve. For example, if the real wage is $4 per hour and employment is 40,000 hours, then the wage bill is 160,000.

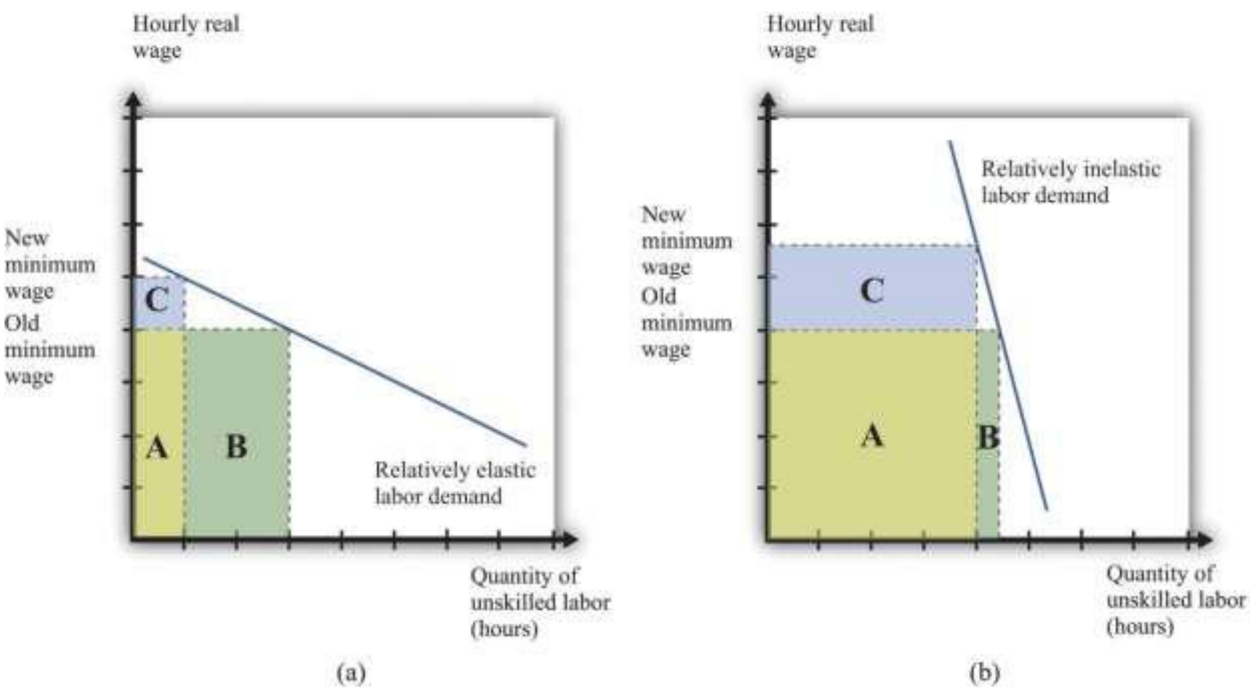
From this equation it can be shown that [1]

percentage change in total wages =percent change in real minimum wage × (1 – [–(elasticity of demand)]).

The elasticity of demand is a negative number: if wages increase (that is, the change in the wage is positive), then hours worked decreases (that is, the change in hours worked is negative). It is therefore easier if we use –(elasticity of demand) because this is a positive number. The equation tells us that the change in the total wage is positive if the percentage increase in wages is greater than the percentage decrease in hours worked—in other words, if −(elasticity of demand) is less than 1.

If the demand for labor is relatively sensitive to changes in the wage, employment will decrease significantly following an increase in the minimum wage. Total wages paid will decrease. This is shown in part (a) of Figure 10.14 "Effects of an Increase in the Minimum Wage on the Wage Bill". Before the increase, total wages are given by the sum of areas A and B. After the increase, total wages are given by the sum of areas A and C. We get the opposite conclusion if labor demand is inelastic. In this case, an increase in the wage increases the total wages paid. The conclusion is intuitive: if employers do not change their hiring very much when wages increase, then total wages will increase. But if an increase in the minimum wage leads to a big decrease in the demand for labor, total wages paid will decrease.

Figure 10.14 Effects of an Increase in the Minimum Wage on the Wage Bill



If labor demand is relatively elastic (a), a change in the minimum wage leads to a reduction in the wage bill: the original wage bill is A + B, and the new wage bill is A + C. If labor demand is relatively inelastic (b), the same change in the minimum wage leads to an increase in the wage bill.

Seller Surplus

The wage bill tells us how much workers are paid in total. A better measure of the benefits obtained by workers is the total sellersellers' surplus#8217;s surplus in the market. We cannot measure this exactly unless we know exactly what the labor supply curve looks like, but we can conclude that just looking at the wage bill understates the benefits to workers of the increased wage. The reason is simple and does not even need any diagrams. Following an increase in the minimum wage, workers work fewer hours in total. Everything else being the same, people prefer leisure time to working. For example, suppose that total wages increase following an increase in the minimum wage. Then workers gain twice: they are being paid more, and they are working less.

Even if total wages decrease, workers might still be better off. They might be more than compensated for the lower wages by the fact that they don't have to work as many hours. We are not saying that having a job is a bad thing; those who are working prefer having a job to being unemployed. But those who are working also prefer working fewer hours.

Expected Wage

So far, we have looked at the minimum wage through the lens of a competitive labor market. This is not a bad approach: as we have argued, the unskilled labor market is probably a reasonably good example of a competitive market. It is, however, a static way of thinking about the labor market, when the labor market is in fact highly dynamic. [2] People move in and out of jobs: they quit or are laid off from old jobs, and they search for new jobs. A worker who is employed this month may find herself unemployed next month; a worker with no job this month may be hired next month.

Earlier we claimed that an increase in the minimum wage has no direct effect on the unemployed. This is true in the static labor market picture, but once we take a more dynamic view of the labor market, it no longer makes very much sense to draw a hard-and-fast distinction between the employed and the unemployed. Over time, they will include many of the same people. So when we look at the distributional effects of the minimum wage, it is better to draw the distinction between unskilled workers (that is, the employed and unemployed together) and the relatively rich. With this in mind, let us now consider whether the unskilled as a group are likely to benefit from the minimum wage.

We might expect that unskilled workers will spend some of their time employed and some unemployed. When employed, they earn the minimum wage, but when unemployed, they receive much less. To keep things simple, suppose these workers earn nothing when unemployed. On average, the fraction of time that workers spend employed rather than unemployed is given by

$$\text{fraction of time employed} = \frac{\text{quantity of labor demanded}}{\text{quantity of labor supplied}}.$$

We can think of this as the **probability** that a typical unskilled worker will be employed at any given time. Combining this with the idea of **expected value**, we can calculate the expected wage of such a worker. If a worker earns nothing when unemployed, then the expected wage is as follows:

$$\text{expected real wage} = \text{fraction of time employed} \times \text{real minimum wage} = \frac{\text{quantity of labor demanded}}{\text{quantity of labor supplied}} \times \text{real minimum wage}.$$

Toolkit: Section 17.7 "Expected Value"

You can review probability and expected values in the toolkit.

How does this expected wage change when there is an increase in the minimum wage? The answer, as you might expect by now, depends on the elasticities of demand and supply. Specifically, it turns out that the expected wage will increase if

$$-(\text{elasticity of labor demand}) + \text{elasticity of labor supply} < 1.$$

If both demand and supply are sufficiently inelastic, the average wage will increase. Conversely, if they are both relatively elastic, then expected wages will decrease.

There are some things missing from this story. In a more careful analysis, we would take into account the fact that workers are probably risk-averse and dislike the randomness of their earnings. We would likewise take into account that unemployed workers obtain some income—perhaps from unemployment insurance. This actually makes it more likely that an increase in the minimum wage will increase the expected wage. These are details, however. We can draw two big conclusions from our discussion so far:

1. Under some circumstances at least, an increase in the minimum wage will have the effect of redistributing income from the relatively rich to unskilled workers. Policymakers may wish to

reduce inequality in society, and the minimum wage is one possible tool that they can use. At the same time, the minimum wage comes at a cost to society: it distorts decisions in the labor market and leads to deadweight loss. This is an equity-efficiency trade-off.

2. The distributional impact of a change in the minimum wage cannot be deduced from economic theory. It depends on the elasticities of labor demand and supply in the market for unskilled labor and is an empirical question. In the next section, therefore, we turn to the evidence on minimum wages.

KEY TAKEAWAYS

- The minimum wage affects buyers and sellers of labor services in the markets where the minimum wage is binding. It also affects the owners of these firms.
- Buyers and sellers of labor services in low wage (unskilled) labor markets are directly affected by changes in the real wage. Workers in this market who keep their jobs are better off. Those that lose their jobs are made worse off, at least in the short run. Owners of the firms are hurt because of reduced profits from the minimum wage. The magnitude of these effects depends on the elasticities of labor demand and supply.
- The equity-efficiency trade-off means that policies that increase equity can create inefficiencies. The minimum wage provides an example of that trade-off.

CHECKING YOUR UNDERSTANDING

1. Why does an increase in the minimum wage reduce the profits of a firm?
2. If the real minimum wage is reduced by inflation, can the real wage bill paid to workers increase?

[1] To derive this equation, we first apply the rules of growth rates to obtainpercentage change in total wages = percentage change in real minimum wage + percentage change in hours worked.Then divide both terms by the percentage change in the minimum wage and use the definition of the elasticity of demand:

$$\text{percentage change in total wages} = \text{percentage change in real minimum wage} \times (1 + \text{percentage change in hours worked}/\text{percentage change in real minimum wage}) = \text{percentage change in real minimum wage} \times (1 - [-(\text{elasticity of demand})]).$$

[2] Chapter 8 "Growing Jobs" contains more discussion.

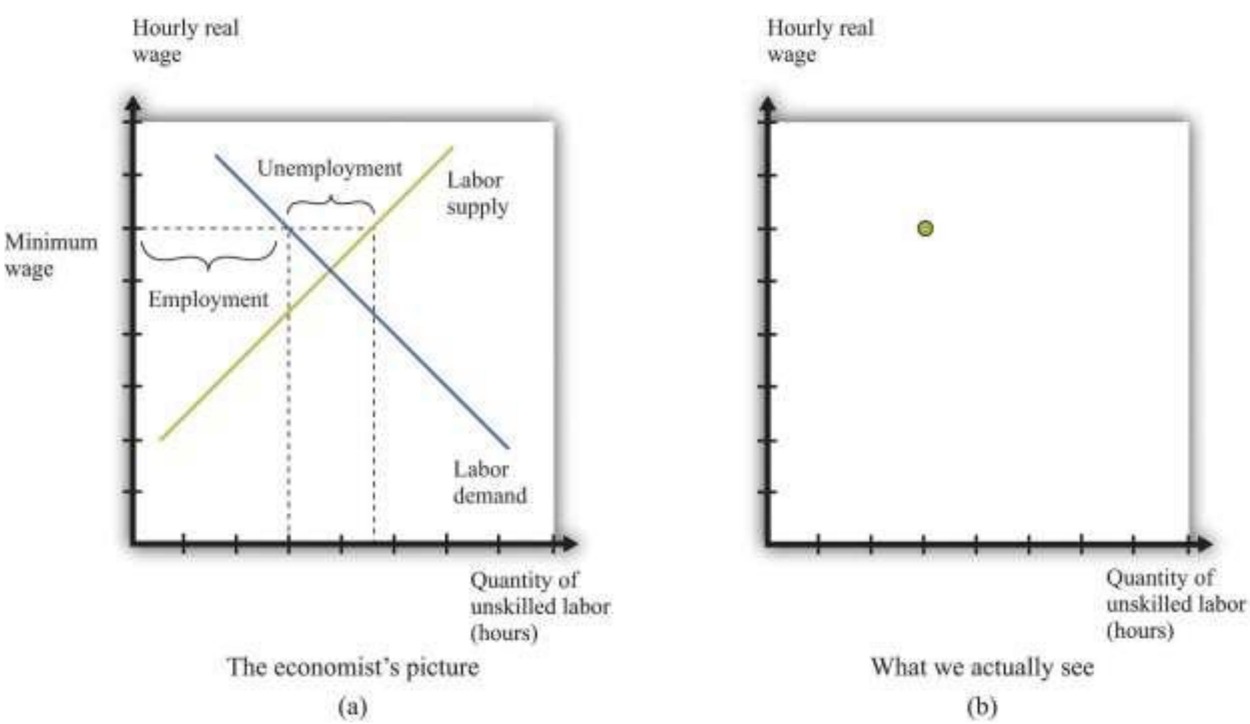
10.5 Empirical Evidence on Minimum Wages

LEARNING OBJECTIVES

1. How do economists determine the elasticities of labor supply and labor demand?
2. What are the estimates of these elasticities?
3. What are the estimated effects of an increase in the minimum wage?

For the most part, economists cannot carry out experiments to test their theories and must use much more indirect methods. They must rely on observations that are generated by the everyday experience of individuals in an economy. In a textbook like this, we constantly draw demand and supply curves, and we get so used to seeing these diagrams that we might be fooled into thinking that we can just go out and observe them in the real world. In fact, all we observe are the market outcomes—the equilibrium price and quantity that are traded. Our conception of the labor market might look like part (a) of Figure 10.15 "Models and Data", but the data that we actually gather look like part (b) of Figure 10.15 "Models and Data".

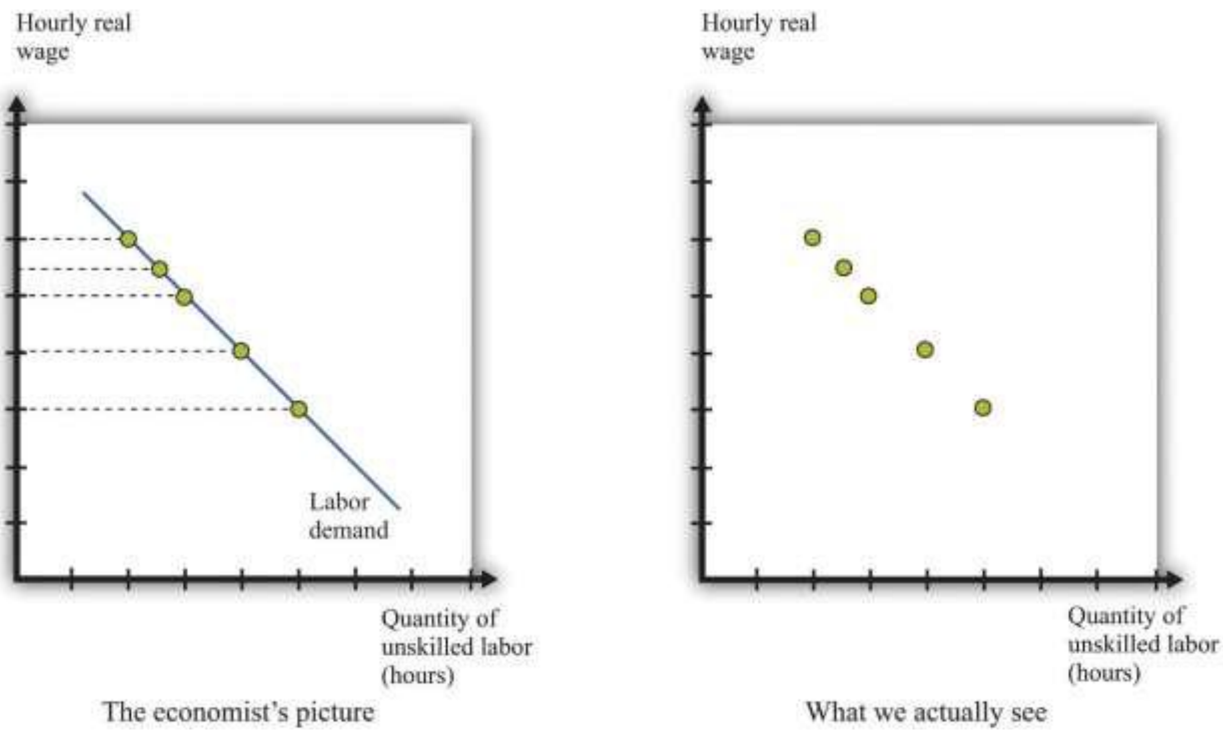
Figure 10.15 Models and Data



We construct an entire framework based on the supply and demand curves for labor (a), but at any time we observe only a single data point: the wage and the level of employment (b).

If we want to *estimate* a demand curve, we need much more than part (b) of Figure 10.15 "Models and Data". We need more data points. We need different observations. In the case of the labor market, we might be able to use the fact that the minimum wage changes over time. Figure 10.16 "Inferring Labor Demand from Data" shows an example. Minimum wage changes allow us to observe different points on the labor demand curve. Given enough observations, we might be able to get a good idea of what the demand curve looks like and come up with an estimate of labor elasticity.

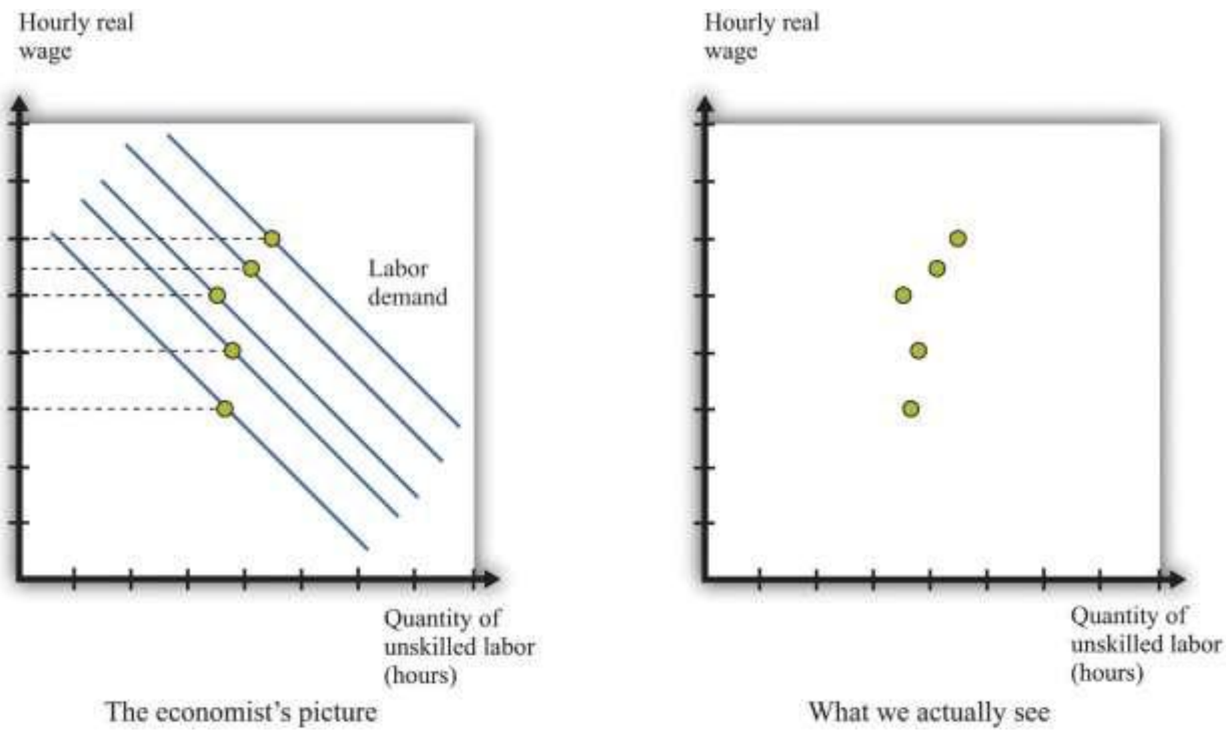
Figure 10.16 Inferring Labor Demand from Data



We may be able to infer a demand curve for labor by looking at what happens to the quantity of hours worked as the minimum wage changes.

Reality is messier. In actual labor markets, many things are going on at once. At the same time that the wage is changing, firms might be facing changes in the demand for their products, changes in the costs of other inputs, changes in their technology, or changes in their competitive environment. All of these changes would cause the labor demand curve to shift. As an example, look at Figure 10.17 "Difficulties Inferring Labor Demand from Data". The story here is as follows: The minimum wage is increasing over time, but—perhaps because of increased product demand—the demand for labor is also increasing over time. What we observe is shown in Figure 10.17 "Difficulties Inferring Labor Demand from Data". The unwary analyst, looking at these data, might conclude that minimum wages have little or no effect on the demand for labor.

Figure 10.17 Difficulties Inferring Labor Demand from Data



When the demand curve for labor is shifting at the same time as the minimum wage is changing, it is more difficult to see the effects of the minimum wage.

Economists who analyze data are forever trying to distinguish effects of interest (for example, minimum wage changes) from those caused by changes in other variables (for example, product demand). The key to successful empirical work is obtaining informative sources of variation and excluding irrelevant sources of variation. In the case of the minimum wage, we might look at differences in the minimum wage at different times, or we might look at differences in the minimum wage in different places (such as specific states in the United States).

Estimates of Labor Demand Elasticity

Labor economist Daniel Hamermesh summarizes the findings from numerous studies in his book, *Labor Demand*. [1] Based on his review of all these studies, Hamermesh argues that a good estimate of −(elasticity of labor demand) is about 0.3. This means that if we increase the minimum wage by 10 percent, employment will decrease by about 3 percent. With these, we would conclude that labor demand is relatively inelastic, and the employment and welfare implications of the minimum wage are not that large. Labor supply also tends to be relatively inelastic, partly because the income effects and substitution effects of changes in the wage tend to be offsetting. It certainly seems possible, then, that an increase in the minimum wage will raise expected wages.

We must be careful, though. Hamermesh notes that labor demand becomes more elastic as the skill level of workers decreases. It is difficult to find substitutes for workers with specialized skills. If you need an electrician, then you must hire an electrician, even if plumbers are much cheaper. If you need an airline pilot, you can't hire a computer programmer. If you are running a store, however, and you find that labor is becoming more expensive, you might be able to upgrade your cash registers and inventory control systems and get by with fewer employees. We might expect the elasticity of demand at the minimum wage to be significantly higher.

It is therefore useful to look at studies that focus directly on minimum wage changes. One approach is to look at the relationship over time between the minimum wage and the employment experience of groups that are most directly affected by the minimum wage. Alternatively, one can look across groups that are differentially affected by the minimum wage to gauge its effect.

A reminder before we proceed. The point of going through this material is not only to help you understand the effects of a change in the minimum wage but also to provide you with a glimpse of how economic research proceeds so that you can do a better job of evaluating evidence that economists compile in all sorts of areas.

Economists have focused particular attention on teenage workers because that group is typically unskilled and is likely to be subject to the minimum wage. (This is not to say that the *effects* of the minimum wage are primarily on teenagers. Indeed, about two-thirds of minimum wage earners are adults.)Figure 10.18 "Teenage Unemployment in the United States" graphs the teenage unemployment rate over the period from 1956 to 2007 along with the real minimum wage for this period. We would like to know if the minimum wage has an impact on the teenage unemployment rate. This figure is suggestive of a relationship, particularly in the last couple of decades. For example, the real minimum wage fell during the 1980s, and the teenage unemployment rate also fell during that time. The teenage unemployment rate also often seems to increase around the time that the minimum wage increases. On the other hand, teenage unemployment also fell substantially during the 1960s at a time when there were several increases in the minimum wage.

If the only cause of changes in teenage employment were minimum wage changes, Figure 10.18 "Teenage Unemployment in the United States" might give us lots of answers. But this is a very big "if." In terms of our analysis of supply and demand, it would amount to saying that the supply and demand for labor didn't change over the entire period. In fact, you can make an enormously long list of things that might have shifted the supply curve, the demand curve, or both. Examples include whether the economy was in a boom or a recession, changes in tax rates, technological advances, and population growth. As we saw in Figure 10.17 "Difficulties Inferring Labor Demand from Data", it is difficult to disentangle the effects of a changing minimum wage from the effects of other changes. Thus we cannot isolate the effects of the minimum wage just by looking at diagrams like Figure 10.18 "Teenage Unemployment in the United States".

We need some way to take into account these other factors so that we can focus on the effects of the minimum wage. This is a complicated statistical problem that arises time and again in economics, and the ways of dealing with it go far beyond this textbook. Indeed, this problem is one of the hardest things about studying economic data. But even though the statistical details are complex, there are three simple ideas that are important to understand:

1. The basic idea of these statistical techniques is that economists attempt to include in their analysis—as best they can—all the factors that shift the demand and supply curves. Then they can

predict *what would have happened if there had been no change in the minimum wage*. Once they have done this, they can then determine the effect of minimum wage changes.

2. When different economists study the same problem, they do not always reach the same conclusion. The reason that they disagree is usually not because they have different ideas about economic theory. Almost all economists have the same general framework for understanding labor markets, for example. Economists disagree because they use different approaches to take into account all the other factors.
3. If economists could conduct direct experiments, such as those performed in the physical sciences, they would not have this problem. They would be able to completely control the *stimulus* (minimum wage changes), and thus get an accurate estimate of the *response* (changes in employment). Instead, the best we can do is to look at past minimum wage changes and try to determine what happened.

Studies from the 1970s and early 1980s found some evidence that increases in the minimum wage reduced the employment of 16- to 19-year-olds. According to economists David Card and Alan Krueger, the average estimate is that a 10 percent increase in the minimum wage would reduce employment by about 1.5 percent. [2] So the implied value of –(elasticity of labor demand) is about 0.15. With an average employment rate of about 50 percent, this means that a 10 percent increase in the minimum wage would reduce the rate of teenage employment by about 0.75 percentage points.

It is striking that this estimate is *lower* than the estimate from looking at labor demand as a whole. What is more, Card and Krueger note that more recent studies produce even smaller estimates of these effects. [3] In their own work, they find that the apparent negative effects of the minimum wage on employment are statistically insignificant. This result holds not only for all 16- to 19-year-olds but also when the sample is split by race and sex.

The evidence from teenage employment studies suggests that, as an empirical matter, the effects of the minimum wage on employment may be very small. This result has surprised many economists. Although economic theory did not suggest actual magnitudes for the labor demand elasticity, the existence of substitutes for unskilled labor did suggest that labor demand would be at least somewhat elastic.

Not surprisingly, there have been many other studies of the minimum wage. Some researchers have found larger employment elasticities for teenage employment than those reported by Card and Krueger. Sometimes we are simply unable to give a definitive answer to empirical questions in economics. This can be frustrating for both students of economics and practitioners—but we are not going to pretend that the world is simpler than it actually is.

Cross-Section Studies

The studies we have just discussed analyzed the minimum wage by looking at minimum wage changes over time. Another approach to analyzing the effects of the minimum wage is to take advantage of differences over individuals rather than variations over time. Economists call these *cross-section studies* because they look at a cross section of different individuals or firms at a point in time. Many of these studies look at the effects of minimum wage changes at the level of an individual worker. Others exploit differences in minimum wage laws across states. Such differences across states give rise to a natural experiment because they can substitute, at least in part, for economists' inability to conduct experiments in which only one thing changes at a time.

Here is an example. Recall that in 1938 a minimum wage of $0.25 per hour was put into effect under the Fair Labor Standards Act. In the United States, this minimum wage was about 40 percent of the average manufacturing wage. However, the law also applied to Puerto Rico, which was much poorer, where this minimum wage was about twice the average factory rates. In a book published more than 25 years ago, researchers John Petersen and Charles Stewart noted that the increase in the minimum wage led to numerous factory closings, and there was a dramatic decrease in output and employment. [4] In the case of Puerto Rico, the introduction of the minimum wage apparently had a large adverse effect on employment.

Peterson and Stewart also provide an extensive account of early studies of minimum wages that looked at employment and wages at individual production sites. These studies compared employment before and after a change in the minimum wage in an attempt to infer the effects of the policy. A study of the seamless hosiery industry from 1938 to 1941 is of particular interest. This period saw the introduction of the $0.25 per hour minimum wage, followed by an increase to $0.325 per hour in September 1939. A researcher named A. F. Hinrichs looked at 76 different plants and divided them into two groups: those that paid high wages and those that paid low wages. We would expect the minimum wage to have a much bigger effect at the low-wage plants. Between September 1938 and September 1939, the low-wage plants had employment losses of 12 percent. Employment at high-wage plants actually expanded by 23 percent, perhaps in part because workers who lost jobs in the low-wage plants became part of the labor supply for the previously higher-wage plants. A similar pattern was noted for the period from 1938 to 1940.

A much more recent study by Card and Krueger is another example of this approach. They studied employment patterns in fast-food restaurants in New Jersey and Pennsylvania. The key to their research was that, during the period of study, the minimum wage was increased in New Jersey but not in Pennsylvania. (Remember that individual states sometimes set minimum wages above the federal minimum.) From this natural experiment, Card and Krueger found that the increased minimum wage in New Jersey actually seemed to have *increased* employment.

The evidence from other countries (both cross-section studies and studies over time) is likewise mixed. [5] One study of Greek labor markets, for example, found a negative effect for men but a positive effect for women. Another study found negative effects for Mexico but not for Colombia. Different researchers in France have come to different conclusions about the effects of the minimum wage there; researchers in New Zealand likewise disagree; and so on.

Beyond Employment Effects

We started with what seemed to be some simple questions about the minimum wage. The answers turned out to be quite complex. Empirical research does not deliver a definitive answer about whether minimum wages have a big effect on employment. This leads to some disagreement among economists, particularly because the minimum wage is a politically charged issue. From the perspective of policymaking, the lack of a consensus creates difficulty in formulating good policy. On the other hand, the lack of a consensus provides a stimulus for continued work on these important issues.

Though we have emphasized the employment effects of minimum wage changes, there are other effects of minimum wages as well. First, remember that the main argument in favor of minimum wages is that they are a vehicle for redistributing income toward the working poor. Card and Krueger present a detailed analysis of the types of individuals most likely to be directly affected by minimum wage changes. Although empirical work often focuses on the employment of teenage workers, young workers are not the only group in the labor market that is paid close to the minimum wage. About 50 percent of the workers affected by the April 1990 increase in the minimum wage were older than 24 years old, for example.

How much income then flows to these workers as a consequence of an increase in the minimum wage? Card and Krueger conclude that the increase in the minimum wage during 1990 and 1991 had only a tiny effect on the distribution of income. They calculate that the minimum wage increase from $3.35 to $4.25 transferred about $5.5 billion of income to low-wage earners. This amounts to about 0.2 percent of family earnings. The host of transfer programs in place in the United States swamps the effects of the minimum wage on the redistribution of income. The evidence from other studies and countries is broadly in line with this conclusion: several studies find some effects of minimum wages on income distribution, but these effects are typically small.

Second, we can think about the effect of the minimum wage on firms. An increase in the minimum wage increases firms' marginal cost of production. As a consequence, firms will increase their prices and sell less output. Because of this, increases in the minimum wage reduce profits, so we might expect to see this reflected in the share prices of firms that employ minimum wage workers. Relative to the large empirical literature on employment effects, the implications for employers have been largely neglected. Card and Krueger survey the evidence and find relatively small effects on the stock market value of firms.

KEY TAKEAWAYS

- Economists use data from labor market outcomes (wages and employment) to infer the shapes of labor supply and demand curves. A key part of this inference is to isolate economic variation to trace out one of the curves.
- Based on many studies, −(elasticity of labor demand) is about 0.3. So if we increase the minimum wage by 10 percent, employment will decrease by 3 percent.
- Studies that look directly at the effects of minimum wage changes find minimal effects of minimum wage changes on employment and unemployment.

CHECKING YOUR UNDERSTANDING

1. Why is it so difficult to determine the effects of minimum wage changes on unemployment?
2. What is a natural experiment?
3. Why do changes in the real minimum wage allow researchers to trace out the labor demand curve?

[1] Daniel S. Hamermesh, *Labor Demand* (Princeton, NJ: Princeton University Press, 1996).

[2] David Card and Alan Krueger, *Myth and Measurement: The New Economics of the Minimum Wage* (Princeton, NJ: Princeton University Press, 1995).

[3] David Card and Alan Krueger, *Myth and Measurement: The New Economics of the Minimum Wage* (Princeton, NJ: Princeton University Press, 1995), Table 6.2 "Calculating Revenues".

[4] John M. Peterson, and Charles T. Stewart Jr., *Employment Effects of Minimum Wage Rates* (Washington: American Enterprise Institute, 1969).

[5] A recent Organisation for Economic Co-operation and Development report has a summary of minimum wage studies from different countries: "Making the Most of the Minimum: Statutory Minimum Wages, Employment and Poverty," accessed March 14, 2011, http://www.oecd.org/dataoecd/8/57/2080222.pdf.

10.6 End-of-Chapter Material

In Conclusion

The minimum wage is a public policy that is debated the world over. It has widespread public support because there is something very appealing about the notion that those who work hard will be rewarded with a reasonable standard of living. Economists, living up to their reputation as dismal scientists, point out that this is all very well, but there may be unintended consequences of such a policy:

- A minimum wage leads to a reduction in employment and to unemployment.
- A minimum wage leads to fewer trades than in a competitive market and therefore to inefficiency.

The theory of the minimum wage is straightforward and convinces most economists that, even if the minimum wage has some benefits in terms of the distribution of income, it carries costs with it as well. There is less consensus among economists about whether the redistribution brought about by the minimum wage is desirable. Although the working poor benefit, others are made worse off—including those who are unemployed, who are perhaps even more in need of help than the working poor. Individuals differ in their beliefs about how society’s resources should be distributed, and there is no right answer to the question “what is fair?”

When we looked at the *evidence* on minimum wages, however, we found that the picture is much less clear. Although some studies are in line with the predictions of theory, many studies suggest that, in practice, the effect of the minimum wage on employment is minimal. At the same time, the effect of the minimum wage on the distribution of income is small as well. In the end, it is difficult to resist concluding that the minimum wage is much less important—in terms of both benefits and costs—than one would think from the rhetoric of the debate.

When you read the newspapers or watch television, you will frequently hear economists offer different viewpoints on economic policies. These disagreements are typically not because economists differ on the theory. The disagreements often come down to different opinions about how to analyze and interpret economic data. Remember as well that television and print journalists go out of their way to find differing points of view because that makes a better story, so the disagreement you see in the media is usually not representative of economists as a whole. That said, economists also have different political viewpoints, and they are sometimes guilty of letting their political preferences cloud their economic analysis.

Having gone through the arguments in this chapter, you should be better able to assess debates and discussion on the minimum wage the next time it comes to the forefront of public debate. This chapter has a much broader purpose, however. We have been studying the effect of government intervention in a market, and we have shown how we can use our tools of supply and demand to understand the likely effects of that intervention. There are many other examples of government interventions that can be investigated using very similar reasoning.

Finally, we have learned something about how empirical work is conducted in economics. Because economists cannot conduct experiments, they are forced to trawl through messy data in an attempt to test their theories. It is difficult to be sure that the variables in which we are interested are indeed changing enough to be useful, and it is even more difficult to disentangle those changes from all the other irrelevant changes that affect the data that we observe.

Key Links

- Bureau of Labor Statistics: http://www.bls.gov
- Department of Labor: http://www.dol.gov
- US minimum wage laws:http://www.dol.gov/esa/minwage/america.htm

EXERCISES

1. List three jobs you think probably pay minimum wage and three jobs that you think do not.
2. Illustrate an increase in the minimum wage when both demand and supply are (a) relatively inelastic and (b) relatively elastic. Explain why the change in unemployment is smaller when the curves are inelastic.
3. Explain why the deadweight loss from the minimum wage is larger if labor demand is relatively elastic.
4. How does the elasticity of labor supply affect the deadweight loss from the minimum wage? Specifically, if labor supply is more elastic, is the deadweight loss smaller or larger? What is the economic intuition behind your answer?
5. (Advanced) Draw a version of Figure 10.9 "Deadweight Loss from Minimum Wage" for the case where a single individual controls access to scarce jobs. Suppose that she is able to charge job searchers a fee (the same fee for all searchers) equal to the difference between the minimum wage and the wage that workers would be willing to accept. What area of the figure does she obtain?

6. In our discussion of the evidence of the effects of minimum wage changes, we said, “If economists could conduct direct experiments, such as those performed in the physical sciences, they would not have this problem.” Exactly what problem were we referring to?
7. What is the difference between cross-sectional and time-series studies? Does one hold “more things fixed” than the other?
8. Suppose the government imposed a maximum wage in the market for some high-paying job. Draw a diagram to illustrate this market. What would be the consequences of this maximum wage?
9. Explain why, when we analyze the minimum wage, the elasticity of labor supply affects the unemployment rate but not the employment rate.
10. Why does the government not set a minimum wage for corporate lawyers and airline pilots?
11. If the rate of inflation is 10 percent higher than expected, and –(elasticity of labor demand) is 5, what will happen to the employment level in jobs that pay the minimum wage?
12. What happens to the rate of unemployment of minimum wage workers if the rate of inflation is lower than expected?
13. Does the elasticity of labor supply have an effect on the change in the wage bill when there is an increase in the minimum wage? Does this elasticity have an effect on the unemployment rate when the minimum wage changes?
14. (Advanced) Using the discussion of estimation of labor demand, if you could conduct an experiment to see the effects of a minimum wage increase, what exactly would you do?
15. (Advanced) Using supply and demand curves in the market for fast food, what are the effects of an increase in the minimum wage in this market? Think about shifts in both the supply curve and the demand curve. Explain your predictions.
16. Why isn’t an increase in the minimum wage just a redistribution from firms to workers?
17. A politician is in favor of getting rid of the minimum wage entirely. How would you argue against that proposal?

Economics Detective

1. Pick three countries and find the minimum wage in each country.
2. Find a country that does not have a minimum wage. Do you think the lack of a minimum wage means that workers are badly treated in that country?

3. Find some recent discussion of minimum wage legislation in either the United States or some other country. What arguments were made to support the minimum wage? What arguments were made against the minimum wage?

Spreadsheet Exercise

1. (Advanced) Find data on the minimum wage and the price level for another country. Construct a real minimum wage series for that country.

Chapter 11

Barriers to Trade and the Underground Economy

An Unusual Shopping List

You come downstairs one morning and find a note on the table.

Please go to the store today and buy the following:

- *A bag of sugar*
- *Two pints of milk*

So far there is nothing unusual about this. You plan to go to the grocery store on your way home that evening. Then you read on.

- *One carton of cigarettes*
- *One bottle of whiskey*

These are a bit trickier. If you are like many readers of this book, you may not be allowed to purchase alcohol or possibly even cigarettes. In the United States, you must be 21 or over to buy alcohol and over 18 (or 19 in some states) to purchase cigarettes. Depending on where you live, it may also be quite inconvenient to purchase alcohol. In some places, by law, alcohol is sold only at certain times of day. In some places—certain states in the United States and certain countries in Europe, for example—it is sold only in government-run stores.

Many goods, like alcohol, are restricted in terms of who can buy them, when they can be purchased, and where they can be purchased. Alcohol laws differ from country to country. In most European countries, for example, you can buy alcohol at the age of 18. The laws also change over time. Thirty years ago, 18-year-olds could buy alcohol in the United States as well. Ninety years ago, it was illegal for anyone to buy alcohol in the United States.

Next on the list is the following.

- *Two tickets for the sold-out rock concert in town tomorrow night*

This may also be difficult. You know that you can probably find someone who has tickets and is willing to sell them, but you know that local laws say that this, too, is illegal. So-called scalping of tickets is forbidden. Still, if you go to eBay, you'll probably be able to find some tickets for sale.
Then the list gets stranger:

- *Six Cohiba cigars*
- *One French raw milk camembert*
- *Four ounces of marijuana*

At this point (at least if you are living in the United States), you begin to seriously worry. You search the Internet for "Cohiba" and discover that these cigars are manufactured in Cuba, but you vaguely remember that it is illegal to import goods from Cuba to the United States. You know that camembert is a French cheese, but "raw milk" sounds strange. More online investigation informs you that it is also illegal to import cheeses into the United States unless they are made from pasteurized milk. Apparently, raw milk cheeses may carry dangerous bacteria. As for marijuana, you already know that it is illegal in the United States.

You read on.

- *Also, please hire a cleaning person (an undocumented migrant worker would probably be the cheapest)*

This is another transaction that you know is illegal. That said, you know that there are many illegal immigrants working in your town. It would be easy to find someone to hire if you were willing to break the law. With some foreboding, you turn the list over and read the other side.

- *Finally, please buy one human kidney (suitable for transplant).*

Most of the things that were on the list up to this point were goods or services that you would probably be able to find if you had to. Even though some of them could not be purchased legally, it would not be too hard to find out where to purchase most of them. (Oddly, it would probably be easier to get the marijuana than the cheese.) A human kidney is a different proposition, however. You're pretty sure, even without research, that buying and selling human organs is illegal, and you would have no idea where to go to buy a kidney even if you were willing to break the law.

We know that the market interaction of buyers and sellers creates value in an economy. [1] In a market, sellers supply a good or a service, and buyers demand that good or service. Because each transaction is voluntary, the value that the buyer places on the good is always greater than its value to the seller. This means that each trade creates some value. In addition, if the market is competitive, all value-creating trades occur in the market; there are no disappointed buyers or sellers.

This logic suggests that governments should be doing everything in their power to encourage and facilitate trade. Yet, in practice, there are several ways in which governments do the opposite: they actively intervene to restrict trade. We have just listed a large number of examples, and you can surely think of many more. We would like to understand all the restrictions that are deliberately put in place to impede trade.

Our main aim here is *not* to analyze the rationales behind these restrictions, although we do briefly explain some of them. In other chapters, we provide more insight into precisely why governments impose these and other limitations on our ability to transact with one another. [2] Our goal in this chapter is to explore *what happens* when governments interfere with trade in different ways.

One message of this chapter is a reiteration of the gains from trade, together with the recognition that they provide a powerful incentive for people to get together and transact with one another. It seems that whenever the government steps in to try to prevent them from trading, people still try to find a way around these restrictions. The gains from trade are a powerful motivator. Indeed, people continue to trade even when this is an illegal act that carries a significant risk of fines or imprisonment. We use the term *underground economy* to describe where these trades occur. The question we want to answer in this chapter is as follows:

> *What are the consequences of government restrictions on trade?*

Road Map

In this chapter, we will see many different ways in which governments intervene. For most of our analysis, we use the supply-and-demand framework. We analyze different kinds of government policy and examine the following questions:

- What happens to prices and quantities?
- What happens to welfare and the distribution of income?
- What happens to incentives? Are there any resulting unintended consequences?

We organize our discussion by looking at different categories of restrictions on trade. First, we look at the sale of goods and services in domestic markets. Then we turn to restrictions in international markets for goods and services. Finally, we turn to restrictions not on goods and services but on labor, both within and across countries.

[1] We discuss this in detail in Chapter 5 "eBay and craigslist", and Chapter 7 "Why Do Prices Change?".

[2] See in particular Chapter 12 "Superstars", and Chapter 14 "Busting Up Monopolies".

11.1 How the Government Controls What You Buy and Sell

LEARNING OBJECTIVES

1. What tools does the government use to control market transactions?
2. Why might the government restrict trades?
3. What are the effects of these restrictions on market outcomes and welfare?

Looking at your shopping list, there are some items that you simply cannot buy. For example, marijuana and the raw milk cheese from France are not available for purchase in stores in the United States. And depending on your age as well as the time and day of the week, you may not be able to buy the cigarettes and whiskey. We begin by discussing these types of market interventions.

Closing Down the Market

The most fundamental intervention in a market occurs when the government closes down trading completely—that is, the government simply says that it is illegal to trade certain goods or services. Examples are numerous and stem from many different motivations.

Health and safety. Most governments ban addictive drugs, such as heroin, cocaine, and marijuana. The primary reason is that these and similar drugs are deemed to be harmful to those who use them. A secondary reason is that governments may think—rightly or wrongly—that the trade of such drugs also has other harmful implications, such as increased crime.

Governments also ban trade in other products for similar reasons of health and safety. One of the functions of government in most countries is to oversee the safety of products, both generally and more specifically in terms of health risks. In the United States, the Food and Drug Administration certifies factories and food processing and also oversees the approval of pharmaceuticals. Meanwhile, the Bureau of Consumer Protection (http://www.ftc.gov/bcp) is charged with ensuring that goods meet certain legislated safety standards. Goods that do not meet these standards cannot be legally traded. For example, it is illegal to sell a new car without seatbelts and airbags in the United States.

Ethics, morality, and religion. The exchange of some goods and services is banned for ethical or moral reasons. Examples include the trading of human organisms, the sale of alcohol, and various forms of prostitution.

The ban on the trading of human organs is rooted primarily in an ethical belief that buying and selling body parts is wrong. Many people argue that the moral case for banning organ selling is very shaky, and the world would be a better place if such trades were allowed. It is true that many find the idea of trading body parts for dollars to be repulsive. We have a sense that people would sell a kidney only if they were truly in desperate financial straits, and there is something terrible about the image of, say, a mother selling a kidney to feed her children. Yet there are people who die every day because doctors are unable to find a suitable organ donor in time; that, too, is a sad image.

In many places, the consumption and sale of alcoholic beverages is forbidden, often for religious reasons. The sale of alcohol is prohibited in some Muslim countries, such as Saudi Arabia and Kuwait. Religious pressure also led to a 13-year ban on alcohol in the United States under the 18th Amendment to the Constitution; this state of affairs was known as *Prohibition*. Indeed, in many counties in the United States, the sale of alcohol is still prohibited. Likewise, many other countries in the world have regions that are "dry."

Not surprisingly, there is often disagreement about which trades should be ruled immoral or unethical. Different laws in different countries regarding the sale of alcohol are one illustration of this. Another example is prostitution, which is illegal in many places yet legal in others. For example, prostitution is legal (although heavily regulated by the government) in the Netherlands and in parts of Nevada.

Fairness. Sometimes, the government simply takes the view that certain trades are unfair. For example, scalping—the reselling of tickets to concerts and exhibitions—is frequently prohibited for this reason. The following story illustrates that people often see the reselling of tickets as unfair.

> *Perry Loesberg wanted to surprise his 10-year-old daughter Amy with tickets to "Hannah Montana," the sizzling-hot concert tour featuring 14-year-old TV star Miley Cyrus.*

Instead, he was the one surprised. Though he bought a $30 fan club membership to get access to tickets ahead of the general public, and then logged on to the Ticketmaster Web site before the general public sale began, Loesberg still came up empty-handed….

Tickets to each of the 54 shows on the "The Best of Both Worlds: Hannah Montana and Miley Cyrus" tour…sold out within minutes of going on sale. Almost immediately, online marketplaces such as StubHub and craigslist were offering dozens of seats, many selling for more than $2,000 each. Tickets were originally priced at $22–$66.

What kind of ignited parents is I think they thought it should be more fair," said Debra Rathwell, senior vice president for AEG Live, the tour's promoter…"We would like fans to sit in these seats. But everything you do, [scalpers] find a way to skirt around it."

Many of the purchasers are parents unfamiliar with the post-Internet ticket market. They were amazed at the availability of tickets—not to mention the high prices—on the re-sale market….

Many have pointed to computer software programs that allow users to, in essence, cut in line on the Ticketmaster Web site….The outcry over Hannah Montana is unusual for other reasons. Ray Waddell, senior editor at Billboard magazine, said parents and children are being disappointed, and their complaints have found sympathetic ears, including the attorneys general of Missouri, Arkansas, Connecticut and Pennsylvania…

It has been like the Wild West out there," said Waddell. "Things are going to tighten up, (there will be) more regulation about who's selling, who's buying and how they are getting their tickets." [1]

Restrictions on Who Can Trade

There are many products that can be legally traded, but the government places substantial conditions on the terms of those trades. For example, several legal goods and services cannot be purchased by minors and can be sold only by licensed sellers. Obvious examples are alcohol and cigarettes, but there are many others. Casino gambling is restricted to adults. Many pharmaceuticals can be sold only by licensed pharmacists and bought only with a doctor's prescription.

These restrictions vary a lot by time and place, which again tells us that there is no simple right or wrong where these laws are concerned. Different states have different laws. Not all stores can sell liquor. In Sweden, for example, alcohol is sold only in state-run stores. The legal drinking age in Europe is different from the legal drinking age in the United States. Some drugs require a prescription in some countries yet are available over-the-counter in others.

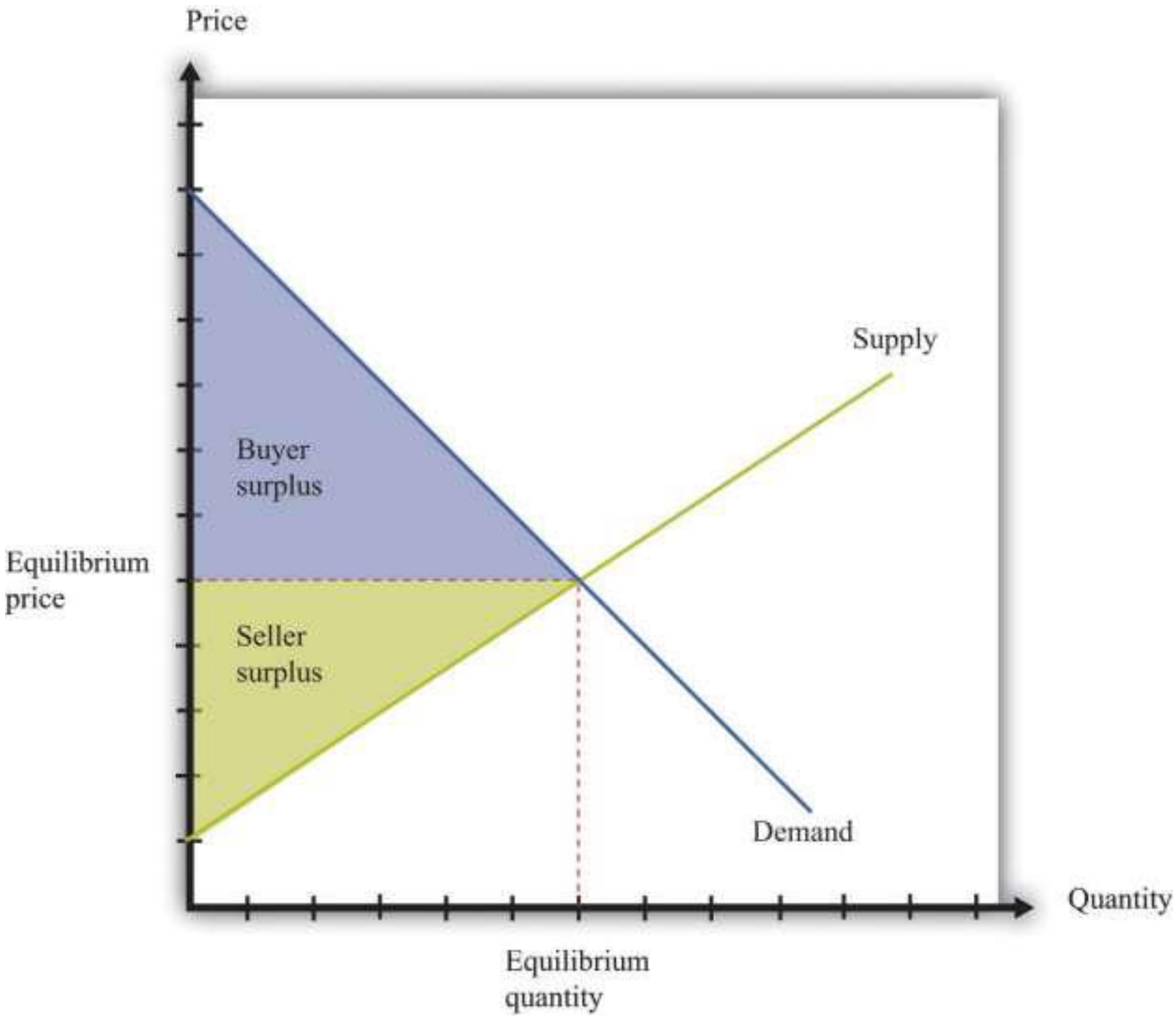
Implications

Figure 11.1 "Supply, Demand, and the Gains from Trade" shows the **buyer surplus** and **seller surplus** in a competitive market and reminds us that the **gains from trade** in a competitive market are at a maximum. All mutually beneficial trades have been carried out. Government interventions in markets typically have the effect of eliminating some or all these gains from trade.

Toolkit: Section 17.10 "Buyer Surplus and Seller Surplus"

You can review the different kinds of surplus and the gains from trade in the toolkit.

Figure 11.1 Supply, Demand, and the Gains from Trade



The area below the demand curve and above the price is the buyer surplus; the area above the supply curve and below the price is the seller surplus.

The economic analysis of the closing of a market is very simple. If the government successfully prevents trade, then the quantity traded is zero. All producer and consumer surpluses in Figure 11.1 "Supply, Demand, and the Gains from Trade" are lost. An economist's first response to the closing of a market—any market—is that it brings a loss because some potential gains from trade go unrealized. The question then becomes whether any benefits from closing down a market justify the lost gains from trade.

As our examples reveal, there are many reasons for closing a market, so there is no simple answer to the question, "Is it good to shut down a market?" Each argument must be looked at on a case-by-case basis, and the particulars of specific examples are beyond the scope of this book. Entire books have been written, for example, on the market for human organs or the legalization of prostitution.

When you read the shopping list at the beginning of the chapter, you might also have been struck by the fact that the government's success in blocking trade is often limited. You probably would find it difficult to buy a heart for transplant on the open market. But if you know where to go, you could almost certainly buy marijuana. Even if you are underage, you may be able to get a fake identification card and buy alcohol. And buying scalped tickets to a concert or a sports event is usually easy, if you have the money. The economic message is simple and fundamental. When there are gains from trade, people will try to realize those gains. When trades are illegal, economic activity moves into the so-called underground economy but is unlikely to disappear completely.

Rationing

Another way in which governments intervene in markets is not by banning trade outright but by placing a restriction on the quantity traded. In most modern economies, such restrictions are little used in a domestic context but are much more prevalent in international trade. If we look back in history, though, we can find instances of rationing in the domestic economy. Rationing means that the quantity available on the market is less than the equilibrium quantity. Some surplus goes unrealized because willing buyers and sellers are prevented from trading. During and after World War II, many basic goods were rationed in the United States, Britain, and elsewhere.

The following excerpt by journalist Joelle Kirch Preksta comes from oral histories of World War II collected by the Carnegie Library.

> *Ruth showed me several of the ration books she was issued during World War II….She explained that staples such as sugar, butter, and eggs were rationed in order to help supply our troops overseas and therefore were difficult to obtain in stores….The following excerpt [is] taken from the "Instructions" section of the books…*
>
> 1. *This book is valuable. Do not lose it.*

2. *Each stamp authorizes you to purchase rationed goods in the quantities and at the times designated by the Office of Price Administration. Without the stamps, you will be unable to purchase those goods.*

Rationing is a vital part of your country's war effort….Any attempt to violate the rules is an effort to deny someone his share and will create hardship and discontent. Such action, like treason, helps the enemy. Give your whole support to rationing and thereby conserve our vital goods. Be guided by the rule: "If you don't need it, DON'T BUY IT."

The books also contained a warning which indicated that someone who violated the rules for the ration books could be imprisoned for as long as 10 years or fined as much as $10,000. [2]

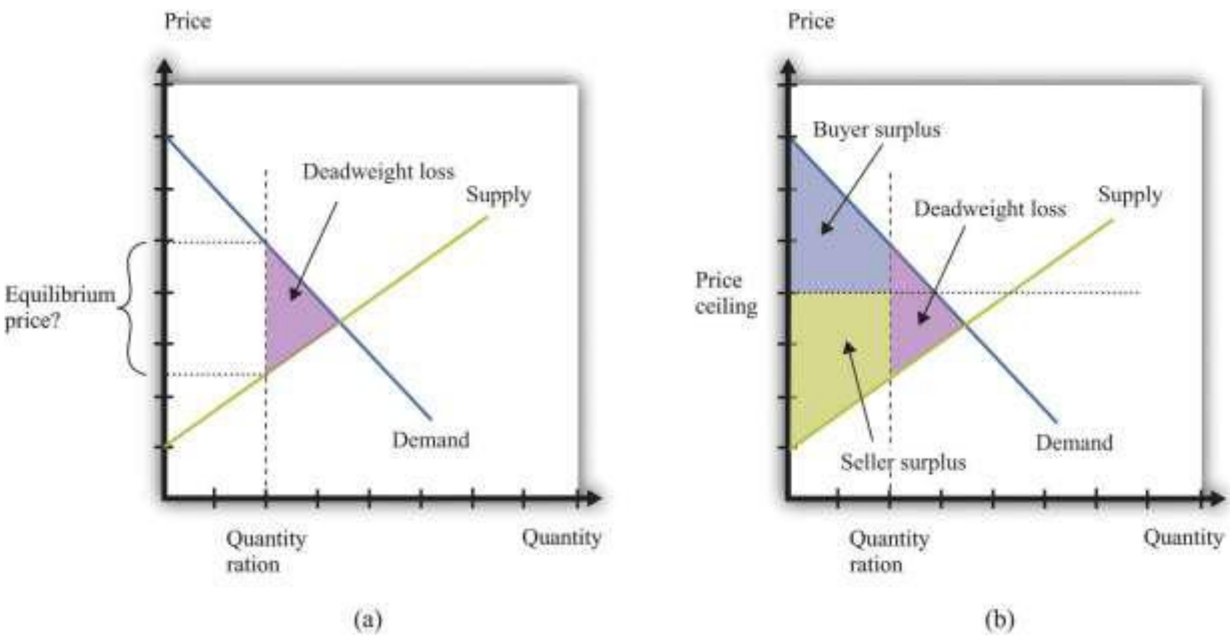
Despite these strong moral and legal sanctions—comparing black market trading with treason, no less—there was a substantial underground market for all sorts of rationed goods. For example, the Carnegie oral histories describe a young woman in her twenties named Mary: "She somewhat embarrassingly recalled that she was able to dishonestly procure an extra carton of cigarettes every month for herself because her aunt worked at the drug store where they could be purchased. To this day she says she feels somewhat guilty over this unpatriotic indiscretion." [3]

Figure 11.2 "The Implications of Quantity Rationing" shows the implications of quantity rationing. Part (a) of Figure 11.2 "The Implications of Quantity Rationing" shows that there is a deadweight loss. We see that a quantity ration does not tell us what the price will be. It could be anywhere between the minimum price that the marginal seller will accept (the price found on the supply curve) and the maximum price that the marginal buyer will pay (the price found on the demand curve). In the absence of any other mechanism, the price is determined by bargaining among buyers and sellers. In the case of World War II rationing, sellers were often in stronger bargaining positions, which pushed the price toward the higher end of the range. For this reason, quantity rations were often supplemented by a maximum price, called aprice ceiling (part (b) of Figure 11.2 "The Implications of Quantity Rationing").Figure 11.3 "A World War II Poster" shows a poster from this period.

Toolkit: Section 17.11 "Efficiency and Deadweight Loss"

You can review the concepts of efficiency and deadweight loss in the toolkit.

Figure 11.2 *The Implications of Quantity Rationing*



A quantity ration leads to deadweight loss but by itself does not tell us what the price will be.

Price Ceiling

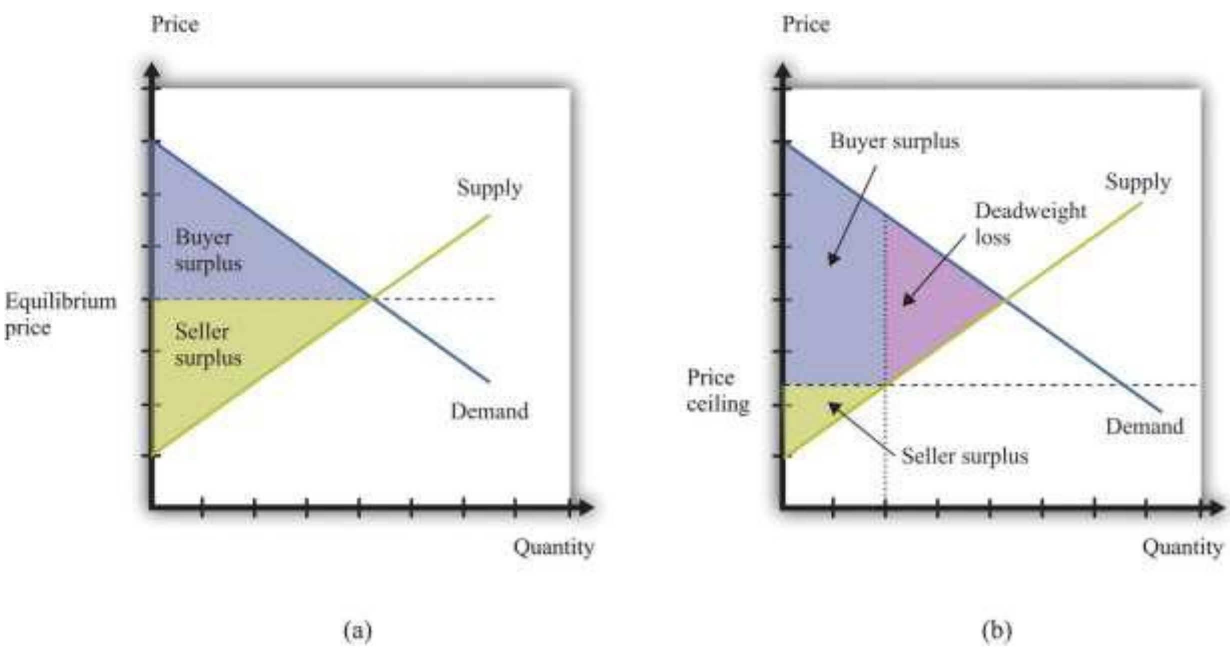
It is more common for governments to intervene in an economy by using price tools rather than quantity tools. In particular, governments sometimes intervene using restrictions on how high the price in a market can go. This is called a *price ceiling*. A classic example of a price ceiling is rent control. In New York City and some other places, there are restrictions on how much landlords can increase the rent on apartments.

Figure 11.4 "The Effects of a Price Ceiling" illustrates a price ceiling. Notice that (unless there is also a quantity ration in place) the price ceiling must be below the equilibrium price; otherwise the policy is irrelevant. The main economic implications of a price ceiling can be readily seen from this figure.

- Because no one can force you to sell if you don't want to, the quantity traded is determined by the supply curve.

- Because the quantity traded is below the equilibrium quantity, there is an inefficiency (deadweight loss).
- Because the quantity demanders wish to buy exceeds the quantity suppliers wish to sell, there must be some kind of rationing in the market to determine who actually buys the good or the service in question.

Figure 11.4 The Effects of a Price Ceiling



With no price ceiling (a), all the possible gains from trade in the market are realized. With a price ceiling (b), some gains from trade are lost because there are fewer transactions.

Rent controls keep the price of an apartment rental below its equilibrium level. Not surprisingly, lots of people would like to live in rent-controlled apartments. The quantity demanded is greater than the quantity supplied. Because the price is, by law, not allowed to undergo the adjustment that would restore equilibrium in the market, some other kind of rationing must take place instead.

Rent controls are enacted with distributional goals in mind. The aim is to ensure that people with lower incomes are not priced out of the rental market. Put differently, the goal is to redistribute income from sellers to buyers—that is, from landlords to those who are renting apartments. A difficulty with price ceilings is that people have an incentive to try to get around the restrictions in creative ways. There is often more to a transaction than a simple exchange of money for a good or a service. There may be nonmonetary aspects of the transaction that governments find harder to regulate. When apartments are covered by rent controls, landlords often ask for "key money." This is an off-the-books, up-front payment that renters must agree to pay before renting the apartment. In other words, it is a polite term for a bribe. In addition, some landlords may not put much money or effort into the upkeep of rent-controlled apartments, thus compensating for the low rent by reducing the quality of the apartment.

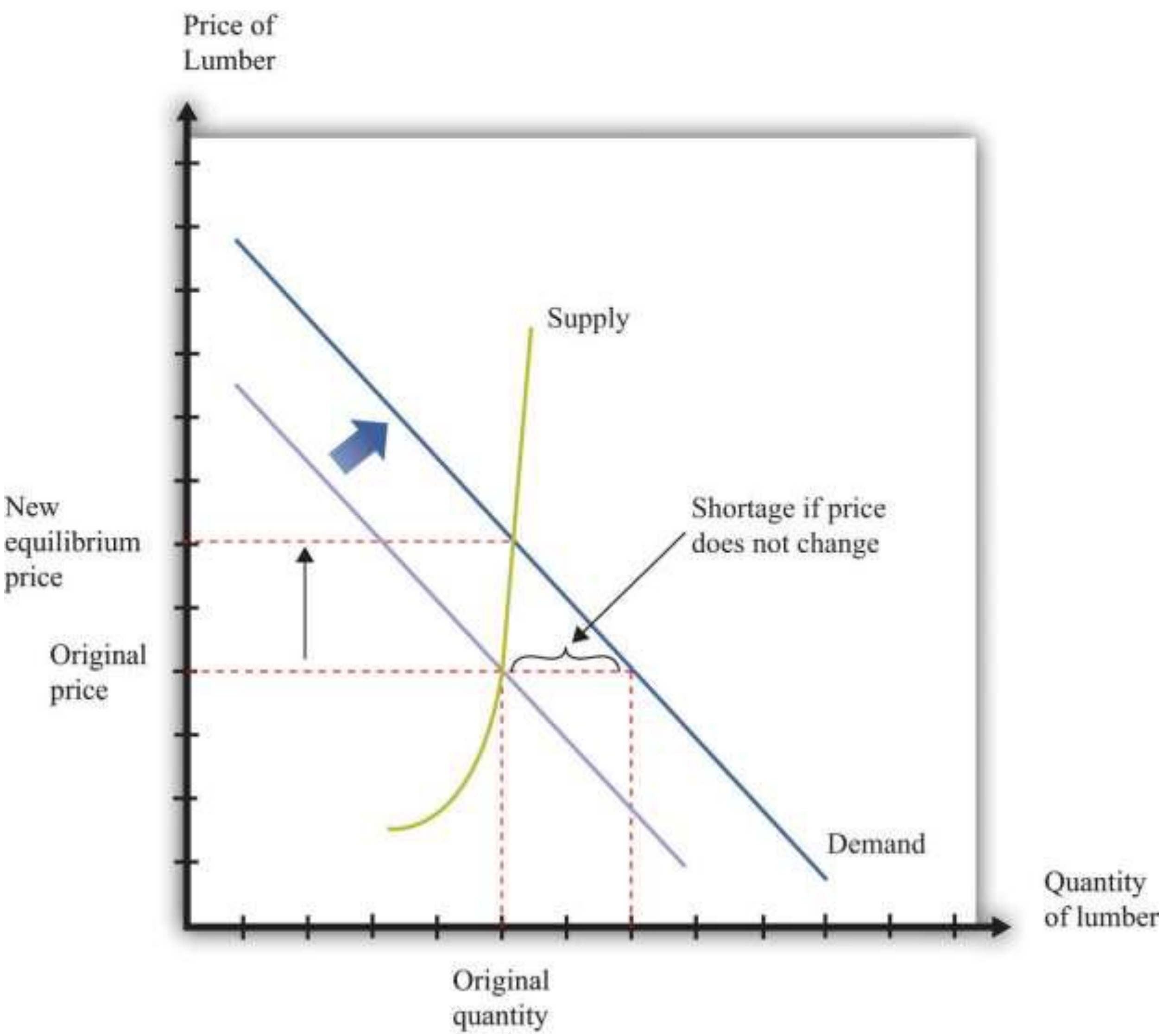
In emergency circumstances, temporary price ceilings may be put into effect. These take the form of laws that prevent so-called *price gouging*. For example, in the aftermath of a hurricane, some goods and services are typically very hard to come by. Basic necessities like food and water may be in limited supply. In the weeks and months after such a disaster, building supplies and similar products may be almost completely unavailable.

After Hurricane Katrina, price-gouging laws applied to states affected by the storm.

> *While there is no federal price gouging law, many states have enacted some type of prohibition or limitation on price increases during declared emergencies. All of the affected states—Louisiana, Mississippi, Alabama, and Florida—have price gouging laws that are triggered by the declaration of an emergency in the state. Generally, the laws prohibit the sale of goods and services in the designated emergency area at prices that exceed the prices ordinarily charged…*
>
> *However, there exists a general exemption for increased prices that are the result of additional costs incurred for procuring the goods or services in question.*
>
> …
>
> *In Alabama,…evidence of unconscionable pricing exists "if any person, during a state of emergency declared pursuant to the powers granted to the Governor, charges a price that exceeds, by an amount equal to or in excess of 25% the average price at which the same or similar commodity or rental facility was obtainable in the affected area during the last 30 days immediately prior to the declared state of emergency."* [4]

Think about the market for lumber (wood for building purposes) in the first few weeks following a hurricane. Were we to apply supply and demand reasoning to this situation, we would get a diagram like Figure 11.5 "The Market for Lumber after a Hurricane". Because there is a great deal of new construction going on, there is a rightward shift in the demand for lumber. The supply of lumber is likely to be fairly inelastic, at least until it is possible to start bringing supplies in from other states. Thus the shift in the demand will lead to a large increase in the existing price. If the price is allowed to increase to its new equilibrium, existing suppliers will obtain a big gain. Price-gouging laws, however, prevent suppliers from raising their prices in this way.

Figure 11.5 The Market for Lumber after a Hurricane



If the market were allowed to work, the price of lumber would increase substantially, but there would not be much more wood supplied. If suppliers are not allowed to increase prices, then demand exceeds supply.

This presents two problems. First, suppliers no longer receive the price signal that tells them to bring more wood to market. In the short run, this may not matter so much. After all, Figure 11.5 "The Market for Lumber after a Hurricane" shows that, with inelastic supply, the shift in the demand curve would not in fact lead to a big increase in the quantity supplied, even if the price were allowed to adjust. In the longer run, though, this is more of a problem because there is less incentive for suppliers from further away to bring in additional lumber.

The second problem with forcing sellers to keep their price fixed is that the increase in demand will lead to a shortage. This is also shown in Figure 11.5 "The Market for Lumber after a Hurricane". Because demand now outstrips supply, the limited supply will have to be rationed in some way. Most likely, what will happen is that demanders will have to queue to get the lumber that they need. The time that they must spend standing in line has an opportunity cost; they would rather spend that time doing something else. We can think of the time spent in line as increasing the effective price that they have to pay.

These arguments do not necessarily mean that price-gouging laws have no merit. In the aftermath of a hurricane, many things may be happening. Lumber firms may see a temporary increase in their market power. Such an increase in market power gives them an incentive to increase prices, so price-gouging laws may serve as a way to limit the abuse of monopoly power.

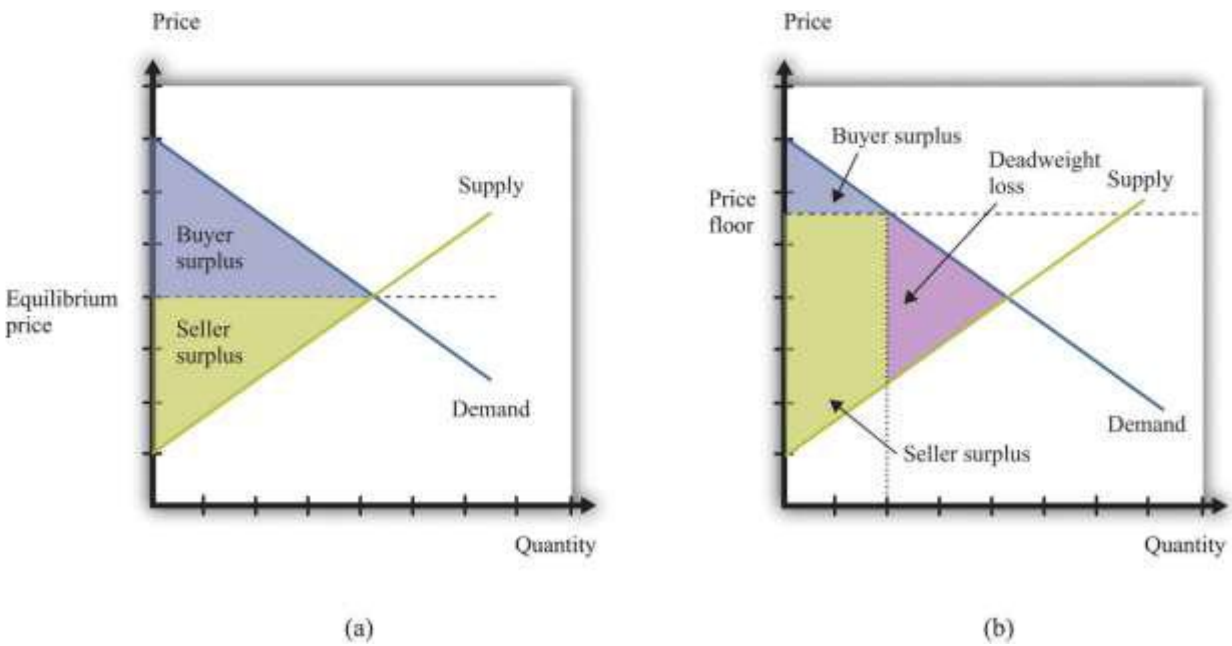
Price Floor

A price floor is closely analogous to a price ceiling. The difference, as the name suggests, is that it is a government-imposed *minimum* price rather than a government-imposed maximum price. The government says that all transactions must be at or above this minimum price. The minimum wage is the most important example of a price floor. [5]

Figure 11.6 "The Implications of a Price Floor" illustrates a price floor. The main economic implications of a price floor can be seen from this figure.

- Because no one can force you to sell if you don't want to, the quantity traded is determined by the demand curve.
- Because the quantity traded is below the equilibrium quantity, there is inefficiency (deadweight loss).
- Because the quantity suppliers wish to sell exceeds the quantity demanders wish to buy, there must be some kind of rationing in the market to determine who actually sells the good or the service in question.

Figure 11.6 The Implications of a Price Floor



With no price floor (a), all the possible gains from trade in the market are realized. With a price floor (b), some gains from trade are lost because there are fewer transactions.

Just as renters use key money and other devices to get around rent control, firms (and workers) sometimes devise ways to get around minimum wage requirements. Employers who are forced to pay a minimum wage may provide worse working conditions than those who pay a market wage. Or, if you want to work at a company and are willing to work at less than the minimum wage, you can negotiate a deal with your employer so that you are paid the minimum wage for reported hours but then work additional hours for nothing. The minimum wage regulations in the United States stipulate that this is illegal, punishable with fines of $1,100 per violation. [6]

Sometimes individuals work their way around such restrictions even more blatantly. In the former Soviet Union, price ceilings were put in place in an attempt to keep the prices of basic goods down for households. Martin Walker, a journalist in Moscow, wrote of his experiences with these price ceilings in the food markets outside Moscow. [7] A butcher offering to sell Walker a side of beef assured him that the price per kilogram was fixed. "However," said the butcher, "the weight is subject to negotiation."

In the Soviet Union, the limited supply of goods led to long lines for those who wanted to purchase basic commodities, such as bread. You can think of these lines as an additional component of the price: you pay money *plus* the value of the time that you spend standing in line.

Although price ceilings and price floors have different implications for the price in the market, they both imply that *the quantity traded in the market will be less than the equilibrium quantity*. The reason is simple: neither buyers nor sellers can be forced to trade if they do not want to. If the price is above the equilibrium price, the quantity is determined by the amount of the good or the service that people are willing to buy. Some would-be sellers are disappointed because they cannot find someone to buy from them. With a minimum wage, for example, not everyone who wants a job can find one. If the price is below the equilibrium price, the quantity is determined by the amount of the good or the service that people are willing to sell. Some would-be buyers are disappointed because they cannot find someone to sell to them. With rent control, for example, not everyone who wants a cheap apartment can find one.

Taxes and Subsidies

Price floors, price ceilings, and quantity restrictions are important but relatively rare policies. The government intervenes regularly in almost every market in the economy in a different way—by the imposition of taxes. Had you purchased the milk and sugar on our shopping list, for example, you would very likely have paid a sales tax. Sometimes cities levy their own sales taxes as well. On certain goods, such as alcohol or gasoline, you may pay additional taxes.

Taxes

A tax is a payment made to the government that is associated with an economic transaction. Although the details of taxes can differ substantially, most taxes come down to one simple point: the price paid by the buyer is higher than the price received by the seller.

Suppose you want to purchase a book for its list price, say, $20. In the United States, if you take this book to the cash register, you will typically be charged a sales tax. If the sales tax is 5 percent, you will have to pay $21 for the book. The store collects the $1 tax on behalf of the government. So who is paying this tax? On the one hand, the amount of the tax is marked right there on the receipt as an amount you have to pay. Yet it is the store that actually sends the money to the government.

Imagine, by contrast, that you had to give the bookstore only $20 but then were personally responsible for sending the sales tax to the government. You would have to file a sales tax declaration each year for every item you bought. That would be both inconvenient and difficult for the government to monitor; for this reason, sales taxes are funneled through the seller. But we are interested in a more fundamental question: would this make a difference on who pays the tax? The answer is no. You would still pay $21, the government would still get $1, and the bookstore would still get $20.

In other words, *it does not make any difference whether the tax is imposed on buyers or sellers*. This is one of the most surprising results that economics teaches us. In our book example, the conclusion may seem obvious. Yet people often to fail to appreciate the far-reaching significance of this insight.

For example, social security taxes in most countries are imposed on both workers and employers. Suppose the government changed its policy and declared that the portion of social security that was previously paid by the employer now had to be paid by the worker instead. Looking at this as employed workers, we might think that we had just been hit with a huge tax increase. Indeed, if nothing else changed, the policy change would make workers worse off. Fortunately, the logic of supply and demand would quickly come to our rescue. At existing wages, firms would no longer be able to hire all the workers they wanted. Wages would be bid up, and before long we would expect to see workers and firms no better and no worse off than they were previously.

Who Pays the Tax?

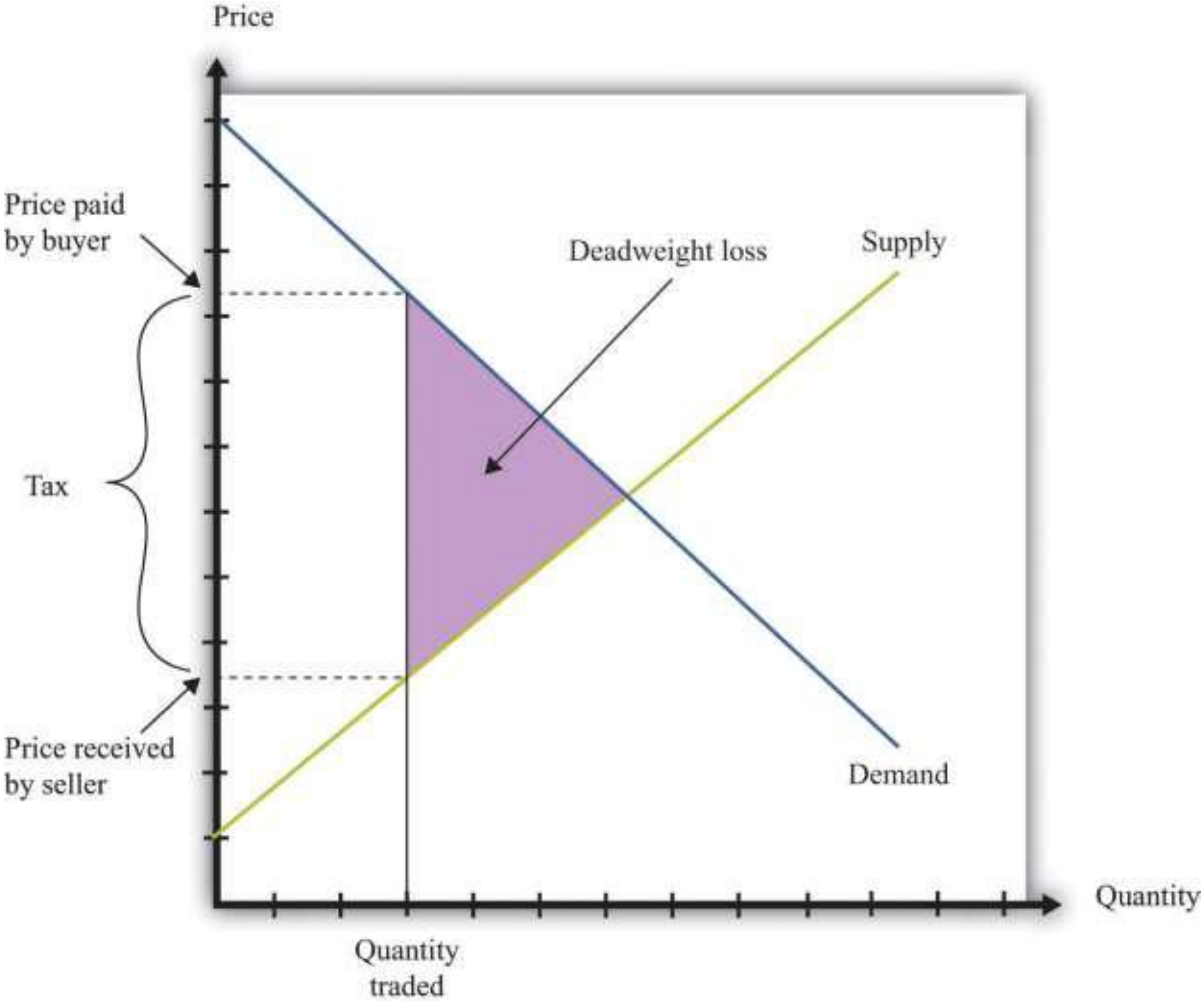
The key question, then, is not who sends the money to the government. The key question is, *What happens to prices when a tax is imposed?*

To answer this, imagine that the government increases taxes on gasoline by 50 cents a gallon and consider two extreme cases. First, suppose the price of gas increases by 50 cents a gallon. Households are evidently paying the tax; the amount they must pay per gallon has gone up by the full amount of the tax. Now suppose that the price of gasoline at the pump does not change at all. Then firms are paying the tax: they are receiving 50 cents less per gallon once they pay the tax to the government. Most often, we expect to see the price of gasoline increase but by less than 50 cents. Therefore, the burden of the tax is shared between the gas station and the household. It is the change in the price that tells us who *really* pays the tax.

Figure 11.7 "The Deadweight Loss from a Tax", and Figure 11.8 "The Loss in the Buyer Surplus and the Seller Surplus from a Tax" illustrate the effects of a tax.

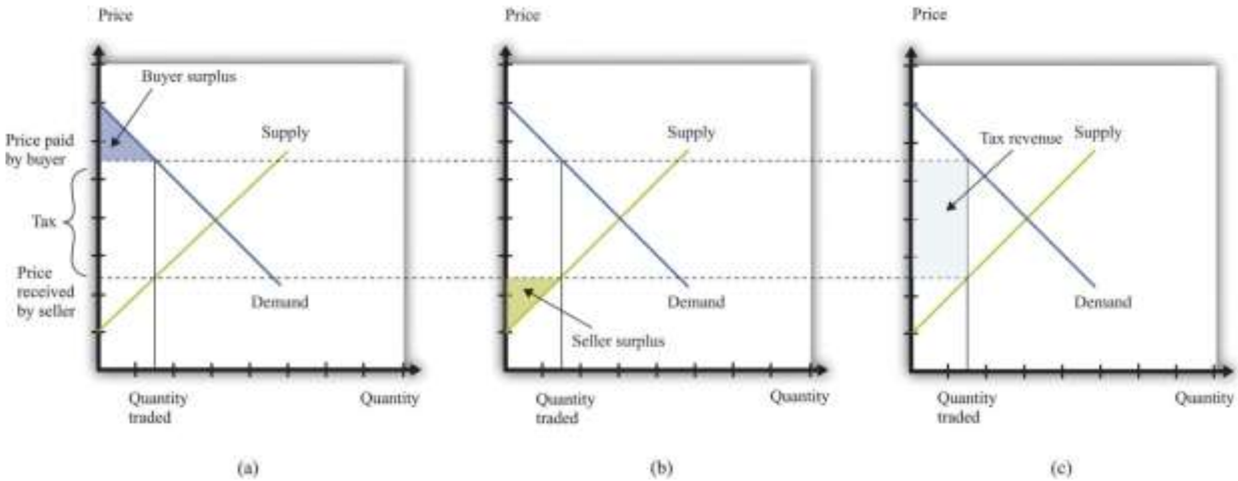
- The gap between the buyer's price and the seller's price means that the quantity sold is less than the market equilibrium quantity (Figure 11.7 "The Deadweight Loss from a Tax").
- There is a deadweight loss: some mutually beneficial trades go unrealized. This is again visible in Figure 11.7 "The Deadweight Loss from a Tax". There are potential trades where the buyer's valuation exceeds the seller's valuation. However, because the difference in valuations is less than the amount of the tax, these trades are not worthwhile once the tax must be paid.
- There is a reduction in both the buyer surplus and the seller surplus, as can be seen in Figure 11.8 "The Loss in the Buyer Surplus and the Seller Surplus from a Tax". The buyer surplus is the area under the demand curve and above the price paid. The seller surplus is the area above the supply curve and below the price received.
- Figure 11.8 "The Loss in the Buyer Surplus and the Seller Surplus from a Tax" also shows that some of the surplus generated by these trades now goes to the government in the form of tax revenues. Government tax revenues equal the amount of the tax multiplied by the quantity traded. Graphically, they are equal to the rectangle shown in part (b) of Figure 11.8 "The Loss in the Buyer Surplus and the Seller Surplus from a Tax".

Figure 11.7 The Deadweight Loss from a Tax



A tax means that there is a wedge between the price paid by the buyer and the price received by the seller.

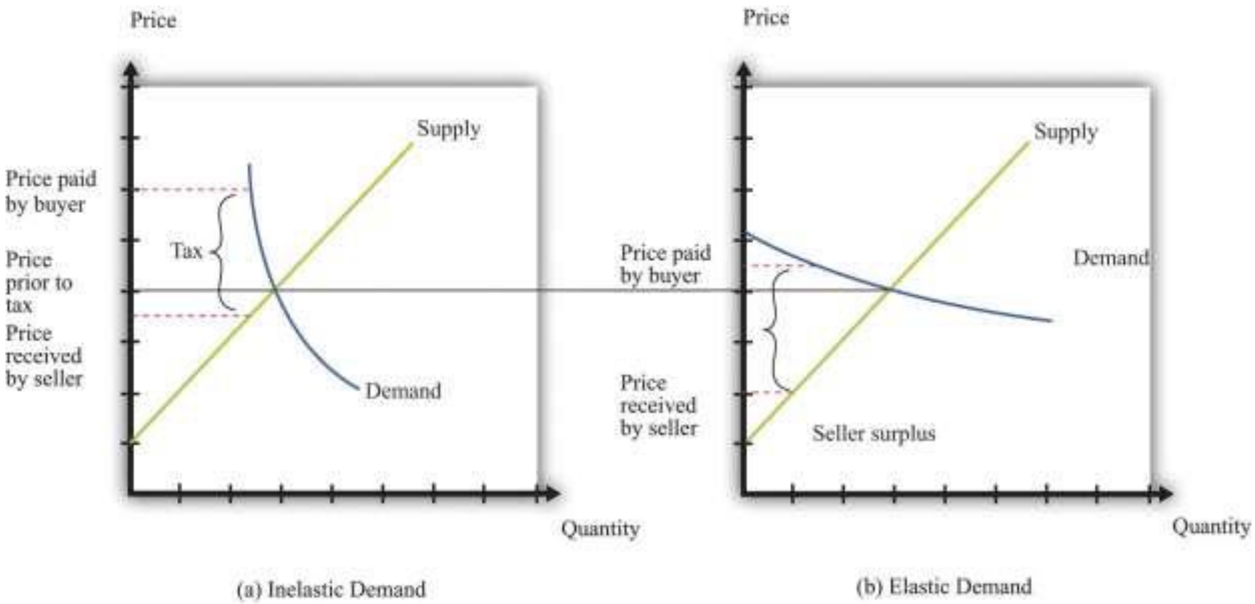
Figure 11.8 The Loss in the Buyer Surplus and the Seller Surplus from a Tax



The total surplus is the sum of the buyer surplus (a), the seller surplus (b), and the tax revenue received by the government (c).

Tax incidence is the way in which the burden of a tax is divided between buyers and sellers. In general, the incidence of a tax depends on theprice elasticity of supply and the price elasticity of demand. Figure 11.9 "Tax Incidence with Inelastic and Elastic Demand" shows why tax incidence depends on the elasticity of demand. That figure has two parts. In both parts, we start from the same initial competitive equilibrium and impose a tax of the same size. This means that the gap between the price paid by buyers and the price received by sellers is identical.

Figure 11.9 Tax Incidence with Inelastic and Elastic Demand



When demand is inelastic (a), most of the burden of the tax is borne by buyers, while the opposite is true when demand is elastic (b).

In part (a) of Figure 11.9 "Tax Incidence with Inelastic and Elastic Demand", demand is inelastic. Buyers are not very price sensitive, so even if the price increases, their quantity demanded does not change a great deal. The result is that the price paid by buyers increases a lot. Most of the burden of the tax is borne by buyers. In part (b) of Figure 11.9 "Tax Incidence with Inelastic and Elastic Demand", demand is elastic. As the price increases, the quantity demanded decreases a great deal. In this case, the price paid by buyers increases much less, and the price received by sellers decreases by more. Most of the burden of the tax is borne by sellers.

Keep in mind also that the distortion induced by the tax is smaller when demand is inelastic. The key indicator of the distortion is how much change there is in the quantity traded. When demand is inelastic, the quantity traded changes by less. As a consequence, there is a much smaller deadweight loss in part (a) of Figure 11.9 "Tax Incidence with Inelastic and Elastic Demand" than in part (b) of Figure 11.9 "Tax Incidence with Inelastic and Elastic Demand".

Why Do Governments Impose Taxes?

Given our analysis so far, you might think that governments should not impose taxes at all. After all, taxes reduce the surplus received by buyers and sellers. However, there are several reasons why governments tax households and firms, despite the adverse consequences for the gains from trade. [8]

Raising revenue. Governments perform certain essential functions, such as maintaining a legal system and defending the borders. Governments also typically supply various goods and services (such as roads, schools, and streetlights) as well as paying out subsidies to certain industries and transfers to individuals. All of these require government revenues. We are not interested right now in which of these things governments *should* do nor with the question of whether governments intervene too much or too little in the economy. It is simply a fact that governments incur a lot of expenses, and these expenses must be paid for through taxation. One key reason for taxes is therefore to raise revenue to fund government activities.

In fact, governments sometimes finance their expenses through borrowing rather than current taxation. But borrowing is the same as deferred taxation: the debt obligation must eventually be paid through taxes levied in the future.

Redistributing income. Taxes are a means by which governments can take money from one group of people and give it to another. Governments often use progressive taxation, meaning that the rich are taxed proportionately more than the poor. Taxation then serves to make the distribution of income more equal. [9]

Externalities. In some circumstances, an individual's actions have an influence, either positive or negative, on others in the economy. Economists call such an effect an externality. [10] In the presence of externalities, distortions in the market and some type of government intervention may be warranted. Often, that intervention takes the form of taxes and subsidies that alter individual incentives to encourage behavior that promotes economic efficiency.

Sometimes externalities are adverse; these are known as negative externalities. The effect of second-hand smoke is an example. Other times there are positive externalities associated with an action. An example is education, which has benefits to society as well as to the individual who obtains the education. When there are negative externalities, the government can impose a tax to discourage the activity in question. When there are positive externalities, the corresponding government response is a subsidy.

Uninformed choices. Economists generally presume that informed individuals will make informed choices. Not everyone agrees with economists about this. One often hears the argument that governments ought to intervene so that individuals do not make the "wrong decisions." Take, for example, the decision to smoke cigarettes. It has been known for a long time that cigarette smoking is harmful to one's health. One reasonable view is that smoking should be purely a matter of individual choice: people can make their own choices about the enjoyment of smoking versus the adverse health effects. As long as individuals make informed choices, there seems to be little basis for government intervention.

But another view is that people are not always capable of informed choice. Perhaps people are not good at making decisions that involve their health 30 years from now. Perhaps people are not good at making decisions about addictive substances. Perhaps it is not appropriate to think of rational individuals making informed choices when many people start smoking as children. An argument can then be made that governments should step in and alter incentives, through taxes and subsidies, to help people make better choices.

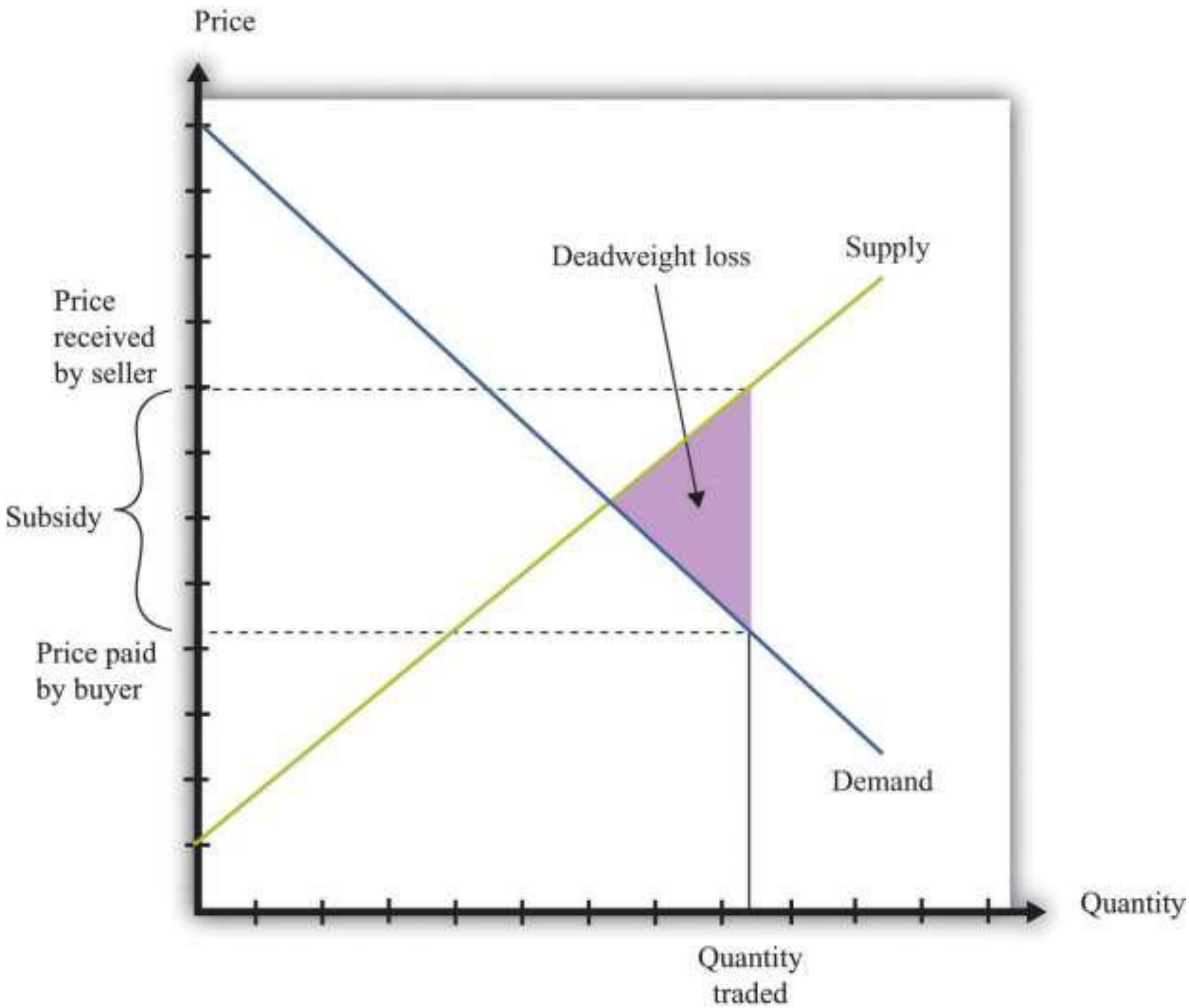
Toolkit: Section 17.2 "Elasticity"

You can review the definition and calculation of elasticities in the toolkit.

Subsidies

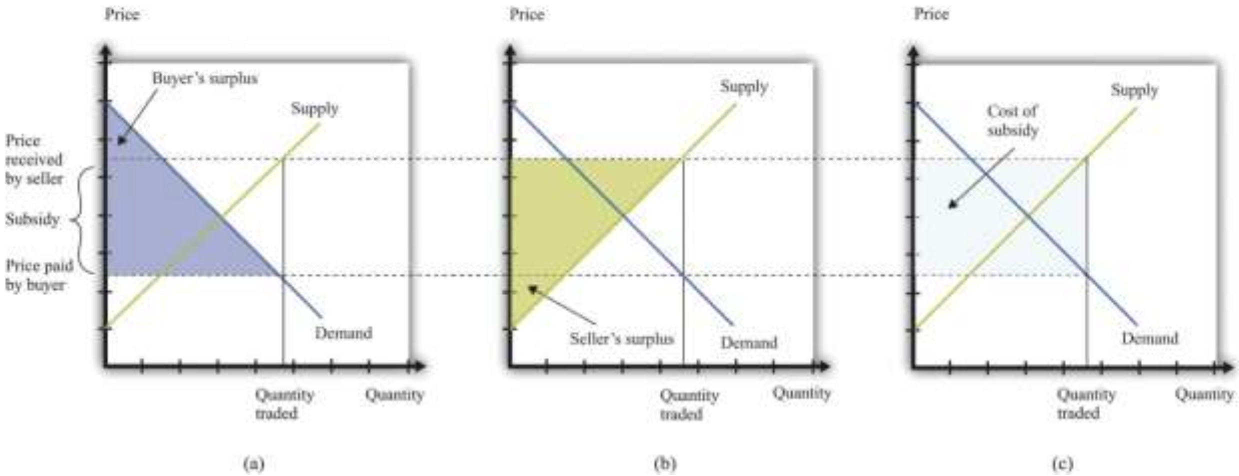
A subsidy is the opposite of a tax. It is a payment made to a producer to encourage production. A subsidy means that the price paid by the buyer is lower than the price received by the seller. Figure 11.10 "The Deadweight Loss from a Subsidy" shows the deadweight loss from a subsidy. Subsidies distort markets not by leading to too small a quantity being traded but by causing too large a quantity to be traded. The deadweight loss lies to the right-hand side of the competitive equilibrium quantity because some trades occur where the cost exceeds the benefit. Figure 11.11 "The Buyer Surplus and the Seller Surplus after the Imposition of a Subsidy" shows the buyer surplus and the seller surplus in the presence of a subsidy. Both are increased by the subsidy. However, subsidies mean that the government spends resources rather than taking them in. The figure shows that the cost of the subsidy is greater than the increased surplus received by the buyers and the sellers. The difference between the cost and the increases in surplus is the deadweight loss.

Figure 11.10 The Deadweight Loss from a Subsidy



A subsidy means that some transactions are now carried out even though they actually destroy value.

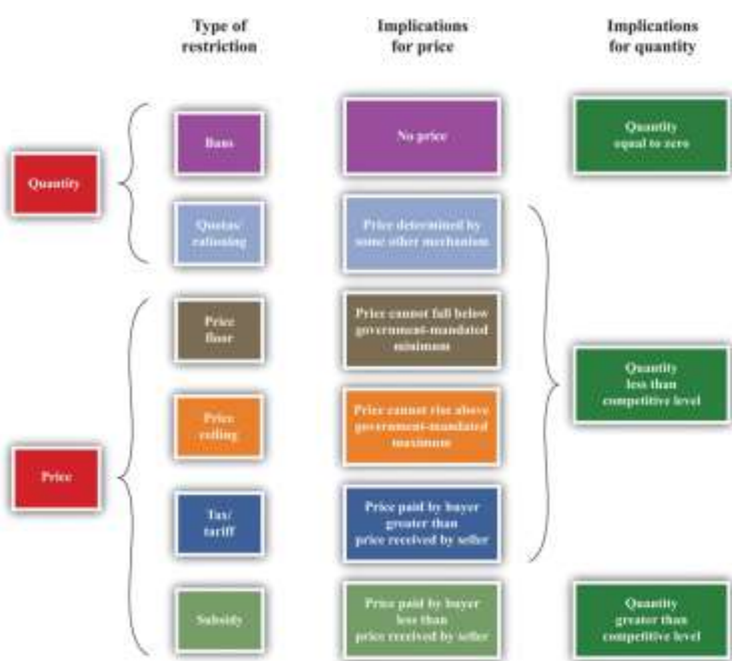
Figure 11.11 The Buyer Surplus and the Seller Surplus after the Imposition of a Subsidy



The buyer surplus and the seller surplus are shown in (a) and (b), and the cost of the subsidy is shown in (c). The total surplus is obtained by adding together the buyer surplus and the seller surplus and then subtracting the subsidy paid by the government.

Figure 11.12 "The Different Ways in Which Governments Intervene in Markets" summarizes the different kinds of trade restrictions that we have looked at.

Figure 11.12 *The Different Ways in Which Governments Intervene in Markets*



KEY TAKEAWAYS

- Government restrictions take a variety of forms, including bans on trades, controls on prices, and the imposition of taxes and subsidies to change incentives. These are summarized in Figure 11.12 "The Different Ways in Which Governments Intervene in Markets".
- Some of the reasons governments restrict trades are to protect individuals and society from unsafe and unhealthy products, for moral reasons, and for fairness.
- Through these restrictions, some gains from trade may be lost. For example, in the presence of taxes, there are deadweight losses due to the lost gains from trade. If a market is shut down entirely, then all the gains from trade are lost. In some cases, individuals find a way to circumvent government restrictions to realize these gains from trade.

CHECKING YOUR UNDERSTANDING

1. In what sense is the closing down of a market like a tax?
2. If the government sets a tax rate, how is the quantity of revenue collected determined?
3. Explain why the allocation of the tax burden does not depend on who pays a tax to the government.

[1] Peggy McClone, “Parents Are Angry about ‘Hannah Montana’ Ticket Sales,” *The Star Ledger*, October 29, 2007, accessed January 29, 2011,http://www.nj.com/news/index.ssf/2007/10/parents_are_upset_about_hannah.html.

[2] Ruth L. Baxter, interview by Joelle Kirch Preksta, May 21, 2001, Carnegie Library of Pittsburgh, http://www.carnegielibrary.org/research/pittsburgh/history/ww2/ww27.html.

[3] Mary Hresko and Mary Vincher Shiner, interview by Mark Kernion, May 21, 2001, Carnegie Library of Pittsburgh,http://www.carnegielibrary.org/research/pittsburgh/history/ww2/ww29.html.

[4] Angie A. Welborn and Aaron M. Flynn, “Price Increases in the Aftermath of Hurricane Katrina: Authority to Limit Price Gouging,” *Congressional Research Service Report for Congress*, September 2, 2005, accessed January 29, 2011,http://www.fas.org/sgp/crs/misc/RS22236.pdf.

[5] We devote a whole chapter to the analysis in Chapter 10 "Raising the Wage Floor". With the minimum wage, the aim is to redistribute income from buyers to sellers—that is, from firms to suppliers of unskilled labor.

[6] US Department of Labor, “Wages: Minimum Wage,” accessed March 14, 2011,http://www.dol.gov/dol/topic/wages/minimumwage.htm.

[7] Martin Walker, *Guardian*.

[8] Many of these arguments for taxation are also discussed in other chapters.

[9] In Chapter 12 "Superstars", we look in detail at the arguments for redistribution in society.

[10] Chapter 13 "Cleaning Up the Air and Using Up the Oil" is all about such externalities.

11.2 Limits on Trade across Borders

LEARNING OBJECTIVES

1. What is the source of gains to trade across two countries?
2. Why do governments put restrictions on trade across borders?
3. How do governments restrict international trade?

Restrictions do not appear only within a country. We see restrictions on trade across countries as well. In our shopping list at the beginning of the chapter, we mentioned several goods that are imported from other countries, such as Cuban cigars and French cheese. We begin by reviewing the motivations for trade between countries. Just as individuals are motivated to trade by the fact that it can make them better off, countries can also benefit from trading with each other.

Comparative Advantage

The principle of **comparative advantage** provides one reason why there are gains from trade among individuals. [1] Because different individuals have different skills and abilities, everyone can benefit if people specialize in the things that they do relatively well and trade with others to obtain the goods and services that they do not produce. Such specialization is a cornerstone of our modern economy, in which people are specialists in production but generalists in consumption.

The idea of comparative advantage also provides a basis for trade among countries. In the absence of trade, countries end up producing goods and services that they can provide only very inefficiently. When countries trade, they can instead specialize in the goods and the services that they can produce relatively efficiently. All countries can take full advantage of their different capabilities.

We illustrate comparative advantage in a simple way, with a story about trade between Guatemala and Mexico. If you understand this story, you should also be able to see that we could make the example more complex and yet still keep the same basic insight. Table 11.1 "Beer and Tomato Production in Mexico and Guatemala" provides information about the technology in each country: how much a typical individual can produce in a 36-hour workweek. The table shows how much time is required in each country to produce two goods: beer and tomatoes.

Table 11.1 Beer and Tomato Production in Mexico and Guatemala

	Hours of Labor Required	
	Tomatoes (1 kilogram)	**Beer (1 liter)**
Guatemala	6	3
Mexico	2	2

In both Mexico and Guatemala, people like to consume beer and tomatoes in equal quantities: 1 liter of beer to accompany each kilogram of tomatoes. In Guatemala, it takes 6 hours of labor to produce 1 kilogram of tomatoes, and 3 hours of labor to produce 1 liter of beer. In 9 hours, therefore, it is possible to produce 1 kilogram of tomatoes *and* 1 liter of beer. In a 36-hour week, the worker can enjoy 4 kilograms of tomatoes accompanied by 4 liters of beer.

Mexico is much more efficient at producing both tomatoes and beer. It takes only 2 hours to produce 1 kilogram of tomatoes, and it takes only 2 hours to produce 1 liter of beer. In 36 hours, therefore, a Mexican worker can produce 9 kilograms of tomatoes and 9 liters of beer.

Because Mexico is better at producing both tomatoes and beer—it has an**absolute advantage** in the production of both goods—it would be natural to think that Mexico has nothing to gain from trading with Guatemala. But this conclusion is wrong. Mexico is a bit better at producing beer but a lot better at producing tomatoes. Guatemala has a comparative advantage in the production of beer. One way to see this is through opportunity cost. In Guatemala, the opportunity cost of producing 1 kilogram of tomatoes is 2 liters of beer. In Mexico, the opportunity cost of producing 1 kilogram of tomatoes is only 1 liter of beer. Thus Guatemala should specialize in the production of beer.

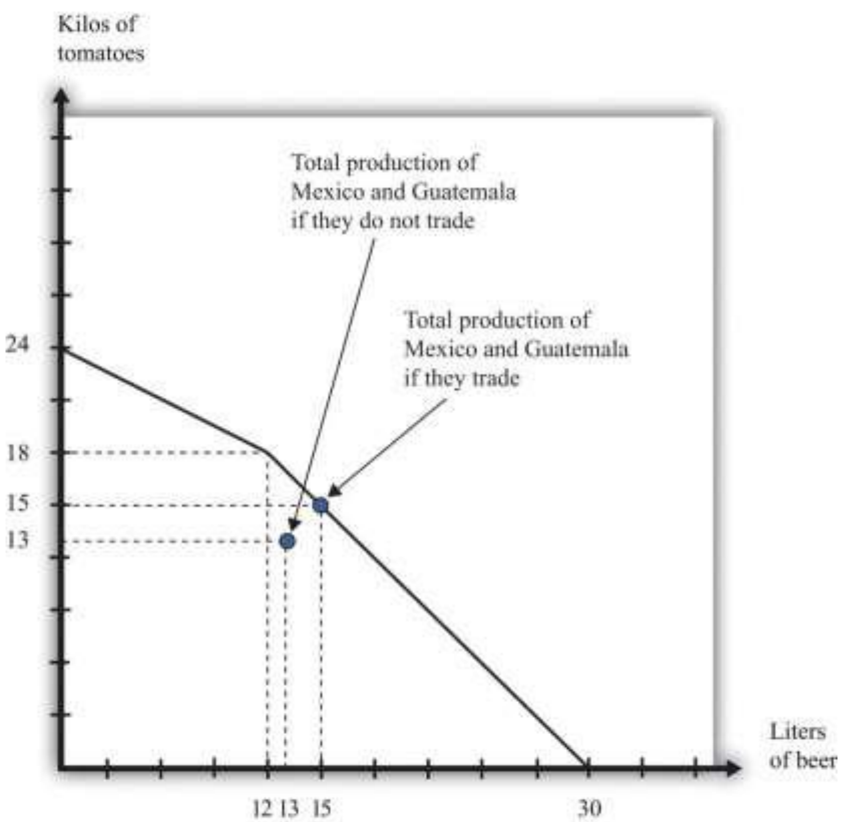
In a 36-hour week, Guatemala produces 12 liters of beer. Now suppose Mexico devotes 30 hours to producing tomatoes and only 6 hours to producing beer. Then Mexico will produce 15 kilograms of tomatoes and 3 liters of beer. The two countries produce, in total, 15 kilograms of tomatoes and 15 liters of beer. Previously, they were producing 13 kilograms of tomatoes and 13 liters of beer. Both countries can be better off if they trade and take advantage of comparative advantage. We illustrate this in Figure 11.13

"The Production Possibilities Frontier". [2] It shows the joint **production possibilities frontier** for Guatemala and Mexico. When they produce individually and do not trade, they end up at a point inside the production possibilities frontier. If they specialize and trade, they end up on the production possibilities frontier instead.

Toolkit: Section 17.12 "Production Possibilities Frontier", and Section 17.13 "Comparative Advantage"

You can review the idea of the production possibilities frontier and the concepts of comparative and absolute advantage in the toolkit.

Figure 11.13 The Production Possibilities Frontier



Individually, Mexico and Guatemala produce 13 kilograms of tomatoes and 13 liters of beer. Jointly, they can produce 15 kilograms of tomatoes and 15 liters of beer by exploiting comparative advantage.

Why Do Governments Intervene in International Trade?

To economists, the logic of comparative advantage is highly compelling. Yet noneconomists are much less convinced about the desirability of free trade between countries. We see this reflected in the fact that countries erect a multitude of barriers to trade. Where economists see the possibility of free trade and mutual gain, others often see unfair competition. For example, many countries in the developing world have very low wages compared to the United States, Europe, and other relatively developed economies. Economists see this as a source of comparative advantage for those countries. Because labor is cheap, those countries can produce goods that require a large amount of labor. Countries like the United States, by contrast, can specialize in the production of goods that require less labor. The logic of comparative advantage suggests that both countries would be made better off. To noneconomists, however, the cheap labor looks like "unfair" competition—how can workers in rich countries compete with workers in poor countries who are paid so much less?

This concern has some merit. Comparative advantage tells us that a country as a whole is made better off by trade because that country can have more goods available for consumption. Yet comparative advantage, in and of itself, says nothing about *who* gets those benefits or how they are shared.

Hypothetically, it is possible to share these goods so that everyone is made better off. [3] As a practical matter, even if the country as a whole has more goods and services to consume, some individuals within the country are made worse off. There are winners and losers from trade, and there is frequently political pressure to limit international trade from or on behalf of those who lose out.

Another reason governments intervene in international trade is because of political lobbying. Generally, the beneficiaries from trade barriers are a small and identifiable group. For example, the United States provides sugar subsidies that increase the price of sugar. Sugar producers are the clear beneficiaries of this policy and have an incentive to lobby the government to ensure that the subsidies stay in place. The losers from this policy are those who consume sugar—that is, all of us. But there is no lobby representing sugar consumers.

Whatever the reasons, governments frequently intervene in international trade. Sometimes they completely close certain markets. Sometimes they impose limits on how much can be imported from abroad. And sometimes they impose special taxes on imports. We look at each in turn.

Sanctions and Bans

In some cases, governments close down certain categories of overseas trade completely. They may do so in an attempt to further international political goals. An example from our shopping list is the Cohiba cigars. You cannot buy these directly from Cuba due to a ban on the import of Cuban goods into the United States. [4] This policy is designed to make it harder for Cuba to function in the world economy and thus puts pressure on the Cuban government.

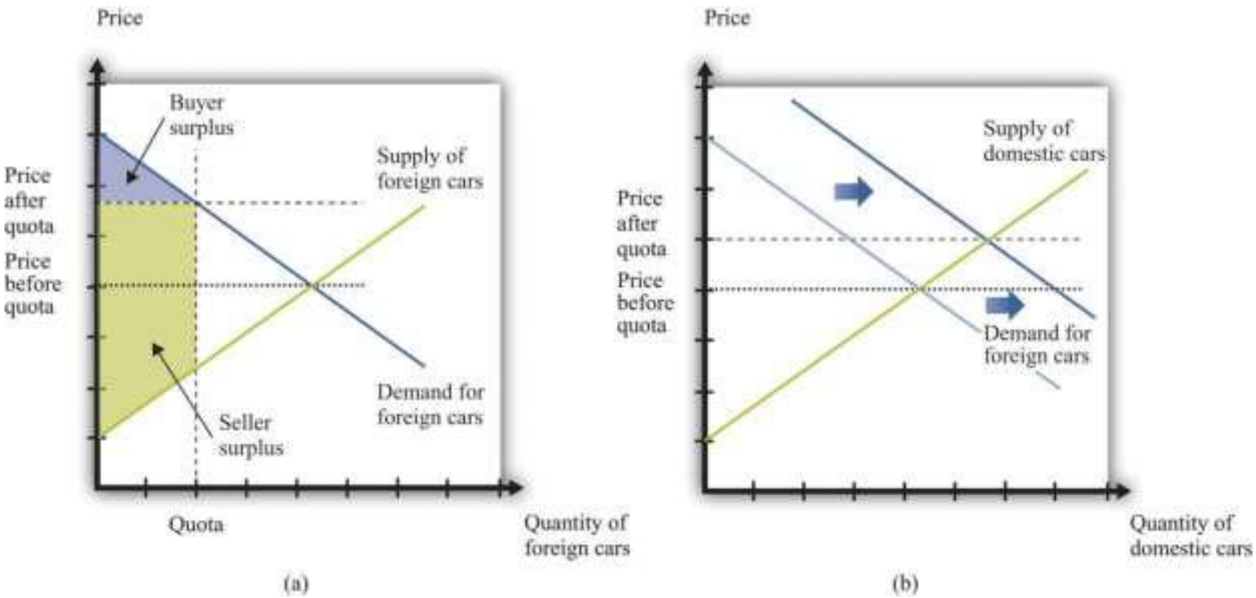
Governments quite often use international sanctions in an attempt to achieve political goals. These measures can be enacted by individual governments or by international bodies such as the United Nations. Currently, the international community is putting pressure on Iran because of concerns about the development of nuclear capabilities in that country. [5] From 1990 to 2003, there were international sanctions placed on trade with Iraq.

Governments also ban certain products from overseas for reasons of health and safety. The United States does not allow the importation of cheeses made with raw milk because it argues that such cheeses pose a health risk; thus it is difficult to find the French raw milk camembert on our shopping list. When the United Kingdom had an outbreak of bovine spongiform encephalopathy (better known as mad cow disease), many countries banned the import of beef from that country. More generally, countries have their own health and safety laws, so foreigners who wish to compete in these markets must ensure that they satisfy these standards.

Quotas

Another way in which governments frequently intervene in international transactions is by means of a quota—that is, a quantity restriction on imports. Figure 11.14 "The Effects of a Quota" gives an example of how quotas work. Suppose that Australian consumers buy both domestically produced cars (Holdens) and cars imported from the United States (Fords). These cars are not perfect substitutes for each other, so we draw a market for each kind of car. To begin with, both markets are in equilibrium where demand equals supply. Suppose that Australia were then to impose a quota on the import of Fords. The price of Fords is determined by consumers' willingness to pay at the quantity set in the quota—that is, we can find the price by looking at the demand curve. Australian consumers must pay more for Fords. Meanwhile, the fact that fewer Fords are being sold means that Australian households will demand more Holdens. The demand curve shifts to the right. This increases the price of domestic vehicles.

Figure 11.14 The Effects of a Quota



After the imposition of the quota, the price of cars increases in the market for foreign cars (a) and the demand for domestic cars increases (b).

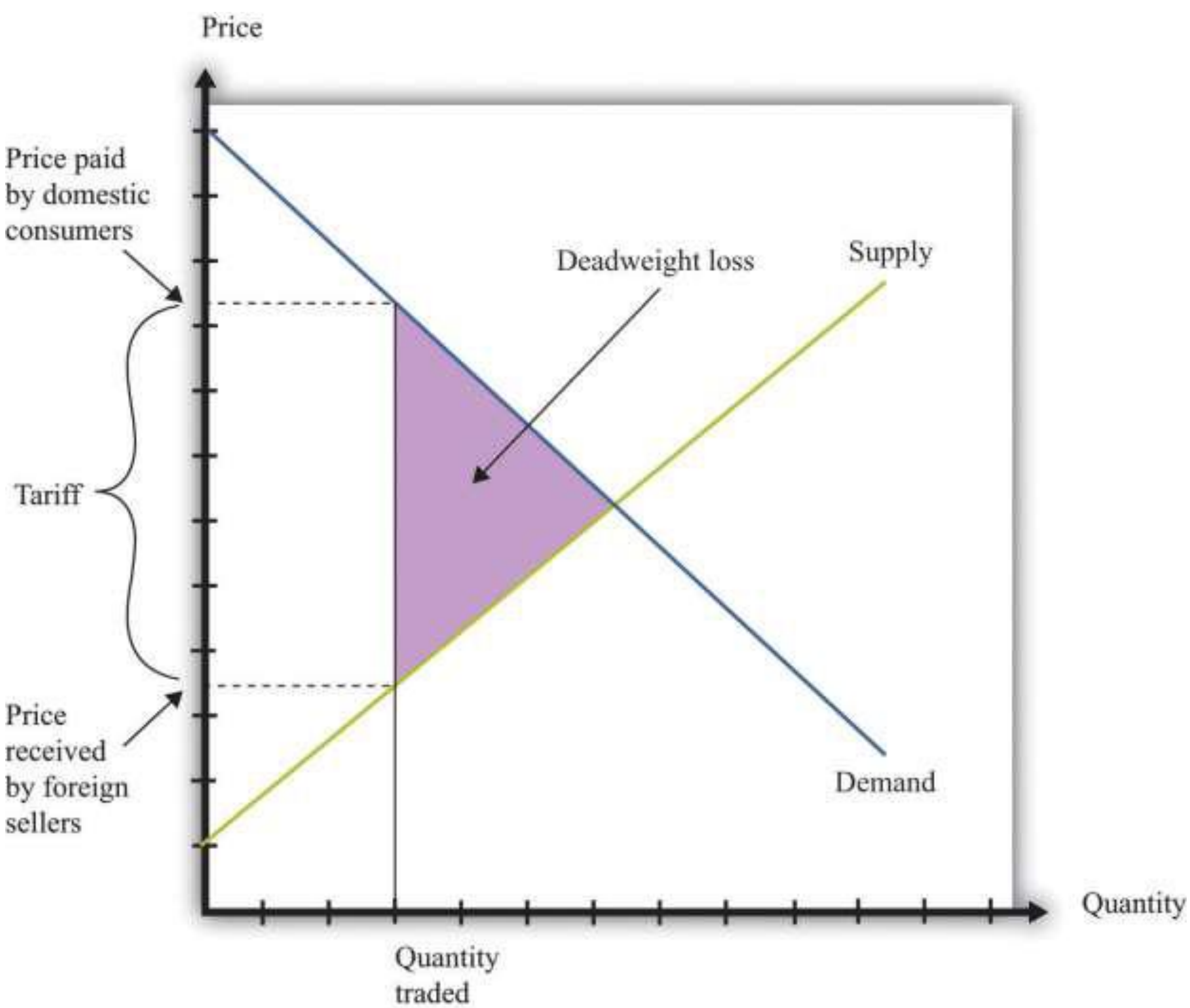
Who are the winners and the losers in this process? The clear winners are domestic producers of automobiles. They get to sell more cars at a higher price, and their surplus increases. Australian consumers, meanwhile, are losers. We cannot see this immediately by looking at the buyer surplus because the buyer surplus decreases in the market for foreign cars but increases in the market for domestic cars. However, we can tell that consumers are worse off because both domestic and foreign cars have become more expensive. Finally, the effects on foreign producers are in general ambiguous. They sell fewer cars but at a higher price. American producers might even benefit from the Australian quota.

In general, governments that impose quotas are transferring resources from domestic consumers to domestic producers. This illustrates the point we made earlier: the beneficiaries of this kind of policy are typically a small group—in this case, producers of Holdens. The losers are everyone who wants to buy a car. The producers are likely to have much more political influence than the consumers.

Tariffs

Where quotas are the equivalent of quantity restrictions, applied in the context of international trade, tariffs are the equivalent of taxes. A tariff on a good is an amount that must be paid by someone who wishes to import that good from another country. The main implication of a tariff is that the price received by foreign sellers is less than price paid by domestic consumers. We illustrate a tariff in Figure 11.15 "The Effects of a Tariff".

Figure 11.15 *The Effects of a Tariff*



A tariff means that there is a wedge between the price paid by domestic buyers and the price received by foreign sellers. Just as with a tax, the quantity traded is lower because some transactions are no longer worthwhile. There is a deadweight loss.

The main implications are very similar to those of a tax. Consumers are made worse off, as are foreign producers. There is a deadweight loss, as indicated inFigure 11.15 "The Effects of a Tariff". As with quotas, tariffs are often designed to protect domestic producers. Thus, as we saw when looking at a quota (Figure 11.14 "The Effects of a Quota"), a tariff on foreign goods induces a substitute toward goods produced in the domestic country. This is the law of demand at work: when the price of a good increases, the demand for substitute products will be increased.

One element that distinguishes a tariff from a quota is the collection of government revenue. When a quota is imposed, trade is limited at a particular quantity, but the government collects no revenue. Instead, as shown in Figure 11.14 "The Effects of a Quota", the surplus from the trade is distributed among buyers and sellers. When a tariff is imposed, the government collects revenue equal to the product of the tariff and the quantity traded. Comparing Figure 11.14 "The Effects of a Quota" with Figure 11.15 "The Effects of a Tariff", part of the surplus that was shared by buyers and sellers under the quota is now captured as revenue by the government. This parallels the results that we saw when looking at domestic quotas and taxes earlier in the chapter.

KEY TAKEAWAYS

- There are gains to trade across countries due to comparative advantage.
- Governments place restrictions on trade for political reasons, to protect jobs, and to increase revenue by taxing trade.
- Governments may impose outright bans on trade, place limits on the quantities traded, or put taxes on trade.

CHECKING YOUR UNDERSTANDING

1. Looking at Table 11.1 "Beer and Tomato Production in Mexico and Guatemala", we said that Mexico had an absolute advantage in both goods, but still there were gains from trade. Change two numbers in the

table so that Guatemala has an absolute advantage in both goods. Explain how each country still has a comparative advantage in your new example.

2. Do the economic effects of a tariff depend on who pays it?

[1] We discuss this in Chapter 5 "eBay and craigslist".

[2] If you have read Chapter 5 "eBay and craigslist", then this figure should look familiar. A similar figure shows up there for trade between two individuals.

[3] Actually, even this statement carries an implicit assumption that it is possible to share out these goods without distorting economic activity too much. In Chapter 12 "Superstars", we explain that redistribution typically involves some loss in efficiency.

[4] The sanctions began in 1962 in response to the takeover of US property by the Cuban government of Fidel Castro. The Cuban Democracy Act of 1992 extended the sanctions. US Department of State, "Title XVII—Cuban Democracy Act of 1992," accessed March 14, 2011, http://www.state.gov/www/regions/wha/cuba/democ_act_1992.html.

[5] An embargo on trade with Iran was imposed by the United States in 1987. Details can be found at US Department of the Treasury, Resource Center, "Iran Sanctions," accessed March 14, 2011, http://www.treasury.gov/resource-center/sanctions/Programs/Pages/iran.aspx.

11.3 Government and the Labor Market

LEARNING OBJECTIVES

1. What are the forms of government restrictions in labor markets?
2. What are the effects of government restrictions on migration?
3. Who bears the burden of a tax on labor income?

Some of the most important sets of markets in the economy are those for different kinds of labor. There are many ways in which governments intervene in these markets.

Toolkit: Section 17.3 "The Labor Market"

You can review the labor market in the toolkit.

Licensing

Some occupations cannot be carried out without licensing or accreditation of some kind. You cannot set up in business as a doctor or a lawyer without any training. Here, the government's reason for intervening is because of information problems: we do not have the knowledge to determine if someone is indeed trained in medicine or law. [1]

Migration Restrictions

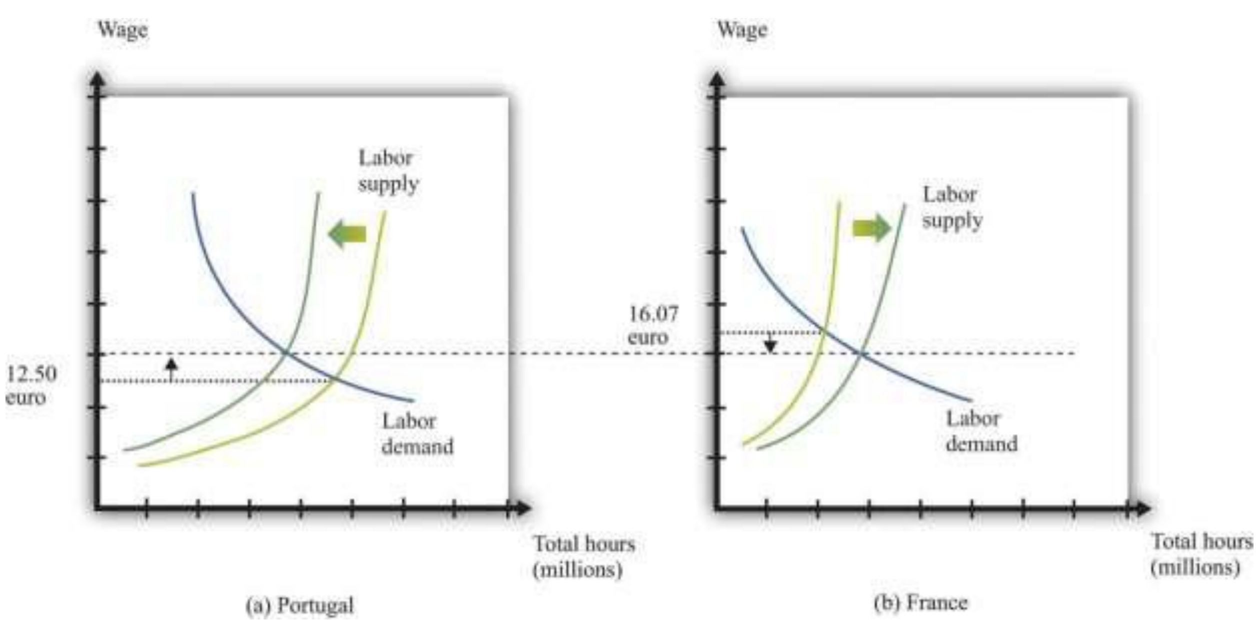
Other things equal, people want to move to where they can earn a high wage. Within the United States, people are free to move from state to state in search of good jobs and good wages. Workers are likewise free to move among the countries of the European Union. In both places, we see many examples of people moving to where wages are higher. Young Polish students move to the United Kingdom in search of work; workers in Louisiana move to Washington state because wages are higher there. Obviously, many factors influence where people choose to live and work, but wages are one of the most important.

If a firm is willing to pay a worker $15 per hour in New Jersey but firms in Idaho will pay that same worker only $12 per hour, then this is an indication that the worker's time is more valuable in New Jersey than it is in Idaho. The market, through the higher wage, sends a signal to the worker that it is desirable to move. The movement of workers from Idaho to New Jersey will cause the supply curve of labor in Idaho to shift to the left, so wages in Idaho will increase, and the supply curve of workers in New Jersey to shift to the right, so wages in New Jersey will decrease. The movement of workers thus also serves to make wages more equal.

Workers in the United States are permitted to move anywhere in the country. The same is true for workers in New Zealand, Mexico, and most other countries. In some places, however, laws enacted by national or local governments make such migration harder. In China, certain government benefits are highly localized, making it difficult for a worker to move from one town to another.

In a world with no restrictions on labor movement, workers would move across countries as they do within a country. Consider the market for labor within the European Union. Figure 11.16 "Migration Eliminates Wage Differences" shows the markets for workers in Portugal and France. If labor is unable to migrate, then the equilibrium wage in France is higher than the wage in Portugal. Once labor mobility is allowed within Europe, workers naturally move to the labor market with the higher wage. This forces wages to decrease in France and increase in Portugal.

Figure 11.16 Migration Eliminates Wage Differences



Workers move from Portugal to France in search of higher wages.

If workers care only about wages, then migration would completely equalize wages in France and Portugal. In practice, some differences in wages persist. For example, if most people think living in Portugal is better than living in France, then the wage rate in Portugal will be lower than that in France. Despite this wage differential, individuals living in Portugal will not move to France. The higher wage is France is an example of a compensating wage differential: it is the difference in wages needed to compensate individuals for living and working in France.

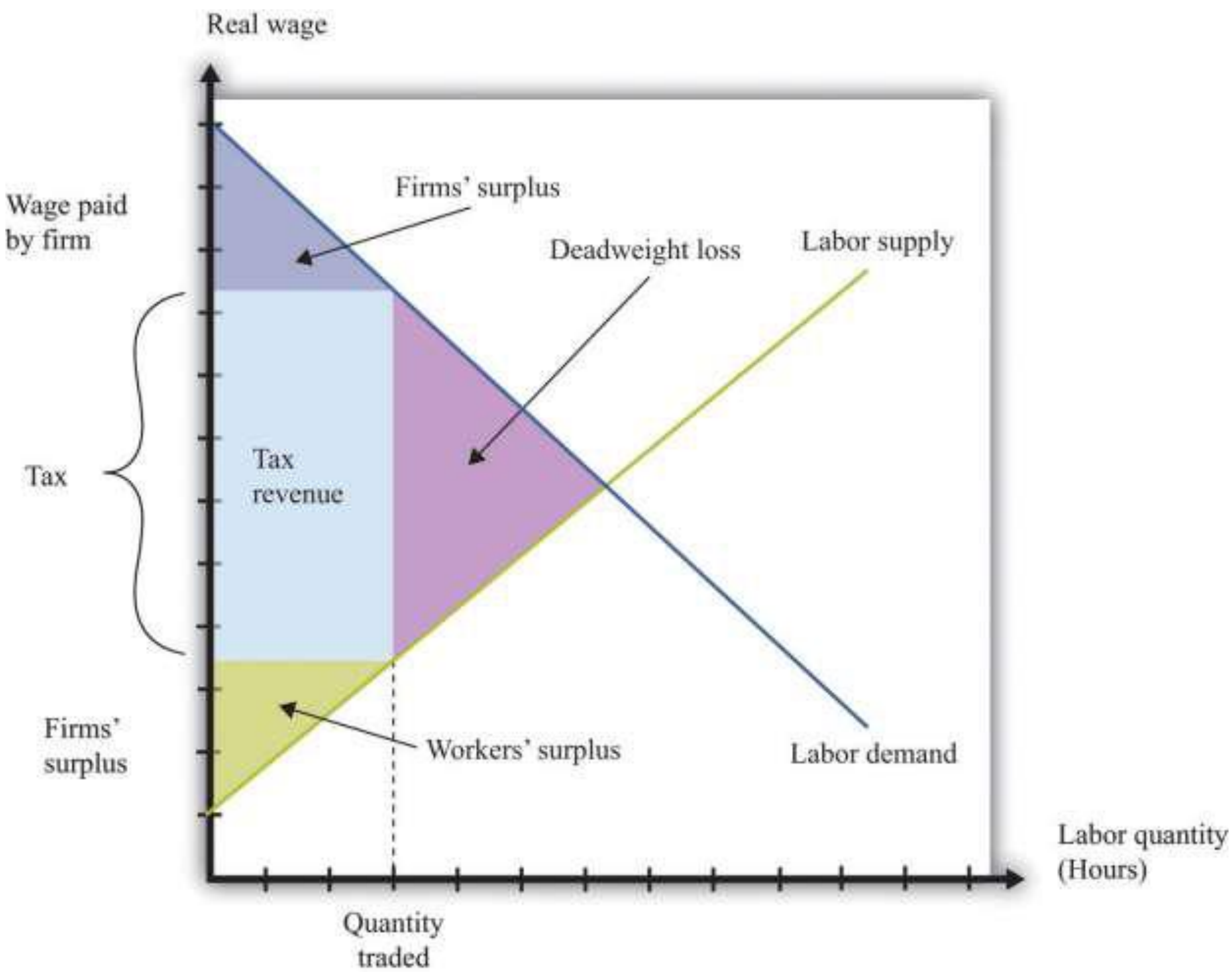
Free migration across countries exists in the European Union, but international migration is typically much more restricted. One of the items on our shopping list at the beginning of the chapter was the hiring of an illegal domestic worker. This is yet one more example of a restriction on trade because people are not allowed to work wherever they want. Most countries restrict the amount of immigration permitted into the country; some countries restrict emigration as well.

Income Taxes

Governments also affect the labor market through the imposition of taxes. In most countries, there is an income tax. In some cases, income taxes may also be imposed more locally: some individual states within the United States have an income tax in addition to the federal tax. An income tax works like the taxes we saw earlier. Fundamentally, it means that the amount paid by the employer exceeds the amount received by the worker. Exactly as before, this gives rise to a deadweight loss. Some workers will choose not to work or work fewer hours because of the tax. It follows that some mutually beneficial transactions go unrealized.

Because the market for labor is so fundamental to the economy and because the income tax is, in most economies, a major source of revenue for the government, economists and politicians pay a lot of attention to this market. Figure 11.17 "The Effect of an Income Tax in the Labor Market" summarizes the effects of a tax on wages using a diagram of the labor market. When there is a tax on wages, there is a gap between the wage paid by the firm and the wage received by the worker. As shown in the figure, the effect of the tax is to reduce the quantity of labor traded. The wage paid by the firm is higher than the wage in the original equilibrium, and the wage received by the worker is less than the wage in the original equilibrium.

Figure 11.17 *The Effect of an Income Tax in the Labor Market*



An income tax means that there is a wedge between the wage paid by the firm and the wage received by a worker.

As in our earlier example, the incidence of the tax will depend on the elasticity of labor supply and labor demand. We can understand incidence by looking at how the tax on labor income affects the wage. If the wage paid by the firm increases by the amount of the tax, then the firm is paying the tax. This will happen if the demand for labor is very inelastic. If the wage received by the worker decreases by the amount of the tax, then the worker is paying the tax, not the firm. This will happen if the supply of labor is very inelastic.

Tax Evasion in the Underground Economy

In some countries, governments have difficulty collecting income taxes from their citizens. This is partly an enforcement issue: if many people in a country misrepresent their income, it is difficult to hire enough people to enforce the tax laws. In this case, income taxes become ineffective, and governments resort to other forms of taxation, such as sales taxes.

Another form of tax evasion is to conduct trades in the underground economy. When income taxes are very high, small business owners and other individuals may offer to do work "under the table." They will ask to be paid in cash, so there is no record of the transaction and no basis for collecting income tax. In return, they will do the work for a cheaper price. This is illegal, but the likelihood of getting caught is low enough that many people decide that avoiding the income tax is worth the crime. The magnitude of this underground activity can be substantial: "In a report to the Senate in May, Deputy Finance Minister Vincenzo Visco said that the hidden, untaxed economy accounted for around 27 percent of Italy's gross domestic product of nearly $2 trillion." [2]

Thus there are two different aspects of the underground economy. There is the exchange of goods and services that cannot be traded legally (drugs, scalped tickets, etc.). And there are trades that are legal but not reported to the tax authorities (illegal).

The underground economy tends to be larger when income taxes are higher and where tax enforcement is difficult, but it exists everywhere. If your neighbor pays you $20 to mow his lawn, and you do not declare this on your taxes, you are participating in the underground economy. Besides allowing you to avoid income taxes, working in the underground economy has an additional benefit. If you do not work a regular job, then you can collect unemployment insurance. This means that you can work and earn income without paying taxes in the underground economy and also collect unemployment insurance.

A recent study by the International Monetary Fund (IMF) concluded, "In the European Union in the late 1990s, 20 million people engaged in shadow [underground] economy activities. In all European OECD countries combined, about 35 million people did so. In some individual countries, the shadow economy labor force was very large: in Italy, 30–48 percent of the total labor force (1997); Spain, 12–32 percent (1997–98); and Sweden, 20 percent (1997–98). In many countries, these high shares coexisted with high official rates of unemployment." [3] According to this study, the underground economy is between 35 percent and 44 percent of gross domestic product (GDP) in developing countries and around 15 percent in the advanced Organisation for Economic Co-operation and Development (OECD) countries.

Finally, people sometimes barter goods and services rather than trade them. If you are a web designer and your next-door neighbor is a plumber, you might agree to build a website for her in exchange for her installing a new shower for you. Again, if you fail to report this "income in kind," you are evading taxes. Bartering schemes can be very sophisticated, involving the creation of groups that set up their own special money to pay for transactions.

KEY TAKEAWAYS

- Governments often limit the movement of people across borders and tax labor income.
- Restricting the flow of labor across international borders can lead to an inefficient allocation of labor across its productive uses.
- The burden of a labor income tax, like other taxes, depends on the elasticities of supply and demand.

CHECKING YOUR UNDERSTANDING

1. If the government required licenses to sell fruit in outdoor markets, who would benefit from this restriction and who would lose?
2. If there are no migration restrictions between two states in the United States, must wages be equal in the two states?
3. Use a version of Figure 11.17 "The Effect of an Income Tax in the Labor Market" to show that if labor supply is very inelastic, then a worker is bearing most of the burden of the income tax.

[1] Chapter 15 "A Healthy Economy" has more to say about this.

[2] Elisabetta Povoledo, "Italy Changes Rules of Tax-Evasion Game," *New York Times*, June 13, 2007, accessed January 29, 2011,http://www.nytimes.com/2007/06/13/world/europe/13iht-taxes.4.6129007.html?_r=1.

[3] Friedrich Schneider and Dominik Enste, "Hiding in the Shadows: The Growth of the Underground Economy," *International Monetary Fund*, March 2002, accessed January 29, 2011, http://www.imf.org/external/pubs/ft/issues/issues30.

11.4 End-of-Chapter Material

In Conclusion

The underground economy is not new; it has been around for as long as rulers have been levying taxes and banning trades. If you read about the prohibition of alcohol in the United States, for example, you will quickly learn that there was still a thriving market for alcohol and alcoholic beverages, despite the illegality of these trades. This was partly due to the fact that the production of alcohol was legal in nearby countries, such as Canada. Alcohol produced in Canada and elsewhere was imported and sold in the United States.

The establishments that served alcohol at that time were called speakeasies. Today you can find local bars that advertise themselves as having started as speakeasies during the Prohibition years. Of course, while Prohibition was in force, the speakeasies did not advertise so loudly. They were generally run by gangs that were willing to take the risk of being arrested to get the profits from selling alcohol.

Associated with Prohibition are several infamous individuals, such as Al Capone and his competitor, Bugs Moran. They were leaders of gangs in Chicago that provided alcohol to speakeasies. But you can, if you like, think of them as managers of firms that were involved in the importation, manufacturing, production, and sale of a consumer good. In many ways these firms operated according to the same principles as firms in this textbook. They were interested in producing efficiently and maximizing their profits.

Capone was eventually indicted and convicted. But the legal action against Capone was not directed at his violation of Prohibition. Instead, the federal government indicted him for tax evasion. Even if you are a leading producer in the underground economy, you still have to pay your taxes.

This story of Prohibition reminds us that the government does more than simply restrict trades in the economy. The government also provides the framework that allows trades. It provides a system of laws that allows people to enter into contracts, and it provides courts as a mechanism for enforcing these contracts.

Capone and Moran could not turn to the government to enforce their contracts and agreements. The firms in the industry had to create their own mechanisms for settling disputes. You won't be surprised to hear that these mechanisms were not pretty. One famous incident was the Saint Valentine's Day Massacre in 1929 when the Capone gang engaged the rival gang led by Bugs Moran. This was like a strategic interaction between rival producers. In this case, their respective competitive strategies left seven people dead. When the government is not there to enforce contracts, agreements will be enforced by other, often violent, means.

Key Links

- Food and Drug Administration: http://www.fda.gov
- US Citizenship and Immigration Services:http://www.uscis.gov/portal/site/uscis
- Jeffrey Miron's blog: http://jeffreymiron.com/2010/04/drug-prohibition-and-immigration

EXERCISES

1. List three additional examples of government restrictions on your ability to buy or sell something.
2. The sharing of the burden of a tax also depends on the elasticity of supply. Draw diagrams like Figure 11.7 "The Deadweight Loss from a Tax" looking at the case of elastic and inelastic supply—that is, draw two diagrams with identical demand curves but different supply curves. How does the elasticity of supply affect the changes in the buyer surplus and the seller surplus? Can you explain why?
3. (Advanced) In some countries, there are restrictions on the length of a contract to rent an apartment. Suppose the restriction is that contracts must last for five years. In response, some people sign private agreements to rent for shorter durations, such as a year. What are the problems that might arise from signing these private agreements? What happens if there is a dispute? What role might reputations play in the enforcement of these private agreements?
4. Suppose there is a forecast that a hurricane will hit in a day. Everyone expects the government to ration the supply of coffee. What will likely happen to the price of coffee once the forecast is announced?
5. How does rent control affect the incentives for an owner to invest in upgrades of an apartment?

6. The payout from a social security system depends on years worked. How would an increase in social security payments affect the choice of workers between jobs in the formal and informal (underground) parts of the economy?
7. If the underground part of an economy is large due to tax evasion, could a tax cut increase tax revenue?
8. If two states have different rates of labor taxation, what can you say about wages before and after taxes in the two states?
9. Can you think of a good or a service that the government *requires* you to consume? Why do you think the government has this policy?
10. There are substantial differences in food and product safety standards across countries. Can you think of reasons why this might be the case?
11. Liquor sales are state controlled in Pennsylvania but not in New Jersey. What effects do you think this has on the buying and selling of liquor near the border of the two states?
12. Suppose that Principles of Economics is a very popular course at your school. More people want to take the course than there are seats available. Do you think it would be a good idea if those initially enrolled in the class were able to sell their seats to those who didn't get a spot? What would be the advantages of such a system? What would be the problems?
13. Sometimes armies are raised by a draft, while in other times armies are volunteer. Which way of raising an army do you think is most efficient in terms of getting the best people to participate in the army? Which way of raising an army is most "fair"?
14. (Advanced) When Question 13 talks about "the best people to participate in the army," does it make a difference whether we are talking about comparative advantage or absolute advantage?
15. Explain how the incidence of a new tax on textbooks, collected at the point of sale, will be determined.
16. Which type of trade barrier creates more revenue for the government—a tariff or a quota? Why would a government ever impose a quota?
17. (Advanced) One benefit of working in the formal labor market in some developing countries is eligibility for both unemployment insurance and retirement pensions. All else being the same, would you predict that wages are higher in the formal or the informal sector of the economy? In addition, workers in the informal sector do not pay income taxes. What is the effect of this on wage levels in the two sectors?

18. Due to mobility restrictions, the labor markets in China are not fully integrated. If restrictions on mobility of workers in China were relaxed, what would happen to wage differences across regions? What predictions would you have for the flow of workers across parts of China?
19. If you are a member of a professional union, would you be in favor of licensing requirements to join that profession? How might you defend the need to have a license?

Economics Detective

1. Suppose you live in Mexico. If you wanted to get a job in Canada, what would you have to do to obtain permission to work? What if, instead, you wanted to work in the United States? Does your answer depend on your occupation?
2. Try to find estimates of the size of the underground economy in two different countries (for example, Portugal and Sweden). Is the underground economy of very different sizes in the two countries? Why?

Chapter 12
Superstars

Rich and Richer

Table 12.1 "Wealthiest Individuals in the United States" shows the top 10 wealthiest people in the United States in 2006 and 2010. These names come from lists compiled each year by *Forbes* magazine of the 400 wealthiest individuals. [1] You almost certainly recognize some of the names, such as Bill Gates and Michael Dell from your dealings with the computer industry. Other names may be less familiar to you.

Whether or not you know their names, you surely have difficulty conceiving of their wealth. Bill Gates's net wealth in 2010 was estimated at $54 billion, which is $9 billion more than the wealth of financier Warren Buffett. To give some idea of what this means, if Gates were to receive no further income for the rest of his life but wanted to use up all his wealth before he died, he would need to spend it at a rate of about $5 million a day. The person at the bottom of the *Forbes* list—that is, the 400th wealthiest person in the United States—had a net worth of a mere $1 billion.

Comparing the two lists, you can see that some of the names and rankings changed between 2006 and 2010. The top two names are the same in both years, but the rest of the list is different. Sheldon Adelson, Paul Allen, and Michael Dell were in the top 10 in 2006 but not in 2010. In 2010, Charles and David Koch joined the top 10. Even among the very rich, there is some instability within the distribution of wealth.

The *Forbes* list was of the wealthiest Americans. Only the top 3 from the 2010 list are on the list of the world's wealthiest individuals. In 2010, the wealthiest individual in the world was Carlos Slim Helu, a Mexican businessman who made his fortune from real estate speculation and the telecom industry. Others in the world top 10 come from India, France, Brazil, Spain, and Germany. *Forbes* also publishes many other lists, including a list of the most powerful celebrities. At the top of that list in 2010 was Oprah Winfrey, who earned $315 million. (Notice that this is her *income*—the amount she earned in the year—while Table 12.1 "Wealthiest Individuals in the United States" is based on the total *wealth*accumulated.) Also on the list were Beyonce Knowles, Lady Gaga, Tiger Woods, Johnny Depp, and others from the entertainment industry.

When *Forbes* published its 2007 list, it also published an article by economist Jeffrey Sachs discussing the other extreme of the wealth distribution: the world's poorest households. Sachs pointed out that there are about a billion households in the world living on about $1 a day. He calls this group the *Forbes* One Billion. Sachs calculates that the richest 946 households have the same earnings as the*Forbes* One Billion. The discussion in *Forbes* and the calculations by Sachs make it clear that there are immense differences in income and wealth across people in the world. This is true both if we look across countries, comparing the richest to the poorest nations, and if we look within countries.

These differences are *persistent*, meaning that an individual's place in the income or wealth distribution is not likely to change significantly from one year to the next. If you are poor this year, you will probably be poor next year. It is not impossible for people to become rich overnight, but it does not happen often. In fact, such differences persist not only from year to year but also from generation to generation. This doesn't mean that everyone is completely stuck in the same place in the economic hierarchy. There are opportunities for children to become much richer—or much poorer—than their parents. But when we look at the data, we will see that the income level of parents is an important indicator of the likely income of their children.

One goal of this chapter is to document some facts of inequality. This is not a straightforward task. For one thing, it is not even clear what measure of a household's economic success we should look at. Is it more useful to look at inequalities in income, wealth, consumption, or some other variable altogether? We also get a different picture if we look at these differences at a point in time or across time.

Data on inequality matter for discussions about taxation and redistribution. Governments throughout the world levy a number of different taxes, including taxes on the income people earn and the purchases that they make. Some of the revenues from these taxes are transferred to poorer households in the economy. The taxation of some households and the transfer of the resulting revenue to other households make up the redistribution policies of the government. We are interested in documenting facts about inequality in large part because we need these facts to have a sensible discussion about how much redistribution we—as a society—would like.

In this chapter, we therefore consider the following questions.

What determines the distributions of income, wealth, and consumption?

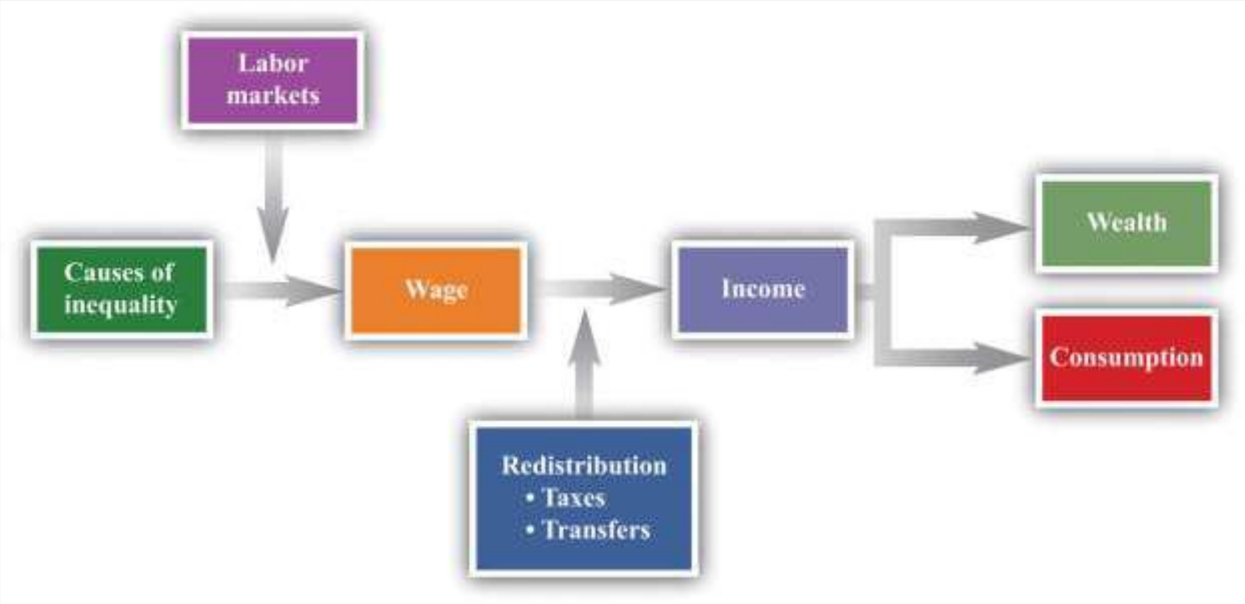
Is the market outcome "fair" or is there a need for government intervention?

What are the consequences of government redistributions of income and wealth?

Road Map

A road map for this chapter is shown in Figure 12.1 "Road Map". We begin with some facts about inequality and introduce some techniques to help us describe the amount of inequality both in a country and across countries. Then we consider some explanations of why we observe inequality in society. We observe first that people have different abilities, which translate into differences in income. Then we consider how individual choices—about education, training, and effort—are a further source of difference.

Figure 12.1Road Map



This figure shows a plan for this chapter. We investigate the different underlying causes of inequality and explain how these translate, through labor markets in the economy, into differences in wages. We then explain how government policies affect the distribution of income in the economy. We also look at what determines the distribution of income, consumption, and wealth.

We then turn to a more abstract discussion of some different philosophical views of inequality. These different views influence current thinking about the distributions of income, wealth, and consumption and help us understand why people have such different opinions about equality and redistribution. We consider how redistribution might affect people’s incentives to work, study, and cheat. Finally, we turn to economic policies that affect inequality.

[1] Forbes has many such lists available for your study (http://www.forbes.com/lists).

12.1 Facts about Inequality

LEARNING OBJECTIVES

1. What is a Lorenz curve?
2. What is a Gini coefficient?
3. What has happened to income distribution in the United States?

There is no single, simple measure of the amount of inequality in a society. For example, we could study the distribution of consumption, income, or wealth, but each will tell us something different about the amount of inequality in our economy. These differences matter for the debate about inequality and our evaluation of policy.

The Lorenz Curve and the Gini Coefficient

Suppose you want to document the distribution of income in an economy. You could begin by asking every household its level of income. In many countries, the government already collects such data. In the United States, for example, this investigation is carried out by the US Census Bureau (http://www.census.gov). If everyone on the list had exactly the same level of income, you would conclude that income was equally distributed. If all but one person on the list had zero income and the remaining person had all the income, then you would conclude that income was very unequally distributed. In reality, of course, you would find that different households have all sorts of different levels of income.

The Lorenz curve provides a useful way of summarizing the distribution. It plots the fraction of the population on the horizontal axis and the percentage of income received by that fraction on the vertical axis. We construct a Lorenz curve as follows.

1. Take the list of incomes and order them from the lowest to the highest.
2. Calculate the total income in the economy.
3. Calculate the income of the lowest 1 percent of the population. Then calculate the income of the lowest 1 percent of the population as a percentage of total income.
4. Calculate the income of the lowest 2 percent of the population. Then calculate the income of the lowest 2 percent of the population as a percentage of total income.
5. Continue for all income levels.

6. Plot these points on a graph with fraction of the population on the horizontal axis and fraction of income on the vertical axis.

We know that 0 percent of the population earns 0 percent of the income, so the Lorenz curve starts at the origin. We also know that 100 percent of the population earns 100 percent of the income, so the other end of the Lorenz curve is at that point. If income were exactly equally distributed, then any given fraction of the population would earn that same fraction of income. The lowest 28 percent of the population would earn 28 percent of the income, the lowest 74 percent of the population would earn 74 percent of the income, and so on. In this case, the Lorenz curve would be a 45-degree line connecting the two endpoints. The closer the Lorenz curve to the 45-degree line, the more equal the distribution of income.

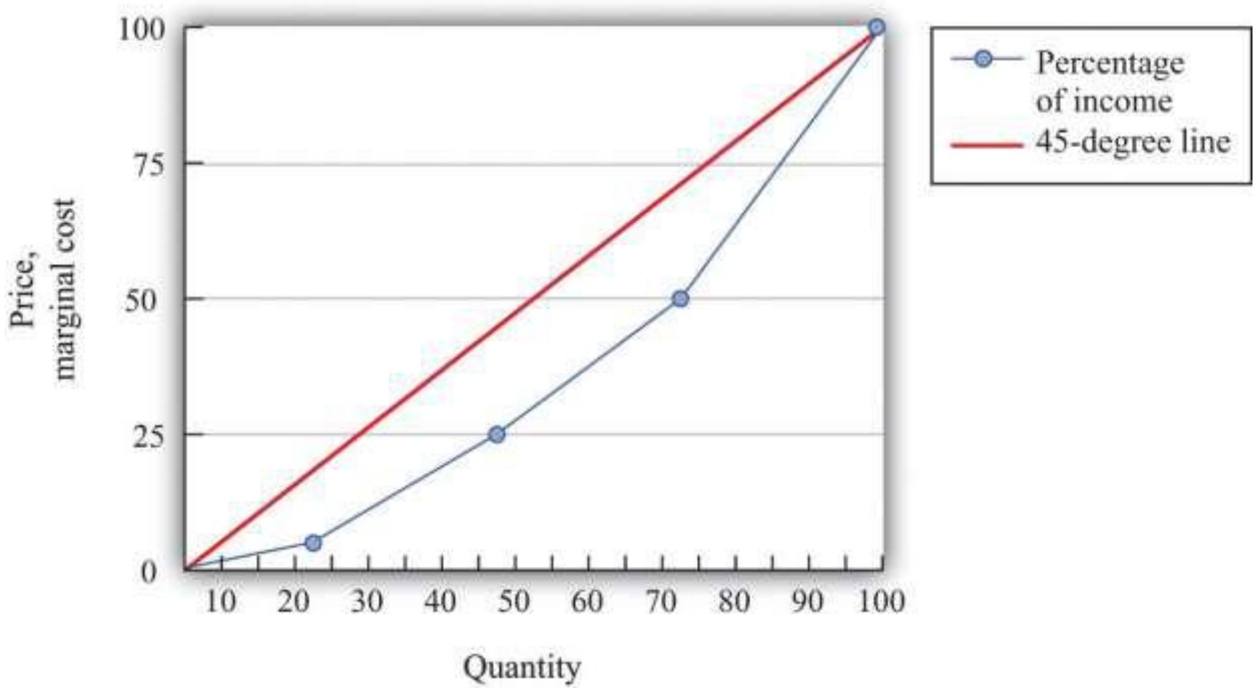
Table 12.2 "Example of Income Distribution" illustrates how to calculate the points on a Lorenz curve. The table shows four households, ordered by their income levels. The total income earned is $2,000. The lowest household (25 percent of the population) earns 5 percent of the total income because $100$2,000 = 5 percent. If there were complete equality, this number would be 25 percent. So the lowest income household accounts for one quarter of the population but only one twentieth of the income. The first and second households together account for 50 percent of the population (see the last column of the table). They earn $500 in total, which is 25 percent of the total income. The first, second, and third households account for 75 percent of the population and 50 percent of the total income. Finally, if we look at all four households (100 percent of the population), this group earns $2,000, which is, of course, 100 percent of the total income. This Lorenz curve is illustrated in Figure 12.2 "The Lorenz Curve".

Table 12.2 Example of Income Distribution

Household	**Income Level ($)**	**Percent of Total Income Earned by Household**	**Percent of Total Income Earned by All Households with This Income or Lower**	**Percentage of Population with This Income or Lower**
1	100	5	5	25
2	400	20	25	50
3	500	25	50	75

Household	Income Level ($)	Percent of Total Income Earned by Household	Percent of Total Income Earned by All Households with This Income or Lower	Percentage of Population with This Income or Lower
4	1,000	50	100	100

Figure 12.2 The Lorenz Curve



The more equal the distribution, the closer is the Lorenz curve to the 45-degree line.

We explained that the Lorenz curve coincides with the 45-degree line if there is complete equality. There is also a Lorenz curve for the case of complete inequality—in which a single person earns all the income. In this case, the Lorenz curve lies along the horizontal axis until the final household (that is, at 100 percent on the horizontal axis). At that point, the Lorenz curve lies along the vertical line at the right of the figure because the last person has all the income. Real economies exhibit neither complete equality nor complete inequality; a typical Lorenz curve lies below the 45-degree line and above the horizontal axis.

If we want to compare inequality over time or across countries, then we need something even simpler than the Lorenz curve. For this, we use the Gini coefficient, which is equal to the area between the 45-

degree line and the Lorenz curve divided by the area below the diagonal. Figure 12.3 "The Lorenz Curve and the Gini Coefficient" shows how the Gini coefficient is related to the Lorenz curve.

If the Lorenz curve is exactly the same as the 45-degree line, then the Gini coefficient is zero. In this case, there is no area between the Lorenz curve and the 45-degree line. At the other extreme, if the Lorenz curve coincides with the horizontal axis until the final household, then the area above the Lorenz curve and the area below the diagonal are exactly the same. With complete inequality, the Gini coefficient is one. A higher Gini coefficient therefore means more inequality in the distribution of income.

Data on Inequality

We now use the Gini coefficient and other data to look at some facts about the distributions of income and wealth.

The Distribution of Income

Table 12.3 "Household Income by Quintile" presents data from the US Census Bureau on the distribution of various measures of income from 2003 to 2005. There are three measures of income given for each of the three years:

1. **Market income.** A measure of income earned from market activity, such as labor income and rental income.
2. **Postinsurance income.** Market income plus transfers received from the government.
3. **Disposable income.** Market income less taxes paid to the government plus transfers received from the government.

These measures of income for each of the three years create the columns of the table. The rows of the table are quintiles (fifths) of the population. As in the construction of the Lorenz curve, the population is ordered according to income. This means the first quintile is the bottom 20 percent of the population in terms of income. The fifth quintile is the top 20 percent of the population in terms of income. To see how these quintiles are created, imagine taking 100 people and arranging them by their income, starting at the lowest level. Then create five groups of 20 people each where the first 20 people in the income distribution are in the first group, the second 20 in the income distribution are in the second group, and so on. Each group of 20 is a quintile of this population.

For each measure of income and for each year, there is an entry in the table showing the fraction of income in that year for a particular quintile. For example, looking at disposable income in 2004, the third (middle) quintile had 16.1 percent of the disposable income, and the highest quintile had 44.9 percent.

There are two striking features of this table. First, there is substantial inequality in the US economy. Looking at market income, the lowest 20 percent of the population receive about only 1.5 percent of the total market income. Contrast this with the highest quintile, which receives more than 50 percent of the total market income. This inequality is reflected in the Gini coefficient of about 0.49. If we look at the very top of the income distribution, the inequality is even more marked: the top 5 percent of the population in 2005 received about 30 percent of income after taxes and transfers, and the top 1 percent received about 16 percent of income. [1]

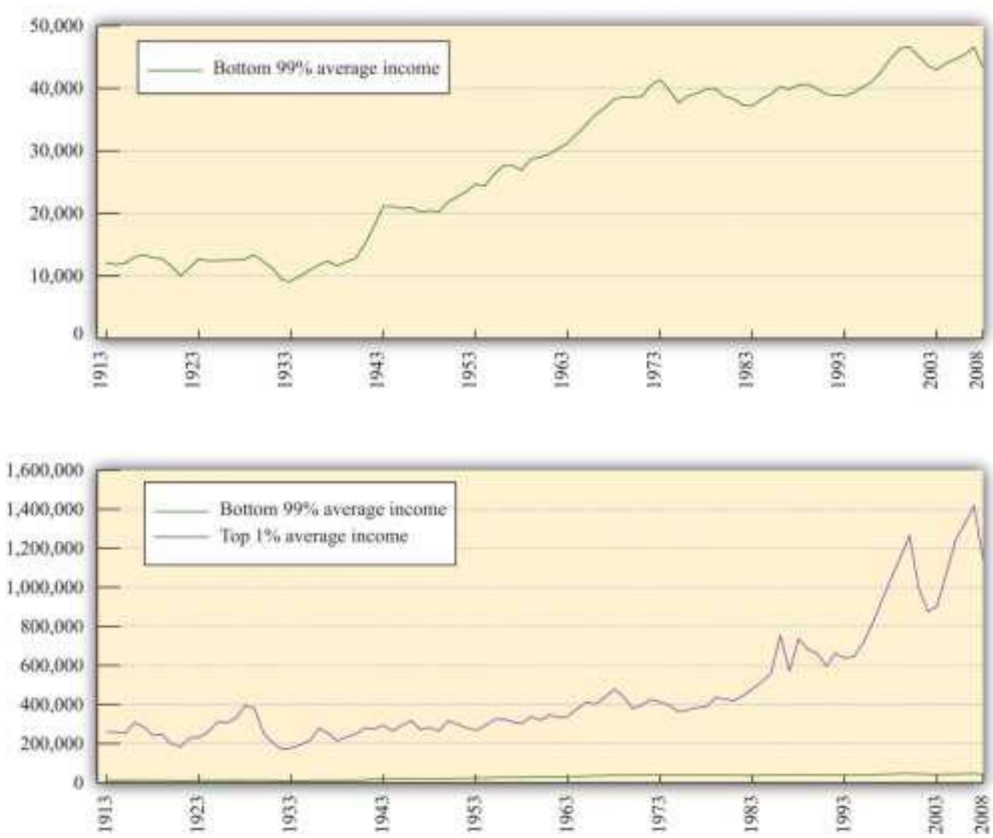
Second, the Gini coefficient decreases if we look at postinsurance income relative to market income and at disposable income relative to postinsurance income. This is because transfers represent—on average—a flow from richer to poorer households, and taxes are *progressive*: they redistribute from the rich to the poor. Government policies bring about some redistribution from richer households to poorer households. That said, there is still substantial inequality even after this redistribution: the lowest quintile receives less than 5 percent of total income, while the highest quintile receives about 45 percent.

Table 12.4 "Gini Coefficient over Time" shows changes in the Gini coefficient over time. (The data are on household incomes and come from the Census Bureau. [2]) This table shows that inequality in the United States, as measured by the Gini coefficient, has increased steadily over the last few decades. In fact, if you go back to the end of World War II, the end of the 1960s represents a turning point in the income distribution. [3] From 1940 through the 1960s, the income share of the top 10 percent fell from about 45 percent to about 33 percent. But starting in the 1970s, the pattern reversed, so that by 2007, the share of the top 10 percent exceeded 45 percent of total income.

Figure 12.4 "The Distribution of Income from 1913 to 2008" focuses on the top of the income distribution: the top 1 percent. In part (a) of Figure 12.4 "The Distribution of Income from 1913 to 2008", we can see that the real income of the bottom 99 percent of the population increased dramatically between the 1930s and the 1970s, increasing from $9,000 in 1933 to over $40,000 in 1973. (These numbers are adjusted for inflation and are in 2008 dollars.) Income over this period, for this group, grew an average of 3.7 percent per year. Over the next 35 years, the real income of this group hardly grew at all: the average growth rate was 0.2 percent per year. By contrast, the income of the top 1 percent grew only 1.7 percent per year on average between 1913 and 1973 but grew at an average 2.8 percent from 1973 to 2008. As a consequence, the top 1 percent of the income distribution roughly doubled their share of total income over this period.

At the very top of the income distribution, we have the true superstars: rock stars, movie stars, sports stars, top CEOs, and so on. The top 0.01 percent of the population—that is, the richest 30,000 or so people—has seen their share of income increase sevenfold since 1973.

Figure 12.4 The Distribution of Income from 1913 to 2008



(a) The average real income in 2008 dollars for the bottom 99 percent of the population rose substantially between the 1930s and the 1970s but has been much flatter over the past few decades. (b) The top 1 percent has seen substantial income growth in recent decades.

The Distribution of Wealth

Table 12.5 "Gini Coefficients for Net Worth" looks at wealth data for a cohort of individuals between 1989 and 2001. At the beginning of the study, this group was between 34 and 43 years old. [4] We can see that the Gini coefficients for wealth are considerably larger than the ones we saw earlier for income. There is more equality in income than in wealth.

Income is a *flow*, meaning that individuals receive labor income on a weekly or monthly basis. Wealth is a *stock*: it is a measure of the assets that an individual or a household has accumulated and is measured at a particular point in time. Wealth comes partly from what people inherit and partly from decisions they make about allocating income between consumption and saving. The table also shows that wealth inequality increased for this group. There are two reasons that this could happen: (1) it may reflect greater inequality as a whole in society and (2) it may be due to inequality increasing as people become older.

Dynamics of Inequality

The position of a household in the income distribution is not static. A household in the lowest quintile of income one year will not necessarily be there the following year. Households can move up and down in the income distribution. For example, suppose you are fortunate enough to win the lottery or publish a hit song. Your income and thus your position in the income distribution will change quickly. For others without a hit song or luck with the lottery, changes in income can take more time. Perhaps you invest in a college education; after graduation and with a new job, you begin a climb through the income distribution. Bad luck can send you in the opposite direction. If your skills become less valuable, perhaps because of changes in technology, you may find that you have to move from a higher-paying to a lower-paying job, or you may become unemployed. There are many routes from rags to riches and from riches to rags.

One reason for mobility is the changes in income that most people experience in their lifetimes. For most people, the income they earn in their first job after school pays a lot less than the job they retire from. Thus most individuals experience a profile of income over their lifetime that takes them from one part of the income distribution to another. For most people, income also decreases in retirement.

Table 12.6 "Dynamics of Income in the United States" illustrates these dynamics over a five-year period. The top part of the table refers to earnings and the lower part to wealth. The data come from looking at distributions of earnings and wealth in two years: 1989 and 1994.

Under "Earnings," there are five rows indicating the quintiles of the distribution in 1989. Along the top, there are five columns indicating the quintiles of the distribution in 1994. The entries refer to the percentage of people who go from one quintile in 1989 to another quintile in 1994. For example, 27 percent of the households in the second highest quintile in 1989 were in the top quintile in 1994, while 34 percent of the households in the second highest quintile in 1989 stayed there. A similar interpretation is given for the wealth part of the table.

The two parts of this table give a sense of income and wealth mobility through the distribution. If there were no mobility over time, so that households stayed in the same income and/or wealth quintiles), then the table would have 100 on the diagonal and 0 everywhere else. Mobility is indicated by the fact that the numbers along the diagonal are less than 100. From the part of the table referring to earnings, 90 percent of the people in the top income group in 1989 were there in 1994 as well. This means that very high income is extremely persistent. In contrast, only about two-thirds of the people in the lowest income class in 1989 remained in that group in 1994, while 17 percent moved up one quintile. As time passes, those who moved up will then move on to other parts of the income distribution.

Table 12.6 "Dynamics of Income in the United States" shows income and wealth dynamics over a relatively short period of time. It is also useful to look at dynamics across generations, though data are more difficult to obtain. One approach that researchers use over longer periods of time is to follow families. If your family was in the middle income group, we can see the likelihood that you will be in that same income group or in another income group. These dynamics take a longer amount of time because they are affected by things like parents' choices about the education of their children.

One way to study intergenerational income mobility is to take a group of individuals at a point in time and see how much of their current income can be "explained" by the income of their parents. (*Explained* is in quotation marks because it is difficult to disentangle the effects of family income from other influences. There are many factors associated with parents' income, such as the quality of schools and schoolmates, which are correlated with family income.)

One study reports an elasticity of 0.5 on the relationship between family and child income. This means that if parents' income is 1 percent higher, the child's income will be higher by about 0.5 percent. So if two families have an income difference of $100,000, then the prediction is that their children will have a difference of $50,000. [5] This number is higher for the United States than for almost all the other (mostly European) countries studied. This same elasticity in Denmark is only 0.15, for example.

Toolkit: Section 17.2 "Elasticity"

You can review the concept of elasticity in the toolkit.

The same study also looked at the mobility of families across the quintiles of income. A child whose family was in the middle quintile income had about a 40 percent chance of moving down the income distribution to a lower quintile and a 36.5 percent change of moving up. But 47 percent of the children born to a family in the lowest quintile remained there.

Inequality in Other Countries

Table 12.7 "Gini Coefficients in Different Countries" presents some evidence on the distribution of income in different countries. There are some significant differences across countries in income inequality. Eastern European countries, such as Hungary and Albania, and Western European countries, such as Sweden and France, have relatively equal distributions of income. At the other extreme, countries like Namibia and Brazil are highly unequal. The United States is about in the middle of these distributions.

When we compare countries, remember that some countries have much higher income than others. Looking at Table 12.7 "Gini Coefficients in Different Countries", low-income countries generally seem to have more inequality than high-income countries. This is suggestive of a link between inequality and stages of development. Economist Simon Kuznets suggested that inequality would increase in the early stages of the development process but decrease in later stages. This became known as the *Kuznets hypothesis*. One story was that as a country grows, the labor force is split between a relatively high-income industrial sector and a relatively low-income agricultural sector. As a country grows, more labor is allocated to the more productive manufacturing sector, and thus inequality is reduced over time.

Whatever the mechanism, world inequality appears to be decreasing significantly. A recent study found that the Gini coefficient for the world had declined from about 0.58 in the 1970s to about 0.51 in the late 2000s. [6]

There are also some fascinating differences in the dynamics of inequality. The decline in inequality in the middle of the 20th century was common throughout much of the developed world. The more recent increase in equality that we have documented in the United States is also visible in some other countries, such as Australia, New Zealand, and the United Kingdom. By contrast, most of Western Europe has not seen the same kinds of increases in inequality.

KEY TAKEAWAYS

- The Lorenz curve shows the distribution of income in an economy by plotting the fraction of income on the vertical axis (after households have been ranked by their income) and the fraction of the population on the horizontal axis. The closer the Lorenz curve to the 45-degree line, the more equal the distribution of income.
- The Gini coefficient is a statistic that indicates the degree of inequality by looking at how far the Lorenz curve is from the 45-degree line.
- A given household's position in the distributions of income, wealth, and consumption changes over time. This is partly due to education and work experience and partly due to luck. Another dynamic element of the income distribution comes from transfers across generations of a household.

CHECKING YOUR UNDERSTANDING

1. If you have two countries, what does it imply about the Lorenz curves for the two countries if the Gini coefficient on income is higher in the first country compared to the second?

2. Is it possible for disposable income to be distributed more equally across households in a country than market income? How could this happen?
3. How do taxes influence the distribution of disposable income?

[1] These figures come from Congressional Budget Office, *Historical Effective Federal Tax Rates, 1979 to 2005*, table 4C, accessed March 14, 2011,http://www.cbo.gov/ftpdocs/88xx/doc8885/Appendix_wtoc.pdf; the definitions of income therefore differ slightly from the US Census Bureau numbers in the table.

[2] See http://www.census.gov.

[3] Thomas Piketty and Emmanuel Saez, "Income Inequality in the United States, 1913–98,"*Quarterly Journal of Economics* 118 (2003):1, together with their updated data set available at Emmanuel Saez's faculty home page,http://www.econ.berkeley.edu/~saez/TabFig2008.xls.

[4] Arthur B. Kennickell, *A Rolling Tide: Changes in the Distribution of Wealth in the U.S., 1989–2001* (Washington, DC: Federal Reserve Board, 2003). Wealth is defined as assets minus liabilities.

[5] The estimate is reported in Thom Hertz, *Understanding Mobility in America* (American University, Center for American Progress, April 26, 2006).

[6] See Maxim Pinkovskiy and Xavier Sala-i-Martín, "Parametric Estimations of the World Distribution of Income" (National Bureau of Economic Research Working Paper 15433, October 2009), accessed March 14, 2011, http://www.nber.org/papers/w15433.pdf.

12.2 The Sources of Inequality

LEARNING OBJECTIVES

1. Where do differences in income come from?
2. Why might the marginal product of labor differ across people?
3. What is the skill gap?
4. What is a winner-takes-all market?

We have provided some facts about differences in income across households. We now turn to a discussion of where those differences come from.

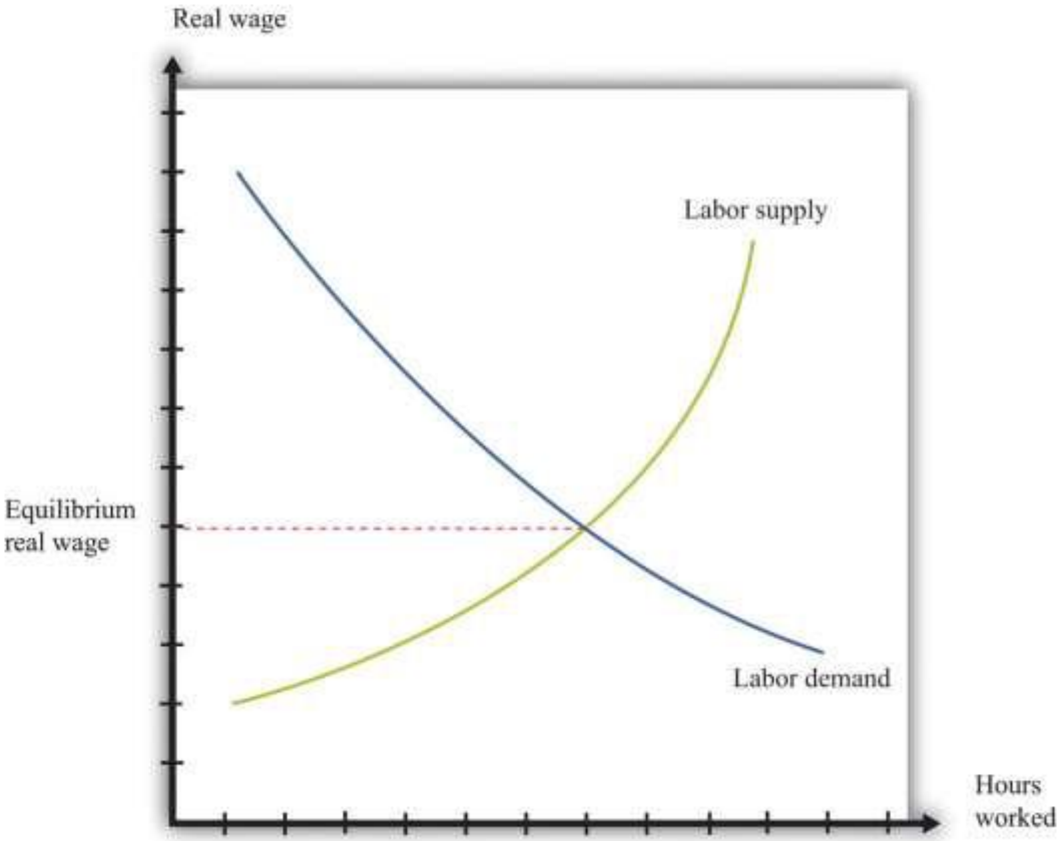
From Ability to Earnings

We begin by looking at earnings, by which we mean the income that households obtain from their work in the labor market. Figure 12.5 "Labor Market Equilibrium" shows the labor market. The real wage is on the vertical axis, and the number of hours worked is on the horizontal axis. The labor demand curve indicates the quantity of labor demanded by firms at a given real wage. As the real wage increases, firms demand less labor. The labor supply curve shows the total amount of labor households want to supply at a given real wage. As the real wage increases, the quantity of labor supplied also increases. [1] Here we are interested in what the labor market can tell us about how much people earn.

Toolkit: Section 17.3 "The Labor Market"

You can find more details about the labor market in the toolkit.

Figure 12.5 Labor Market Equilibrium



When firms are deciding how many hours of work to hire, they use this decision rule: hire until

real wage = marginal product of labor.

The left side of this equation represents the cost of purchasing one more hour of work. The right side of this equation is the benefit to the firm of one more hour of work: the marginal product of labor is the extra output produced by the extra hour of work. If the marginal product is higher than the real wage, a firm can increase its profits by hiring more hours of work.

We use this equation as a starting point for thinking about distribution and inequality. Different individuals in the economy are paid different real wages. This reflects, among other things, the fact that there is not a single labor market in the economy. Rather, there are lots of different markets for different kinds of jobs: accountants, barbers, computer programmers, disc jockeys, and so on. We can imagine a diagram like Figure 12.5 "Labor Market Equilibrium" for each market. In all cases, the firms doing the hiring will want to follow the rule given by the equation. And if firms follow this hiring rule, then two individuals who earn different real wages must differ in terms of their marginal product. The worker who earns the higher wage is also the worker who is more productive.

But why would workers have different marginal products? One reason is that people differ in terms of their *innate abilities*. For any individual, we could come up with a long list of the skills and abilities that he or she is born with—natural talents. Some are good at mathematics, some are particularly strong, some are good at music, some are good at building things, some are very athletic, some are good at managing other people, and so on. Abilities that tend to make someone have a high marginal product allow that person to earn higher real wages. Differences in innate abilities, then, are the first explanation we can suggest for why there are differences in earnings when we look across individuals.

The possession of innate ability is not enough to guarantee someone a high marginal product; the market must value the individual's talents as well. The demand for particular abilities or skills is high if they can be used to produce something that people want to buy. Think about a talented quarterback: his talents translate into an ability to draw paying customers to games, which in turn translates into a willingness to pay a lot for his labor. Or think about a skilled manager: her ability to make good business decisions translates into higher profits for a firm, which in turn translates into a willingness to pay for her labor. If an ability is valued in the market, then there will be high demand for the labor of people with that ability.

What is valuable changes over time and from place to place. Being a skilled quarterback is valued in the modern-day United States. The same innate talent was worth much less 50 years ago in the United States and is still worth little today in a village in the Amazon. Rock stars who can earn hundreds of millions of dollars today would have had very little earning power in 19th-century Australia. The same holds for more mundane skills. The innate abilities that make for a good software designer are more valuable than in the past; the innate abilities that make for a good clockmaker are less valuable than in the past.

Labor supply matters because the value of your innate abilities also depends on how many other people have similar talents. Another reason that highly talented quarterbacks command such high earnings is because their abilities are in short supply. Being a good taxi driver also requires certain skills, but these are much more common. As a result, the supply of taxi drivers is larger, so the real wage earned by taxi drivers is smaller.

Education, Training, and Experience

Star quarterbacks have innate abilities that most of us don't possess. But they also have more training and experience in this role. Just about every one of us could be a better quarterback than we are now, if we were willing to train several hours a day. Indeed, most occupations require some skills and training. Computer programmers must learn programming languages, engineers must learn differential equations, tennis players must learn how to play drop shots, and truck drivers must learn how to reverse an 18-wheeler.

As well as such specific skills, an individual's general level of education is usually an indicator of his or her marginal productivity and hence the wage that can be earned. Basic literacy and numeracy are helpful—if perhaps not absolutely necessary—for nearly any job. A high school education typically makes an individual more productive; a college education even more so. So the distribution of labor income is affected by the distribution of education levels. People also learn on the job. Sometimes this is through formal training programs; sometimes it just comes from accumulating experience. Generally, older and more experienced workers earn higher wages.

Education and experience affect both labor demand and labor supply. More highly skilled workers are typically more valuable to firms, so the demand curve for such workers lies further to the right. At the same time, experienced and trained workers tend to be in more limited supply, so the supply curve lies further to the left. Both effects lead to a higher real wage. Just as a worker's real wage depends on how valuable and scarce are her abilities, so also does it depend on how valuable and scarce are her education and training.

The influence of experience on earnings is a reminder of an observation that we made when discussing the data. Even in a world where everyone is identical in terms of abilities and education, we would expect to see some inequality in earnings and income simply because people are at different stages of life. Younger, inexperienced workers often earn less than older, experienced workers.

The Skill Gap

In recent years, economists have looked closely at the differences in wages among skilled and unskilled workers. Loosely speaking, skilled workers are more educated and in occupations that rely more on thinking than on doing. So for example, an accountant is termed a skilled worker, and a construction worker with only a high-school diploma is an unskilled worker. Data on wages suggest that the return to skill, as measured by the difference in wages between skilled and unskilled workers, has widened dramatically since the mid-1970s. Many economists think that this is an important part of the explanation for the increasing inequality in the United States.

One way to measure the increased return to skills is to look at the financial benefit of education, given that more educated workers are typically skilled rather than unskilled. Table 12.8 "Relationship between Education and Inequality in the United States" summarizes some evidence on the distributions of earnings, income, and wealth from 1998. The table indicates that there is a sizable earnings gap associated with education. According to this sample, completing high school increased earnings by nearly $20,000, and a college degree led to an additional $34,000 in average annual income. Education is an important factor contributing to inequality. One way to decrease inequality is to improve access to education.

Effort

So far we have said nothing about how hard people choose to work, in terms of either the number of hours they put in on the job or their level of effort while working. Those who are willing to work longer hours and put in more effort will typically obtain greater earnings.

Effort is a matter of individual choice. Some other factors that can influence your earnings are likewise under your own control. Training and education are largely a matter of choice: you can choose to go to college or take a job directly out of high school. By contrast, the abilities you are born with are, from your point of view, a matter of luck. We have more to say about this distinction later when we evaluate the fairness of the distribution of income.

The Gender Gap

Study after study indicates that the gender of a worker also influences real wages. Figure 12.6 "Labor Market Outcomes for Women" shows the wage gap and the participation rates for married women in the United States. [2] The participation rate for married women—the fraction of married women in the labor force—has increased from slightly above 20 percent in 1950 to about 70 percent in 2000. Meanwhile, the ratio of wages paid to married women relative to married men displays an interesting pattern over this period. From 1950 to 1980, the ratio fell from 65 percent to 60 percent—that is, the wages of married women fell relative to married men. Thereafter, the ratio rose substantially, to about 80 percent in 2000. At the end of the 20th century, in other words, married women were earning about four-fifths of the wages of married men.

Economists and other social scientists are interested in understanding these facts. What was the source of the increased participation in the labor force by women and what factors increased their wages relative to men? One tempting approach is to use a supply-and-demand diagram like Figure 12.5 "Labor Market Equilibrium", thinking specifically about women's labor. For example, we could explain the overall shift between 1950 and 2000 by a rightward shift of the labor demand curve. A shift to the right in the demand curve increases the real wage. The higher real wage would also induce women to supply more hours: this is the corresponding movement along the labor supply curve. More women would be induced to move away from work at home and toward work in the market, given the higher return for market work. To explain the increase in women's wages *relative* to men's, we would need to see a larger increase in the demand for women's labor than for men's labor.

But this is a somewhat odd story. There is no reason to think that there should be a separate labor market for women and men. Women and men can and do perform the same jobs and thus compete in the same labor market. Any supply-and-demand explanation needs to be subtler. One possibility is that there has been a shift in the kinds of jobs that are most important in the economy and hence a shift in the kinds of skills needed. Suppose, for example, that women are more likely to be accountants than construction workers. A shift in labor demand toward accountancy and away from construction will increase wages in accountancy relative to construction work and will therefore increase women's wages, on average, relative to men's. Researchers looking closely at the data see some evidence of such effects when they look at wages and employment patterns across jobs that require different skills.

There is another, perhaps even more basic question: why are women's wages consistently lower than men's wages? Researchers have also devoted a great deal of effort to this problem, looking to see in particular if differences in education and skills can account for the difference in wages. Typically, these studies have found that such differences can explain some—but not all—of the gap between wages for men and women. The remaining difference in wages is very possibly due to discrimination in the labor market. If this is the case, then recent increases in women's wages relative to men's wages could be due to a reduction in discrimination.

Of course, women are not the only group that has been subject to discrimination in the labor market. In the United States, African Americans and other minority groups have suffered from discrimination. In many other countries, there are similarly different groups that have been unfairly punished in the labor market. Economists point out that supply and demand is actually a positive force for combating discrimination. Discrimination against women workers, for example, means that women are being paid less than their marginal product. Nondiscriminatory employers then have an incentive to hire these workers and make more profit, which in turn would tend to increase women's wages.

Economic forces can mitigate discrimination, but this is not an argument that discrimination is not or cannot be a real problem. First of all, discriminatory attitudes might make employers incorrectly perceive that the marginal product of women (or other groups) is lower than it actually is. Second, even if employers are not actively discriminating against women, coworkers may be discriminatory, and this could lead to lower productivity among women in the workforce. Research in social psychology tells us that such discrimination—by employers or colleagues—can occur even if people have no explicit discriminatory intent.

Winner-Takes-All Markets

There are some markets where compensation reflects ability in a very extreme way. These are often called winner-takes-all markets. In such a market, the person with the highest ability captures the whole market, and everyone else gets nothing. You can think of this as a race where the winner of the race gets all the prize money. The phrase *winner takes all* is not meant literally. The idea is more that a small number of people earn very large returns. Think, for example, of the professional golf or tennis circuits, where perhaps a few hundred people obtain the winnings from the tournaments—and the bulk of the winnings go to a small number of top players.

In these markets, we cannot assume that the wage equals the marginal product of labor. In a winner-takes-all market, you get a wage that depends not on your productivity in isolation but on how your productivity compares with that of others. If you are the most productive, you win the entire market.

Many markets have at least some aspects of a winner-takes-all market. Think of the market for rock musicians. If there were one group that everyone liked more than all the others, then that group would sell CDs and MP3s, give concerts, and completely dominate the music scene. Other groups would disappear. The actual music market is not this extreme. There are many groups who produce songs, give concerts, and so on. But there is a clear ranking between the first-class groups and the others. So even though there is not a single winner who takes all the market, there are a relatively small number of big winners who together take most of the market.

Why does the market for rock musicians have winner-takes-all characteristics? A good way to understand the phenomenon is to think about the market for musicians centuries ago—before recording technologies. Good musicians might still be rewarded well—perhaps they would play for the king or queen—but there was room for, relatively speaking, a large number of good musicians because each would be serving only a relatively small local market. Today, though, the very best musicians can record their music and sell it all around the world. A single group, at relatively low marginal cost, can serve a very large market. (This is particularly true for CDs or MP3 files. It is less true for concert appearances because these do not have such low marginal cost.)

In winner-takes-all markets, there is a very skewed distribution of income relative to ability. Small differences in ability can translate into substantial differences in income. Moreover, winner-takes-all forces may be becoming stronger as a result of technological advances. The most popular rock stars, sport stars, and movie stars are now worldwide celebrities. Lady Gaga is famous in Thailand and Toledo; Brad Pitt is known from Denver to Denmark. This is perhaps one reason the very rich are getting relatively richer.

From Income to Consumption and Wealth

We are interested not only in the distribution of income but also in the distribution of consumption and wealth. To connect these three, we use the following equation: [3]

wealth next year = (wealth this year + income this year − consumption this year) × interest factor.

The first term on the right-hand side is the wealth you have at the start of a given year. To this wealth you add the income you earn in the current year and subtract your consumption. Because income − consumption = savings, this is the same as saying that you add your savings to your wealth. You earn interest income on your existing wealth and your new savings. Your initial wealth plus your savings plus your interest income gives you the wealth you can take into next year.

Suppose you currently have $1,000 in the bank. This is your wealth this year. You receive income of $300 and spend $200 of this income. This means that you save $100 of your income. So wealth this year plus income this year minus consumption this year equals $1,100. With an interest rate of 5 percent, your wealth next year would be $1,100 × 1.05 = $1,155.

This equation tells us several things.

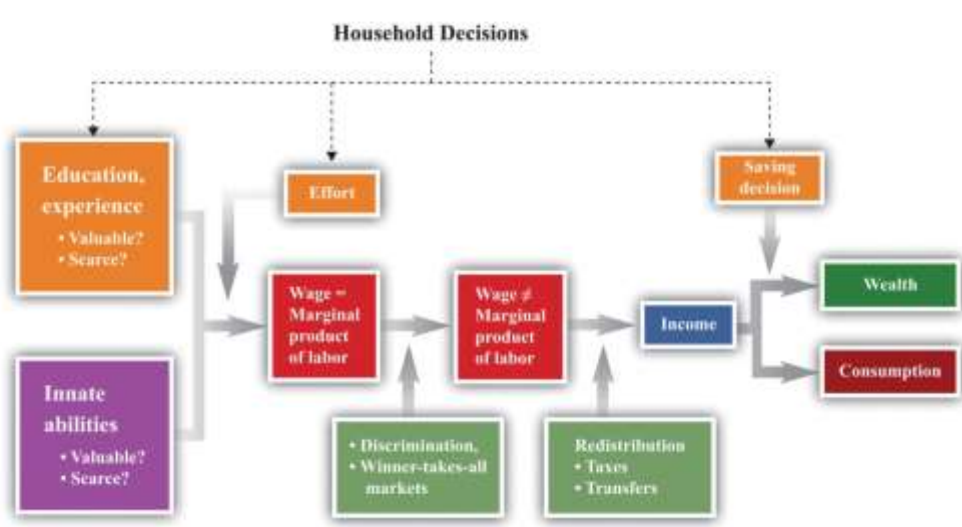
- Wealth, income, and consumption are interconnected. A household's decisions about how much it wants to save and how much it wants to consume determine what its consumption and wealth will look like. Imagine two otherwise identical households with different preferences about consuming this year versus the future. The impatient household consumes a lot now and saves little. It has high consumption early in life, low consumption later in life, and relatively low wealth. A more patient household has a very different pattern of wealth and consumption. It has lower consumption early in life, higher consumption later in life, and higher wealth on average.
- Differences in earnings cumulate over time to generate a distribution of wealth. High-ability households are more productive and thus earn more income. Some of this income is saved, and the rest consumed. Higher-income households thus tend to have higher wealth than lower-income households because the higher-income households have higher levels of saving each year.

- Inherited wealth can be a source of differences in income and consumption. Some individuals start out their working lives as beneficiaries of inheritances from their parents (or others). These people can enjoy higher consumption. They also obtain more income in the form of interest earnings on their wealth.

The equation also *conceals* at least one relevant fact for inequality: wealthier households typically enjoy higher returns on their wealth. The interest rate is not the same for all households. There are several reasons for this, such as the fact that richer individuals find it worthwhile—and can afford—to hire professionals to manage their portfolios of assets or the fact that richer people may be able to purchase assets that are riskier but offer higher returns on average. It is not surprising that, as we saw, the wealth distribution is more unequal than the income distribution.

Figure 12.7 "The Different Sources of Inequality" brings together all the ideas we have discussed so far. It shows us three things. (1) Discrimination and winner-takes-all situations can break the simple link between the marginal product and the wage. (2) Government policies can break the simple link between wages and income. (3) Household decisions about how much to consume and save affect the observed amounts of income, consumption, and wealth. The figure also makes it clear that some of the forces leading to inequality are under the control of the individual, while others are outside the individual's control.

Figure 12.7 The Different Sources of Inequality



KEY TAKEAWAYS

- Differences in income can reflect, among other things, differences in ability, education, training, and gender.
- Wage differences across people reflect differences in marginal products across people.
- The skill gap shows the differences in earnings from differences in education. This gap has widened in recent years.
- In a winner-takes-all market, the most talented individual captures all (or almost all) of the market.

CHECKING YOUR UNDERSTANDING

1. Draw two versions of the labor market: one for lawyers and one for taxi drivers. How would you use these labor markets to explain the differences in labor income between lawyers and taxi drivers? Are these two labor markets related in any ways?
2. Does Figure 12.6 "Labor Market Outcomes for Women" imply that as more women participate in the market, there are increases in the ratio of wages earned by women relative to men?
3. Where do differences in wealth come from?

[1] See Chapter 3 "Everyday Decisions", Chapter 7 "Why Do Prices Change?", and Chapter 8 "Growing Jobs" for more discussion.

[2] We are grateful to Michelle Rendell for this figure. The discussion in this section is drawn in part from her PhD dissertation research.

[3] See Chapter 4 "Life Decisions" for more discussion.

12.3 Distributive Justice

LEARNING OBJECTIVES

1. What is the evidence from economic experiments about "fairness"?
2. What are some of the leading theories about "fairness"?

So far we have described some facts about inequality in the United States and the world, and we have offered some explanations of why we observe these inequalities. In this section, we take a more philosophical perspective on the distribution of income and wealth. We ask questions of a kind that economists generally ignore, such as the following: "Is the distribution of income fair?"

As you might expect, questions like this are extremely contentious. Different people have very different ideas about what is fair and just, and this topic is highly politicized. It is not our job, nor is it our intention, to tell you what is and is not fair. What we can do is give you a (very brief) introduction to some of the ways that philosophers, economists, political scientists, and others have thought about these very hard questions. More particularly, we can give you some "thought experiments" to help you determine your own views on these topics. Hundreds of books have been written on these issues, however, so we simply scratch the surface here.

Experimental Evidence on Fairness

Noneconomists frequently speak about a "fair wage" or a "fair price" for a particular product. To economists, this language is unfamiliar, even confusing. Economics provides no theory about what is fair or unfair; it gives us no basis to ask whether particular prices in the economy are fair.

Yet ideas about fairness motivate people in many economic transactions. As one example, some people are willing to pay extra for "fair trade" goods, such as coffee or chocolate bars. The idea of these goods is that the seller makes some guarantees about payments to producers, working conditions, or other variables that are not intrinsic to the good itself. As another example, people are often willing to take part in boycotts, meaning that they voluntarily forgo a good that they like to send a message to the producer of the good.

Experimental economists have conducted many studies to try to understand some of these ideas of fairness. Sometimes they have used a dictator game. This game has two players. Player A, the dictator, is given a sum of money and decides *how much of that money to give* to player B. Player B keeps the money he is given, and player A keeps the rest. From the perspective of economic reasoning, this game is completely trivial. Suppose you are the dictator, and you are given $100 to allocate. The self-interested thing to do is to keep all the money for yourself and give nothing to player B.

Yet study after study has shown that people typically give away some of their money, often dividing it up in equal shares. You may be able to think of several reasons why people behave this way. Perhaps they are worried about what the other person will think about them. Perhaps they are worried about what the *experimenter* will think about them. Researchers have gone to great lengths to design studies where no one except player A can possibly know her decision. Even in this case, most people do not keep all the money. It is hard to dismiss the view that people's decisions are motivated in some way by what they think is the fair thing to do.

A related but slightly richer game is known as the ultimatum game. It also has two players. Player A is given a sum of money and then decides *how much of that money to offer* to player B. Player B then decides whether to *accept or reject* player A's offer. If player B accepts that offer, he keeps the amount offered, and player A keeps the rest. If player B rejects the offer, then both player A and player B receive nothing.

Toolkit: Section 17.18 "Nash Equilibrium"

You can read more about these games and others in the toolkit.

The difference between the ultimatum game and the dictator game is that player B has the right to veto the offer. If he vetoes the offer, then both players get nothing. Economic theory again has a simple prediction about what completely self-interested players will do. Player B is better off accepting *any* positive offer than he is rejecting the offer. Suppose player A starts with $100 and offers $1 to player B. If player B accepts, he gets $1. If he rejects, he gets $0. Because $1 is better than nothing, player B should accept the offer. Knowing this, player A should offer the smallest amount possible. For example, if player A has $100 to allocate, she should offer $0.01. Player B should accept the offer ($0.01 is bigger than $0.00), and player A will then end up with $99.99. In fact, this is not what usually happens. People in the role of player A typically offer much more than the minimum amount. One reason is the risk that if player B is made a stingy offer, he will reject it out of spite. Another reason, like in the dictator game, is that people may care about fairness when making their offers. The evidence suggests that both factors seem to matter in this game.

Hundreds of studies have been conducted using different variants of these two games. The big conclusion from all these studies is that people seem to be motivated by more than just narrow self-interest when they play games such as these in the laboratory. Instead, they care about allocating the rewards from the experiment in a way that is fair. Understanding exactly what underlies these ideas of fairness is an exciting area of research in experimental and behavioral economics, as well as in psychology and other disciplines.

Meritocracy

We begin with a very simple framework. Imagine an economy in which there are two kinds of people: high ability and low ability. Half the people in the economy are high ability: they can produce 100 chocolate bars in a year. The other half are low ability: they can produce only 50 chocolate bars in a year. In this economy, productivity and ability are the same thing. High-ability people are more productive than low-ability people. We use this simple economy to think about different approaches to the allocation of society's resources.

Libertarianism

One view of distribution is summarized by the statement "you are entitled to whatever you earn." In this world, the distribution of income and consumption will be the same as the distribution of output. High-ability people have income of 100 chocolate bars. If our fictional economy were to last for only one year, their consumption would also be 100 chocolate bars. Similarly, low-ability people will have income and consumption of 50 chocolate bars. This economy has an unequal distribution of income and consumption.

If we were to associate this position with a particular philosophical school of thought, it would be *libertarianism*. Libertarians generally believe that people are entitled to whatever they can earn, the state should intervene as little as possible, and the state should not actively seek to redistribute resources. The fact that there is inequality in this society is simply a reflection of differing abilities, which is not any reason for the government to get involved. (To be clear, libertarians have no objection to people making charitable contributions. If the high-ability people in the economy wanted to voluntarily give money to the low-ability people, libertarians would have no complaint about this.)

Consumption, Saving, and Insurance

Now let us consider a slight variant on this economy. Suppose the economy lasts for two periods: in each period, every individual has a 50-50 chance of being either high or low ability. If we measured income in either period, we would see the same amount of inequality as before.

Consumption, however, is a different story. Suppose you are a high-ability person in the first period. You know that you face a risk of being low ability in the second period. Should you eat your entire 100 chocolate bars in the first period? Most people prefer to keep their consumption at least somewhat smooth, so they will "save for a rainy day." We expect that high-income people in this economy will consume less than their income.

Similar reasoning applies to low-ability people. They earn only 50 bars in the first year but have a 50-50 chance of higher income next year. By the same consumption-smoothing argument, they would like to somewhat increase consumption today. Thus low-ability people will consume more than their income in the first period. There will be a credit market (or loan market) in which high-income people lend money to low-income people in the first year, and those loans are repaid the following year.

Toolkit: Section 17.6 "The Credit Market"

You can review the idea of the credit market in the toolkit.

This example of borrowing and lending driven by the desire for smooth consumption affects the distributions of income and consumption. Economic theory tells us that consumption will be more equal than income. This is consistent with the evidence: consumption is indeed more evenly distributed than income. Again, believers in a libertarian philosophy would see no reason for any intervention in this economy.

If this economy were more sophisticated, it might even develop an insurance market. All the individuals in the first year would recognize that their future income was uncertain. If they are risk-averse, then they would all prefer to eliminate this uncertainty. Being risk-averse means you prefer the average of a gamble to the gamble itself. Suppose a person is faced with a choice between

- 100 chocolate bars with a **probability** of 0.5 and 50 chocolate bars with a probability of 0.5
- 75 chocolate bars

A risk-averse person prefers the option that delivers 75 chocolate bars with certainty. The first option also yields 75 chocolate bars on average (more technically, it has an **expected value** of $0.5 \times 100 + 0.5 \times 50 = 75$), but this option has uncertainty that risk-averse people will want to avoid. In this economy, there would be some redistribution of income in the second year. However, it would be a voluntary redistribution based on the insurance contract that everybody agreed to in the first year. Again, there would be no role for government.

Toolkit: Section 17.7 "Expected Value"

You can review the concepts of probability and expected value in the toolkit.

The Rawlsian Veil of Ignorance

One of the most famous approaches to the questions of fairness and justice was pioneered by the philosopher John Rawls in his celebrated book, *A Theory of Justice*. [1] Rawls introduced a powerful thought experiment to help people decide how they feel about different distributions of society's resources.

It is difficult for any of us to think about redistribution without framing it in terms of our own personal circumstances and interests. Rawls's idea was designed to help us shed those considerations. He proposed thinking about redistribution from behind a veil of ignorance. Behind this veil, you know what the distribution of resources and abilities will look like in society, *but you do not know where you will be in this distribution*. You might be born rich, or you might be born poor. You could end up as Bill Gates, or as a homeless person in New York. If you want to play this game globally, you might end up as a member of a royal family in Europe or as someone scavenging for food on a garbage heap in Cambodia. If we frame this in terms of our previous example, then, behind the veil of ignorance, you know that 50 percent of the people will be high ability, and 50 percent will be low ability, but you do not know which you will end up being.

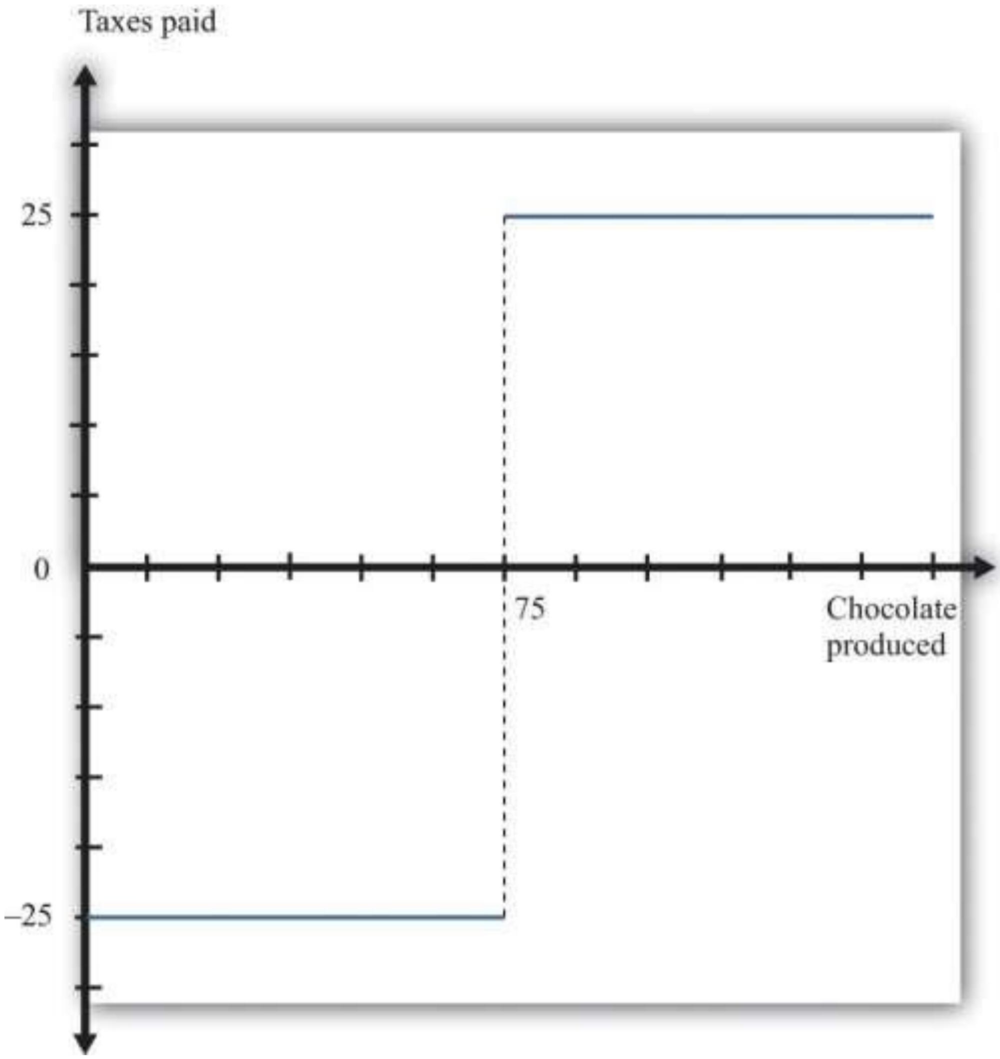
Now suppose that decisions on how to allocate chocolate bars across households are made before people know whether they are high or low ability. Rawls suggested that people behind the veil would adopt a social contract in which they agree to the following.

- Once born, they will produce chocolate bars according to their ability and then put the chocolate bars they produce into a big pile.
- Each individual will take out an equal share of 75 bars.

This contract involves taxation and redistribution. The high-ability people are each taxed 25 bars, and the low-ability people receive a transfer of 25 bars. The taxes are sufficient to finance the transfers.

Figure 12.8 "Taxes and Transfers in a Rawlsian Social Contract" shows a taxation and transfer scheme that could be used with this social contract. On the horizontal axis is production, which is income. On the vertical axis is the tax paid by each income group. With this scheme, anyone with income above 75 bars pays a tax of 25 bars. Anyone with income below 75 bars gets a transfer of 25 bars. Because there are an equal number of high- and low-ability households, taxes collected equal transfers. The government's budget balances.

Figure 12.8 Taxes and Transfers in a Rawlsian Social Contract



With this tax scheme, households producing more than 75 chocolate bars pay a tax of 25 bars, and those producing fewer than 75 bars receive a transfer (negative tax) of 25 bars.

Because everyone is risk-averse, all will prefer this deal to the allocation that gave the high-ability people 100 bars and the low-ability people 50 bars. Though additional chocolate bars are not produced, the redistribution of the contract is preferred to everyone *before they know their ability*. The key, emphasized in the previous sentence, is that the contract is agreed on before people know their ability. Because of this timing, the risk sharing through the redistribution of the chocolate bars makes everyone better off, compared to the—imaginary—initial condition.

You have almost certainly noticed that this Rawlsian social contract very closely resembles the insurance contract that we described in Section 12.3.2 "Meritocracy". In effect, Rawls suggested that people behind the veil would want to write the same kind of insurance contract that they would write in a similar situation in real life. But because we obviously can't write contracts before we are born, Rawls thought that we should agree to government policies that would mimic these kinds of insurance contracts. Notice that, in the Rawlsian world, the distribution of income has a higher Gini coefficient than does consumption. In fact, in this example, there is no inequality in consumption.

From Each According to His Ability and to Each According to His Needs

Karl Marx, the famous philosopher and social theorist, suggested that society should distribute its resources as follows: "From each according to his ability, to each according to his needs." Marx's prescription recognizes that individuals differ in their ability to produce and in their consumption needs. He said that workers should produce at a rate commensurate with their ability, so high-ability individuals would be expected to produce more output than low-ability individuals.

In the Marxian view, there is a complete disconnect between production and consumption. There is no sense that those who *produce* more of society's resources should be entitled to *consume* more of those resources. It stands in complete contrast to the libertarian view that individuals have a right to whatever they produce. The distribution of production is independent of the allocation of income and consumption.

How would the Marxian view work in our chocolate bar economy? "From each according to his ability" means simply that the high-ability individuals should produce 100 chocolate bars and the low-ability individuals should produce 50 chocolate bars. Meanwhile, "to each according to his needs" means that the total number of chocolate bars produced in the economy ought to be allocated in a way that reflects the needs of the individuals. In our simple example, individuals do not differ in their valuation of a chocolate bar. All individuals like chocolate bars the same amount. Therefore, the allocation that satisfies the Marxian prescription is that everyone should have the same number of chocolate bars.

In our simple example, Marx and Rawls agree on how to allocate chocolate bars. We can imagine, however, ways in which individuals might differ in terms of their needs. For example, some people are fortunate enough to be healthy and fit, while others suffer from illness or disease. A Marxian prescription would allocate more of society's resources to the sick, on the grounds that their needs were greater. (It is also possible, of course, that people behind the Rawlsian veil of ignorance would make a similar allocation.)

Luck versus Merit

In all of our examples so far, we have supposed that people differed only in terms of their abilities, which are—by assumption—completely outside their control. In our earlier discussion of the sources of inequality, however, we listed many different possible reasons why people might have different earnings. Some of these factors were outside any individual's control; others were not. Table 12.9 "Luck versus Merit" provides a partial listing.

Table 12.9 Luck versus Merit

Outside an Individual's Control	Within an Individual's Control
Innate abilities	Effort and hours worked
Demand for these abilities	Education (in part)
Supply of these abilities by others	Experience and training
Discrimination	Consumption/saving decisions
Inherited wealth	

An individual does not control his or her basic abilities. Some are lucky, possessing the abilities that allow them to be great basketball players, pianists, authors, or scientists. Abilities that are scarce are likely to be more valuable. The value associated with a particular set of abilities is also heavily dependent on time and place—for example, being a great rock drummer would not have been worth much in the Roman Empire, and an ability to throw a spear hard and accurately is not especially valuable in modern-day San Francisco. All of these come down to luck when viewed from the perspective of any individual.

We have hinted at many other factors that are also a matter of luck. Those born of wealthy parents in wealthy countries are likely to attend high-quality schools and receive inherited wealth. They may also be able to earn higher real interest rates on their savings. Meanwhile, those who are subject to discrimination will earn lower incomes.

There are also many factors that influence the distribution of income, consumption, and wealth that *are* under the control of an individual. Individuals can choose how hard to work and how many hours to work. They can choose whether to sacrifice current earnings to go to college. They can decide to go back to school to earn a master's degree. They can choose careers that allow them to develop skills and experience on the job.

Why does this distinction matter? Most people would agree that there is little or no problem with inequalities that result from people's choices. There is nothing self-evidently unfair about one person having a higher income than another because he works harder or chose to take time off from work to pursue a graduate degree. But opinions differ much more about the fairness of inequalities that result from luck or chance. Tiger Woods is an immensely talented golfer, but is it fair that he should earn so much on the basis of his genetic luck? Is it fair that someone who struggles in school and possesses little in the way of valuable skills should earn only minimum wage? These are not questions that we can answer, but thinking about these questions should help you form your own opinions on what is a fair and just distribution of society's resources.

Equality of Opportunity versus Equality of Outcome

The distinction between luck and merit gives us a more nuanced view of equality. It is closely related to another distinction that is often made when discussing the distribution of society's resources: equality of opportunity versus equality of outcome. Here is an example to help make the distinction clear.

At major soccer tournaments, such as the World Cup, teams often line up behind banners proclaiming "fair play." The international soccer association, FIFA (Fédération Internationale de Football Association), places a lot of emphasis on this idea. Fair play means that players should always play within the rules, and these rules provide equality of opportunity on the soccer field. At the start of any game, both teams line up with the same number of players, try to score in the same sized goal, and enjoy the benefits of impartial referees. This does not mean that soccer games always end in a tie: FIFA's rules do not mean that there is equality of outcome. The outcome depends on the two teams' abilities. So although the opportunity to win is shared equally by the teams, the outcome is not: the winner takes all.

Equality of opportunity without equality of outcome is pervasive in the economy as well. Institutions exist to enhance equality of opportunity with no guarantees about outcomes. For example, going to a public school is an option for everyone (though there are significant differences across schools in terms of their quality). But there is no guarantee that two people graduating from the same school will have the same outcome. When you apply for a job, you have an opportunity to compete along with anyone else for that job, but the outcome is different for the person who is hired compared to those who are not.

It is tempting to identify equality of opportunity with the view that merit should be rewarded but luck should not. There is certainly a connection. Both imply that discrimination should not affect the distribution of income in the economy. But equality of opportunity still allows those with high abilities to get higher rewards, even though those abilities are a matter of luck. If your college soccer team were to play Real Madrid, either team would have the chance to win the game, according to the rules. That equality of opportunity would be of little consolation to your team's goalkeeper as he picked the ball out of the net for the 20th time.

Yet there is one very good reason why equality of opportunity is so important. Imagine what would happen if FIFA started instructing referees to ensure that every soccer game ended in a draw. To ensure equality of outcome, the referee would alter the rules of the game to help the side that was losing. Fair play would be gone, together with lots of other things: teams would have no incentive to play hard, they would have no incentive to find quality players, and fans would not enjoy the game as much. We get the best from a team because it knows that if it performs well, under the rules, it will win and receive financial and emotional rewards. These provide the incentives for team members to train and play hard, within the rules of the game. Combining equality of opportunity with the ability to compete for a prize strikes the right balance—at least for soccer—between equality and incentives. In the next section, we will examine why incentives matter so much for decisions about redistribution.

KEY TAKEAWAYS

- In experimental bargaining games, players seem to be motivated by more than narrow self-interest. In many cases, they give money to the other player.
- Equality of opportunity argues that everyone should have an equal chance of succeeding without guaranteeing that success. It contrasts with the view that everyone should work as hard as they can, and goods and services should be allocated according to need.

CHECKING YOUR UNDERSTANDING

1. Does the grading in your economics class exhibit equality of opportunity? Why or why not?
2. If you think about the allocation of resources within a household, which of the theories of distributive justice best applies?

[1] John Rawls, *A Theory of Justice* (Cambridge: MA, Harvard University Press, 1971). Rawls's work is rich, complicated, and much debated, and the presentation here is very simplified and stylized. For example, Rawls focused more on the institutions that people behind the veil would want, rather than on the actual distribution of income.

12.4 Government Policy

LEARNING OBJECTIVES

1. What actions does the government take to influence the distributions of income, wealth, and consumption?
2. What is the rationale for these government interventions?
3. What limits the effects of government redistribution?

Governments play a significant role in the distribution of income, consumption, and wealth. The argument for government intervention usually takes the form that the market outcome is too inequitable, relative to, for example, a Rawlsian view. We now look at various forms of redistribution through government actions, paying particular attention to their effects on incentives.

Incentives

Redistribution is more than setting taxes and transfer payments to give money from one person to another. The problem is that redistribution can affect people's incentives in various ways.

The Incentive to Be Truthful

Go back once more to our chocolate bar economy. We proposed a scheme whereby high-ability individuals would be taxed 25 chocolate bars, with this being paid to low-ability individuals. A tax-and-transfer scheme of this kind would allow us to achieve the equitable outcome mandated by the Rawlsian or Marxian view.

Low-ability households evidently have an incentive to participate in this scheme: they give up 50 bars and get back 75 bars. The redistribution is in their favor. The story is different for high-ability people. They give up 100 bars and get 75. Before abilities are known, everyone likes this social contract. But once ability is known, high-ability people prefer not to participate. If they can produce and then hide some of their chocolate bars, they have an incentive to

- produce 100 chocolate bars,
- pretend to be a low-ability person and declare production of 50 bars, and
- take the transfer of 25 bars and consume 125 chocolate bars.

High-ability people can get away with this if chocolate bar production cannot be monitored. They have an incentive to rip off the system by pretending to be low ability. Because all high-ability people behave this way, the contract will fail: no one will pay taxes, and everyone will demand a transfer.

In this extreme example, the incentive problem completely destroys the redistribution policy. In reality, there is some redistribution through taxes and transfers because the government, acting through the taxation authority, is able to tax households at different rates: low-income households face lower tax rates than higher income households. In addition, low-income households receive transfers from the government. Governments can carry out such policies because they have access to information about the income households earn. Yet incentive problems like the one we have outlined pose very real difficulties for governments. Rich people have an incentive to hide their true income and do so through legal and illegal means. For example, a recent story in the *New York Times* began as follows: "In the wealthy, northern suburbs of [Athens, Greece], where summer temperatures often hit the high 90s, just 324 residents checked the box on their tax returns admitting that they owned pools. So tax investigators studied satellite photos of the area—a sprawling collection of expensive villas tucked behind tall gates—and came back with a decidedly different number: 16,974 pools." [1]

The Incentive to Work

As the real return to working increases, households will generally work more. Labor supply is upward sloping: increases in the real wage lead to more people participating in the labor market and individuals' choosing to work more hours. Households care about the real wage *after taxes*—that is, they decide how much to work based on the wage they receive after paying tax. Everything else being the same, an increase in the tax rate on labor income reduces the real wage received by households, and they will work less in response.

Contrast high-ability and low-ability workers. High-ability workers are more productive. From society's point of view, it is better for them to work more. But if tax rates are higher for higher-income people, then these people will have an incentive to work less, so total output for the economy will be lower. This lost output is the efficiency loss from the progressive tax system.

The Incentive to Train

Redistribution can also affect the incentive to study and acquire additional skills. Once again, we use our chocolate bar example. We still have two types of individuals: high ability and low ability. Which type you are when you are born is completely beyond your control; it is just a matter of luck. But the actions you take, given your ability, are something you control.

Suppose that high-ability people can only produce 100 chocolate bars if they first go through some training. Further, assume that this training is not fun: everything else being the same, people would prefer not to spend time training. Instead, they would prefer to use their leisure time in other ways. Under the social contract, the efficient way to organize society would be for high-ability people to incur the cost of training to produce more output.

If the tax-and-transfer system completely equalizes incomes, however, high-ability people will not think it worthwhile to train. This highlights a problem with the Marxian view of "from each according to his ability, to each according to his needs." The incentives needed to induce people to produce according to their ability may be inconsistent with allocating goods according to need.

Assuming that a little inequality is better than a lot of lost chocolate bars, the social contract needs to be amended to create an incentive for high-ability people to train. The solution is to give them some extra chocolate bars as an inducement to train and thus produce more for society. The result is inequality in consumption.

The Leaky Bucket

The incentive problems that we have discussed so far result in an equity-efficiency trade-off. Arthur Okun, a famous economist in the 1960s, proposed a very useful thought experiment for thinking about such trade-offs. He imagined that redistribution from the rich to the poor is like carrying a bucket of water from one person to another. Unfortunately, the bucket leaks. So the process of transferring water from one person to another also means that there is less total water available.

At one extreme, if the bucket does not leak, then there is no trade-off. You can redistribute water evenly in society without any loss in efficiency. At the other extreme, all the water gets lost in the transfer. The only way to achieve equality in this society is simply by destroying the wealth of the rich. Okun invited his readers to contemplate how much leakage they would be willing to tolerate to make society more equal. If you are in favor of a more equal society, then you too should think about the extent to which you think it is worth sacrificing some of our output to share the rest out more equally.

The Inheritance Tax

At the beginning of this chapter, we listed the wealthiest people in the United States in 2006 and 2010. Do you think that 50 years from now, the families of these people will appear on the *Forbes* list of the wealthiest people in the United States in 2060? The answer to this question partially depends on the choices of these wealthy people: how much of an estate will they decide to leave to their families? It also depends on how much of the estate the government will tax.

When we talked earlier about the dynamics of inequality, we noted that there were links across generations of a family. Some of those links come directly from expenditures on children. Everything else being the same, richer families have more income to spend on their children's education, and thus their children are likely to be more productive. The transfer of wealth is a second link that leads income (earned on financial investments) to be higher for children of wealthier families.

According to the current tax code in the United States, the tax rate applied to an estate appears to be progressive, with higher tax rates levied on larger estates. But there is an exclusion of $5 million, and only estates above this level are taxed at a 35 percent tax rate. So if you were left an estate valued at $6 million, you would pay a tax of $350,000 (= 0.35 × [$6,000,000 − $5,000,000]). Not surprisingly, the inheritance tax is hotly debated. Opponents of the tax argue that individuals ought to have the right to spend their lifetime income on whatever they want, including their children. Proponents of the tax see it as a way to increase mobility within the wealth distribution and argue that it promotes equality of opportunity.

Transfers

The government redistributes across households using taxes and transfers. This redistribution is reflected in the difference between the Gini coefficient for market income and postinsurance income in Table 12.3 "Household Income by Quintile". Transfers arise through unemployment insurance payments to unemployed workers, government-financed health care to the poor and the elderly, and other government schemes. [2]

Transfers, like taxes, can affect incentives. Suppose the government makes transfers of $100 to everyone in the economy with income less than or equal to $1,000. Think about an individual who works 40 hours at a wage of $25 per hour to earn a weekly income of $1,000. What are the gains to working 41 hours? If the individual works an hour more, then her income (before taxes and transfers) will increase by $25 to $1,025. But by working an extra hour, she no longer qualifies for the transfer of $100. So she would lose $100 in transfers: the extra hour's work would reduce her income by $75.

Not all transfers are public; some are private. Many of the wealthiest people in the world are also some of the most generous in terms of setting up private foundations. For example, the Bill and Melinda Gates Foundation (http://www.gatesfoundation.org/Pages/home.aspx) was created in 2000 "to help reduce inequities in the United States and around the world." The reported value of the trust endowment is $34.6 billion, which includes $1.6 billion from Warren Buffett, the number two person on the 2006 and 2010 *Forbes* lists. Another common form of private transfers comes from tuition reductions from private universities. As a leading example, Princeton University replaced student loans, which had to be repaid, with outright grants to qualified students. Other universities provide both grants and subsidized loans.

KEY TAKEAWAYS

- Governments use a variety of tools, such as income taxes and inheritance taxes, to influence the distributions of income and wealth.
- Governments are motivated by the view that market outcomes are not equitable enough.
- Actions to redistribute are limited by the adverse incentives created by taxes and transfers.

CHECKING YOUR UNDERSTANDING

1. Does a progressive income tax lead to a more or less equitable distribution of disposable income?
2. Will an inheritance tax create an incentive for people to work more or less?

[1] See Suzanne Daley, "Greek Wealth Is Everywhere but Tax Forms," *New York Times*, May 1, 2010, accessed January 30, 2011,http://www.nytimes.com/2010/05/02/world/europe/02evasion.html?hp.

[2] Chapter 15 "A Healthy Economy" returns to the topic of government transfers associated with health care.

12.5 End-of-Chapter Material

In Conclusion

Most of the time in the study of economics, we focus on efficiency. We ask if there are better or worse ways for society to organize its production of goods and services, and we ask if society has institutions in place that allow people to obtain all the available gains from trade. Although these can be complex questions, there is broad agreement among most people that efficiency is a desirable goal.

In this chapter, we tackled a rather different and more contentious set of issues: what is fair and just? Economists can (relatively) easily explain how society ends up distributing its resources, but the tools of economics do not allow us to say whether a given distribution is fair or not. People certainly seem to care about fairness and hold strong opinions about how society should share out its resources. Unfortunately, different people have very different ideas about what is fair.

The questions we address here go beyond economics; they vex philosophers and political scientists as well. They go to the heart of what we think of as right and good. They also force us to think about the appropriate role of the state and how that matters for the distribution of resources. Is the role of the state simply to provide an environment where people are free to pursue their own self-interest and to keep what they earn? Or does the existence of the state mean that we are all in a social contract, so all have some rights to the output of society as a whole?

As we have said previously, we cannot and do not want to answer these questions for you. Indeed we, as authors of this book, do not even agree among ourselves on the answers. Instead, we have given you some tools so you can think about these questions—which are some of the most important you will ever confront—yourself.

Key Links

- US Census Bureau: http://www.census.gov
- *Forbes* lists: http://www.forbes.com/lists

- The World Bank on poverty:http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/0,,menuPK:336998~pagePK:149018~piPK:149093~theSitePK:336992,00.html
- Gates Foundation: http://www.gatesfoundation.org/Pages/home.aspx

EXERCISES

1. Draw a Lorenz curve for the data given in Table 12.2 "Example of Income Distribution".
2. Often income data are reported by household. How does the US Census Bureau define a household? Is this the same as a family?
3. Draw the Lorenz curve for the wealth of the top 10 people in the United States for 2006 and 2010 using the data in Table 12.1 "Wealthiest Individuals in the United States".
4. Can you think of two other markets with significant winner-takes-all elements?
5. During the past 100 years, there has been tremendous technical progress in creating machines to run in the household, such as dishwashers, washing machines, clothes dryers, and so on. How do you think these inventions have affected the labor participation decisions of women and the wages they are paid?
6. Suppose that the cost of training is 20 chocolate bars. Assume high-ability people produce 100 bars if they get training and 50 bars if they don't. Low-ability people produce 50 bars regardless of training. If under the social contract you decide to provide an incentive for high-ability people to train, what is the distribution of consumption in the economy? Is society better off with inequality in consumption or is it better to have equal consumption and no training by high-ability people?
7. Start with the example of the social contract given in Question 6 but suppose that 75 percent of the people are high ability and 25 percent are low ability. What does the social contract look like for this economy? How much is produced by high- and low-ability people? What is the total amount of output per capita? What is the consumption per capita?
8. Identify two institutions that provide equality of opportunity but not outcome. Identify two institutions that favor equality of outcome over equality of opportunity.
9. Use college admissions to illustrate the difference between "equality of opportunity" and "equality of outcome."

10. Suppose a household holds a share of stock in a particular company and receives a dividend from that share. Which of these is a stock, and which is a flow? Which is part of income, and which is part of wealth?
11. (Advanced) Consider two countries: one has a higher Gini coefficient and the other has less mobility across income groups over time. Which country has greater equality?
12. If the return to education depends on innate ability, then what is the point of going to college?
13. Do you think that trading in the stock market exhibits equality of opportunity? Why or why not?
14. Can you come up with your own example of a trade-off between equity and efficiency?

Economics Detective

1. Pick one of the wealthiest people in the United States. How did this person get his or her wealth? How much do you think this person earns each year from his or her assets?
2. Find a list of the world's wealthiest people. What countries are these people from? Pick one person and see how the person got his or her wealth. Are the wealthiest people in the world distributed across lots of countries or isolated in a just a few?
3. The Rockefeller family was one of the wealthiest in the United States around 1900. How did the family accumulate its wealth? Where did the wealth go??
4. Try to find data on the share of income of the bottom 20 percent of the income distribution in two different countries. Also try to find the Gini coefficients for the two countries. How might you explain the differences in income distribution between the two countries you chose?
5. Go to the website of the Internal Revenue Service. Find the tax rates currently in effect for different income levels in the United States. Are these progressive?

Spreadsheet Exercise

1. Create (or find on the Internet) data on income. Input the data into a spreadsheet and plot the Lorenz curve.
2. Create a spreadsheet to follow the income and wealth of two households. Suppose the first household earns 50 chocolate bars each year, and the second household earns 100 chocolate bars each year. Suppose that each household saves a fixed fraction of its income (you can vary this in the spreadsheet). Follow these households for 50 years. Calculate familial wealth year by year using the equation at the

beginning of Chapter 12 "Superstars",Section 12.2 "The Sources of Inequality".6. To do this, you will have to specify the interest rate (which you can also vary). In what sense is the distribution of wealth more unequal than the distribution of income? What if the high-income households also had a higher return on saving? What if households sometimes produced 50 chocolate bars and other times produced 100 bars? As a very advanced topic, can you build this uncertainty into your spreadsheet program? What happens to wealth?

Chapter 13
Cleaning Up the Air and Using Up the Oil
Dirty Travels

Here are some places you probably would have difficulty finding on a map:

- Sumgayit, Azerbaijan
- Linfen, China
- Tianying, China
- Sukinda, India
- Vapi, India
- La Oroya, Peru
- Dzerzhinsk, Russia
- Norilsk, Russia
- Chernobyl, Ukraine
- Kabwe, Zambia

These 10 places have the dubious distinction of being the world's most polluted cities, according to a nongovernmental organization called the Blacksmith Institute. [1] Figure 13.1 "The 30 Most Polluted Cities in the World" shows these cities plus 20 more, giving us the 30 cities in the world with the worst pollution. In some of these places, mining and smelting industries have contaminated the air or the groundwater. In some, dangerous chemicals have been improperly disposed of—often illegally. In some, there is radioactive contamination. In some, garbage and sewage pollute the groundwater, or automobile emissions pollute the air. In sum, you would not want to live in any of them.

The consequences of such pollution are tragic. Pollution of this magnitude causes severe health problems, birth defects, and high mortality rates. For example, according to the Blacksmith Institute, life expectancy in Dzerzhinsk, Russia, is 42 for men and 47 for women. Lead pollution is directly linked to a reduction in children's intelligence and has also been linked to increased violence.

The map reveals that many of these highly polluted cities are found in India, China, and the countries of the former Soviet Union. The richer countries of the world, such as the United States, Canada, Japan, Australia, New Zealand, and all of Western Europe, are not featured. Neither, for the most part, are the very poorest countries of the world, such as much of Africa. Severe industrial pollution seems to be at its worst in middle-income, developing countries. This does not mean that rich countries did not recently have—or do not still have—pollution problems of their own. Only a few decades ago, the Cuyahoga River in Ohio was so polluted that it caught fire; indeed river fires were once relatively commonplace in the United States. The US Environmental Protection Agency lists well over 1,000 sites as eligible for Superfund cleanup (http://www.epa.gov/superfund) because environmental contamination is judged hazardous to health. Little more than 50 years ago, air pollution killed an estimated 4,000 people in London, England, during the so-called Great Smog.

We emphasize in many places in this book that market transactions generate value in an economy. Firms produce things that people want to buy, so both firms and consumers benefit. People voluntarily work for companies, earning money they can then use to purchase goods and services while simultaneously allowing firms to produce the products that people want. These claims are correct, yet the citizens in Linfen, China, or La Oraya, Peru, could be forgiven for thinking that this is a very rosy view of how economies function in practice. Those who live in these communities around the world obviously do not like living in such polluted environments. So what is going wrong? How is it that voluntary trades made by individuals and firms can sometimes lead to such unpleasant and dangerous outcomes?

To begin our answer to this question, let us take a particular example: Mexico City, a city that also makes the list of the worst 30. The air in Mexico City contains particulate matter (think of this as soot and smog) that can cause lung disease and other bronchial problems. This pollution largely comes from automobile emissions, which are a severe problem in part because so many cars in Mexico City are old. According to the Blacksmith Institute report, [2] even a moderate reduction in the amount of particulate matter in the Mexico City air could save thousands of lives each year.

But if everyone dislikes the pollution, why is it being produced? After all, no one is forcing the residents of Mexico City to drive their cars. They could all decide to drive much less, and if they did so, the result would be a cleaner city. Indeed, not everyone is a polluter. Particularly in richer countries, more and more people are driving electric cars or hybrids, which use a combination of electricity and fossil fuels. Such cars emit less pollution from their tailpipes. According to one study, the main reason that people purchase these cars is because they "want an environmentally friendly car." This reason was cited by 66 percent of the respondents. The next most popular response was "I want to save money on gas," which was cited by 16 percent of the respondents. The survey also found that half of all hybrid owners also donated money to environmental causes. [3]

Some of these reasons are clearly self-motivated: when gasoline costs $4 or more a gallon, fuel-efficient vehicles look very attractive. However, the desire to behave in an environmentally conscious way is rather different. People like to feel that they are behaving responsibly, even if they understand that their impact on total pollution is negligible. But there is another aspect of this desire to be green that is even more intriguing. If you interview one of these individuals, you will typically learn that he sees two ways in which hybrids are a good choice for the environment: (1) they generate less pollution, and (2) they consume less oil.

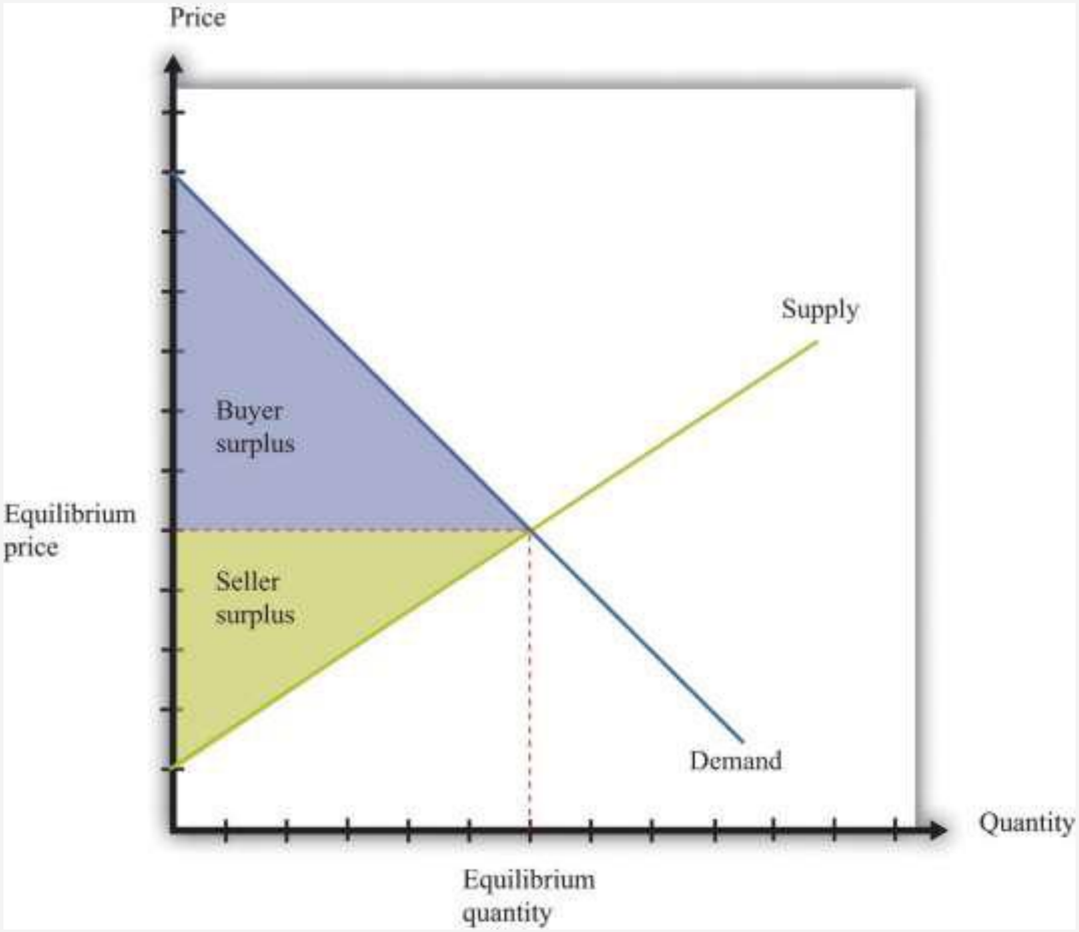
Both are environmental concerns. Both address how we use up different natural resources: clean air and fossil fuels. Yet they are very different problems. In this chapter, we consider both pollution and our consumption of natural resources—including but not limited to oil—and ask,

Can we rely on markets to deal with pollution and natural resources?

Road Map

We start our analysis with a familiar idea: the gains from trade. Figure 13.2 "The Gains from Trade", which also appears in other chapters in this book, illustrates one of the biggest insights of economics: voluntary transactions create value. In every voluntary transaction, both buyer and seller obtain surplus from trading. Even more striking, if these transactions take place in a competitive market, then buyers and sellers reap *all* the gains from trade.

Figure 13.2 The Gains from Trade



In a competitive market, total surplus (the sum of the buyer surplus and the seller surplus) is maximized.

The first section of the chapter looks at the use of clean air. To start off, we tackle this in a small-scale situation: we consider what happens if a smoker and a nonsmoker share an office. We ask under what circumstances they might be able to resolve their disagreement without outside assistance. We then explain that air pollution in Mexico City is really the same problem, albeit much larger. We show that the problem of pollution has two related aspects: (1) we cannot easily force polluters to pay for their "use" of clean air, and (2) as a result, there is a divergence between the cost of an action to an individual polluter and the cost to society as a whole.

We discuss different kinds of policies that are used to address these problems. Then we turn to our use of natural resources such as oil. We consider various kinds of resources and consider what economic theory can teach us about how these resources are likely to be used. Finally, we consider the implications for economic policy.

[1] See http://www.worstpolluted.org. As a child, one of the authors of this book lived for a year in one of these towns.

[2] Blacksmith Institute, "The World's Worst Polluted Places," September 2007, accessed March 14, 2011, http://www.blacksmithinstitute.org/wwpp2007/finalReport2007.pdf.

[3] Jonathan Klein, "Why People Really Buy Hybrids," *Topline Strategy Group*, accessed January 31, 2011, http://www.toplinestrategy.com/green_form.htm. Of course, people can have multiple motivations for purchase. The same study concludes that only about 27 percent of hybrid users do not have a financial motivation for their purchase.

13.1 The Economics of Clean Air

LEARNING OBJECTIVES

1. What is the Coase theorem?
2. Why is the Coase theorem important?
3. What is a social dilemma?

If we lived in a world where all economic transactions took place in competitive markets and in which there were "enough" markets, then we would obtain all the possible gains from trade. This logic falls down in reality because markets sometimes fail, for various reasons.

- Something prevents the economy from reaching the competitive outcome.
- Something prevents trade altogether, so the market is missing.
- Some people other than the buyer and seller are affected by the transaction.

We will get to Mexico City shortly. We begin, however, by thinking about a more isolated case of air pollution: cigarette smoke in an office.

Smokers, Nonsmokers, and the Coase Theorem

Cigarettes are sold and smoked almost everywhere. Yet in most countries around the world, you are not able to smoke when and where you please. Governments around the world place limitations on who can buy cigarettes, where they can be bought, and where they can be consumed. From an economic point of view, governments are deliberately restricting the ability of individuals to engage in voluntary transactions. [1] Why do governments restrict an individual's ability to smoke where and when that person wants?

Our answer to this question begins by imagining two people who must share an office. One is a nonsmoker who dislikes the smell of cigarette smoke, while the other likes to smoke while working. On the way to work one day, the smoker purchases a pack of cigarettes that she plans to smoke at work. We can reasonably deduce from this that her valuation of these cigarettes is greater than the price she has to pay. She gets buyer surplus from the purchase. To be concrete, suppose a pack of 20 cigarettes costs \$4 and her valuation of a pack of cigarettes is \$10. Her surplus is then \$6.

We can also reasonably assume that the seller's cost is less than the price, otherwise he would not choose to make the sale. He gets seller surplus. For example, if his wholesale price for a pack of cigarettes is $2, then he earns $2 surplus (= $4 – $2) on every pack that he sells. The total surplus from the sale is the buyer surplus plus the seller surplus—that is, $8. So far so good.

The problems begin when the worker smokes her pack of cigarettes in the office. She obtains her $6 worth of enjoyment. However, a third party has now been affected by her decision to consume cigarettes: her office mate. The office mate dislikes the smell of smoke and may even face health risks from second-hand smoke. Thus even though the smoker and the store that sold the cigarettes are both better off, the office mate has been made worse off.

We should not automatically assume that the best thing is to ban smoking in the office just because the office mate is adversely affected. We need to know *how much* the nonsmoker is inconvenienced. Suppose the most the nonsmoker would be willing to pay for a smoke-free office is $2 per day. In this case, the $6 gain to the smoker exceeds the $2 loss to the nonsmoker. It seems like it should be easy enough for the two individuals to find a way for everyone to be happy. For example, imagine that the smoker agreed to pay the nonsmoker $4 a day for the right to smoke in the office. Both would then get $2 surplus per day.

On the other hand, suppose a smoke-free office is worth $10 per day to the nonsmoker. In this case, his valuation of clean air ($10) exceeds the smoker's gain ($6). The smoker would be unwilling to pay the nonsmoker enough to compensate for dirtying the air in the office. It would be better not to allow smoking in the office.

We have assumed here that the default situation is that the office should be smoke-free. In the language of economics, the nonsmoker owns theproperty rights to the clean air in the office. Property rights over a resource mean that, by law, the owner can make all decisions regarding the use of the resource. Because of this, the smoker must pay the nonsmoker compensation if she wishes to be allowed to smoke in the office.

We could imagine the opposite situation, where the smoker starts off with the right to smoke in the office. Would we expect a different result from their negotiations? If the smoke-free office was worth only $2 to the nonsmoker, then he would not be willing to pay enough to persuade his office mate not to smoke: the most he would pay is $2, which is less than the smoker's surplus. If, on the other hand, the nonsmoker valued the smoke-free office at $10, then he values a smoke-free office more than the smoker values smoking in the office. The nonsmoker could pay the smoker not to smoke. For example, imagine he pays her $8 per day to not smoke. He pays $8 for the clean air, which is worth $10 to him, so he gets $2 of surplus. The smoker receives $8, which exceeds the surplus she would get from smoking in the office. Again, they would both be happy.

Thus if the smoker "owns" the clean air, the nonsmoker must pay the smoker if he wants a smoke-free office. If the nonsmoker has the property rights, it is the smoker who must pay. In either case, the basic outcome will be the same: there will be smoking in the office if the smoker's valuation exceeds the nonsmoker's valuation; there will be no smoking if the nonsmoker's valuation exceeds the smoker's valuation. But the property rights are valuable. It is the owner of the property rights—whoever that may be—who gets compensation from the other.

As long as they know who has the property rights, it seems likely that the two individuals will be able to come to an agreement that benefits them both. You might imagine, though, that they would find it far harder to come to an agreement if it was ambiguous who had the property rights in the first place. The smoker would likely claim the right to smoke in the office, while the nonsmoker would assert his right to clean air. If they could not settle this basic question, it is unlikely that they would be able to reach a more complicated agreement involving compensation payments.

Let us imagine a further twist. Suppose the nonsmoker has property rights but values clean air at only $7 per day. This is still greater than the smoker's surplus, so, as before, we expect that they would agree to a smoke-free office. But the nonsmoker's valuation for clean air is *less* than the total surplus of the smoker and the store that sold her the cigarettes (recall that the smoker gets $6 surplus and the store gets $2 surplus). If the storekeeper, the smoker, and the nonsmoker all got together, they should again be able to find an arrangement that benefits everyone. For example, the storekeeper and the smoker could jointly give the nonsmoker $7 and still have $1 of surplus to bargain over.

It seems perfectly reasonable to imagine that two people who share an office could come to a mutually beneficial agreement about smoking in the office. It seems much more far-fetched, though, to imagine that they would come to an agreement together with the storekeeper who sold the cigarettes. Economists say that the difference between the two cases is due to transaction costs—the costs of making and enforcing agreements.

We began this section by observing that a transaction may affect not only the buyer and the seller but also third parties. When this is the case, we cannot be sure that trade benefits everyone. Even if the buyer and the seller are both made better off, third parties may be made worse off. However, if property rights are clearly established and transaction costs are low, then we can expect that private negotiations could solve these problems. This idea was first articulated by the Nobel prize–winning economist Ronald Coase (http://www.coase.org) and is called the Coase theorem: if property rights are clearly established and transaction costs are low, private bargaining will lead to efficient outcomes.

It is notable that in reality, we do not see office workers buying and selling the right to smoke in an office. Instead, blanket bans on smoking have been enacted throughout the United States and in many other countries throughout the world. Why do we get this government response? One argument for these bans is that smoking poses health risks. In this case, antismoking campaigns are based on an idea that individuals are not always capable of making good choices for themselves. [2] But another reason is a recognition of the transaction costs involved in these private negotiations. Even if you think two coworkers could reach an agreement, imagine an office with 10 people—perhaps there are 3 smokers and 7 nonsmokers—all of whom place a different valuation on clean air in the office. If we knew everybody's true valuations, then in theory it would be possible to create a system of payments that made everybody better off. In practice, however, people might lie about their valuations. Finding the right system of payments would be very hard indeed and would take a lot of time and effort.

Perhaps you can now see the parallel with Mexico City. The major pollution problem is the emissions from the cars that people drive. Individual residents of Mexico City make decisions to buy gasoline and drive. These transactions create value for the drivers and the sellers of gasoline—but third parties are adversely affected. The Coase theorem can work when the parties involved are easily identifiable and small in number. In contrast, it is impossible to imagine the 20 million residents of Mexico City all meeting and coming to some kind of private agreement to limit their collective driving behavior.

Social Dilemma

Air pollution and second-hand smoke have a common structure, which we explain in this section. Once you understand these common elements, you will probably be able to think of many other examples. We show the interactions between an individual and the rest of society in Table 13.1 "The Payoffs in a Social Dilemma Game". This is called a social dilemma game.

Table 13.1 The Payoffs in a Social Dilemma Game

	Everyone Else Drives (Air Is Polluted)	**Everyone Else Takes Public Transportation (Air Is Clean)**
You Drive	\$0	\$2
You Take Public Transportation	−\$1	\$1
Regardless of which action you choose, your payoffs are higher if everybody else takes public transportation (\$2 > \$0; \$1 > −\$1). Regardless of the actions of others, your payoffs are always higher if you drive (\$0 > −\$1; \$2 > \$0).		

In Mexico City, it is very possible that people would agree that they would prefer a situation where everybody drives less. Yet this is not what happens. To see why, look at Table 13.1 "The Payoffs in a Social Dilemma Game" and imagine you are a car owner in Mexico City. You have to decide how to get to work—by driving or taking public transportation. There are two rows in the table. One is labeled "You Drive," and the other is labeled "You Take Public Transportation." The rows represent your possible choices. The columns refer to everybody else's choices. Everyone else who owns a car similarly chooses between driving and taking public transportation. To keep things simple, we suppose that everyone else makes the same choice. If everybody chooses to drive, then the air is polluted. If everybody chooses to take public transportation, then the air is clean. The current situation in Mexico City is that you and all other car owners are driving to work. You enjoy the convenience of driving rather than taking public transportation, but you suffer from the polluted air.

The numbers in the table refer to your payoffs from the different possible combinations. As in our smoking example, we can think of these as the valuations per day that you place on different outcomes. What matters is how these different possibilities compare with the status quo, where you and everyone else drive. We therefore begin by setting your payoff at the status quo (the number in the top left cell) at \$0. Suppose also that the following are true.

- You value clean air rather than dirty air at \$2 per day—that is, you would be willing to pay \$2 per day to have clean air rather than dirty air.
- After taking into account the relative costs and inconveniences of driving versus public transportation, you value driving compared to public transportation at \$1 per day—that is, you would need to be compensated \$1 per day to make you just as happy to take public transportation rather than drive.

  Based on these conditions, we can calculate the payoffs in the other three cells of the table:
- The top right cell is the case where you drive and everyone else takes public transportation. Compared to the status quo, you are better off because you get to enjoy clean air. This is worth \$2 to you, so your payoff is \$2.
- The bottom left cell is the case where you take public transportation and everyone else drives. You still have to breathe polluted air, and you no longer get the convenience of driving (which is worth \$1). Compared to the status quo, you are worse off. Your payoff is −\$1.
- The bottom right cell combines the two previous cases. In this case, you and everyone else take public transportation. You get the benefit of clean air (worth \$2), but you lose the benefit of driving (worth \$1). Your payoff is \$2 − \$1 = \$1.

What would you do in this situation? Suppose you think that everyone else is going to drive. You are better off if you drive (payoff is \$0) rather than take public transportation (payoff is −\$1). What if everyone else takes public transportation? Then you still prefer to drive (payoff is \$2) rather than take public transportation (payoff is \$1). We conclude you will drive regardless of what others in society choose to do.

Here is the crux of the problem: *the situation looks the same to everybody else as it does to you*. As you evaluate these relative payoffs and choose to drive, so too does everyone else. We therefore expect that everyone will follow their individual incentives and choose to drive. We end up in the top left cell, where your payoff is $0.

What is striking, though, is that you would prefer the outcome where everybody—including you—uses public transportation. Your payoff in the bottom right cell is $1, which is better than the current situation. Everyone else would prefer this outcome as well. Society ends up in a bad situation, with everybody driving, even though everyone agrees that there is a better option out there. This is the essence of the social dilemma.

You are one of many people. Although you may value clean air, you are powerless as an individual to keep it clean. And because you are only one person, your decision has a tiny effect on the overall quality of the air. Thus you choose to drive, without paying attention to your effect on the environment. But because everybody makes the same decision, the cumulative effect is that there is a lot of air pollution.

The social dilemma is also sometimes known as the "tragedy of the commons." This refers to the time when cattle farmers had access to common grazing land. Because every farmer had the right to graze his cattle on this land, no one was in a position to ensure that the land was well managed. Every farmer paid attention only to the health of his own cattle and did not worry about the effect of his cattle on the overall quality of the grazing land. Because every farmer made the same decision, the result was overgrazing, which destroyed the land for everyone.

How do we solve problems such as this? To avoid the bad outcome of the social dilemma, we must find some way of changing the payoffs of the game. We have already seen that, if transaction costs are low, people may be able to negotiate privately. In the case of Mexico City smog, however, such negotiation is impractical. In this case, one possible solution is for the government to alter the payoffs. For example, in the run-up to the 2008 Olympics in Beijing, the Chinese government wanted to improve air quality in the city. It therefore allowed cars to drive in the city only every other day (whether a car was permitted on a given day depended on whether the last digit of the license number was odd or even). If you tried to drive on the wrong day, you might have to pay a large fine, so the payoff to driving changed. We have more to say about government policy later in the chapter.

Toolkit: Section 17.18 "Nash Equilibrium"

You can review the social dilemma and other games in the toolkit.

Private Benefits and Social Costs

When people face a social dilemma, the actions that are the best for all individuals lead to an outcome that is bad for everyone. Much of the rest of the chapter addresses why this happens.

We begin by remembering our theory of how people make consumption choices. In general, people consume a good up to the point where their marginal valuation from the last unit of that good equals the price of that unit. This is a specific statement of a more general principle for decision making: “consume until marginal benefit equals marginal cost.”

In the social dilemma of Table 13.1 "The Payoffs in a Social Dilemma Game", you had two choices: driving or not driving. Let us now expand that to think about a situation where you are deciding how much to drive. Driving your own car brings private benefits (that is, benefits that are obtained only by you), such as comfort and convenience. Driving also brings private costs, such as the costs of gasoline and maintenance of your vehicle. If this were all that was going on, there would be no problem: you would drive up to the point where your marginal (private) benefit from driving was equal to your marginal (private) cost of driving.

When you drive, though, you also impose costs on other people. Your decision to drive one more mile has a marginal social cost as well as a marginal private cost. Marginal social cost is the cost to society of consuming or producing one more unit of a good or a service. By “society,” we simply mean “you and everybody else.”

When you choose to drive your car, you contribute to air pollution. This is a cost to the rest of society. However, you have no incentive to worry about this. You care only about the extra cost to you. The same is true when the smoker smokes one more cigarette in the office: She imposes a cost on her office mate, but—unless they make an agreement otherwise—she does not pay this cost. She takes into account the marginal private cost to her (that is, how much she must pay for one more cigarette), but she ignores the cost to other people.

Ignoring for a moment some tricky questions about how to measure the cost of your actions to society, we can set out a principle for how much you *should* consume if your concern were the overall well-being of society: "consume until marginal benefit equals marginal social cost." The marginal social cost of your driving is the extra cost if you drive more. It is the cost both to you *and* the cost that you impose on others in society. There can be many components to this cost. One component is the marginal private cost to you. In addition, you pollute the air, contributing to public health problems. You add greenhouse gases to the atmosphere, contributing to the risks of climate change. You make the roads more congested, thus wasting the time of other drivers on the roads. You cause wear and tear on the roads, which will ultimately be paid for by taxes on all drivers.

You might be puzzled at this point. If you drive one extra mile, does that really have any appreciable cost on society? The extra emissions from your driving obviously have a tiny effect on pollution and greenhouse gases. The wear and tear you impose on the roads is minimal. How can these minuscule effects possibly matter? The first and more obvious answer to this question is that the quality of the atmosphere and the roads is affected by everybody's decisions—not only yours. Hundreds of thousands of cars around the world are polluting the atmosphere and damaging the roads. The second, more subtle, answer is that though your individual influence on air and road quality is very small, you are affecting a very large number of people. If you drive an extra mile in Mexico City, you are affecting the air that is breathed by 20 million people. The marginal social cost of your driving includes the effect on every single one of these people. Likewise, one more mile of driving may add only a tiny amount to the greenhouse gases in the atmosphere, but that tiny effect must be added up over the entire population of the world.

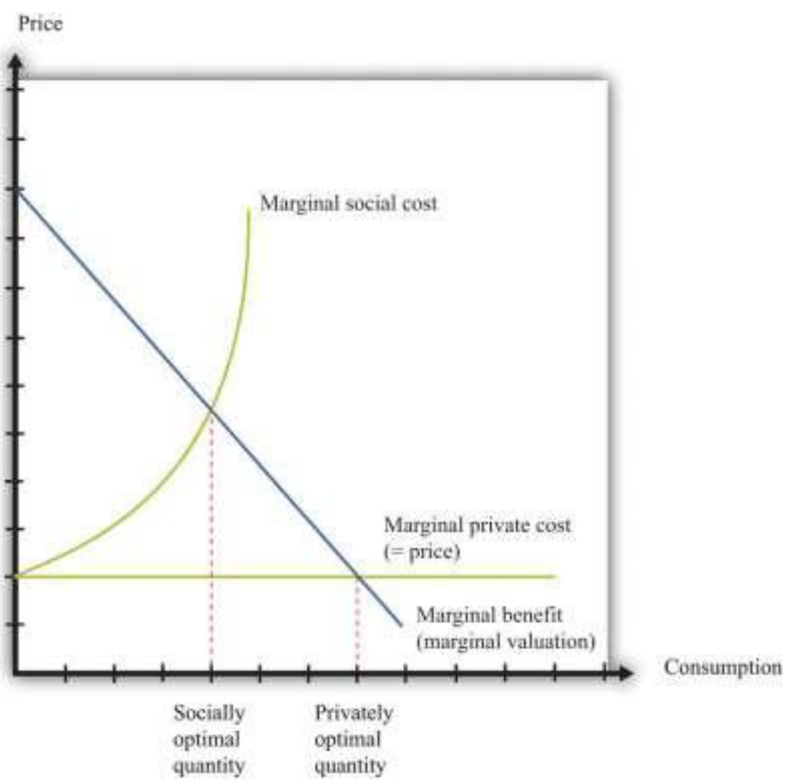
As a result, the marginal social cost of your driving is greater than your marginal private cost. Figure 13.3 "A Divergence between Marginal Private Cost and Marginal Social Cost" shows the implications of this. Think of consumption in this diagram as referring to the amount of driving you do. The marginal private cost of your driving is the cost of fuel, depreciation of your car, and so on. But the marginal social cost also includes the pollution of the air and the congestion of the roads. Because the marginal social cost is greater than the marginal (private) cost, you will drive too much, from the perspective of society as a whole. This is exactly what we saw in the social dilemma. People choose to drive and pollute the air, even though all members of society could be happier if everyone were to take public transportation and generate less pollution.

The gap between private costs and social costs means that too much driving is undertaken, from the perspective of society as a whole. The outcome is inefficient because people only have an incentive to take account of the private costs of their actions.

Toolkit: Section 17.11 "Efficiency and Deadweight Loss"

You can review the concept of efficiency in the toolkit.

Figure 13.3 A Divergence between Marginal Private Cost and Marginal Social Cost



The gap between marginal private cost and marginal social cost is a measure of the impact that an individual has on the rest of society.

Figure 13.3 "A Divergence between Marginal Private Cost and Marginal Social Cost" also illustrates another point. Economic analysis tells us that there is "too much" pollution, from a social point of view. This observation probably comes as no surprise. Economic analysis also makes it clear, though, that it is possible to have too little pollution as well as too much. There is an optimal amount of driving for each individual, to be found where marginal social cost and marginal benefit are equal. At this amount of driving, there will be some pollution: the optimal amount of pollution for society as a whole. If we were to ban driving altogether, we would have less pollution, but we would also lose all the benefits from driving.

Our discussion here has been about decisions of consumers, but firms also are sources of pollution. Firms use trucks and other vehicles that, like cars, impose costs on the rest of society. Some firms pollute the air or the water. Exactly the same principles still apply. Any individual firm has no incentive to take into account the costs that it imposes on the rest of society. As a result, firms pollute too much from a social point of view.

KEY TAKEAWAYS

- The Coase theorem states that if property rights are well defined and transaction costs are low, then bargaining will lead to an efficient outcome.
- The Coase theorem provides the rationale for a market solution to pollution and other similar social problems.
- A social dilemma arises when there are many individuals each making choices that are in their self-interest but leading to an outcome that is bad for society.

CHECKING YOUR UNDERSTANDING

1. Why does the Coase theorem require that transaction costs be low?
2. Consider Table 13.1 "The Payoffs in a Social Dilemma Game". Suppose your payoff from taking public transportation when everyone else drives is −$2 rather than −$1. Would this change the basic message of this example?
3. Look at Figure 13.3 "A Divergence between Marginal Private Cost and Marginal Social Cost". How would you modify this figure to make the difference between social and privately optimal quantities larger? Explain your reasoning.

4. Suppose you are in the library and there are two people making out between the shelves. Describe this situation in terms of social versus private costs. How would you use the Coase theorem to find an efficient allocation?

[1] There are many different ways in which governments intervene in market transactions. Chapter 11 "Barriers to Trade and the Underground Economy" contains more discussion.

[2] Chapter 4 "Life Decisions" has more to say about whether individuals are good judges of their own actions, particularly when making decisions with long-term consequences.

13.2 Externalities

LEARNING OBJECTIVES

1. What is an externality?
2. What are the ways in which problems caused by externalities can be solved?
3. What are some of the difficulties in designing policies to deal with externalities?

At the heart of the social dilemma is a divergence between private costs and social costs. Individuals and firms take into account only private costs when making decisions. But the social costs should matter as well. Thus actions that are individually optimal are damaging to society as a whole. Now that we have diagnosed the problem, how do we fix it? The economist's answer is that we must change people's incentives. A social dilemma arises when individual incentives are not well aligned with the interests of society as a whole. Economic policies focus on how to adjust those incentives so that there is a better match between individual and social aims.

Before discussing these policies in detail, we look again at the problem of the social dilemma, focusing now on the *actions* that people choose to take. In our Mexico City example, people decide whether or not to drive. If they choose to drive, this action affects the well-being of others. Economists say that there is an externality associated with the action of driving.

Toolkit: Section 17.19 "Externalities and Public Goods"

An externality occurs when one person takes an action that directly affects another's welfare, but the effect does not operate through prices.

An externality must come from an *action*—something that somebody does. Good weather is not an example of an externality. Nor is an earthquake. The action could be taken by an individual (say, smoking a cigarette) or a firm (dumping toxic waste into a river). In most cases, the action is associated with production by a firm or consumption by a household.

In addition, the action must directly affect another individual's well-being or a firm's profits. It could be something that affects the health or happiness of an individual. It could be something that affects the profits of a firm. ("Directly" here means that the effect doesn't come about because of an induced change in behavior. Suppose, for example, that a firm offers you a job, but to get to that job you now must spend a longer time commuting. The extra commute is *not* an externality imposed on you by the firm.)

Finally, the effect must not operate through prices. Whenever we take part in market transactions, we have effects (usually tiny effects but effects nonetheless) on market prices. These changes in prices make others in the market better or worse off. But they are not externalities.

In our earlier example of driving, the marginal social cost was larger than the marginal private cost. The gap between the two is a measure of the size of the externality. Because the action of driving imposes a cost, we call this anegative externalities. Pollution is the classic example of a negative externality, but there are others. Congestion of public roads or public parks is another instance of a negative externality.

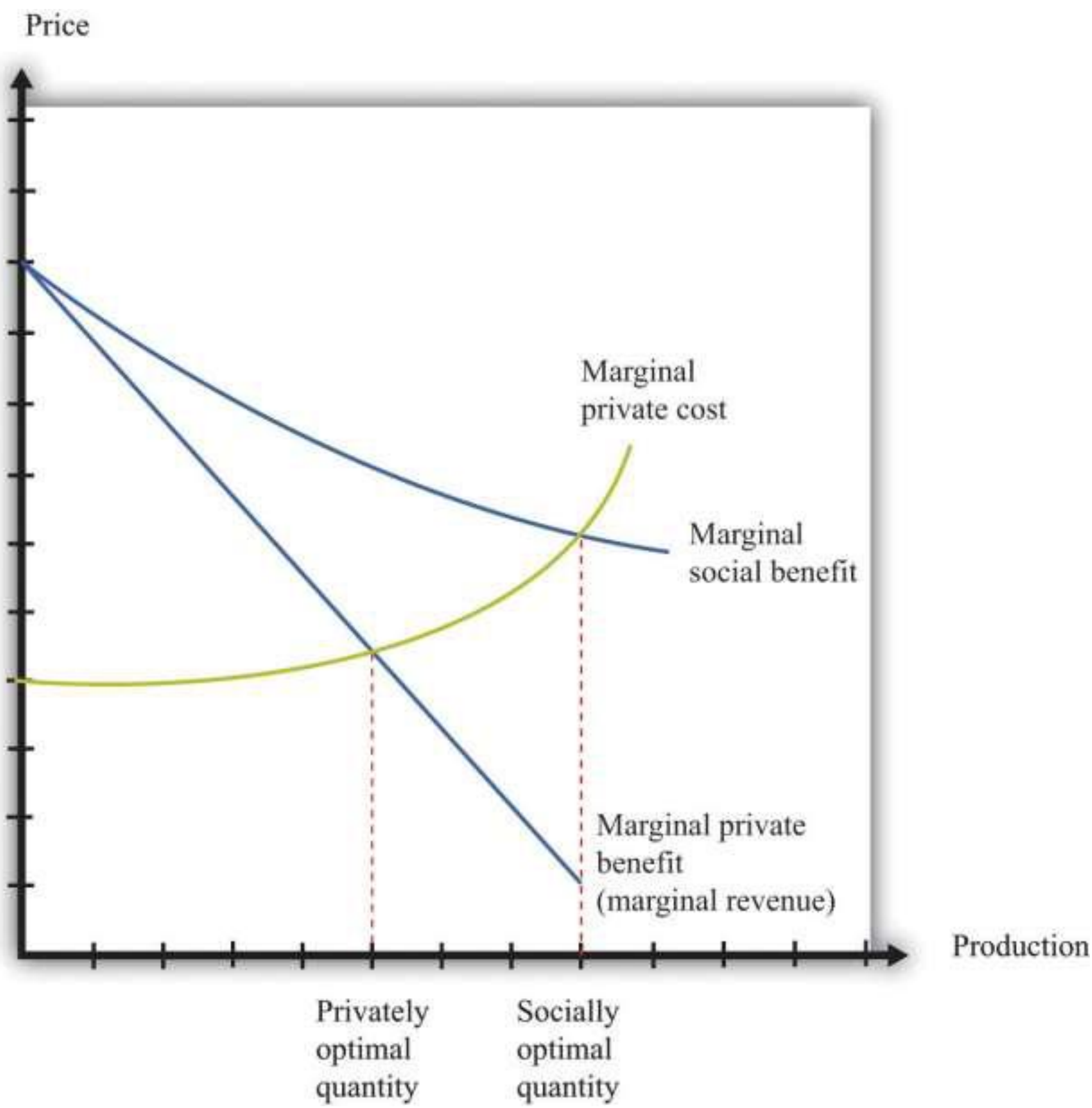
By contrast, there are also occasions when an action bestows an external benefit on third parties. We call this a positive externalities. For example, writers of open-source software create a social benefit that is in excess of the private benefit that they personally obtain. As another example, suppose that a firm engages in research and development and creates new knowledge. If others are also able to benefit from that knowledge without paying for it (for example, after the expiry of a patent), they are beneficiaries of a positive externality.

Toolkit: Section 17.19 "Externalities and Public Goods"

- A positive externality arises when one person's or firm's action bestows benefits on others. When there is a positive externality, too little of the action is undertaken.
- A negative externality arises when one person's or firm's action imposes costs on others. When there is a negative externality, too much of the action is undertaken.

Although negative externalities sound bad and positive externalities sound good, positive externalities are also a source of inefficiency. The logic exactly parallels the case of negative externalities. Suppose a firm is deciding how much output to produce. To maximize its profits, it sets marginal cost equals to marginal private benefit (that is, marginal revenue). But if the firm's production generates a positive externality, the marginal social benefit exceeds its marginal private benefit. The firm produces insufficient output from a social point of view, as illustrated in Figure 13.4 "A Divergence between Marginal Private Benefit and Marginal Social Benefit". The principle for socially efficient production is for the firm to produce up to the point where marginal cost equals marginal social benefit.

Figure 13.4 *A Divergence between Marginal Private Benefit and Marginal Social Benefit*



From a social point of view, however, the firm should produce up to the point where marginal social benefit equals marginal cost.

Solutions to Externality Problems

The definition of an externality makes it clear that the fundamental problem is one of behavior—actions by a firm or a household. The behavior reflects a difference between private costs or benefits and social costs or benefits. These observations also point us to a solution. We need to change incentives so as to align private costs or benefits and social costs or benefits. For example, if the private marginal cost of pollution to a firm were somehow equal to the social marginal cost, then a firm acting in its own self-interest would produce the socially optimal amount of pollution. The challenge for policymakers is to find a way to adjust the incentives so that the firm takes into account social marginal costs in addition to private marginal costs.

From this perspective, inefficiency arises because there are no market signals that force the polluter to take into account how its actions are affecting others. The goal of government policy in the presence of externalities is to provide incentives for firms and households to *internalize* their effects on others. These policies include direct restrictions on what people can do (for example, banning smoking in public buildings), taxes and subsidies that affect prices in an economy, and the introduction of markets that force polluters to pay for the right to pollute. Because externalities involve a divergence between private costs and social costs (or private benefits and social benefits), the goal in all cases is to adjust the incentives so that the actor internalizes the externality.

Creating Markets

We said that externalities are a source of inefficiency, but we should be more precise: externalities are a source of inefficiency *unless they are compensated for*. Think back to the smoker and nonsmoker who shared an office. The smoker's actions impose a negative externality on the nonsmoker. Without any compensating payments, we end up with an inefficient outcome. But when the smoker pays the nonsmoker for the right to use up the clean air, we end up with an efficient outcome. Negotiation between the smoker and the nonsmoker in effect creates a market for the clean air. Once this market is in place, the inefficiency disappears.

Building on this insight, governments can actively try to create markets to solve pollution and other externality problems. A good example of this is the 1990 Amendments to the Clean Air Act in the United States. Much of the air pollution in the United States is caused by utility companies (think of power stations), particularly those that generate electricity from coal. Such power stations pump sulfur dioxide into the atmosphere, which causes acid rain and other environmental problems. The amendments to the Clean Air Act created tradable emission permits that were allocated to utility companies.

Such permits are licenses to emit a specified amount of pollution. A firm must own or purchase a permit if it wishes to emit pollutants into the atmosphere. These permits can be traded in a market. A firm that wishes to emit more pollution than allowed by its existing permits can purchase permits from others. A firm with more permits than it needs can sell them to other firms.

The first response of many people to a policy such as this is moral outrage. At first hearing, it may seem odd that the government is granting a license to pollute. In fact, this can be a very effective way to control pollution. To see how such a system works, suppose that there are two power stations.

1. GreenPower has installed pollution reduction measures and is emitting 100,000 tons of sulfur dioxide per year. Because it is already an environmentally friendly power station, it is very costly for it to reduce emissions further. To be concrete, suppose it will cost $20 per ton to reduce its emissions.
2. Atmosfear has no pollution control devices in place and is currently emitting 300,000 tons of sulfur dioxide per year. Because it has not yet installed any pollution reduction devices, it is able to reduce its emissions relatively cheaply, at a cost of $10 per ton.

Suppose the government decides that it wants to restrict the amount of pollution to a total of 200,000 tons per year, down from the current 400,000. It doesn’t matter which power station emits the pollution; either way, the sulfur dioxide ends up in the atmosphere.

One approach is for the government to simply instruct each power station to cut its emissions by 50 percent. The trouble with this is that GreenPower already has an environmentally friendly system in place. It will cost $1 million (= 50,000 tons × $20 per ton) to reduce its emissions from 100,000 tons to 50,000 tons. Atmosfear has to reduce its emissions by 150,000 tons, which costs $1.5 million (= 150,000 tons × $10 per ton). The total cost of reducing emissions to meet the target is $2.5 million.

There is a better alternative. Suppose the government gives each power station a license to emit a certain quantity of pollution. For example, suppose it gives GreenPower a license to emit 50,000 tons and Atmosfear a license to emit 150,000 tons. So far, this is identical to the previous situation. Crucially, though, the government also allows the power stations to trade these licenses. If GreenPower can buy the right to emit sulfur dioxide for less than $20 per ton, it will prefer to do this rather than reduce its own emissions.

Because it costs Atmosfear only $10 per ton to reduce its emissions, the cheapest way to achieve a 200,000-ton reduction is for Atmosfear to carry out the entire reduction in emissions, down to 100,000 tons. If it does so, then it will have 50,000 unused permits that it can sell to GreenPower. The total cost of emissions reduction in this scenario is only $2 million. For example, suppose they agree on a price of $15 per permit. Then the cost to GreenPower is $50,000 × $15 = $0.75 million (instead of $1 million). The cost to Atmosfear is $1.25 million: $200,000 × $10 = $2 *minus* the $0.75 million it collects from GreenPower.

In some ways, this is like our smoking example. The power stations are able to trade, so they can both be better off. Emissions are reduced to the required level of 200,000 tons, and this reduction is achieved in the most cost-effective manner. Notice also that the government gets to decide on the total amount of acceptable pollution; the approach is consistent with very tight or very lax environmental standards. However, once the government has determined its desired level of pollution, the trading of permits allows the necessary reductions in pollution to be achieved at the lowest possible cost to society.

With tradable permits, pollution is controlled in the most efficient way, without regulators' needing detailed knowledge on different power stations. The trading system has some other advantages as well. Firms have an incentive to pollute less because they can sell excess permits for a profit if they do not need to use them. Environmental groups can even purchase emissions permits and take them out of circulation to reduce pollution below the level mandated by the government.

A more contentious question has to do with how the permits are allocated in the first place. One approach is to do what we did in our example: the government can simply allocate permits based on the existing level of emissions. A problem with this policy is that it effectively punishes firms that have already engaged in environmentally responsible changes. It also creates an incentive for lobbying by firms: because permits are valuable, firms will invest resources in trying to persuade policymakers to give the permits to them rather than to their competitors. Instead of giving away the permits, the government can instead auction them. This increases costs for the firms but has the advantage that it generates funds for the government. [1]

Taxes and Subsidies

Remember that the problem with externalities is that private incentives do not reflect social costs and benefits. Another approach to fixing these incentives is through taxes and subsidies.

Consider first the case of a negative externality. The problem in this case is that the marginal social cost exceeds the marginal private cost. As we saw in Figure 13.3 "A Divergence between Marginal Private Cost and Marginal Social Cost", this leads to overconsumption of a good. One way to fix the problem is to impose a tax that equals the difference between the marginal private cost and the marginal social cost. In effect, this converts the condition for private optimality (consume until marginal benefit equals marginal private cost) into the condition for social optimality (consume until marginal benefit equals marginal social cost).

The case of a positive externality is entirely analogous. The problem is that the marginal social benefit exceeds the marginal private benefit. As we saw in Figure 13.4 "A Divergence between Marginal Private Benefit and Marginal Social Benefit", this leads to underproduction of a good. One way to fix the problem is to impose a subsidy that equals the exact difference between the marginal private benefit and the marginal social benefit. This converts the condition for private optimality (consume until marginal private benefit equals marginal private cost) into the condition for social optimality (consume until marginal social benefit equals marginal cost).

We see such policies in practice. Taxes on gasoline exist in part to compensate for externalities such as the pollution caused by automobiles and the wear and tear on roads. Subsidies to universities and think tanks exist in part to encourage the production of knowledge, which is a good with positive externalities.

Command and Control

There are also other kinds of environmental policies. The government can simply mandate that certain levels of pollution must not be exceeded. In some cases, such command and control regulation may be easier to implement and monitor.

Under most circumstances, economists favor either taxes or the creation of a permit market to command and control. The reason is that command and control is relatively inflexible and requires a lot of knowledge of how much pollution is generated by each individual firm. Taxes, subsidies, or permit markets are more flexible and do a better job of changing the incentives faced by polluting firms.

Encouraging Altruism

A rather different approach to externalities is to appeal to people's altruism. In economics, we typically assume that people consider only their own self-interest when making decisions. Moreover, we usually think of this self-interest in fairly narrow terms. This may seem a rather embittered view of human nature. People often do things that are directed toward others' happiness rather than their own—that is, people are sometimes altruistic rather than selfish. In the environmental context, people sometimes purchase products from environmentally responsible firms even if those products are more expensive. People sometimes purchase carbon offsets to compensate for the carbon generated by their driving or air travel. Companies often find it worthwhile to advertise the fact that they are environmentally responsible.

Governments also provide environmental information. The European Union has recently proposed measures requiring that new cars display their impact on global climate change, just like health warnings on cigarettes. [2] Unlike health warnings on cigarettes, this labeling does not help people live healthier lives. And unlike efficiency ratings on appliances, this labeling does not help people make better decisions that will save them money. Climate change labels appeal purely to people who want to act in a way that will lessen their environmental impact. As we saw earlier in the chapter, purchases of hybrid cars seem to be primarily motivated by people's desire to be more environmentally responsible.

Psychologists and economists have studied why people sometimes behave in such ways. Perhaps it is because they are genuinely altruistic—they care about the well-being of others as well as themselves. Perhaps it is because of persuasion and peer pressure—if your friends behave in an environmentally conscious way, then there is pressure for you to do the same thing. Perhaps it is because of the feeling of satisfaction (sometimes called a “warm glow”) associated with such behavior. For our purposes in this chapter, however, the exact reasons why people behave this way are less important than the behavior itself. There are several ways of thinking about altruistic behavior, but in the environmental context, altruism really amounts to people deciding to internalize some of the externalities that they impose. They incorporate some of the social cost into their own private costs. Governments, through advertising campaigns, can encourage such altruistic behavior.

Difficulties with Environmental Policy

Our discussion of externalities makes it seem that it is straightforward for policymakers to design effective environmental regulation. Government regulators simply need to calculate the difference between marginal private costs and marginal social costs and between marginal private benefits and marginal social benefits. Then they can put in place the appropriate taxes, subsidies, or both.

In practice, a major difficulty is knowing how to place values on externalities. Environmental policies to combat air or water pollution require the government to monitor the amount of emissions effectively and accurately. If emissions cannot be monitored, then tax or permit schemes are impossible to implement. Effective environmental policies also require the government to measure the damage incurred by the victims of the pollution.

As with many economic policies, questions also arise concerning the distribution of resources. In an environmental context, the key point of debate is often whether compensation should and will be paid. If a factory pollutes the river running through a town, imposing negative externalities on the town’s residents, then is it enough to adjust the incentives so that there is the “right” amount of pollution—that is, so that the marginal benefit to the firm equals the marginal social cost? Or should the firm also be required to compensate the residents of the town? This is again a question of property rights, for we are really asking who has the initial right to the clean water in the river. It is also closely related to the question of whether pollution permits should be given away (implying that firms have the property rights) or sold at auction.

Whenever government steps in and enacts policies to affect behavior, it also must worry about whether there will be perverse responses to those incentives. As an example, the disposal of solid waste is a significant environmental problem in many countries. It is in part an economic problem: people do not usually pay directly for the removal of their garbage, so they do not have an incentive to recycle or avoid waste in other ways. A solution—favored by many economists and adopted by some municipalities—is to charge people a fee per bag for garbage removal. A study by Don Fullerton and Tom Kinneman revealed that this policy can bring its own problems. [3] When such a scheme was introduced in Charlottesville, Virginia, citizens responded by putting much more garbage in each bag than they used to (Fullerton and Kinneman called this the "Charlottesville stomp"), so the reduction in the waste stream was lower than the authorities had anticipated. More seriously, some people also responded by dumping their trash illegally.

Valuing the Environment

The measurement of environmental harm is complicated. Suppose, for example, that pollution brings with it increased risk of disease or death. How do we place a cost on these risks? We can perhaps attempt to value health in terms of the costs of treatment and lost working hours, but it is much harder to place a value on the distress and suffering caused by ill health. Economists, lawyers, and others have even come up with varying ways to place a dollar value on human life, but—as you can surely imagine—these techniques are contentious. One reason is that all such approaches tend to rely, at least in part, on estimates of lost earnings. This means that the lives of skilled and high-paid individuals may end up being valued more than the lives of unskilled, lower-paid individuals.

A second set of issues has to do with how we value damage to the natural environment. Take, for example, the oil leak in the Gulf of Mexico in 2010, which caused substantial harm to birds, fish, and ocean ecosystems. Most of us are upset by the sight of seabirds with their feathers clogged with oil. But how should we assess the value of such damage? Similarly, what is it worth to ensure the survival of a particular endangered species?

Where possible, economists look to market prices to provide some indication of the value that society places on goods and services. In the case of environmental goods, though, we typically cannot look to markets. In such cases, we may need to use surveys and other methods for inferring household valuations. This is known as contingent valuation.

For several reasons, it is very difficult to carry out reliable contingent valuation surveys. In such a study, a household might be asked, "What would you be willing to pay to ensure that the whooping crane (for example) does not go extinct?" Such surveys often give implausible answers. For example, following an oil spill, people in Washington state and the province of British Columbia were supposedly willing to pay over $11,000 for each seabird that was saved—even though the seabird population would recover naturally in a decade or so.

The main problem, as you can perhaps guess, is that people do not face a real budget constraint when asked such questions, so they have no particular incentive to give a truthful answer. A related issue is that people are typically presented with an issue in isolation. A household that claims to be willing to spend $50 to save one species from extinction might not be willing to spend $10,000 to save 200 species at a cost of $50 each. A third issue is that people may not have the information they need to make good decisions. We are not used to making purchases of environmental quality, so it is harder for us to give our valuation of a clean river than it is for us to give our valuation of, say, a bar of soap.

Yet, for all these objections, we do need some way of placing a value on environmental resources. The fact that we find it difficult to measure such things does not mean that they have no value. Contingent valuation studies are now very sophisticated, and researchers are very active in the search to make such studies better and more accurate.

Whose Welfare Should Be Included?

Another difficulty with environmental policymaking is that it is not always clear whose opinions should be taken into account when making environmental policy. For example, suppose there were a proposal to allow a major resort development at the Grand Canyon. Should that be the concern of residents of the area, the state of Arizona, or the entire United States? For that matter, should residents of other countries be entitled to a voice? After all, the Grand Canyon is one of the most spectacular sites in the world, visited by thousands of foreign tourists every year.

Perhaps you think it self-evident that foreigners should have no say in US environmental policy. Yet environmentalists in the United States and Europe have often voiced their opinion on the environmental consequences of policies in other countries, such as the construction of the Three Gorges Dam in China. Similarly, much of the world was outraged when, in 2001, the Taliban in Afghanistan destroyed two giant Buddhas that had been carved in the sixth century.

An even bigger problem has to do with the treatment of future generations. When you drive your car today, you are imposing costs not only on the living but also on those as yet unborn. The people most likely to be adversely affected by global climate change are not yet alive. How should we take into account their welfare and well-being? Scientists are largely in agreement that carbon emissions from the burning of fossil fuels will have an effect on global climate. Most environmental economists are convinced by this evidence, yet there continues to be disagreement among economists about the appropriate policy response. Most of that disagreement in the end comes down to different views about how to account for the welfare of future generations.

Uncertainty

One of the biggest difficulties with designing environmental policy is uncertainty. Global climate change is the clearest example. Although there is widespread agreement that we face some risk of climate change from carbon emissions, there is debate about the size of the effect. It is possible that we are facing only a small change in global climate, in which case it might not be worth spending a lot of resources now on reducing emissions. It is also possible that there could be catastrophic effects on global climate, in which case we should spend a lot of resources right now on reducing carbon emissions. And it is probable that we are facing something in between these extremes. Policymakers must try to get good estimates of how likely these scenarios are and then must decide how risk-averse they want to be when setting policy.

International Cooperation

Because environmental problems are often not confined to a single country, international agreements are sometimes needed for effective environmental policy. The *Montreal Protocol on Substances That Deplete the Ozone Layer* [4] is an example of successful international cooperation on environmental policy. In the

1970s, scientists recognized that certain chemicals known as chlorofluorocarbons (CFCs) were leading to a reduction in the atmospheric ozone layer. The ozone layer filters out dangerous radiation, so its destruction was linked to increases in skin cancer and other problems. The Montreal Protocol came into force in 1989 and has been signed by almost every country in the world. It mandated a gradual phaseout of CFCs, and current research shows that the atmospheric concentrations of CFCs have decreased as a result.

To date, however, there has been much less progress on the even bigger problem of global climate change. Although there has been much negotiation, there is still no international agreement on climate change that is comparable to the Montreal Protocol. In 1997, countries signed the Kyoto Protocol, [5] which was the first major international agreement to address the accumulation of greenhouse gases in the atmosphere. Some countries—notably Canada, Australia, New Zealand, and most of Europe—committed to specific targets, while others made more general commitments. However, the United States did not ratify the agreement. Meanwhile, several of the largest emitters of greenhouse gases, such as China, India, and Indonesia, did not have any specific commitment to greenhouse gas reductions. As a result, the impact of the protocol is greatly limited. One interesting feature of the Kyoto Protocol is that it included provisions for emissions trading.

In 1999, the United Nations held another summit in Copenhagen, in an attempt to make more progress on this topic. The outcome of the summit was the Copenhagen Accord: [6] a declaration that climate change was a problem. However, no binding commitments on greenhouse gas emissions resulted. A follow-up meeting in Cancún, Mexico, in 2010, delivered the Cancún Agreement, [7] which—although still not a binding treaty—was judged by many observers to represent significant progress over the Copenhagen Accord. Still, despite many meetings and fine-sounding commitments, the world is a long way from having an agreement on greenhouse gas emissions to match the Montreal Protocol.

Other Kinds of Externalities

Our discussion here has been about externalities arising from either the production or the consumption of goods and services. However, the idea of externality is used much more broadly in economics because it is a very helpful way of diagnosing inefficiencies.

One example comes from the working of labor markets. In general, both firms and workers spend time and resources trying to make a good “match.” Firms have human resource departments that spend resources advertising jobs, going through résumés, interviewing job applicants, and so on. Workers spend time preparing their résumés, interviewing at different firms, and so on. Both are willing to do this because a good match can be very beneficial: the firm gets a highly productive worker for which it is willing to pay a good wage. [8]

Now, the more effort the worker and the firm expend, the more likely they are to come up with a good match. And though each benefits individually from this effort, some of the benefit also flows to the other side of the market. In other words, if a worker tries harder to make a good match, this bestows a positive externality on a firm, and if a worker tries harder to make a good match, this bestows a positive externality on the worker.

As another example, is it a good thing if a new fast-food restaurant opens in your town? To answer this, we can think about the externalities that arise because of the decision to enter. When a new restaurant opens, it is providing a product that is similar but not identical to some of the offerings already in the market. It thus steals business from existing restaurants and reduces their profits. We can think of the entry of a new restaurant as imposing a negative externality on existing restaurants. Conversely, the entry of a new restaurant bestows a positive externality on consumers because they benefit from the increased choice. Thus the entry of a new restaurant leads to both positive and negative externalities.

We said previously that if an action has positive externalities, then there will be too little of that action from a social point of view. If an action imposes negative externalities, there will be too much of that action. In this example, the action is the entry of a new firm or product. If the positive externalities are more important, then there is too little entry, from a social point of view. If the negative externalities are more important, then there is too much entry. In general, we cannot draw any conclusions about whether there are too many firms (or products) or too few.

KEY TAKEAWAYS

- An externality arises when an action taken by one person directly affects another’s welfare. These operate outside of markets.

- One solution to an externality problem is to create a market so that the effects of one person's actions on others will be reflected in the market price of taking that action. Another solution is to put in place taxes or subsidies so that private incentives are aligned with social goals.
- One challenge for policy design is that the valuation of environmental goods is difficult to measure. Moreover, external effects do not respect borders, so international agreements are often required.

CHECKING YOUR UNDERSTANDING

1. If someone living nearby you is playing loud music, is that a positive or negative externality?
2. In a system with tradable permits, how is the total quantity of permits determined?
3. Suppose you invent a new product. After the patent expires, others can freely copy your design. Is your act of invention an example of an externality?

[1] See Chapter 5 "eBay and craigslist" for a discussion on auction mechanisms.

[2] See Ian Traynor and David Adam, "It's Lean and Mean, but Is It Green? EU Plans Clampdown on Car Ads," *Guardian*, June 5, 2008, accessed February 28, 2011,http://www.guardian.co.uk/environment/2008/jun/05/carbonemissions.carbonfootprints.

[3] Don Fullerton and Thomas Kinneman, "Household Responses to Pricing Garbage by the Bag," *American Economic Review* 86, no. 4 (1996): 971–84.

[4] United Nations Environment Programme, "Chapter 1 "What Is Economics?", Section 1.1 "Microeconomics in a Fast-Food Restaurant"," *Handbook for the Montreal Protocol on Substances that Deplete the Ozone Layer—7th Edition (2006)*, accessed March 14, 2011,http://ozone.unep.org/Publications/MP_Handbook/Section_1.1_The_Montreal_Protocol.

[5] United Nations Framework Convention on Climate Change, *Kyoto Protocol to the United Nations Framework Convention on Climate Change*, 1998, accessed March 14, 2011,http://unfccc.int/resource/docs/convkp/kpeng.pdf.

[6] United Nations Framework Convention on Climate Change, "Copenhagen Accord, Draft decision -/CP.15," Copenhagen, December 7–18, 2009, accessed March 14, 2011,http://unfccc.int/resource/docs/2009/cop15/eng/l07.pdf.

[7] United Nations Framework Convention on Climate Change, "Outcome of the work of the Ad Hoc Working Group on long-term Cooperative Action under the Convention, Draft decision -/CP.16," accessed March 14, 2011,http://unfccc.int/files/meetings/cop_16/application/pdf/cop16_lca.pdf.

[8] Chapter 8 "Growing Jobs" discusses search and matching in the labor market in more detail.

13.3 Renewable, Nonrenewable, and Accumulable Resources

LEARNING OBJECTIVES

1. What is the Hotelling rule for the use of resources?
2. What is a nonexcludable good?
3. What is an accumulable resource?

Some individuals take deliberate actions to limit or reduce their environmental impact. For some people, the choice of a greener lifestyle can affect almost all aspects of their consumption, including housing, transportation, and the food that they eat. There are, of course, many different ways in which our actions affect the environment. Our choices affect the amounts of pollutants that are emitted into the air and the water. They affect the amount of greenhouse gases that enter the atmosphere, which in turn has an impact on climate change. And they affect the rate at which we use up natural resources such as oil.

We explained that air pollution from automobiles is an example of an externality, which means that there is a divergence between marginal private costs and marginal social costs. Because of this externality, decisions to drive in a city can generate a social dilemma. It is not obvious, though, that consumption of gasoline generates an analogous problem. Some people argue that, if you are willing to pay for gasoline, you have the right to buy as much or as little as you please. If you want to drive a gas-guzzling Hummer or sport utility vehicle (SUV) rather than a fuel-efficient Prius, then—in their view—that is your right. Others make a different claim. They argue that because oil is a scarce resource, we ought to be conservative with its use. Driving a gas guzzler—in their view—wastes the earth's limited natural resources and thus is environmentally irresponsible. We now evaluate these two views.

Nonrenewable Resources and the Hotelling Rule

Oil is a nonrenewable (exhaustible) resource—that is, a resource that does not regenerate over time. Obviously, if we keep using up a nonrenewable resource, we will eventually run out of it, which is why it is called exhaustible. We most often hear this concern voiced about oil. For example, as oil prices increased to record levels in 2008, there was a great deal of discussion in the media about "peak oil"—the point at which world oil production is at its maximum.

Given that total oil resources are limited, we will eventually go beyond the peak oil point. Many commentators believe that we are at or close to this point already and have expressed concern about the implications of this for the world's economy. Consider, for example, the following quotation from a 2005 report:

> *World oil demand is expected to grow 50 percent by 2025. To meet that demand, ever-larger volumes of oil will have to be produced. Since oil production from individual reservoirs grows to a peak and then declines, new reservoirs must be continually discovered and brought into production to compensate for the depletion of older reservoirs. If large quantities of new oil are not discovered and brought into production somewhere in the world, then world oil production will no longer satisfy demand. That point is called the peaking of world conventional oil production.* [1]

When economists read a quotation like this, however, they typically think that something is missing. The quotation says that, at some point "world oil production will no longer satisfy demand." Economists respond that this does not make sense: the price of oil will adjust to ensure that supply equals demand. Similarly, they would say that it is very unclear what "oil demand is expected to grow 50 percent" means. Does this mean a shift in demand (a 50 percent increase in the quantity demanded at every price) or an increase in the equilibrium quantity (the quantity purchased will be 50 percent higher than it is now)? Fundamentally, economists would say that we cannot talk sensibly about the oil market without discussing what will happen to the price of oil. We turn to this question next.

The Hotelling Rule

Suppose you own 1,000 barrels of oil. You want to decide when you should sell them. You could sell them now or hold onto them in the hope that the price of oil will increase. This is a difficult problem, so we begin by making some simplifications. First, imagine you already have the oil in storage, so we can ignore the costs of extracting it from the ground. Second, suppose you can store it for free. Third, suppose you can buy and sell oil in a competitive market.

Even with these simplifying assumptions, it seems as if you are facing a hard decision. In fact, this problem is easier than it looks. If you sell one barrel this year, you get this year's price. You can invest this at the market rate of interest and get (price this year × interest factor). Alternatively, you can store your oil for a year and then sell it next year at next year's price. If this year's price multiplied by the interest rate is greater than next year's price, then you should sell all of your oil this year. If next year's price is higher, you should store all of your oil for sale in the future. By following this rule, you can determine how to make the most money from your stock of oil.

Of course, this decision looks the same to anyone else who is in the same position as you. So if this year's price multiplied by the interest rate is greater than next year's price, then *everyone* would try to sell their oil this year. There would be a huge supply of oil to the market, which would tend to reduce the price this year. If next year's price is higher, then there will be little to no oil supplied to the market this year, which would tend to increase the current price. The only way the oil market will be in equilibrium is if a condition known as theHotelling rule holds. This is an arbitrage condition for the use of resource stocks. In the case of a nonrenewable resource sold in a competitive market, with no costs of extraction, no costs of storage, and no uncertainty, the rule states that

$$\frac{\text{price next year}}{\text{price this year}} = \text{nominal interest factor}.$$

The Hotelling rule tells us that the price of the resource increases at the rate of interest. For example, if the nominal interest rate is 6 percent (thenominal interest factor is 1.06), then the Hotelling rule says that the price of oil will increase 6 percent per year. The rule has a remarkable implication: it does not matter whether you sell your oil or hold onto it. If you sell your oil, you can earn the market rate of interest on your savings. If you hold onto your oil, it as an asset that yields a rate of return equal to the market rate of interest. Either way, you should expect to get the same return. [2]

Complications

The basic Hotelling relationship underlies the pricing of all renewable and nonrenewable resources, but there are many other factors that also come into play. We can in principle build all these other complications into our equation, but this would require mathematics and analysis that go beyond the level of this book. Still, we can give a brief idea of how to incorporate these other factors.

Costs of storage. If it is costly to store the resource from one year to the next, then the price must increase at a rate fast enough to cover the cost of storage as well as the rate of interest.

Costs of extraction. We have ignored the costs of extracting oil from the ground. This cost is not constant in either the short run or the long run. In the short run (say, in a given year), the marginal cost of extracting oil increases when we pump more oil from the ground. (One implication is that the supply of oil at any given time is not perfectly elastic.) As we start to run out of oil, it is likely that the marginal cost of extracting oil from the ground will increase substantially. This, too, must be factored into the arbitrage condition. Marginal extraction costs that increase over time are an additional factor causing prices to increase.

Discovery of new oil fields. If new oil fields are discovered, then the overall supply of oil in the world increases. This increase in supply leads to a decrease in the price. As soon as there is a discovery of a new field, the price of oil jumps to a new Hotelling path. What matters for the pricing of oil is *new information* about existing resources. If the oil companies discover a new oilfield—or come up with a better technique for extracting existing reserves—there will be a decrease in the price of oil as soon as this information becomes known to the market. This is why the price of oil is so sensitive to political changes in the Middle East.

Shifts in the demand curve. The price of oil is also affected by shifts in the demand curve. For example, the increasing prosperity of countries such as China and India is causing the demand for fossil fuels to shift outward. Technological developments are one source of changes in demand. If scientists were to come up with a cheaper source of energy, the demand for oil would decrease.

Market power. The supply of oil is heavily influenced by the decisions of oil-rich countries such as Saudi Arabia and Kuwait. Producers with market power tend to restrict supply to force prices up. The effect of this is to increase prices at all times, which in turn means that existing stocks of oil will last longer. Market power, in and of itself, does not affect the basic conclusion from the Hotelling rule unless the degree of market power changes over time.

Uncertainty. Finally, we took the stock of oil, the demand curve, and the rate of interest as known with certainty. In reality, of course, all of these are unknown. Changes in the information with regard to any of these variables will lead to changes in the price of oil.

Does the Hotelling Rule Work in Practice?

The Hotelling rule is based on a very simple arbitrage idea, so it is highly compelling. Yet it is sometimes difficult to observe the rule in operation in the data for oil or for any other nonrenewable resource. Figure 13.5 "The Price of Oil (in 2008 Dollars)" shows what has happened to the price of oil in the last 60 years or so. The prices are in 2008 dollars. Oil prices were reasonably steady in the 1950s and the 1960s. The 1970s are sometimes called the "oil shock decade," and from the graph you can see why. The price of oil jumped to a level that is equivalent to over $100 a barrel in 2008 dollars. Then oil prices fell again and were relatively low in the 1980s and 1990s. The early years of the 21st century saw another big increase in oil prices, with the price reaching record levels in 2008.

It is not easy to reconcile this figure with the Hotelling rule. We do not see the price of oil increase steadily, as the Hotelling rule seems to suggest. The problem is that the complications that we mentioned in Section 13 "Complications" are quite significant in practice. Over the last several decades, we have seen technological improvements, the discovery of new oil fields, political instability in the Middle East and other oil producing regions of the world, price-fixing by the oil-producing countries, and so on. Most of the time, it seems as if these variables are swamping the pattern we expect from the Hotelling rule.

If we look at other nonrenewable resources, it is likewise difficult to see Hotelling effects. In large part this is because technological improvements have often made resources less valuable, even as they became scarcer. For example, copper became significantly less valuable when scientists developed techniques for transmitting information along fiber-optic cables rather than along copper wire.

For all of these reasons, you should not think of the Hotelling rule as literally describing what will happen to the price of nonrenewable resources in the real world. Instead, you should think of it as explaining one component of the price. If the exhaustion point of the resource is a long way in the future, the Hotelling rule may not play a big role in explaining the price. Other factors that shift the demand and supply curves may explain most of the price variation. Nevertheless, most economists are confident that the Hotelling rule does contribute to resource price changes, and as a resource gets closer to exhaustion, the Hotelling rule will play a bigger and bigger role. James Hamilton, an economist who is an expert on oil, put it as follows in late 2006:

My own view is that, for most of the past century … the resource exhaustion was judged to be sufficiently far off as to be ignored.…I do not infer that the next decade will necessarily be like the previous century. Certainly declining production from U.S. oil reservoirs set in long ago.…

I am not at all prepared to dismiss the hypothesis that [the Hotelling rule has] indeed started to make a contribution to oil prices over the last five years, and will become more apparent over the next five. For example, the announced intention of OPEC producers to cut back production as the price goes below $60 might be most naturally interpreted from that perspective—producers don't see it as being in their interests to sell for less, given what the oil will be worth in the future. [3]

Excludable and Nonexcludable Resources

The Hotelling rule tells us that we expect the price of a nonrenewable resource to rise as we use it up. As its price increases, households and firms have an incentive to substitute other goods for the resource. As the price of oil increases, people switch to other forms of transportation and more fuel-efficient vehicles. More generally, an increase in the price of oil makes other forms of energy—such as wind, solar, or nuclear power—more attractive.

Nothing in this description suggests any failure of the market mechanism. Although we have explained what will happen as we start to run out of a resource, we have not given any reason to suggest that we will use up the resource too quickly. In fact, we have even pointed to one reason we might be using up oil too slowly: if oil producers have market power, they have an incentive to limit the supply to the market and increase the price. The mere fact of using up a nonrenewable resource does not mean that the market is not efficient. To understand the difference between using oil and polluting the atmosphere, we need a new distinction.

Toolkit: Section 17.19 "Externalities and Public Goods"

A nonexcludable good (or resource) is one for which it is impossible to selectively deny access. In other words, it is not possible to let some people consume the good while preventing others from consuming it. An excludable good (or resource) is one to which we can selectively allow or deny access.

Go back once more to our example of a smoker in an office. The smoker actually consumes two things: she consumes cigarettes and she consumes the clean air in the office by turning it into dirty air (economists call this "joint consumption"). Clean air, like a cigarette, is a "good." But it is a good with a special property. Under most circumstances, we cannot allow some people access to clean air while denying others access to clean air. Clean air is nonexcludable.

Drivers in Mexico City likewise consume clean air: driving leads to less clean air and more dirty air. Like the smoker in the office building and the car drivers in Mexico City, polluting firms consume clean air as well. In this instance, we can think of clean air as being another input into the production process. *Air pollution is a process by which the nonexcludable resource that we call "clean air" is consumed by households and firms*. This is a somewhat unfamiliar way to talk about environmental pollution, but it makes sense once you think about it. At the end of the production process, there is less clean air than before because the firm has made some air dirty. The polluting firm uses clean air in its production process.

Partially Excludable Goods

Other examples of nonexcludable goods are fireworks displays, lighthouses, fish in the ocean, concerts in a public space, and broadcast television. These examples prompt the following observations.

- Whether or not something is excludable can change with advances in technology. Early television and radio signals were nonexcludable because they could be accessed by anyone with a suitable receiver. But cable television and Internet radio are excludable because it is technologically possible to restrict access to certain users.
- Excludability is not a hard-and-fast line. Think about roads, for example. Toll roads are excludable; it is possible to control access to toll roads because there are relatively few entrances and exits. By contrast, roads in a neighborhood or in the middle of a city are nonexcludable; it is

almost impossible to imagine allowing selective access to different city blocks. The more different entrances there are to a road, the harder it is as a practical matter to make the road excludable. Similarly, fish in the ocean are neither fully excludable nor completely nonexcludable. Fish in international waters can be fished by anyone, but coastal waters are under the jurisdiction of individual nations that may restrict fishing rights.

Going back to our smoking example, we cannot make the air in a single office excludable, but we can perhaps do the next best thing: within a single building, we could certainly imagine permitting smoking in some offices and not in others. In other words, we could define property rights differently in different offices and in this way make clean air partially excludable. In some cases at least, this might be a practical way to accommodate both smokers and nonsmokers. We see something to this effect in airports and other public spaces. Very often you will find that most of an airport is designated as nonsmoking (meaning that the property rights to clean air have been allocated to nonsmokers), but there are designated rooms or areas where smokers have the right to consume—that is, use up—the clean air.

Externalities and Nonexcludable Goods

There is a connection between nonexcludability and externalities. As our clean air example suggests, goods—like clean air—that are nonexcludable will tend to be overconsumed. When a good is nonexcludable, its marginal private cost is zero. If its marginal social cost is positive, then there is a negative externality from the consumption of a good. As these observations suggest, there are typically negative externalities associated with nonexcludable (or partially excludable) goods.

The flipside is that nonexcludable goods will be underproduced. To see why, think about whether a private firm will want to produce nonexcludable goods. For goods that are completely nonexcludable, the answer is no. There is no return from producing such goods because they cannot be sold for profit. Anyone can consume these goods without paying for them. Goods that are partially nonexcludable may be produced, but they will be produced in insufficient quantities.

Should We Worry about Running Out of Resources?

Because clean air is a nonexcludable resource, there are externalities associated with its use, and we know that this implies inefficiency. Economics provides a clear argument for why we end up overconsuming clean air. Similarly, many people take it as self-evident that we are overconsuming oil and other natural resources. The arguments we made with respect to clean air, however, do not translate to the case of oil. It is true, of course, that consumption of oil may have negative externalities because it leads to pollution. But the problem in this case has nothing to do with the fact that oil is a nonrenewable resource.

It is difficult to buy and sell nonexcludable goods—after all, why would anyone buy something that is free? This makes clean air unlike most of the other goods we study in economics. Excludable goods and services, by contrast, are those that we can prevent someone else from enjoying. Excludable goods—like oil—are easily traded in markets.

When people worry about running out of resources, they are speaking, for the most part, about excludable resources: those that are mined from the ground or grown on farms and privately owned forests. Economic analysis suggests that this worry may be largely misplaced. Unless we have specific reasons to think that markets are misallocating these resources, there is no particular cause for concern or government intervention.

There are some exceptions to this principle. Oil is sometimes drilled from pools that are not uniquely owned. Different companies may be able to access the same oil pool. In this case, the oil in the pool is partially nonexcludable—any company with access to the pool is able to drill it. It follows that there is a negative externality associated with the drilling of the oil in this case, so we expect there to be too much drilling.

Renewable Resources

Oil is a nonrenewable resource. New oil reserves are not going to be created (at least over any period of time relevant to human beings). Over time, the stock of a nonrenewable resource can only decrease, never increase. By contrast, arenewable resource is one that regenerates over time.

Whether the stock of a renewable resource grows, shrinks, or stays constant depends on the balance between how fast we use up the resource and how quickly it regenerates. If a resource is depleted at a rate faster than it regenerates, the stock will decrease. If a resource regenerates faster than it is depleted, the stock will increase. Examples of renewable resources are stocks of fish in the ocean and forests. Clean air and water are also examples of renewable resources: if the source of pollution disappears, then a polluted river or polluted air will tend to improve naturally over time. Of course, this process can take quite a while.

In some cases resources may be renewable only above a certain threshold. For example, we can overfish a particular species of fish to the point of extinction. So although a renewable resource will normally naturally regenerate over time, it can turn into a nonrenewable resource if the stock falls below a critical level.

As with nonrenewable resources, what matters most for efficiency is whether a resource is excludable or nonexcludable. If a renewable resource is excludable, then a modified version of the Hotelling rule applies. The arbitrage condition in this case takes into account the fact that the resource stock gets larger as time goes by. If the resource is nonexcludable, however, then all the problems we saw earlier in the chapter come into play. The marginal social cost of using the resource will typically exceed the marginal private cost, leading to overconsumption of the resource.

Accumulable Resources

Nonrenewable environmental resources get used up over time. With good stewardship, which usually requires well-established property rights, the stocks of renewable resources can stay constant or even increase. Environmental resources, however, typically cannot grow without limit. There is a limit to how many fish there can be in the ocean or how clean the air can be.

Economists also identify a third class of resources, calledaccumulable resources. Such resources can increase (more or less) without limit. The most important examples are physical capital (factories, machines, etc.), human capital (the skills and know-how of workers), and general human knowledge.

Even though there is no reason to think that the depletion of a nonrenewable resource is necessarily a source of economic inefficiency, it does not follow that the depletion of natural resources is without cost for an economy. For example, as oil becomes more expensive, it costs more to produce other goods and services that use energy as an input. From an economic perspective, this is a lot like a worsening of technology.

Crucially, though, we are accumulating resources such as physical capital, human capital, and knowledge at the same time that we are running down our stocks of oil, coal, and other natural resources. In the last couple of centuries, the economy of the world has grown substantially because we have been able to accumulate resources at a rate that far outpaced our depletion of the natural environment. For much of that time, we had natural resources in abundance, so their price stayed low. At the same time, we increased our technological know-how at unprecedented rates.

There is no guarantee, from economics or anywhere else, that this state of affairs will continue. We will continue to run down our stocks of nonrenewable resources. If technological advance (and other accumulation) fails to keep pace, then we might well see the prices of these goods increase. More generally, the fact that we have seen our economies growing in the past is no guarantee that they will continue to grow in the future.

Technological optimists point to the rapid growth of our knowledge and expect this to continue. If they are right, then technological advance will keep our living standards growing even though we are depleting our stocks of resources. Technological pessimists observe that the last two centuries are an anomaly if we look at the broad sweep of human history. If the rate at which we accumulate knowledge, human capital, and other accumulable resources were to decrease significantly, then the drag on the economy from declining resources would begin to seem substantial. Economics does not allow us to predict the future, and we do not know who is right.

KEY TAKEAWAYS

- According to the Hotelling rule, a resource should be extracted so that the rate of price increase in the resource should be the same as the interest rate.
- A good is nonexcludable if it is impossible to deny access to it.
- An accumulable resource is one that can be increased over time with investment. Leading examples include physical capital and human capital.

CHECKING YOUR UNDERSTANDING

1. If the price of oil is decreasing while the interest rate is positive, is the Hotelling rule violated?
2. Is the Internet a nonexcludable good?
3. Will technological advance necessarily offset the depletion of the stocks of natural resources?

[1] Robert L. Hirsch, Roger Bezdek, and Robert Wendling, "Peaking of World Oil Production: Impacts, Mitigation, & Risk Management, " Minnesotans for Stability, February 2005, accessed January 31, 2011,http://www.mnforsustain.org/oil_peaking_of_world_oil_production_study_hirsch.htm; also quoted in James Hamilton, *Limitations of the Hirsch Report on Peak Oil*, *Econbrowser*, August 9, 2005, accessed January 31, 2011,http://www.econbrowser.com/archives/2005/08/limitations_of.html.

[2] Oil can be thought of as an asset. In Chapter 9 "Making and Losing Money on Wall Street", we discuss how the prices of different kinds of assets are determined. The discussion in that chapter also explains how we can take into account the fact that you don't know the future price of your asset—in this case, next year's price of oil—with certainty.

[3] See James Hamilton, "Is Peak Oil Irrelevant?," *Econbrowser*, October 24, 2006, accessed January 31, 2011, http://www.econbrowser.com/archives/2006/10/is_peak_oil_irr.html.

13.4 End-of-Chapter Material

In Conclusion

Economists and environmentalists sometimes do not see eye to eye. Economists think environmentalists often focus on the wrong problems, and environmentalists think economists place too much faith in markets. Yet economics is the science that helps us understand why some environmental problems are among the most important and difficult that we face.

When economists look at excludable resources, for which property rights are well defined, they tend to be less concerned. It is certainly possible that we will run out of oil and other nonrenewable resources. But that in itself does not signal a problem. What matters is whether we are using our resources efficiently or inefficiently. Perhaps the best thing for us as a society to do is to use up our resources quickly. More important, as oil becomes scarce, we know that market prices will force users to economize on oil and look for substitutes instead.

We are assuredly not saying that an economy is better off with fewer resources. We would always like to have more of an exhaustible resource. The most important question, though, is how to best use the resources we have. Markets can sometimes provide a good answer to this question. If markets work properly, sending the correct signals to producers and consumers, then market allocation will be efficient.

Economists worry a great deal more about environmental problems where resources are nonexcludable. Pollution of the air, pollution of rivers and oceans, biodiversity loss, overfishing, and climate change are all examples of environmental problems for which we cannot rely on markets. To an economist, it is not surprising that markets fail in these cases. All of these resources are nonexcludable. When resources are nonexcludable, market allocations will not typically be efficient, and there may be a role for government to try to solve these problems.

Key Link

- James Hamilton blogs: http://www.econbrowser.com

EXERCISES

1. In the example about second-hand smoking in the office discussed in Section 13.1.1 "Smokers, Nonsmokers, and the Coase Theorem", how would you use the Coase theorem to determine how much smoking should occur if there are many people in the office? What difficulties do you see in trying to apply this theorem?
2. (Advanced) Consider Table 13.1 "The Payoffs in a Social Dilemma Game". Suppose your payoff from taking public transportation when everyone else is taking public transportation is $3 rather than $1. Why is it harder to predict what will happen in this situation? See the toolkit. [Hint: Look at what decision you make when others take public transportation and when they don't.]
3. Consider some situations that might arise in your college or university. Which of the following is an example of an externality?
    a. In a class that is graded on a curve, you study harder and get a better grade, so others get a worse grade.
    b. At examination time, lots of people want to study, so it is difficult to find space in the library.
    c. You find it difficult to understand your professor's accent.
    d. Your favorite television show is on the night before you have a big test, and you can't decide whether to watch or study.
    e. Other people ask good questions in class, which makes the class more interesting for you.
    f. Everyone is selling his or her used chemistry textbook, so you can buy one at a cheap price.
    g. Late-night parties in neighboring dorm rooms are preventing you from sleeping.

1. Give three reasons why the marginal social cost of driving is greater than the marginal private cost of driving.
2. There are many endangered species in the world, such as the white tiger and the sea otter. Why are these species endangered whereas cows and sheep are not?
3. Suppose you meet some friends at a pizza restaurant. You are all very hungry, and you know the wait for the pizza may be long. At the next table, you see some people who have just finished one pizza, and another is about to be delivered to their table. Explain how you could conceivably create a market that would make both your table and the other table better off. What is the nature of the transaction costs that prevent such trades from happening in restaurants in real life?

4. One reason people choose to buy large vehicles, such as SUVs, is because they are safer if you are involved in an accident. Explain how this could give rise to a social dilemma game.
5. Copy Figure 13.4 "A Divergence between Marginal Private Benefit and Marginal Social Benefit". Indicate on that diagram how large a subsidy would be required to induce the socially optimal quantity.
6. If you were to graph a measure of pollution on one axis and a level of economic activity (such as the real gross domestic product per person) on the other, what type of relationship do you think you would find? How would you explain this relationship?
7. Think of an externality that arises in a college dorm. What market can you think of that would (or could) eliminate any inefficiencies from that externality?
8. Using Table 13.1 "The Payoffs in a Social Dilemma Game", think of another example of a social dilemma game. What are the choices of the people, the payoffs, and the outcomes?
9. Explain why tradable permits for the right to dump garbage into a river would be more efficient than telling producers how much they are allowed to dump in the river. How is the total quantity of tradable permits determined?
10. Suppose you had a system of tradable permits that allocated permits to those who needed them (that is, the firms that polluted a lot). Would that be a good system to provide an incentive for firms to undertake investment in clean technologies?

Economics Detective

1. In recent years, the European Union has discussed requiring airlines to obtain permits to emit pollution in order to operate. What is the status of this proposal? How might it work?
2. Pick a nonrenewable resource and try to find out what has happened to its price over the last few decades. Do you see any evidence of the Hotelling rule? If not, try to find out what have been the main factors affecting the price of this resource.
3. If you live in the United States, try to find the Superfund site nearest to where you live. What kind of pollution is at the site? (If you live somewhere else, pick a US city of your choice.)
4. What is a "carbon tax"? What is the economic rationale for such a tax?

13.5 Appendix: An Example of the Hotelling Rule in Operation

The Hotelling rule states that *the nominal price of oil will increase at the nominal rate of interest*. This seems a little bit mysterious. After all, we also think that the price of oil is determined by demand and supply in a market. In fact, these two approaches to the price of oil are completely consistent.

To see how, imagine that the worldwide demand curve for oil each year is as follows:

price = 1000 − 0.000001 × quantity,

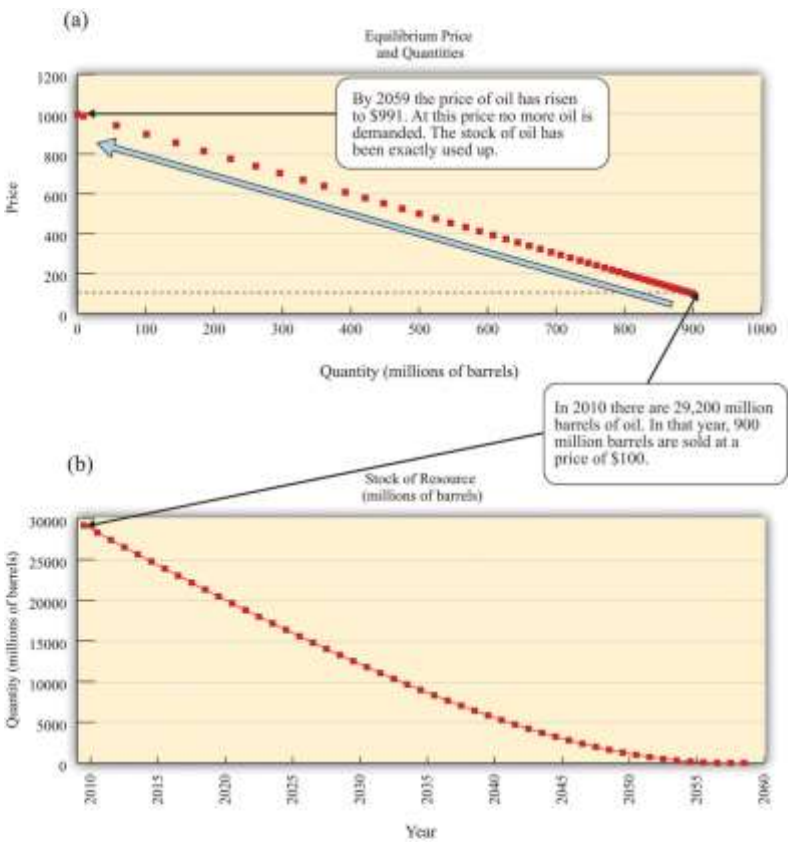
where we measure the price in dollars and the quantity in barrels. Individual suppliers are willing to sell oil at the price determined by the Hotelling rule. At higher prices, they will supply a lot of extra oil to the market; at lower prices, they supply no oil to the market. In other words, the supply of oil will be perfectly elastic at the market price.

But this is not enough to determine the market price. We know the market price this year only if we know the price next year. And we know the price next year only if we know the price the year after that, and so on. Unless we know the price at some future date, we cannot calculate the price today. But here is the trick: we do know that the price will eventually increase to the **choke price** (which is $1,000 per barrel in our example).

Over time, the price increases at the rate of interest, as implied by the Hotelling rule. If the demand curve is unchanged from year to year, the quantity demanded gets smaller. This makes sense: the increasing price reflects the increasing scarcity of the resource. The price increases until it eventually reaches the choke price, where the quantity demanded decreases to zero. But when will that happen? It has to happen when the last barrel of oil is being sold. If there is oil left over at that time, the price would decrease below the choke price. And if we ran out of oil before then, the price would have to hit the choke price earlier. Anything else would cause a mismatch between demand and supply. Thus the price today needs to be just right so that when the last barrel of oil is being sold, the price increases to exactly the choke price.

Figure 13.6 "An Example of Oil Depletion According to the Hotelling Rule" shows an example based on this demand curve. Suppose that the total amount of oil available in the world at the end of 2010 is 29.2 billion barrels, which is owned by many competitive suppliers. How will this oil get released to the market? In 2011, 900 million barrels of oil are sold, so the market price is therefore $100. The price then increases at the rate of interest, which is 5 percent per year in our example. By the year 2044, less than 4 billion barrels remain, and the price of oil has increased to $500 per barrel. In this particular illustration, we run out of oil in slightly under half a century: the last 9 million barrels of oil are sold in the year 2058, at $991 per barrel.

Figure 13.6 An Example of Oil Depletion According to the Hotelling Rule



These figures show an example of resource depletion according to the Hotelling rule. (a) The price and quantity of oil traded each year trace out the demand curve. (b) The stock of oil over time represents the supply curve. Initially, there are slightly under 30 billion barrels of oil available. In the first year, 90 million barrels of oil are sold in the market, at $10 per barrel. The price then increases at the rate of interest. In the year when the price reaches the choke price (2059), the stock of the resource is fully used up.

Chapter 14
Busting Up Monopolies
Working for Antitrust

You have probably never considered working for the Department of Justice, yet here is a job opportunity open to everyone.

> *If you have information about a possible antitrust violation or potential anticompetitive activity, use the following questions as a guideline to describe your complaint:*
>
> - *What are the names of companies, individuals, or organizations that are involved?*
> - *How do you believe they have violated the federal antitrust laws?*
> - *Can you give examples of the conduct that you believe violates the antitrust laws? If so, please provide as much detail as possible.*
> - *What is the product or service affected by this conduct? Where is the product manufactured or sold, or where is the service provided?*
> - *Who are the major competitors that sell the product or provide the service?*
> - *What is your role in the situation in question?*
> - *Who is harmed by the alleged violations? How are they harmed?* [1]

You may never have heard about antitrust laws. You may have only a vague idea of what antitrust is. Yet these laws have a direct impact on your day-to-day life because they can have a significant impact on the prices you pay for goods and services. As the Department of Justice explains,

> *Most states have antitrust laws, and so does the federal government. Essentially, these laws prohibit business practices that unreasonably deprive consumers of the benefits of competition, resulting in higher prices for inferior products and services…*
>
> *When competitors agree to fix prices, rig bids, or allocate customers, consumers lose the benefits of competition. The prices that result when competitors agree in these ways are artificially high; such prices do not accurately reflect cost and therefore distort the allocation of society's*

resources. The result is a loss not only to U.S. consumers and taxpayers, but also the U.S. economy. [2]

Antitrust laws in the United States are principally codified in three acts of Congress—the Sherman Antitrust Act, the Clayton Act, and the Federal Trade Commission Act—and the Federal Trade Commission and the Department of Justice enforce each act. In other countries, similar government agencies perform these same tasks.

How might you become involved in pursuing antitrust violations? Without knowing the provisions of these laws, it is hard to see how you could ever detect and report violations. Fortunately, the Department of Justice provides some further guidance: look for situations where the price of a good is in excess of the marginal cost of producing that good. Remember that marginal cost is the cost of producing one additional unit of output. In a competitive market, firms set prices equal to marginal cost, but when they have market power, firms set prices in excess of marginal cost. [3] The Department of Justice's suggestion reflects this conclusion: price in excess of marginal cost is a likely indicator of market power.

In practice, it is more complicated. For example, the cost of producing one more compact disc (CD) is the cost of the material it is made from, the cost of burning the CD, and the cost of the jewel case in which it is packed. These costs are very small: no more than a few cents. Yet, if you have bought a CD recently, you probably paid between $10 and $20 for it. Should you be reporting your local CD retailer to the Department of Justice?

In fact, the government sometimes actively protects and creates market power, despite the fact that it has entire divisions devoted to encouraging competition. It creates market power through patent and copyright laws, which prevent people from copying inventions and created works (like music or books). In the case of CDs, the government grants copyright protection because the creation of the very first copy of a CD is very costly. When a CD is produced, there are enormous music creation costs incurred by the musicians and the company recording and producing the music. These costs are much more sizable than the actual cost of producing the CD you purchased. From this perspective, the high prices are needed to make it worthwhile for the artist to incur the creation costs.

You can perhaps sense the tension. The costs of producing the very first CD are high, but the cost of producing the thousands (or millions if the artist is successful) of *copies* of the first CD are relatively small. Should the price reflect the marginal cost of that last CD or should it be higher to cover the creation costs as well? This argument was vividly illustrated in 2001 when music file sharing first became popular. A firm called Napster (http://music.napster.com) supplied a technology that facilitated the sharing of music through the Internet. Napster essentially reduced the marginal cost and the price of a song all the way down to zero. Napster's technology was in direct competition with the record companies, and a legal battle ensued. The final ruling forced Napster to block its file sharing and effectively ended its ability to share music. Other peer-to-peer networks have replaced it, however, and a lot of music is still available for free on the Internet.

In this chapter, we investigate these varied aspects of competition policy. This might seem like an arcane topic, but it has a huge impact on the prices you pay for goods and services. It is competition policy that keeps the price of CDs high and the price of airline tickets low. In this chapter, we evaluate the impact of competition policy on the economy and answer the following questions.

- What happens in markets when there are only a few producers?
- What are the different kinds of competition policy carried out by the government?

Road Map

To make sense of competition policy, we need to first understand what firms would do if there were no antitrust laws constraining them. We therefore begin by looking at economic outcomes in the absence of government protection. We first contrast a market in which there is a single seller with a market that is serviced by many sellers—that is, a competitive market. From this comparison, we can understand the basis of antitrust laws.

Next, we look at why the government sometimes actively promotes market power through patents and copyright. Specifically, we show how such laws can encourage innovation and creation of intellectual property. Finally, we look at situations where there are only a few competing firms. We explain how the outcome is different depending on whether firms choose to set prices or to set the quantity that they produce, and we again look at the role of government policy.

[1] See US Department of Justice, “Reporting Antitrust Concerns,” accessed March 14, 2011,http://www.justice.gov/atr/contact/newcase.html.

[2] US Department of Justice, “Antitrust Enforcement and the Consumer,” accessed March 14, 2011, http://www.justice.gov/atr/public/div_stats/211491.

[3] Chapter 6 "Where Do Prices Come From?" contains more information.

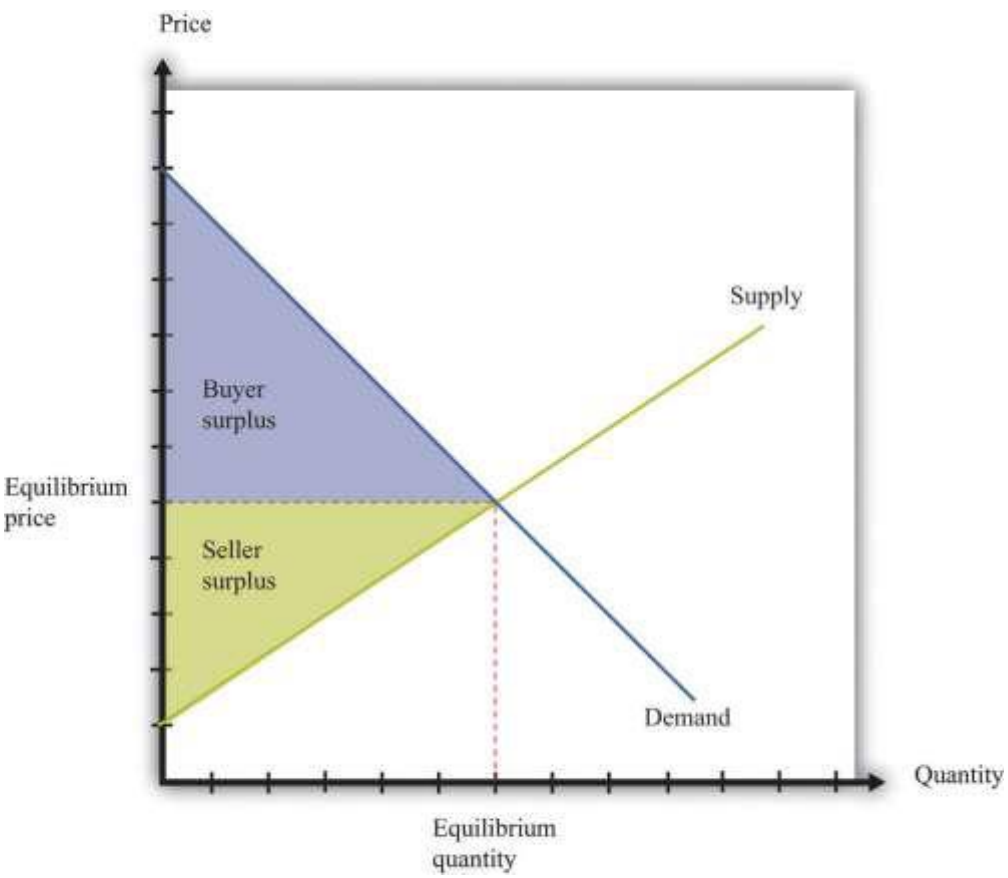
14.1 Market Power and Monopoly

LEARNING OBJECTIVES

1. What is a monopoly?
2. What is the outcome when there is a monopoly?
3. What are the policies taken to deal with monopolists?

When there are many buyers and sellers of a homogeneous product, we have a competitive market (Figure 14.1 "The Competitive Market Outcome"). Equilibrium is at the intersection of supply and demand. At the equilibrium level of output, households enjoy buyer surplus, given by the marked area below the demand curve and above the equilibrium price. The surplus arises from the fact that some buyers are willing to pay more than the equilibrium price for the good.

Figure 14.1 The Competitive Market Outcome



At the equilibrium quantity in a competitive market, all gains from trade are exhausted.

Surplus also flows to firms. Remember that a competitive firm's individual supply curve is equal to its marginal cost curve. In Figure 14.1 "The Competitive Market Outcome", the supply curve slopes upward because marginal cost is increasing. Firms obtain surplus because they can produce output at a marginal cost that is less than the equilibrium price of the good. This is shown as **seller surplus** in the figure.

At the equilibrium quantity, there are no further **gains from trade**. Producing more output would not increase the total surplus. In fact, producing more output would reduce the surplus: the marginal cost of producing more output would exceed the marginal valuation of extra output. Producing less output would likewise lower total surplus because the buyers and the sellers would lose some of their surplus.

The competitive market provides a benchmark because it leads to an efficient outcome. But very few markets are truly competitive. In most markets, firms possess some market power. This means, in particular, that they are able to set a price above marginal cost without losing all of their sales. In a competitive market, the *demand curve facing a firm* is perfectly elastic at the market price, whereas when a firm has market power, its demand curve slopes downward.

Toolkit: Section 17.10 "Buyer Surplus and Seller Surplus"

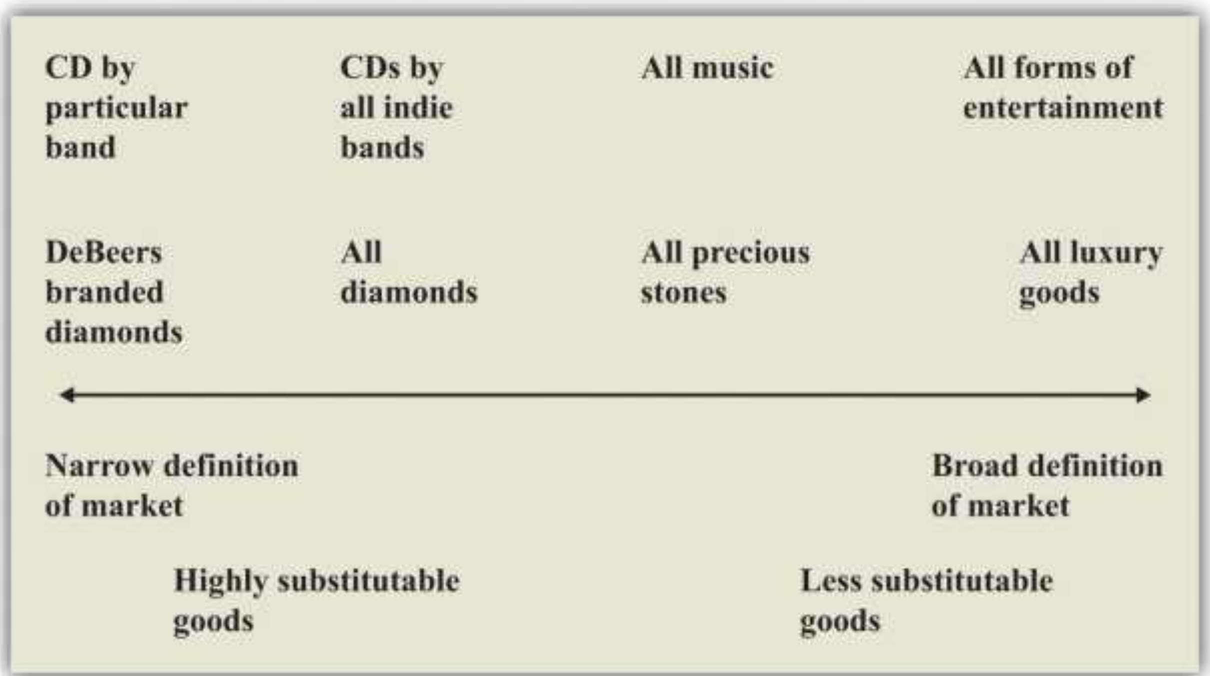
You can review the concepts of buyer surplus, seller surplus, and the gains from trade, in the toolkit.

The Definition of a Market

At the other extreme to the competitive market is the case of monopoly. A monopoly arises when there is a single producer in a market. The demand curve facing a firm is, in this case, the same as the market demand curve.

The definition of a monopoly seems easy, yet it is hard to decide exactly what we mean by "a market." Think about diamonds. It is often said that the De Beers Corporation is a monopolist in the market for diamonds because this company controls most of the world's diamond supply. Yet, depending on how broadly or narrowly we define the market, De Beers has either a lot of competitors or only a few. We could define the market very narrowly as "De Beers-branded diamonds" (De Beers is able to brand its diamonds by using certificates of authenticity). De Beers would then be a monopolist by definition. We could define the market more broadly as “all diamonds,” in which case De Beers has substantial market power but does not have a total monopoly. This is perhaps the most natural definition to use, yet it misses the fact that other precious stones, such as emeralds, rubies, or opals, are also possible substitutes for diamonds. An even broader market definition is “the market for precious stones.” We could go even further still and consider De Beers as part of the market for luxury goods, competing with, say, Louis Vuitton bags and Ferrari sports cars. We illustrate this in Figure 14.2 "The Extent of Competition Depends on the Definition of the Market".

Figure 14.2 The Extent of Competition Depends on the Definition of the Market



There is no hard-and-fast definition of a market, but goods that are highly substitutable for each other are generally taken to be in the same market.

Figure 14.2 "The Extent of Competition Depends on the Definition of the Market" also gives an illustration for the case of music. By definition, a given indie band has monopoly power over its own music. So again, with a very narrow definition of the market, we would say that the band is a monopolist for its own songs. But that band also competes with other indie bands for consumers’ dollars, so another definition of the market would be "CDs by all indie bands." Again, we could define the market still more broadly as "all music" or even "all forms of entertainment."

One of the difficult tasks of antitrust authorities is deciding what definition to use for a market. There is no single correct way to define a market, and the extent of market power depends on where we choose to draw the line. For this reason, the term *monopoly* is somewhat misleading. Almost all firms are monopolists if we adopt a sufficiently narrow definition of the market. The decision about how to set prices in the presence of market power, which is sometimes called the*monopoly pricing problem*, actually applies to nearly every firm in the economy. What matters in practice is determined by the *extent* of a firm's market power.

The extent of a firm's market power depends on two things: (1) the number of firms that potentially compete with it and (2) the extent to which those other companies produce close substitutes for a firm's product. A strip mall at the edge of town might contain several different fast-food restaurants. Each restaurant enjoys some market power because the food it has available is slightly different from that at the other establishments. But many consumers are likely to be willing to substitute reasonably freely among the different restaurants, so the market power of each restaurant is limited. Contrast that with a fancy French restaurant with a famous chef. That restaurant enjoys much more market power because there are almost certainly fewer comparable restaurants nearby and the meals at other high-end restaurants are not such close substitutes.

Market power can stem from many different sources. A firm has market power if it is selling a unique product or service. A firm can also derive market power from its superior or exceptional service. If a firm creates customer loyalty, either through exceptional service or by loyalty programs (such as frequent flyer miles), this is also a source of market power. Retail firms can derive market power from their location: your local corner store has some market power because you would prefer to walk there rather than drive to a more distant supermarket.

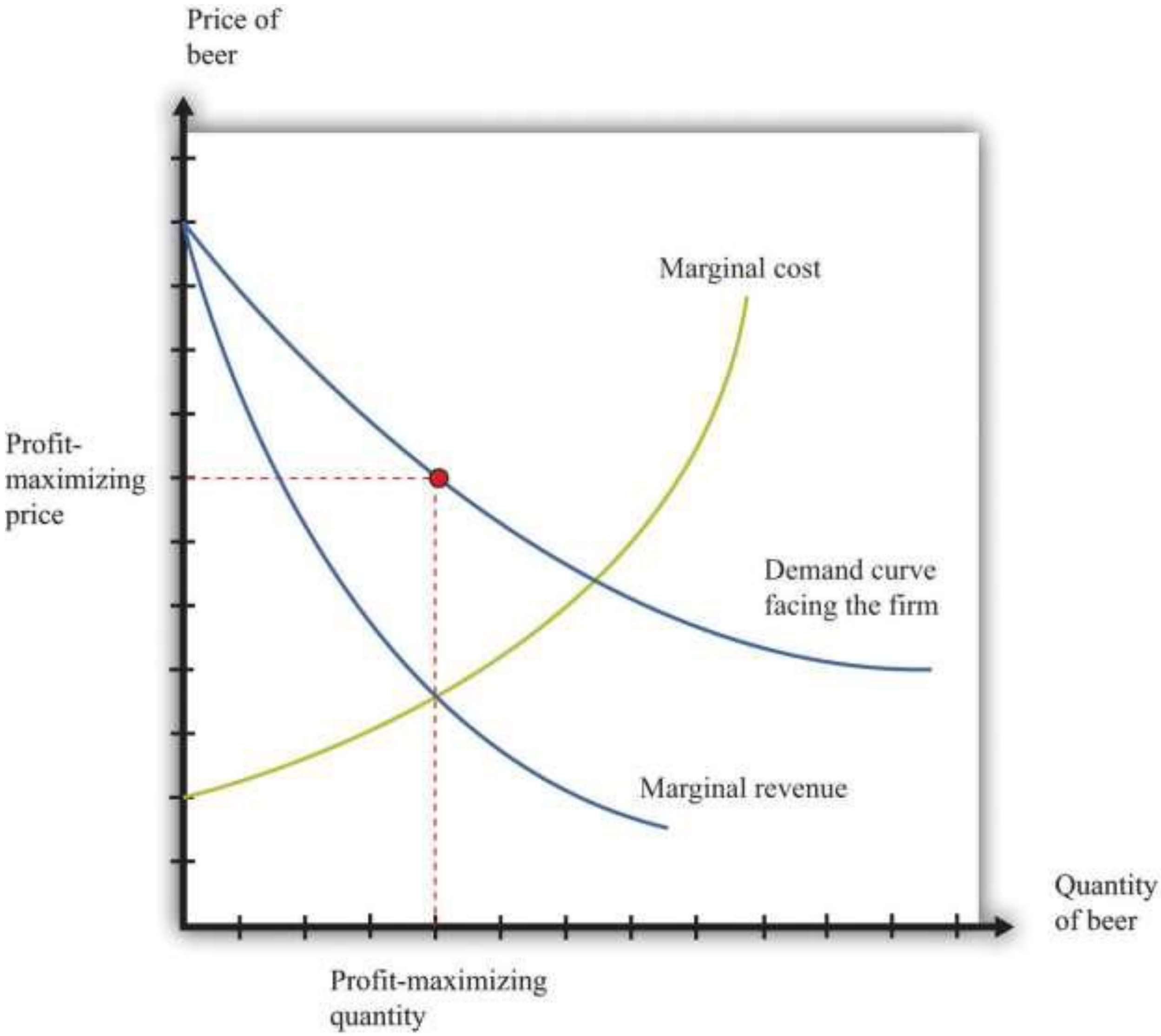
Firms devote a lot of resources to establishing, protecting, and increasing their market power through advertising and other forms of marketing. Many firms spend a great deal of money on developing their brand image and brand awareness. Sporting goods companies such as Nike and Adidas are classic examples. When customers are loyal to brands, firms have market power.

Pricing with Market Power and the Monopoly Outcome

The managers of a monopoly firm must pick the point on the demand curve that will maximize the firm's **profits**—the total revenues of the firm minus its costs of producing its output. We can think about the firm choosing either its level of output or its price. If a firm chooses how much to produce, then the price of its product is determined from its demand curve. (Remember that the demand curve facing the firm and the market demand curve are the same thing for a monopolist.) If a firm chooses a price for its product, then the quantity produced is determined by the demand curve. For now, we tell the monopoly story assuming it chooses the quantity of output.

When a firm is maximizing its profit, producing a little more or a little less output will not increase the firm's profit. This means that the extra revenue from producing one more unit of output is exactly equal to the cost of producing one more unit of output: marginal revenue equals marginal cost. We call this level of output the *profit-maximizing quantity* and illustrate it in Figure 14.3 "The Monopoly Outcome". The price the monopolist sets, termed the *profit-maximizing price*, can then be read from the demand curve. Once a monopolist determines how much to produce, the price of its output is determined by the maximum amount that consumers are willing to pay for the good.

Figure 14.3 The Monopoly Outcome



A monopolist produces a quantity such that marginal revenue equals marginal cost. The price is determined by the demand curve.

Toolkit: Section 17.15 "Pricing with Market Power"

The principles of pricing with market power are explained in more detail in the toolkit.

Distortions Due to Market Power

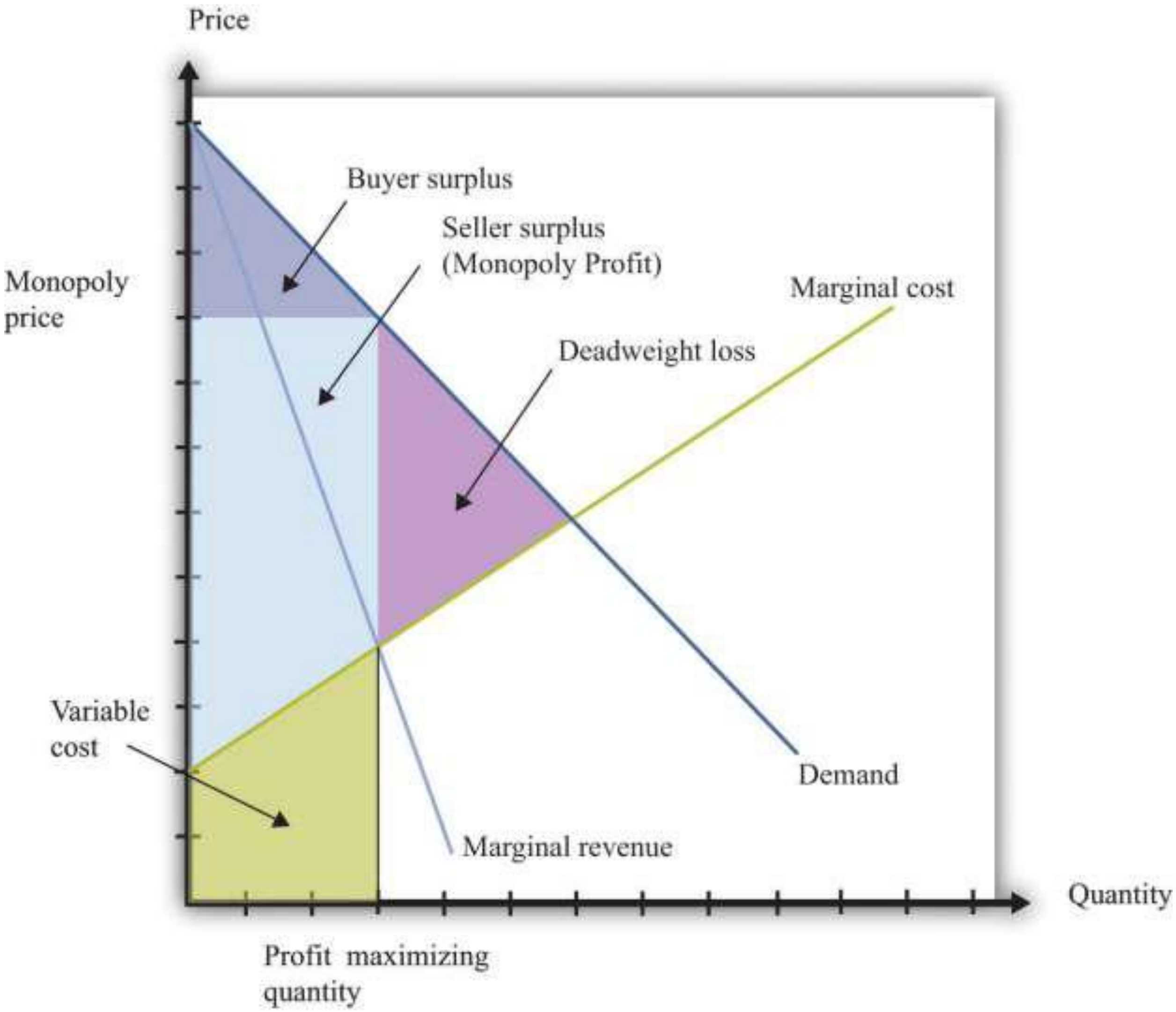
We have competition policies in market economies because market power often leads to an outcome that is not efficient. To understand this distortion, compare the monopoly outcome with the competitive market outcome we saw earlier.Figure 14.4 "Distortions Due to Market Power" is Figure 14.1 "The Competitive Market Outcome" with some areas of the figure labeled.

- We indicate the buyer surplus—the area between the price set by the monopolist and the demand curve. Even though there is a monopolist in this market, buyers can still enjoy some surplus if their **marginal valuation** of the good exceeds the price they pay. So while the price may be too high in the monopoly case, there still remains some surplus flowing to buyers.
- There is an area labeled **variable cost**—the total area under the **marginal cost** curve.
- There is an area labeled seller surplus (monopoly profit)—the difference between revenues at the monopoly price and variable costs. The total revenue received by the monopolist is the price times the quantity sold. This is a rectangle. Monopoly profit, as shown in the figure, equals this rectangular area minus variable costs.
- The difference between the competitive and monopoly outcomes creates a social loss—the triangular area labeled deadweight loss in Figure 14.4 "Distortions Due to Market Power". The demand curve reflects consumers' marginal valuation of the good. Below the competitive quantity, this marginal valuation is greater than the marginal cost of producing the good. This means that there are potential gains from trade that are not being realized.

Toolkit: Section 17.1 "Individual Demand", Section 17.11 "Efficiency and Deadweight Loss" and Section 17.14 "Costs of Production"

You can review in the toolkit (1) the correspondence between demand and marginal valuation, (2) the meaning of efficiency and deadweight loss, and (3) the definitions of variable cost and marginal cost (and the relationship between them).

Figure 14.4 Distortions Due to Market Power



A monopolist produces a quantity that is less than socially optimal: there is a deadweight loss.

Comparing Figure 14.1 "The Competitive Market Outcome" with Figure 14.4 "Distortions Due to Market Power", we see that—relative to the competitive outcome—the monopolist charges a higher price and produces a lower quantity. Since the competitive outcome was efficient, the monopoly outcome is inefficient. Households in the economy would be better off if the monopolist produced at the competitive level of output and sold at the competitive price.

Competition Policy toward Monopolies

When a firm has market power, there is a distortion in the market: prices are too high, and output is too low. Antitrust laws are designed to address precisely this situation.

Antitrust

By virtue of the Sherman Antitrust Act of 1890, the US government can take legal action to break up a monopoly. In 1902, President Theodore Roosevelt used the Sherman Antitrust Act as a basis for trying to break up the monopolization of railway service in the United States. The resulting legal case, known as *Northern Securities Co. v. United States*, involved two key elements: restraint of trade and interstate commerce.

Multiple providers of rail service had joined together in the Northern Securities Company, resulting in a restraint of trade. The Supreme Court decision [1]provides direct discussion of the anticompetitive nature of the creation of this company, noting that the company had been set up to eliminate competition among the different providers. Further, this restraint of trade had interstate implications because the railway lines themselves were not contained within any one state. This was a key point of jurisdiction: the Sherman Antitrust Act spoke directly about the effects of market power on interstate trade. Under the US Constitution, interstate commerce is part of the responsibility of the federal government.

Another famous antitrust case decided by the Supreme Court (http://supreme.justia.com/us/221/1/case.html) centered around the breakup of the Standard Oil Trust in 1911. [2] John D. Rockefeller founded Standard Oil in 1870 soon after the discovery of oil in Ohio. By the late 1870s, the company controlled nearly 90 percent of the refineries in the United States. Shortly thereafter, the Standard Oil Trust was formed with the idea that the individual shareholders in a group of companies would come together to create a single company. This company would then jointly manage all the companies that joined the trust. In doing so, a monopoly was created. (This, by the way, is where the term *antitrust* comes from.)

Once again, the issue of jurisdiction played a role. Standard Oil was originally deemed a monopoly by the Ohio Supreme Court, which ordered it to be broken up. Instead, the company was simply reorganized as a New Jersey–based corporation. Because the company had moved to another state, the Ohio ruling became irrelevant, and the federal government had to step in to prevent unlawful interstate trade. The court ruled that the trust be dismantled. In the modern era, similar concerns arise across national borders: if two companies merge in Europe, for example, US antitrust authorities still take an interest, and vice versa.

These cases may seem like ancient history. Yet they remain very relevant today for at least two reasons. First, the legal system relies heavily on precedent, meaning that cases decided today depend on past rulings in similar cases. The actions of the Supreme Court in these two cases created a significant precedent for the present-day antitrust actions of the federal government. Second, antitrust actions continue to this day. A recent case again alleging a violation of the Sherman Antitrust Act was brought against Microsoft Corporation by the US government in 1998. Microsoft was charged with monopolizing the software market through its Windows operating system. As stated in the case, "Virtually all major PC manufacturers find it necessary to offer Microsoft operating systems on most of their PCs. Microsoft's monopoly power allows it to induce these manufacturers to enter into anticompetitive, long-term licenses under which they must pay royalties to Microsoft not only when they sell PCs containing Microsoft's operating systems, but also when they sell PCs containing non-Microsoft operating systems." [3]

One noteworthy element of the case concerned the definition of the market. After all, there are competing operating systems to Microsoft, most notably Apple Computer's operating system. If the court viewed these products as substitutes for Microsoft's operating system, then the claim that Microsoft was a monopolist would be less clear. Interestingly, the court ruled that Apple was not in the same market as Microsoft.

Price Discrimination

Because there are people willing to pay more than the marginal cost, it seems as if the monopolist in our earlier example is leaving some money on the table. Is there a way for the monopolist to make higher profits? The answer to this question hinges on an assumption that was implicit in our analysis: we suppose that the monopolist has to sell every unit of the good at the same price. In many cases, however, firms sell different units of their product at different prices. This is known as price discrimination, and it can arise either because (1) a firm sells to different customers at different prices or because (2) a firm sells different units at different prices to the same customer. There are numerous examples of both kinds of price discrimination. An example of the first is discounts offered to senior citizens or students. An example of the second is quantity discounts such as "buy two, get one free."

Firms have developed many creative forms of price discrimination, and we could easily fill a whole chapter with them. Our main focus here, however, is to understand how price discrimination operates in a monopoly market. Let us think about a case where the monopolist faces a unit demand curve: each buyer either purchases a unit of the good or does not buy the good at all. The downward-sloping demand curve comes from the fact that different individuals have different valuations of the good.

Toolkit: Section 17.1 "Individual Demand"

You can review unit demand in the toolkit.

Now imagine what the monopolist would do if she knew everyone's individual valuations and was also able to charge a different price to different buyers. In this case, we obtain a remarkable result. If the monopolist could sell each unit at any price she wanted, then she would charge each individual his valuation of the good, with striking consequences.

- The monopolist captures the entire buyer surplus. Compared to the competitive case, there is a redistribution of surplus from buyers to the monopolist.
- The deadweight loss from the monopoly is eliminated. The monopolist produces the same level of output as in a competitive market because it is worthwhile to sell to anyone whose valuation exceeds marginal cost.

Buyers obtain no surplus because each buyer surplus equals his valuation minus the price he pays—which we have just said is equal to his valuation. Each buyer is actually indifferent about buying or not buying the good. (If this seems odd, then you can imagine that the monopolist charges a price slightly below individuals' valuations, so buyers obtain only a tiny amount of surplus. The conclusion is the same.) So where does their surplus go? It shows up as monopoly profits. By price discriminating, the monopolist captures all the gains from trade in this market. And where do these monopoly profits eventually go? They flow to the owners of the monopoly. The buyer surplus that is taken away by the price-discriminating monopolist is eventually returned to the household sector as dividends from the firm. However, it is not the same households who gain and lose surplus. Not everyone is a shareholder in the monopoly. Monopoly power causes a redistribution from the monopolists' customers to its shareholders.

A situation like the one we have described, where a monopolist knows the valuations of all her customers, may seem a bit farfetched. Still, in situations where a monopolist negotiates individual prices with buyers, she will do her best to guess their valuations. For example, think of a used car dealer. When a prospective customer arrives on the lot, the dealer will typically engage the customer in conversation in an attempt to learn where he lives, what kind of job he has, and so on. Such information helps the dealer guess the buyer's valuation.

Other Kinds of Price Discrimination

There are many kinds of price discrimination.

- **Explicit segmentation.** The monopolist may be able to divide the market into identifiable segments and charge different prices to different segments. For example, movie theaters often offer student and senior citizen discounts. As another example, businesses often charge different prices in different countries. The idea is that the monopolist can charge higher prices to segments that are less price sensitive.
- **Self-selection.** If a firm cannot explicitly identify different segments, it can set a menu of prices and allow customers to sort themselves into different groups. When Apple introduced the iPhone, it originally charged a high price, knowing that impatient customers would buy immediately. It then dropped the price to capture the more patient, more price-sensitive customers. Discount coupons are another example. Individuals with more time to spare will clip and redeem coupons—and these individuals will typically also be more sensitive to price.

From the perspective of a firm, the biggest danger of price discrimination is the possibility of arbitrage. If you are selling your product more cheaply to some than to others, you don't want someone to buy at the cheap price and then resell to your higher-value customers. This is a particular issue for pharmaceutical companies that charge different prices in different countries. If you need proof of this, look at the spam in your e-mail. The chances are very good that you have recently received an e-mail offering you pharmaceuticals at Canadian prices.

Competition Policy on Price Discrimination

The Robinson-Patman Act of 1936 directly forbids certain forms of price discrimination. The following is from the act, with our emphasis added: "It shall be unlawful for any person engaged in commerce…to *discriminate in price between different purchasers of commodities of like grade and quality*, where either or any of the purchases involved in such discrimination are in commerce…and *where the effect of such discrimination may be substantially to lessen competition or tend to create a monopoly in any line of commerce*." [4]

The type of price discrimination we have discussed certainly involves discrimination among different purchasers. What is less clear is whether such actions would "lessen competition or tend to create a monopoly." In our earlier example, price discrimination actually makes the market more efficient, and in general, price discrimination can increase or decrease efficiency. There are no simple guidelines, and each case must be examined on an individual basis.

Price discrimination also occurs in business-to-business transactions between firms. In a recent price-discrimination case involving Volvo, a car dealer charged that other dealers had obtained deeper discounts (price concessions) that permitted them to be more competitive. [5] This seems to fit the requirements of the Robinson-Patman Act because Volvo trucks are a homogeneous good (Volvo trucks in this case), and the practice allows one dealer to undercut another.

A jury trial led to an award of $4.1 million to the injured company. Not surprisingly, Volvo appealed the decision, partly because, so it claimed, the other dealers were not in direct competition with the dealer filing the suit. Once again, you can see the critical significance of defining the market. The Supreme Court eventually decided the case in January 2006 in favor of Volvo.

KEY TAKEAWAYS

- A monopoly occurs when there is a single seller, called the monopolist, in a market.

- A monopolist produces the quantity such that marginal revenue equals marginal cost. This is a lower level of output than the competitive market outcome.
- The government has the legal authority to break up monopolies and forbids price discrimination.

CHECKING YOUR UNDERSTANDING

1. Think of two goods or services you have bought recently. Were close substitutes available? Do you think the producer and the retailer of those products had a lot of market power?
2. Looking at Figure 14.4 "Distortions Due to Market Power", why do buyers have any surplus?

[1] Cornell University Law School, Legal Information Institute, "Northern Securities Co. v. United States (No. 277)," accessed March 14, 2011,http://www.law.cornell.edu/supct/html/historics/USSC_CR_0193_0197_ZS.html.

[2] Justia.com, US Supreme Court Center, "Standard Oil Co. of New Jersey v. United States, 221 U. S. 1 (1911)," accessed March 14, 2011,http://supreme.justia.com/us/221/1/case.html.

[3] A full description of this aspect of the case is contained in John Kwoka and Lawrence White, eds., *The Antitrust Revolution* (New York: Oxford University Press, 2004).

[4] Cornell University Law School, Legal Information Institute, "§ 13. Discrimination in price, services, or facilities," accessed March 14, 2011,http://www.law.cornell.edu/uscode/15/13.html.

[5] See Cornell University Law School, Legal Information Institute, "Volvo Trucks North America, Inc. v. Reeder-Simco GMC, Inc. (04-905) 546 U.S. 164 (2006) 374 F.3d 701, reversed and remanded," accessed March 14, 2011,http://www.law.cornell.edu/supct/html/04-905.ZS.html.

14.2 Patents and Copyright

LEARNING OBJECTIVES

1. What is the role of the patent and copyright systems?
2. What factors determine how long patent protection should last?
3. What is the commitment problem associated with the patent system?

In the introduction to this chapter, we mentioned the breakup of Napster, a company that facilitated the sharing of music. Napster provided file-sharing software that allowed computer users to share music files over the Internet. A music file, like any other computer file, is simply information: a collection of bits and bytes. It costs nothing to make a copy of a music file. What Napster did, in other words, was to make it easier for music files to be distributed at marginal cost. You might think that the antitrust authorities would have been delighted. But the argument in this case was that there was an infringement of the rights of the music producers. The courts held in their favor: the law came down on the side of the monopolists. We now consider *why* governments sometimes actively support and promote monopolies.

To understand the Napster case, we begin by recognizing that creating, producing, and recording a new song is a very costly process. First of all, there is the time spent by the band in writing and arranging the song. Professional recordings also need the services of a producer and an expensive recording studio. A great deal of time and resources must also be expended to create an MP3 file of the song. Economists say that the first-copy costs are large. As the name suggests, these are the costs involved in creating the initial version of a good. They are a particular type of entry cost. Goods that involve a large amount of research and development or other intellectual input, such as books, computer software, and pharmaceutical products, have large first-copy costs.

As we just pointed out, though, once a song is produced, it can be *reproduced* at zero cost. The fixed costs of producing a song are very large, whereas the variable costs are zero. Perhaps you can now see the problem. If songs were sold in a competitive market, their price would be zero. Producers of the music would earn no revenues. Composers would earn no money. In this world, no one would have an incentive to produce music unless they were doing it purely for their own pleasure.

Similar tensions can be found in many other industries. Nearly anything that can be stored as a computer file has high first-copy costs accompanied by low variable costs. A newspaper article takes time to research and write but can be copied at zero marginal cost. Computer software can be very expensive to develop but—once created—can be copied at no additional cost. Pharmaceutical compounds are very expensive to develop: they first involve the work of highly trained research scientists in expensive laboratories and then require years of testing on animals and humans. Once a drug has been developed, however, it is often quite cheap to produce.

If the antitrust authorities forced newspapers, software developers, and pharmaceutical companies to sell at marginal cost, these firms would not earn enough revenues to justify their initial investment. Instead, such firms are permitted to sell at above marginal cost. More than that, the government actively bestows monopoly power. It does so through legal protections for inventions and created works, known as patents and copyrights.

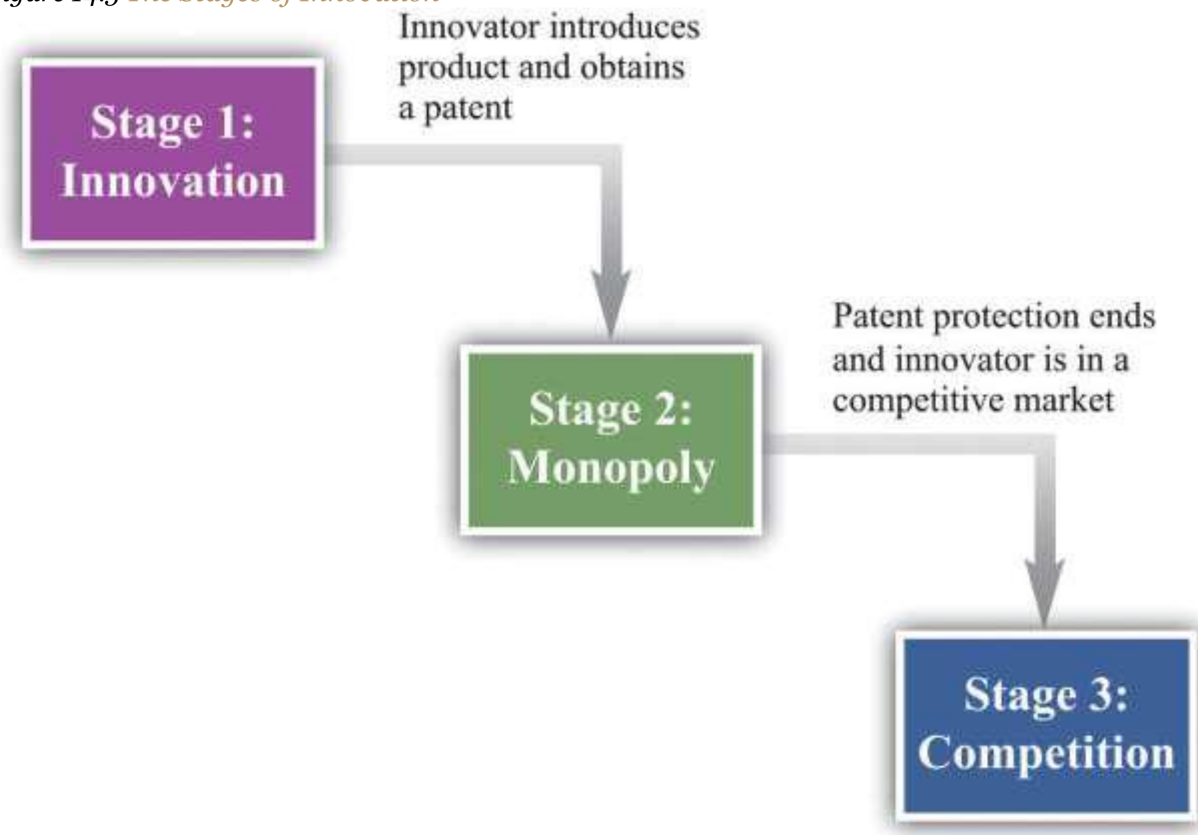
The Decision Problem of an Innovating Firm

To understand how patents and copyrights work, we think about a firm contemplating an *innovation*—the introduction of a new product or a new means of production into a market. The firm's decision involves several stages, as shown in Figure 14.5 "The Stages of Innovation":

1. Do we innovate or not? A decision to innovate is a decision to incur certain costs: the costs of research and development and the costs of entering a market.
2. How much should we produce during the years of patent protection?
3. What do we do once the patent protection ends?

Our ultimate goal is to evaluate the innovation decision at the first stage, but to do so, we must start at the end and work backward.

Figure 14.5 The Stages of Innovation



The Final Stage: Competition

Once a firm's patent expires, other firms can produce a similar or an identical product. The firm will then be operating in a competitive market and can no longer expect to gain any particular advantage from its innovation. When the patent for a pharmaceutical product expires, for example, other companies can step in and produce chemically identical copies of the product, known as*generics*.

In a competitive market, we expect the price of the product will decrease until it equals marginal cost. For this reason, the innovating firm cannot anticipate making very much profit at this stage. For simplicity, we can think of the firm making no profits. Although there may be some advantage in being the original producer of a product, any excess profits that remain after the patent expires are unlikely to be substantial. More precisely, the firm would earn no more than a "normal" level of profit—the same as it could earn in any other activity. Such normal profits would not provide any benefit to justify the initial innovation, so we can ignore them.

The Middle Stage: Patent Protection

If the innovating firm is going to make profits to justify the costs of developing its product, these profits must come in the middle stage when the firm has patent protection. During this period, the firm has monopoly power by virtue of the patent. We know how the firm behaves in this situation.

- It will produce a level of output such that marginal revenue equals marginal cost.
- It will set the price equal to the market's willingness to pay for this output level.

This is exactly what we saw earlier in Figure 14.4 "Distortions Due to Market Power". The monopolist produces less output than is efficient and earns monopoly profits.

Notice that a firm's decision about how much to produce and about what price to set does *not* depend on the costs that it paid for researching and developing its product. After the firm gets to this second stage of its decision, those costs are in the past. They are sunk costs. They have no influence on the marginal cost of production and the price/output decision of the monopolist.

To calculate the total profit that a firm earns in this monopoly stage, we must do two things: (1) calculate the firm's profit in each year and (2) add these profits over the entire time that the firm has patent protection. The firm's profits in any given year are given by

profits = revenues − total cost = revenues − variable cost − fixed operating cost.

The area shown as "monopoly profit" in Figure 14.4 "Distortions Due to Market Power" corresponds to revenues minus variable costs. (Businesspeople and accountants call this a firm's *profit contribution*.) In any given year, the monopoly will also typically incur some costs of operation in any given year that are constant, irrespective of how much output it produces. Examples include rent on a building and other long-term contracts. These are the firm's fixed operating costs, which also must be subtracted from the firm's revenues to calculate its profits.

Because the monopolist earns profit in each year of its patent protection, we add these profits together. We do so using the tool of **discounted present value**. This calculation takes into account that money earned in the future is less valuable than money today whenever the rate of interest is positive. Thus the proper measure of the profits at this stage is the *discounted present value of the sum of the profits made* *during the period of patent protection*. Factors that would increase the discounted value of a firm's profits include the following:

- Lower marginal costs
- More inelastic demand
- More years of patent protection
- Lower interest rates

Toolkit: Section 17.4 "Choices over Time"

You can review discounted present value in the toolkit.

The First Stage: Innovation

We are finally in a position to evaluate whether or not a firm should innovate. The gains from innovation are measured by the present discounted value of the flow of profits. A firm must compare these gains to the costs of innovation to determine whether or not the innovation is worthwhile. These innovation costs are determined by the costs of the research and development (R&D) process together with any other costs of market entry. For example, a firm must pay scientists and engineers, fund research laboratories and R&D departments, and so on. Having done the hard work of analyzing stage two, the decision for stage one is straightforward. The firm should follow this rule: "innovate if the discounted present value of profits is greater than the costs of innovation." The firm should innovate as long as the monopoly profits it will earn in the second stage (appropriately discounted) are greater than the costs of innovation in the first stage.

Think again about a pharmaceutical firm. Such firms spend an enormous amount of money on the research and development phases of new pharmaceutical compounds. This occurs in stage one. Once the product goes to market, however, the costs of development are sunk and have no effect on the firm's profits during stage two. The same point applies to the production of a music CD. The costs of producing, marketing, and distributing a typical CD are estimated to be around $500,000. Of this, at least $100,000 represents the costs of production.

One thing that we have neglected in our discussion is that the payoff from research and development efforts is typically uncertain. Many promising pharmaceutical compounds turn out, on further testing, to be ineffective or have unacceptable side effects. A band recording a new song cannot know for sure if it will sell hundreds of copies, thousands of copies, or millions of copies. The decision about whether or not to innovate must be based on a firm's best estimates of the expected value of its profits.

An Example

Table 14.1 "Calculating the Discounted Present Value of Expected Profits"provides a numerical example of the innovation decision. The first year is the innovation stage: we suppose the cost of innovation is $150. The firm earns no revenues in that year and incurs no costs, so its first year profits are −$150. In the second year, the firm finds out if its innovation was successful. We suppose there is a 50 percent chance that it is successful, in which case the firm has monopoly power for the second and third years. It earns the revenues and incurs the costs listed in the third and fourth columns of the table. If the innovation is unsuccessful, it earns no profits. After the third year, the firm earns no profits. Suppose finally that the interest rate is 10 percent.

Table 14.1 Calculating the Discounted Present Value of Expected Profits

Year	Innovation Cost ($)	Total Revenues ($)	Total Operating Costs ($)	Profit If Successful ($)	Profit If Not Successful ($)	Expected Value of Profits ($)	Discounted Present Value of Expected Profits ($)
1	150	0	0	−150	−150	−150	−150
2	0	200	68	132	0	66	60
3	0	400	158	242	0	121	100

The profits in the second year if the innovation is successful are $132. The expected value of profits in that year is therefore given by (0.5 × $132) + (0.5 × 0) = $66. These must be discounted back one year using the 10 percent interest rate. In other words, we divide the second-year profits by 1.1. The discounted value of the second-year profits in the first year is therefore $60. The expected value of the third-year profits, by a similar calculation, is $121. These must be discounted back two years, all the way to the first year. To do so, we first divide by 1.1 to get the value of expected third-year profits in the second year ($121/1.1 = $110). We then discount this back another year by dividing by 1.1 again. The discounted value of expected third-year profits in the first year is therefore 110/1.1 = $100.

We can now legitimately add together the numbers in the last column, and we find that the discounted present value of the firm's stream of profits is $160. This exceeds the cost of innovation ($150), so the firm should go ahead with the project. It expects to earn $10 from the project.

The Role of Patents

Imagine for a moment that we went through the same analysis in the previous section but without patent protection—that is, suppose that as soon as a firm innovates and introduces a new product into a market, it can immediately be copied and produced and marketed by any other firm. This means that the second stage is completely eliminated: the market goes straight from innovation to competition.

This has an apparent benefit. There is no longer a monopoly in the second stage. We know that monopoly causes inefficiency: the firm sets its price above marginal cost to earn monopoly profits. But there is a problem: the innovation will not occur. The firm innovates only if the discounted present value of profits exceeds the cost of innovation. If we eliminate stage two, then we eliminate the profit flows that justified the innovation in the first place. Without the profits, there will be no innovation. It is the patent protection that provides the *incentive* for innovation.

The trade-off should be clear. Patents

- provide incentives for innovation, but
- create monopoly power and hence distortions.

The gain from patent protection is that it provides the basis for the second stage of the innovation process. Without this protection, the gains from innovation would not exist. Looking back at Figure 14.4 "Distortions Due to Market Power", this gain comes at the cost of an inefficiently low level of output and a consequent loss of gains from trade.

Lawmakers must trade off this cost and this gain. Under current US law (http://www.uspto.gov), patent protection for most products lasts for 20 years. The optimal length of patents remains an active area of research in economics and an active area of policy concern. It is very hard to maintain the right balance between incentives for innovation and eliminating market distortion.

Commitment

Perhaps the government could both encourage innovation and avoid the monopoly distortion. Suppose that, *after an innovation had been introduced*, the government removed the patent protection. As an example, many individuals in the world suffer from HIV/AIDS but cannot afford medication at current prices. This problem is especially severe in much of sub-Saharan Africa. As a matter of social policy, one would like to have the drug companies first develop HIV/AIDS treatments and then sell those products at marginal cost.

Currently, some pharmaceutical companies have developed treatments that are still protected by patents. Eliminating these patents seems like it might be good social policy, given that the lives of millions of people are at stake, but it would come with a significant cost. If the government ignores patent protection for a particular product today, innovators will suspect that the government can no longer be relied on to provide patent protection in the future, which would have a huge impact on innovative activity. Thus by failing to provide patent protection for one product, the government risks destroying its reputation for patent protection in general. This is an example of a *commitment problem*. Prior to stage one of the innovation process, the government promises patent protection to provide an incentive for innovation. After stage one is finished and the product is introduced in the market, the government's incentives change: it wants to remove the market distortions. Because the innovation stage is over, the government could potentially renege on its promise of patent protection.

Pharmaceutical companies like GlaxoSmithKline, which produce the antiretroviral drugs used for treating HIV, have in fact offered to make their products available more cheaply. However, pharmaceutical companies have strongly opposed suggestions that their patents not be honored. A better policy proposed by economists and others is a *patent buyout*. This would work as follows. The government would pay the pharmaceutical company a reasonable market price for its patent and then allow other companies to come into the market and produce generics. A related idea is that governments could offer to buy future patents if drug companies came up with treatments for particular maladies. For example, even though malaria is a major killer in the world, pharmaceutical companies have little interest in researching its treatment. The reason is one of harsh market economics: most sufferers from malaria are poor, so a malaria drug would not be very lucrative. If governments offered a substantial patent buyout for a malaria drug, however, firms might find it worth investing in this disease area after all.

International Dimension of Patents

If you have a US patent for a good that you have invented—for example, a new kind of printer—then you are protected *in the United States*. However, US patent law does not protect you if you are selling in other countries. A firm in another country could take your printer apart, analyze how it works (this is called reverse engineering), and then produce the good itself. If the firm tried to sell the good in the United States, you could take it to court, but if it is selling elsewhere in the world, you have very little protection.

Other countries (such as Japan) and regions (such as the European Union) have patent laws that are similar to those in the United States. (Indeed, one of the earliest biotech patents went to the famous biologist Louis Pasteur in 1873 for his method of producing yeast.) These laws are structured, as are those in the United States, to balance the gains from innovation against the costs of monopolization of the market. Thus producers who sell across the world will typically seek patent protection in many different countries.

Still, to the extent that the benefits of innovation flow to purchasers around the world, the innovation-monopoly trade-off is potentially altered. If we think about the stage two monopoly as “the price we pay for the benefits of innovation,” then residents in countries with strong patent laws are paying in part for benefits that flow to individuals and households in other economies. In some countries—China is a leading example—patents and copyrights are not very well protected. This is good news if you want to buy some cheap DVDs in China, but if you are the owner of the rights to these pirated movies, you surely wish that copyright laws were more thoroughly enforced around the world.

KEY TAKEAWAYS

- Patents and copyrights provide innovators with protection from competition so that there is a return to innovation.
- Although a patent system provides protection, it also creates market distortions by granting monopoly power. A patent system should be designed to balance the incentive to innovate against the losses from these distortions.
- After innovation has taken place, the government may be tempted to take away patent protection to avoid market distortions. This is the commitment problem faced by a government. Governments are aware that if they take away patent protection from firms that have already innovated, they will greatly damage the incentives for future innovation.

CHECKING YOUR UNDERSTANDING

1. Draw the diagram for a competitive market where marginal cost is zero. What does the supply curve look like? Who gets the surplus in the market? Use this diagram to explain the tension between innovation and competition.
2. The subsection on patent protection ended with a list of factors that would increase a firm's profits. Explain *why* each of these would cause the present discounted value of the firm's profit flow to increase.

14.3 Markets with a Small Number of Sellers

LEARNING OBJECTIVES

1. How do we predict the market outcome in a market with a few sellers?
2. How does the outcome depend on whether firms set prices or quantities?
3. What are the main tools of competition policy in markets with a few sellers?

So far we have looked at monopolies: markets with a single seller. But as we pointed out earlier, the extent to which a firm is a monopolist really comes down to how we choose to define the market. The market power of all firms is limited, to a greater or lesser degree, by the presence of firms selling competing products. In this section, we examine how the presence of competitors affects the distortions due to market power.

If there are enough competitors to give us perfect competition, then there is no distortion. But what about intermediate cases, where there are a small number of sellers? The market power of Microsoft Corporation is muted by the presence of competitors producing products that are substitutes for Microsoft software. If Microsoft were to triple the price of its Windows operating system, many buyers would switch to Macintosh computers or even start running the free operating system Linux. To the extent that substitute products are available to well-informed consumers, market power is decreased.

Once firms have to start worrying about the strategies of their competitors, decision making can become surprisingly complex. Each choice that a firm makes concerning what goods and services to produce, how to produce them, how to market them, how to price them, and so on, is now complicated by the fact that all of its competitors are making similar choices. This contrasts with a competitive market, where a single firm does not need to consider the behavior of its competitors at all; it only needs to know the market price for its output.

The task of a firm with market power is to choose the point on the demand curve that maximizes its profit. From the firm's perspective, it doesn't matter whether it (1) sets a price and then lets the quantity demanded come from the demand curve or (2) chooses a quantity and then lets the price be whatever the market will bear. Oddly enough, though, the interaction among firms is very different in a world where firms are setting prices compared to one where they choose their level of output. As we go through this section, we shall see why.

Market Outcomes When Firms Set Prices

We start by thinking about the case where firms set prices, beginning with two firms selling exactly the same product. An example is two gas stations at the same intersection, each of which has to choose what price to set for the gasoline that it sells. We first look at the choice of one of these firms and then study what happens when the two firms interact. We make a simplifying assumption, which is reasonable for gasoline retailers, that marginal cost is constant.

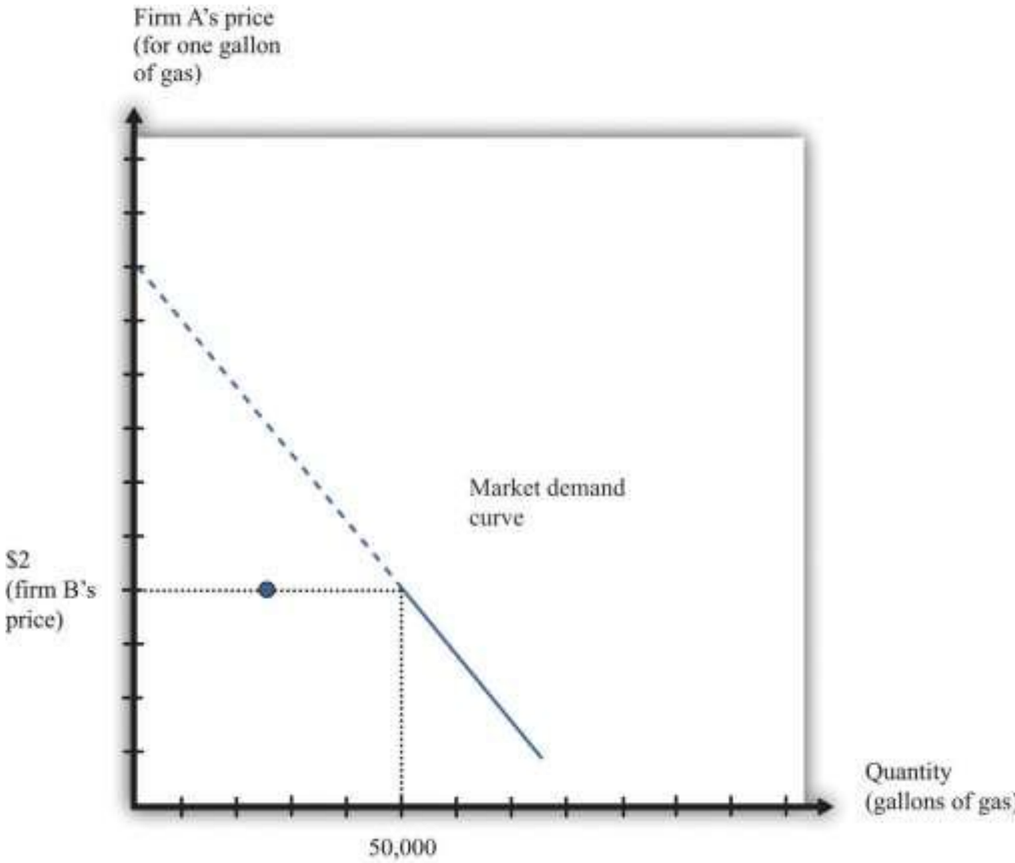
The Pricing Decision of a Single Firm

The first thing we need to know is the shape of the demand curve facing a firm. Let us look at the decision of one gas station (firm A). The owner of this gas station can look across the street and see the price set by its competitor (firm B) across the street. Suppose, for example, that firm B is selling gas at $2 per gallon. What does the demand curve for firm A's gas look like?

- If firm A charges a price greater than $2 per gallon, all its potential customers will go across the street to firm B. The quantity demanded will be zero.
- If firm A charges a price less than $2 per gallon, it can capture the entire market. The demand curve faced by firm A is the same as the market demand curve.
- If firm A sets a price of exactly $2 per gallon, then the two firms will divide up the market. The simplest assumption is to suppose that they will each get half the market.

We illustrate this demand curve in Figure 14.6 "The Demand Curve Facing a Firm, Taking as Given the Price Set by a Competitor".

Figure 14.6 The Demand Curve Facing a Firm, Taking as Given the Price Set by a Competitor



This figure shows the monthly demand curve facing firm A in a market where two gas stations are setting their price per gallon.

If firm B sets a price of $2, what should firm A do? As long as $2 is greater than the marginal cost, firm A makes the most profit if it undercuts firm B a little bit. If it sets a price of $1.99, it can capture the entire market, whereas if it sets a price of $2 it gets only half the market. But exactly the same is true of firm B. For any given price that firm A sets, firm B would do better to undercut it by a penny. Competition provides a strong incentive for firms to cut their prices.

Nash Equilibrium

The discussion so far tells us how one firm will respond to the price of the other, but we don't yet know where the firms will end up. We do not yet know what the equilibrium will look like. Previously, we used the term *equilibrium* in the context of supply and demand, denoting the point where the supply and demand curves intersect. The idea of equilibrium goes beyond this, however: it denotes a situation of balance in which no one has any desire to change.

When we think about strategic situations with a small number of firms, we maintain the idea of an equilibrium where no one wants to change their decision. We use an idea of equilibrium invented by John Nash (a mathematician who won a Nobel Prize in Economics for this and other contributions). The concept of the Nash equilibrium also expresses a sense of balance, but it is applied to strategic situations rather than markets. The key feature of the Nash equilibrium is that no one has any desire to change what is being done.

Toolkit: Section 17.18 "Nash Equilibrium"

The Nash equilibrium is used to predict outcomes in strategic situations, often referred to as "games." In a game, a small number of players (such as firms) interact. Each player chooses an action, and each player receives a payoff (for example, profit). The payoff of a player depends on his chosen action *and the actions chosen by all the other players*. In the Nash equilibrium, two things are true:

1. Each player chooses the action that gives him or her the highest payoff, based on his or her predictions of the other players.
2. Each player's predictions of the actions of the other players are correct.

Nash Equilibrium When Firms Choose Prices (Bertrand Competition)

To see the Nash equilibrium in action, go back to our gas stations. So far, we have seen that—taking firm B's price as given—firm A will want to set a lower price. For example, if firm B sets a price of $2, firm A will set a price of $1.99. But this is not a Nash equilibrium because firm B would then like to do something different. If firm A sets a price of $1.99, firm B will want to set a price of $1.98, and so on. This process will stop only when the firms' prices equal marginal cost. Thus we can make an educated guess: in the Nash equilibrium, each firm sets its price equal to the marginal cost of production. In this equilibrium, both firms earn no profits.

To see that this is indeed an equilibrium, we suppose firm B sets its price equal to marginal cost and then ask if firm A would like to change its price away from marginal cost. Were firm A to set its price greater than firm B, it would get no sales and no profits. This does not make firm A better off. If firm A were to set its price below marginal cost, then it would capture the entire market, but it would make a loss on each sale. It would be selling below the costs of production and thus make negative profit. This certainly does not make it better off. So firm A has no incentive to do anything different. Obviously, the same arguments apply to firm B. If both firms set the price at the marginal cost, neither can change its price and make higher profit.

The game played between two firms producing an identical product and setting prices is called Bertrand competition. The remarkable implication of Bertrand competition is that the predicted outcome (price) will equal marginal cost. With only two firms, in other words, we get the same outcome as with a competitive market. Even with two firms, the market can be very competitive, and there may be no need for antitrust authorities to intervene.

The Prisoners' Dilemma

Households certainly like Bertrand competition because they can purchase goods at marginal cost. The firms, however, clearly do not like the outcome. Both firms understand that they would both be better off if they could only charge more for a good. If they could meet together and collude, then they would want to behave like a monopoly and share the profit between them. But charging the monopoly price is not a Nash equilibrium. If one firm were to set the monopoly price, the other would have an incentive to undercut its rival by charging a slightly lower price.

Here is another way of looking at this problem. Let us again consider a situation where two firms are choosing their prices. To keep things simple, suppose now that the firms must choose between two prices: the (high) monopoly price and the (low) competitive price. We will no longer assume constant marginal cost, so the firms still earn some profit when the price equals marginal cost. There are thus four possibilities: (1) both firms set a high price, (2) both firms set a low price, (3) firm A sets a high price and firm B sets a low price, and (4) firm A sets a low price and firm B sets a high price.

Suppose that Figure 14.7 "The Payoffs (Profits) from Different Pricing Choices" shows us the profits that firm A earns in each case. (There is a similar figure for firm B.) When both firms set high prices, firm A earns profits of $100 (think of this as monopoly profits of $200 that they share.) However, if firm B sets a high price, firm A is better off setting a low price. In this case, firm A gets profits of $120. So if firm B sets a high price, firm A does best by setting a low price. What about if firm B sets a low price? Then firm A gets nothing if it sets a high price and profits of $60 if it sets a low price. Firm A is clearly better off setting a low price in this case as well. No matter what firm B does, firm A should set a low price. The same is true for firm B, so the Nash equilibrium is for both to set a low price. If they could collude, they would both agree to set a high price and earn higher profits. But collusion is not a Nash equilibrium because both firms have an incentive to cheat. This is an example of a prisoners' dilemma game.

Figure 14.7 The Payoffs (Profits) from Different Pricing Choices

Firm A's profits

	Firm B sets high price	Firm B sets low price
Firm A sets high price	100	0
Firm A sets low price	120	60

Firm A and firm B are each choosing to sell at either a high price or a low price. No matter what firm B chooses to do, firm A is better off setting a low price.

Is there any way that the firms might be able to change the incentives so that they can collude? Each firm would like some means of punishing the other if it cheats. If they get to set their prices only once, there is no obvious punishment. But if these firms are competing over a long period of time—as happens in the real world—then more possibilities open up.

Think again about our two gas stations. They might both agree to set a high price for their gas. Then they keep a careful eye on each other. As long as firm B keeps its prices high, firm A is content to do the same. But if firm B ever drops its prices, then firm A can punish it by retaliating. More specifically, suppose the firms adopt the following rules:

- We both set the monopoly price (cooperation).
- However, if one of us ever fails to set the monopoly price (defection), then the other firm will set the competitive price forever after.

Figure 14.8 "The Payoffs (Profits) from Cooperating and Defecting" shows an example of firm A's profits in this case. (This should now be understood as a discounted present value of profits because we are imagining the firms competing for months or years.)

Figure 14.8 The Payoffs (Profits) from Cooperating and Defecting

Firm A's (Discounted Present Value of) Profits

	Firm B cooperates	Firm B defects
Firm A cooperates	500	180
Firm A defects	420	300

Defecting means setting the low price every month. Cooperating means setting the high price as long as the other firm has set the high price in all previous months but switching to the low price every month if the other firm ever sets the low price.

Will the firms have an incentive to follow these rules? Put yourself in the shoes of firm A. If firm B is charging the monopoly price, then you can make a quick profit by undercutting firm B's price. This will generate a gain in that period. But there is a cost: in future periods, firm B will charge the competitive price, and your profit will be driven to $0. For the numbers in the table, it is better for firm A to cooperate if firm B also cooperates. Thus if two firms compete with each other over and over again, they may be able to sustain collusive high prices.

There is something else striking about Figure 14.8 "The Payoffs (Profits) from Cooperating and Defecting". It is a Nash equilibrium for both to cooperate, but it is still *also* a Nash equilibrium for both to defect. Look at the profits of firm A if firm B defects. Firm A earns $300 if it defects but only $180 if it cooperates. If you thought the other firm was going to renege on the agreement, then you would want to do the same thing. Economists say that this game has *multiple equilibria*. Expectations are critical: if each firm expects the other firm to cooperate, then they will indeed both cooperate; if each firm expects the other firm to defect, then they will indeed both defect. This is called acoordination game.

Toolkit: Section 17.18 "Nash Equilibrium"

The prisoners' dilemma game, the coordination game, and other games are discussed in more detail in the toolkit.

One final note: we are showing how firms can (and often do) sustain high prices even in the face of competitive pressures. We are not suggesting that this is what*you* should do if you are ever responsible for setting prices! Conspiring to set high prices is very often a violation of antitrust laws.

Price Competition with Imperfect Substitutes

Up to this point, we supposed that the two firms were producing an identical product. If we think of two firms producing goods that are close—but not perfect—substitutes, we still reach very similar conclusions. For example, suppose there are two pizza restaurants on the same street. If one restaurant undercuts the other's price, it would no longer expect to immediately capture the entire market, but it would still expect to gain a lot of business. In this situation, the arguments that we have just made still apply. Each restaurant would have an incentive in the short run to undercut the other's price. If they compete repeatedly, however, they may be able to sustain high prices.

Market Outcomes When Firms Set Quantities

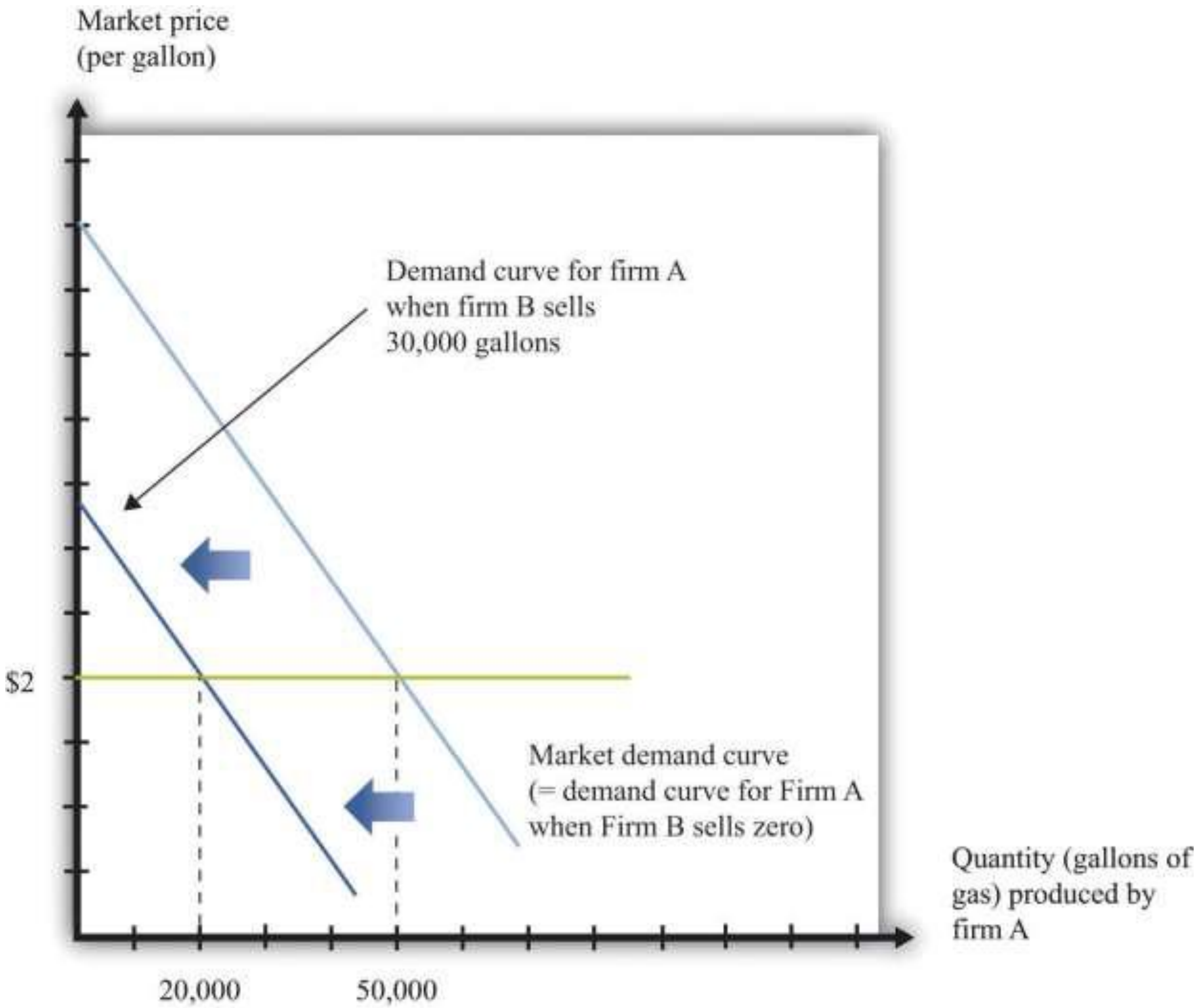
Look again at Figure 14.6 "The Demand Curve Facing a Firm, Taking as Given the Price Set by a Competitor". The demand curve in that figure is based on the idea that if firm A sets its price a little below firm B's price, firm A will capture the entire market. This presumes, though, that firm A can produce enough to supply the entire market. Suppose instead that firm A is unable to supply more than 25,000 gallons of gas per month. Now, if firm B has set a price of $2, then the best that firm A can do is to also set a price of $2. Firm A no longer sees any benefit from cutting its price because it cannot supply any more gas to the market. Similarly, think of the two pizza restaurants. If both restaurants are typically full most evenings, then neither would see a benefit from cutting its price. There is no point in trying to attract your competitors' customers if you cannot then supply them with the goods or services that they want.

This suggests another way in which firms can keep prices high. They can deliberately limit their capacity to change *their own incentives* about price-setting. In effect, this is another way of "changing the game" of Figure 14.7 "The Payoffs (Profits) from Different Pricing Choices". To analyze this kind of behavior by firms, we suppose that they choose their level of output rather than the price they set.

The Capacity Decision of a Firm

We again consider a situation where two firms are competing in the same market. As before, the first step is to determine the demand curve facing an individual firm. Figure 14.9 "The Demand Curve Facing One Firm Shifts to the Left as the Other Firm Increases Its Output" shows our gas station example again. In contrast to our previous analysis, firm B is choosing how much to produce (that is, how much gas to sell) rather than what price to set. As firm B increases its output, the demand curve faced by firm A shifts to the left. If firm B produces nothing, firm A faces the entire market demand curve. If firm B produces (sells) 30,000 gallons, firm A's demand curve is shifted to the left by that amount. You can see that the demand curve faced by firm A has a familiar shape, unlike the odd demand curve in Figure 14.6 "The Demand Curve Facing a Firm, Taking as Given the Price Set by a Competitor".

Figure 14.9 The Demand Curve Facing One Firm Shifts to the Left as the Other Firm Increases Its Output



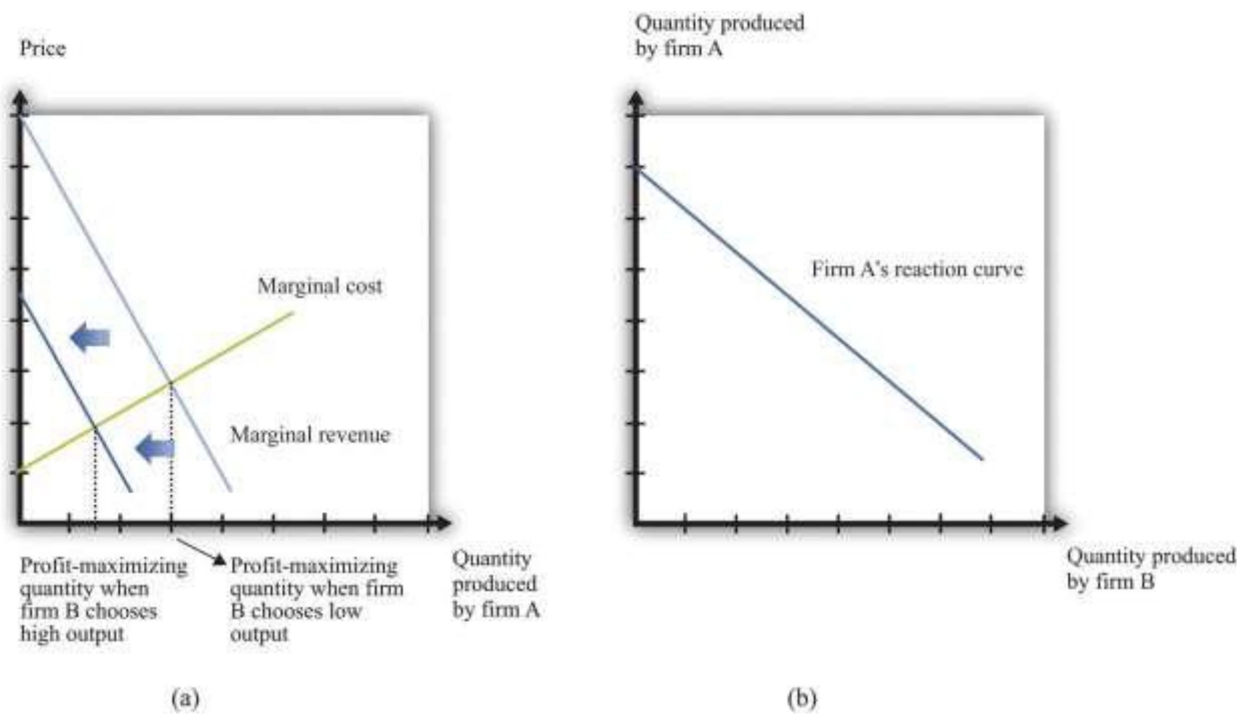
As firm B produces more output, the demand curve faced by firm A shifts to the left.

When its demand curve shifts to the left, firm A's marginal revenue curve also shifts to the left. Figure 14.10 "Firm A's Profit-Maximizing Choice of Output as Firm B Changes Its Level of Output" shows what happens. Note that the downward-sloping curves here are now marginal revenue curves, not demand curves. We omitted the demand curves to keep the diagram from being too cluttered.

Part (a) of Figure 14.10 "Firm A's Profit-Maximizing Choice of Output as Firm B Changes Its Level of Output" shows two marginal revenue curves for firm A associated with different levels of output for firm

B. An increase in firm B's output causes the marginal revenue curve facing firm A to shift to the left. How will firm A respond? As always, we know it will produce a level of output such that marginal revenue equals marginal cost. So as the marginal revenue curve shifts inward, firm A will produce less output. If firm B produces more output, firm A will produce less. This response of firm A to firm B is shown in part (b) ofFigure 14.10 "Firm A's Profit-Maximizing Choice of Output as Firm B Changes Its Level of Output". Here the output of firm B is on the horizontal axis, and the output of firm A is on the vertical axis. The downward sloping curve, sometimes called a reaction curve, shows us the output of firm A for every level of output of firm B.

Figure 14.10 Firm A's Profit-Maximizing Choice of Output as Firm B Changes Its Level of Output



As firm B produces more output, firm A's marginal revenue curve shifts to the left (a), and firm A responds by producing less output (b).

Toolkit: Section 17.18 "Nash Equilibrium"

A reaction curve is used to help find the equilibrium in a strategic situation. It shows what happens to one player's best strategy when the other player's (or players') strategy changes.

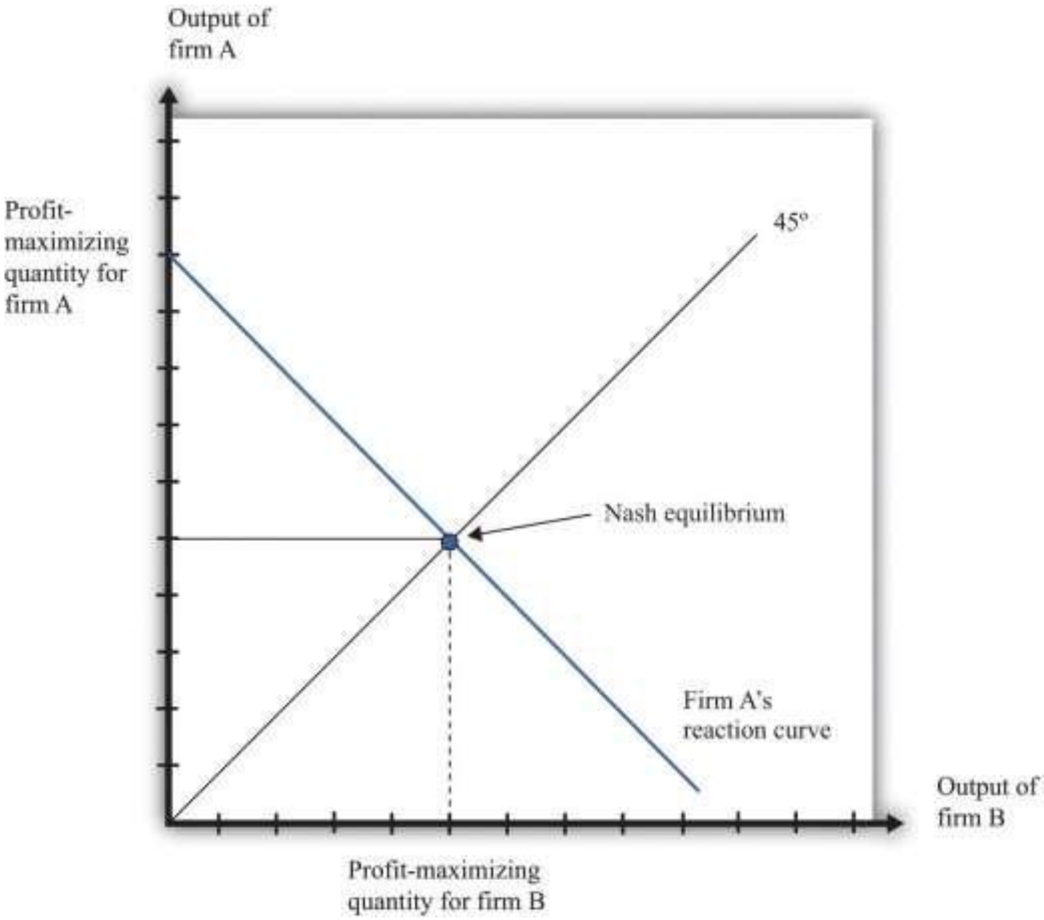
Nash Equilibrium Revisited

We can now predict what will happen in this market. To simplify matters, we assume that the two firms are identical. This will make it easier to find a Nash equilibrium. In a Nash equilibrium, the following things are true.

- Firm A is choosing the level of output that maximizes its profits, which is based on its prediction of how much output firm B is producing.
- Firm A's prediction about firm B's level of output is correct.
- Firm B is choosing the level of output that maximizes its profits, which is based on its prediction of how much output firm A is producing.
- Firm B's prediction about firm A's level of output is correct.

If the two firms are identical, they will produce the same levels of output in the Nash equilibrium. Then, as shown in Figure 14.11 "Nash Equilibrium for Quantity Game", the equilibrium level of output corresponds to the intersection of the reaction curve and the 45-degree line. It is at this point, and only at this point, that all four conditions that we have listed hold. To understand this, put yourself in the position of firm A. You make a forecast about how much firm B will produce. Suppose you correctly forecast firm B's profit-maximizing quantity. Then you will respond with your own profit-maximizing quantity. This is the point labeled as the Nash equilibrium in the figure. But why should you predict that quantity for firm B? That quantity is in fact its profit-maximizing choice, given what you are doing. The beliefs that each firm has about the other's actions are consistent, and indeed they are *self-enforcing*.

Figure 14.11 Nash Equilibrium for Quantity Game

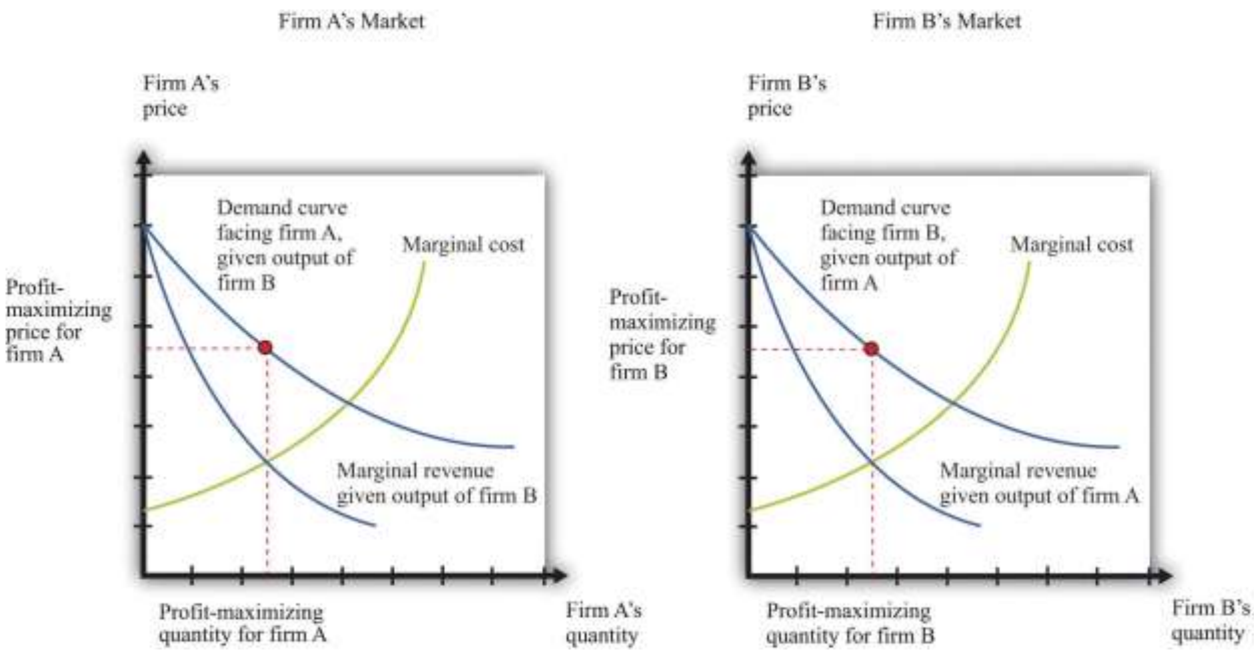


The Nash equilibrium when both firms are identical occurs at the level of output where the reaction curve crosses the 45-degree line.

Determining Prices

Now that we know how firms choose capacity (quantity), how are prices determined? The answer is easy: prices come from the demand curve. If the two firms are producing identical products, the price comes from the market demand curve, given the *total* output of the two firms. This is similar to what we did in the monopoly case: given the output level, we turned to the demand curve to find the price. If the two firms are producing products that are not perfect substitutes, then the analysis is similar. However, there is not one demand curve in this case; there are two. Each firm faces a demand curve that depends on the output of the other firm, as shown in Figure 14.12 "The Markets for Both Firms".

Figure 14.12 The Markets for Both Firms



Firm A correctly predicts firm B's profit-maximizing level of output, and firm B correctly predicts firm A's profit-maximizing level of output.

Inefficiency

When firms are choosing the price to set, it is possible for competition between two firms to drive prices all the way down to marginal cost and eliminate all monopoly inefficiency. This is what we see with Bertrand competition. When firms choose quantity rather than price, the effects of competition are much weaker. Look again at Figure 14.12 "The Markets for Both Firms". You can see that both firms are setting price in excess of marginal cost: there is still a distortion due to market power.

Competition between the firms does matter, however. Increases in firm B's output, for example, shift firm A's demand curve to the left. As a consequence, firm A ends up choosing a lower price than it would otherwise. Competition from firm B helps keep firm A's prices low. By similar reasoning, competition from firm A helps keep firm B's prices low.

In an ideal world (from their point of view), these firms would both limit their output further to get closer to the monopoly outcome. Exactly the same kind of strategies that we discussed earlier could come into play: firms that compete repeatedly over a long period of time might tacitly agree to reduce output further, punishing any defection by increasing output and cutting prices.

Competition Policy with a Small Number of Firms

In the United States, there are two aspects of policy when there are a small number of firms. First, a small number of firms in a market may be able to collude to set high prices. Antitrust laws can be used to punish such collusive behavior. Second, if there are a small number of firms, they may want to merge and become one large firm. Such mergers must be approved by the US government.

Collusive Behavior

The Sherman Antitrust Act is not only used against existing monopolies but is also more generally applied to groups of firms that act to jointly monopolize a market. If multiple firms act collusively to exert market power, they may fall foul of the act. Indeed, in the Standard Oil case, the key issue was the way the Standard Oil Trust brought multiple firms together into a single decision-making unit.

One form of collusive behavior occurs when firms come together to jointly decide on output levels and/or the price of the goods and services they sell. Occasionally, managers of firms are foolish enough to get caught on record making such proposals. In a famous incident a couple of decades ago, the CEO of American Airlines, Robert Crandall, proposed a 20 percent price increase to his counterpart at Braniff Airlines. Unfortunately for Crandall, the conversation was taped. (Interestingly, Crandall was not actually guilty of a violation of the Sherman Antitrust Act because no actual price-fixing took place.) More often, such agreements are likely to be *tacit*. One firm may try raising its prices, to see if others will follow. It can be very hard for the antitrust authorities to determine if price-fixing is actually occurring.

Other countries also have laws and agencies that seek to prevent collusion by firms. For example, in April 2008, the Office of Fair Trading in Britain charged two tobacco companies of colluding with supermarkets to set high prices for cigarettes. Specifically, the Office of Fair Trading said that the companies had set up arrangements "linking the retail price of a manufacturer's brand to the retail price of a competing brand of another manufacturer." [1]

Another form of collusive behavior occurs under the heading of "bid rigging." Suppose only a few dairies provide milk to all schools in a region of the country. The schools set up auctions to decide which dairies will supply milk to different school districts. There is one auction for each district, and several dairies compete to provide milk to the different districts. The situation seems at first glance to be very competitive. We might expect the dairies to compete with each other in all the auctions, with this competition driving down milk prices for the schools. But imagine instead that the dairies agree ahead of time to divide up the districts. So for example, if there are three districts and three dairies, the dairies might agree that dairy 1 would win the auction in one district, dairy 2 in another, and dairy 3 in the third. They do this by each putting in very high bids (meaning they would charge a lot for a service), in the districts where they do *not* want to win the auction, allowing one dairy to win the auction and still charge a high price. Exactly such a scheme occurred in Texas, and an individual was charged with a felony.

Mergers and Acquisitions

Suppose you heard that Apple Computer and Microsoft Corporation were proposing a merger. You can perhaps imagine Steve Jobs and Bill Gates telling us how much we as consumers would benefit from this merger. They would say that Apple and Microsoft could combine the best features of their operating systems. The two companies could avoid costly duplication of research, so they would be able to provide goods more cheaply. They would no longer need to spend so much on advertising, again providing savings that could be passed onto the consumer, and so on. In a nutshell, Jobs and Gates might claim, the merger would bring new exciting products with lower costs of distribution and marketing.

You can be sure, however, that the Department of Justice, European antitrust authorities, and other similar bodies throughout the world would look on a proposed Microsoft-Apple merger with a highly skeptical eye. They would carry out their own studies of the costs and benefits of the merger. Even if the merger were to bring all the advertised benefits, it would also make the computer operating system market much less competitive. If there are originally two firms in a market and they then merge, they become a monopoly. The analysis in this chapter allows us to predict that output would decrease and prices would increase. Because a market with only two firms may still be very competitive, the loss in buyer surplus going from a competitive market to a monopoly can be sizable.

Stemming from the Clayton Act, the Department of Justice and the Federal Trade Commission must approve mergers and acquisitions of larger companies. The guidelines used for those decisions [2] emphasize two general points: (1) the effect of the merger on efficiency and (2) the effect of the merger on market power. Efficiency here refers to *cost efficiency*—the extent to which a merged company will be able to reduce its costs of production. The Department of Justice puts it as follows in its merger guidelines: "Efficiencies generated through merger can enhance the merged firm's ability and incentive to compete, which may result in lower prices, improved quality, enhanced service, or new products." [3]

Efficiency may also relate to the *quality* of the good being produced. If a merger enables a better product to be produced at the same price as before, then the merger increases market efficiency. Put differently, if the merger increases the surplus of buyers, either because the product is improved or because the reduced costs of production lead to a lower price, then the merger has had a positive impact on efficiency. For example, one benefit from airline mergers might be the more efficient use of the information systems that handle travel reservations. Another might be more efficient use of airport landing rights.

With these guidelines in mind, how would the Department of Justice respond to a proposed merger of Microsoft and Apple? They would study the proposed merger with the goal of determining if the merger would create a more efficient market or if it would increase the market power of the sellers. Department of Justice economists would use frameworks like the ones we have presented in this chapter to help them predict the outcome of the merger.

The Department of Justice would almost certainly decide not to approve an Apple-Microsoft merger. However, it does not block most proposed mergers. There is debate among economists about whether the antitrust authorities are too lenient or too stringent. A recent study by economists Orley Ashenfelter and Daniel Hosken looked at five mergers that they suspected of being anticompetitive. [4] They compared prices of the goods produced by the merged firms with prices of goods that were not close substitutes yet had similar costs. They concluded that, in four of the five cases, prices increased between 3 percent and 7 percent, and "given the large amount of commerce in these industries, the implied transfer from consumers to manufacturers is substantial." The evidence from this study thus suggests that the antitrust authorities are too permissive in allowing mergers.

The European Union also watches mergers closely, likewise balancing cost-efficiency and competitiveness considerations. Here is a statement of its policy on mergers: "If the annual turnover of the combined businesses exceeds specified thresholds in terms of global and European sales, the proposed merger must be notified to the European Commission....These rules apply to all mergers no matter where in the world the merging companies have their registered office, headquarters, activities or production facilities. This is so because even mergers between companies based outside the European Union may affect markets in the EU." [5]

KEY TAKEAWAYS

- The market outcome with a few sellers is the Nash equilibrium of the game they play. In the Nash equilibrium, none of the firms has an incentive to change what is being done.
- The market outcome depends on the strategy variable of the firms. If each firm is choosing the price of its output, then the outcome with many firms is the competitive outcome. If each firm is choosing the quantity of its output, then there is a distortion in the output market as price exceeds marginal cost.
- Governments act to regulate markets with a small number of sellers by making sure that firms do not make decisions jointly and evaluating the efficiency gains and market distortions from proposed mergers.

CHECKING YOUR UNDERSTANDING

1. Show that one firm setting its price at marginal cost and the other one setting a price above marginal cost is not the Nash equilibrium.
2. Explain how the interest rate will influence the choice of a firm to cooperate with another one in setting the monopoly price.

[1] Quoted in G. Wearden, "OFT Accuses Tobacco Firms and Retailers of Cigarette Price Fixing," *Guardian*, April 25, 2008, accessed January 31, 2011,http://www.guardian.co.uk/business/2008/apr/25/regulators.retail.

[2] US Department of Justice, "Horizontal Merger Guidelines," accessed March 14, 2011,http://www.usdoj.gov/atr/public/guidelines/hmg.htm.

[3] US Department of Justice and Federal Trade Commission, *Revised Section 4 Horizontal Merger Guidelines*, April 8, 1997, accessed March 1, 2011,http://www.ftc.gov/bc/docs/horizmer.shtm.

[4] Orley Ashenfelter and Daniel Hosken, “The Effect of Mergers on Consumer Prices: Evidence from Five Selected Case Studies” (National Bureau of Economic Research Working Paper 13859), February 2008.

[5] See European Commission, “Mergers: Overview,” accessed March 14, 2011,http://ec.europa.eu/comm/competition/mergers/overview_en.html.

14.4 End-of-Chapter Material

In Conclusion

There are very few examples of truly competitive markets. Most firms in the economy possess a certain amount of market power because their product, service, or location is distinctive. This means that most prices in the economy are in excess of marginal cost.

That said, the degree of market power of most firms is relatively small. Canon has some market power for its cameras because Canon cameras are not identical to Nikon, Olympus, or Sony cameras. But the presence of these other manufacturers severely limits Canon's ability to charge high prices. Your local Thai restaurant has some market power because its food is different from that of other restaurants in the neighborhood. Again, though, this does not mean it can charge very high prices because customers can easily eat at other restaurants instead.

Occasionally, however, firms are so large relative to their markets that they have substantial market power. This distorts prices and output in the economy. Firms with such market power can make a lot of money by restricting their output and charging very high prices. This is where the antitrust authorities come into play. Their task is to identify firms that are abusing their market power in this way. In effect, their job is to try to bring the economy closer to the economists' ideal world, where markets are competitive, there are no distortions, and all possible gains from trade are realized.

In some cases, though, governments have reasons to create and support market power through patents and copyrights. They do so because the benefits from innovation outweigh the distortions associated with monopoly. Policy in this area is highly contentious because the right balance between encouraging innovation and fostering competition is unclear. Economists and policymakers continue to struggle with this and are likely to do so for years to come.

Key Links

- US patent law: http://www.uspto.gov
- Japanese patent law: http://www.jpo.go.jp
- European Union patent law: http://www.epo.org/patents.html

- Department of Justice, Antitrust Division: http://www.usdoj.gov/atr
- Federal Trade Commission: http://www.ftc.gov/bc/index.shtml
- Department of Justice, Microsoft case:http://www.usdoj.gov/atr/cases/ms_index.htm,http://www.usdoj.gov/atr/cases/f0000/0046.htm
- Supreme Court decision in Standard Oil case:http://www.law.cornell.edu/supct/html/historics/USSC_CR_0221_0001_ZS.html
- Texas milk contracts:http://www.usdoj.gov/atr/public/press_releases/1992/211110.htm
- Clayton Act: http://www.usdoj.gov/atr/foia/divisionmanual/ch2.htm
- Guidelines for mergers and acquisitions:http://www.usdoj.gov/atr/public/guidelines/hmg.htm

EXERCISES

1. Suppose you have two types of beverages: a cola and a beer. Are these products in the same market?
2. The table in Question 3 shows data for a monopolist who sells a good to four households, each of which buys at most one unit and each of which has a different valuation for the good. The monopolist can produce the good at a marginal cost of $4. The monopolist can discriminate perfectly in its pricing, charging each household its valuation. Fill in the missing elements in the table. How many units should the monopolist produce? How does your answer change if marginal cost is $6?
3. (Advanced) Looking again at the following table (with marginal cost equal to $4), calculate the marginal revenue. What is its relationship to price? Explain your findings.

TABLE 14.2 PRICE DISCRIMINATION BY A MONOPOLIST

Household	Quantity	Household Valuation	Price	Total Revenue	Marginal Cost	Total Cost	Profit
A	1	12	12	12	4	4	8
B	2	6	6	18	4	8	10
C	3	4	4		4		
D	4	3	3		4		

4. Write an explanation of the monopoly pricing problem assuming the monopolist sets the price rather than chooses quantity. Why is the outcome the same either way?

5. Looking at the table in Question 3, if the interest rate increased to 15 percent, would the firm still have an incentive to innovate?
6. Explain why there is a greater incentive to innovate if the final stage of competition is with a small number of quantity-setting firms rather than price-setting firms.
7. Why might a merger lead to a price reduction? Why might a merger lead to a price increase?
8. Suppose that a firm (the incumbent) produces with constant marginal cost at $10 and has a constant (minus) elasticity of demand of 2. What is its profit-maximizing price? Now suppose that a new firm enters the market. The demand curve facing the incumbent firm shifts inward, but suppose that the elasticity of demand does not change. Should the incumbent firm change its price? What happens to the quantity that it sells? Draw a diagram to illustrate this market.
9. Imagine there is a motorcycle dealer in your neighborhood. You know both the price of the motorcycle set by the dealer and the amount of money the dealer paid for that motorcycle. It turns out that your valuation of the motorcycle is less than the posted price but greater than the cost of the motorcycle to the dealer. Are there gains to trade? Do you think you could convince the dealer to sell the motorcycle to you? If so, is there a deadweight loss? Why might the dealer be unwilling to sell the motorcycle to you?
10. Plane tickets are often sold at different prices to different people. Is this a form of price discrimination?
11. Writers of textbooks sometimes make their products available at a price of near zero. Does this mean they are altruists, or are they earning revenue some other way?
12. If interest rates increase, what needs to happen to patent lengths to maintain incentives for innovation?
13. In Section 14.3.2 "Market Outcomes When Firms Set Quantities", we looked at the situation when two firms chose quantity simultaneously. Describe the game and the outcome if one firm chose its quantity first and the other one followed. Would the outcome be the same as that discussed?
14. If you were a judge looking at a prospective merger between Coke and Pepsi, would you be more inclined to support the merger on efficiency grounds or argue against the merger as being anticompetitive?

Economics Detective

1. If a company invents, patents, and produces a product in the United States and sells the product in China, what type of protection does the company have in China?
2. If a US company operates in Europe, is it subject to European competition policy?
3. What legal authority does the European Union have over US firms?

Spreadsheet Exercise

1. (Advanced) Build a version of Table 14.1 "Calculating the Discounted Present Value of Expected Profits" starting with entries on demand and costs. To do so, use the examples in Chapter 6 "Where Do Prices Come From?" to create demand, revenue, and then marginal revenue. Also use the examples there to create variable cost and marginal cost. Then find the profit-maximizing quantity and price. Using this information, calculate the profit for each year and then calculate the discounted present value of these profits.

Chapter 15
A Healthy Economy
The Cost of Health Care

What do you do when you are ill? You might first go to a drugstore, browse the shelves a bit, and find an over-the-counter medication that you think will make you feel better. Your choice of product could be influenced by many things, including past experience, the advice of friends, or perhaps an advertisement you saw on television.

If the trip to the drugstore doesn't solve the problem, a visit to a doctor usually comes next. The first doctor you visit is likely to be a general practitioner, or GP for short. Even if insurance is picking up some of the cost, a trip to the doctor is often not cheap. Nor is it usually fun: it may involve long waits and unpleasant tests. We go to the doctor not because we enjoy the experience in itself but because of a deeper demand—a desire to be healthy.

A trip to the doctor typically ends with a bill, a prescription, and perhaps a smile along with a "see you again soon." (That last bit, of course, is not quite what you want to hear.) Then you go to the pharmacy to fill the prescription. If you look at the piece of paper the doctor gave you, you might notice a couple of things. First, the doctor's handwriting is often illegible; penmanship is evidently not high on the list of topics taught at a medical school. Second, even if you can read what is written, it probably means nothing to you. The chances are that it probably names some medication you have never heard of—and even if you have heard of it, you probably have no idea what the medication does or how it works.

In other words, though you are the purchaser and the patient, your treatment is largely out of your hands. Health-care purchases do not directly reflect individual choices the way most other spending decisions do. You did not choose to be sick, and you do not choose your treatment either.

Occasionally, your GP might recommend that you visit another doctor, called a specialist. Your GP might try to explain the basis of this recommendation, but you probably lack the expertise and the knowledge to evaluate the decision. Once again, you must trust your doctor to make a good decision for you: the decision to visit the specialist is largely your doctor's rather than your own. You typically follow the doctor's advice for two reasons: (1) you trust the doctor to make decisions in your best interest, and (2) if you have medical insurance, you do not have to pay most of the costs.

We have described this as though you have no control at all over your own health care and treatment. This is an exaggeration. If you are somewhat informed or knowledgeable about what is wrong with you, then you can discuss different treatment options with your doctors. You can become at least somewhat informed by reading articles on the Internet. You can seek out second and third opinions if you do not trust your doctor's diagnosis. If you are having serious treatments, such as a surgical procedure, you will have to sign forms consenting to the treatment. There is a trend these days for people to become more involved and empowered about decisions involving their own health. Yet, unless you have medical training yourself, you will have to rely to some degree, and probably a very large degree, on the advice of your doctors.

If you are seriously ill, you may have to go to the hospital. There you have access to many more specialists as well as a lot of specialized equipment. Whatever autonomy you had about your treatment largely disappears once you enter the hospital. At this point—at least if you are living in the United States—you certainly should hope you have insurance coverage. Hospital costs can be astounding.

If you look back in history, health care was not always provided the way it is today. One difference is that doctors used to visit patients at home. They would traditionally arrive with a small black bag containing their basic tools. (This type of service is still provided in some communities and in some countries, but it is now rare in the United States.) For the most part, that was where your medical treatment ended.

In part, this reflected the state of medical knowledge at the time. It is hard to comprehend how much medical science has advanced in the last century. One hundred years ago, our knowledge about the workings of the human body was rudimentary. There were few treatments available. Antibiotics had not yet been discovered, which meant that the simplest injury—even a scratch—could become infected and be fatal. If you had appendicitis, it would very likely kill you. There were few means of diagnosis and no treatments for cancer.

Today, the story is very different. We visit specialists who have highly advanced (and expensive) training. We have access to advanced diagnostic tools, such as magnetic resonance imaging (MRI) scans, blood tests that identify markers for cancer, and genetic testing. We also have access to expensive treatments, such as kidney dialysis and radiation therapy. Perhaps most strikingly, we have access to a range of pharmaceutical products that have been developed—sometimes at great expense—by scientists and researchers. These products can treat medical conditions from asthma to apnea to acne.

With all these visits to doctors and all these medications, we spend a great deal on health care. Spending as a fraction of gross domestic product (GDP; a measure of the total output of the economy) has been increasing since 1960 (Figure 15.1 "US Health-Care Expenses as a Percentage of GDP"). Figure 15.2 "Global Health-Care Spending" shows total spending on health services per person around the world. [1] The shaded areas indicate the level of spending on health-care services. The United States spends the most on health care per person, with Norway and Switzerland also being high-spending countries. Other rich countries, such as Japan, Australia, New Zealand, South Korea, and countries in Western Europe, likewise spend relatively large amounts on health care. The poorer countries in the world, not surprisingly, spend much less per person on health care. Across countries in the world, as within a country, health-care purchases are related to income.

One reason that we spend so much on health care in the United States is that high-quality care, such as is available in rich countries, is at least in part a luxury good—that is, something that we spend relatively more on as our income increases. Yet even across relatively affluent countries, health care takes very different forms. [2] Compare, for example, the United States and Canada. Canada has a system in which the government pays for health care. The program is financed by the payment of taxes to the government. Doctors' fees are set by the government, which limits competition within the health industry. Furthermore, other developed countries spend much less on health care than the United States but have health outcomes that are as good or even better.

Differences in both the quality and cost of health care mean that, perhaps surprisingly, health care is traded across national boundaries. In some cases, people travel across the globe to obtain care in other countries. Sometimes, people travel to obtain treatments that are unavailable in their home countries. For example, US residents sometimes travel to other countries to obtain stem-cell treatments that are banned in the United States. Or people may seek health care in other countries simply because it is cheaper: people from around the world travel to Thailand, for example, to obtain cheap and reliable dentistry services. There are even tour operators that arrange such "health tourism" trips.[3] Given all these differences in health care costs around the world, we address the following question in this chapter:

What determines the cost of health care?

Road Map

Because we want to talk about the price of health care, supply and demand is a natural starting point. As we use this framework, though, it will rapidly become clear that there are many things that are unique about the market for health care. One indication of this is that there is a whole subfield of economics called "health economics." There is no subfield called "chocolate bar economics," "tax advice economics," or "lightbulb economics." Evidently, there is something different about health care. Another indication is the fact that governments around the world pay an enormous amount of attention to this market. Governments intervene extensively in this market through taxes, through subsidies, and sometimes by being the direct provider of health-care services.

We first study the demand for health care by households. Then we look at the supply of health care, after which we turn to the determination of prices. As we proceed, we will see that health care includes all sorts of different products and services. We will also see that there are many reasons why it is difficult to analyze health care with a simple supply-and-demand framework.

One key reason why health care is such a complicated topic has to do with the fact that, frequently, we do not pay for health care ourselves. Rather, we (or our employers) purchase health insurance, and then the insurance company pays the health-care providers. We therefore discuss health insurance in some detail. The chapter ends with a discussion of the government's role in the health sector, in which we talk about market failures and a variety of proposed government solutions.

[1] World Health Organization, "Total Expenditure on Health per Capita, 2007 (in US$)," accessed March 1, 2011, http://www.who.int/nha/use/the_pc_2007.png.

[2] A comparison of programs is provided by Ed Cooper and Liz Taylor, "Comparing Health Care Systems: What Makes Sense for the US?" *In Context*, accessed March 14, 2011,http://www.context.org/ICLIB/IC39/CoopTalr.htm.

[3] National Public Radio had a March 18, 2008, story of a husband and wife going to China for get stem-cell treatment for their seven-month-old daughter (http://www.npr.org/templates/story/story.php?storyId=88123868). There is also a company that organizes trips to Canada (http://www.findprivateclinics.ca/resources/general/medical-tourism.php).

15.1 Supply and Demand in Health-Care Markets

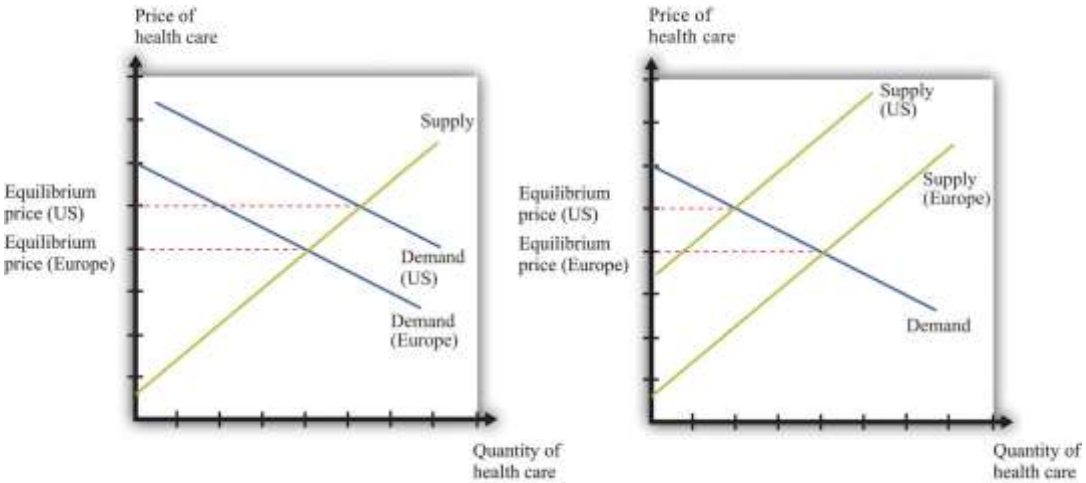
LEARNING OBJECTIVES

1. What factors determine the price and quantity of health care?
2. In what sense is spending on health an investment?
3. What factors determine the demand for health-care services?
4. What is the production function for health?

Suppose we want to explain why health care is more expensive in the United States than in Europe. Then supply and demand seems like a natural starting point. If we imagine a market for health care drawn in the usual way, with the quantity demanded on the horizontal axis and the price on the vertical axis, then the question becomes, "Why is the price of health care higher in the United States than in Europe?"

Supply and demand offers two possible answers (Figure 15.3 "Two Explanations for Why Health Care in the United States Is More Expensive Than in Europe"). The prices can be high because demand is high. For example, if the demand curve is further to the right in the United States compared to Europe (part [a] ofFigure 15.3 "Two Explanations for Why Health Care in the United States Is More Expensive Than in Europe"), this implies—all else being equal—higher prices in the United States. The other reason for high prices is because supply is limited. If the supply curve in the United States lies further to the left than the supply curve in Europe (part [b] of Figure 15.3 "Two Explanations for Why Health Care in the United States Is More Expensive Than in Europe"), then this also would imply—all else being equal—higher prices for health care in the United States. Neither argument seems that compelling, which naturally leads us to wonder if the supply-and-demand framework is really the best framework for analyzing health care. In fact, there are good reasons to think that the supply-and-demand framework is not the best approach to this market.

Figure 15.3 Two Explanations for Why Health Care in the United States Is More Expensive Than in Europe



Supply and demand offers two possible explanations of high health-care costs in the United States: demand in the United States is high (a), or supply in the United States is limited (b). Neither is a very compelling explanation.

Let us think about the demand side first. Our standard approach to demand is based on the idea that each individual will consume a good or a service up to the point where the marginal valuation from one more unit equals the price of that additional unit. [1] Unfortunately, the health-care consumer often has very little idea of the value—let alone the marginal valuation—of the particular treatment being received. The consumer is very often not paying the full price for that treatment because the cost is frequently covered, at least in part, by insurance. Together, these mean that our traditional approach to demand does not work very well for health-care services.

The supply side is also problematic. First of all, some health-care suppliers have significant market power. This does not mean that we can get no insights from supply-and-demand reasoning. But it is trickier to compare the price of health care across countries because we have to consider differences in market power as well. A bigger problem is that some health-care suppliers, such as hospitals, are either government-controlled or not-for-profit institutions. The standard economic approach presumes that firms seek to make as much profit as possible, but government or not-for-profit hospitals may not have profit maximization as their goal.

In addition, health-care prices are not necessarily determined by supply and demand. Again, the government has a significant influence on prices: for example, the governments in some countries set prices for pharmaceutical products. Even if they are not set by the government, prices may be determined by bargaining between, say, hospitals and drug companies rather than by supply and demand. Furthermore, if people need health-care services, then their demand is likely to be very inelastic (the quantity demanded does not respond much to price changes). Inelastic demand is not, in and of itself, a problem for a competitive market. It just means that the equilibrium price could be very high. But if we couple inelastic demand with consumers who lack information and add in some market power by suppliers, then matters become more complicated. Perhaps you already have a sense of why: we have a large group of consumers with very inelastic demand who are relatively uninformed. This sounds like a gold mine for the supplier.

We have so far ignored the issue of what exactly is being traded in this market. "Health-care services" can mean many very different things:

- Labor time of various trained professionals, such as GPs, specialists, nurses, medical technicians, pharmacists, and many others
- Procedures and testing, such as magnetic resonance imaging (MRI) scans and laboratory analyses of blood samples
- Hospital and nursing care services
- Emergency services such as ambulances
- Pharmaceutical products (which itself covers a huge range, from bandages to chemotherapy drugs)

You can probably think of other components as well. So it is more than a little misleading to treat health care as something homogeneous that is bought and sold in a single market.

We hope that by now we have completely muddled your view of the health-care market. Our main point is that the simple framework of supply and demand is not sufficient for understanding health care. There are too many different markets, each with its own peculiarities and unusual features. And those features mean that there are several reasons why we might expect inefficiency. One, as we have already noted, is the presence of market power. Another is the various information problems we have mentioned. A third is that some aspects of health care have the characteristics of a public good.

The Demand for Health Care

Now let us dig a little deeper into the demand side of health care.

Response to Price

The law of demand applies to health care as in other markets: as the price of health care increases, you demand less of it. But we must be careful. What matters is the price of health care to *you*. If you have health insurance, this price may be much lower than the actual cost of providing you with care. Under most health-insurance contracts, the *marginal private cost* of care to a household is less than the **marginal social cost** of providing that care. The household has an incentive to purchase a lot of health-care services because its purchases are, in effect, being subsidized by insurance companies. We take up the topic of health insurance later in this chapter.

Toolkit: Section 17.19 "Externalities and Public Goods"

You can review the distinction between marginal private cost and marginal social cost in the toolkit.

Another key characteristic of health care is that demand is relatively inelastic. If you are sick and require care, you will purchase health-care services at almost any price. Of course, your ability to purchase health care is ultimately limited by your income, but you are likely to trade off spending on many other products to purchase the medical care you need. This is why we often read stories about people without insurance being bankrupted by medical expenses.

Health as Investment

Everyone prefers being healthy to being sick. The demand for health care is in part an expression of this preference. One thing that makes health care different from most other goods and services, though, is that it is simultaneously an investment. Money you spend on being healthy today will also benefit you in the future. There are several different ways in which spending on health care represents an investment.

Mortality. One clear impact of our health-care choices can be seen in terms of mortality rates. Mortality rates measure how likely we are to die at different ages. In 2004, the mortality rate in the United States for people ages 15–24 was about 80 out of 100,000, or 0.08 percent. In contrast, the mortality rate for those over the age of 85 was 13,823 out of 100,000, or 13.8 percent. [2] In other words, the typical young person has about a 1 in 10,000 chance of dying in a given year, whereas the typical old person has more than a 1 in 10 chance of dying.

It is not surprising that the mortality rate increases with age—that is, that young people have a lower probability of dying than older people. (Infants are an exception: a 6-month-old child is more likely to die than an 18-month-old child because very young children are particularly susceptible to certain diseases.) But these average mortality rates disguise a lot of variation, much of which is under our control. There are many behaviors that have predictable effects on our likelihood of dying. Smokers have a higher probability of dying than nonsmokers. Those who are obese have a higher probability of dying than those who are not. Diet, exercise, and risky behaviors (which includes everything from unprotected sex to skydiving) affect mortality rates as well.

Cigarette smoking is linked to lung cancer and thus to mortality. If you compare two similar individuals of the same age, one who is a smoker and the other a nonsmoker, then the mortality rate is significantly higher for the smoker. This does not mean that the smoker will necessarily die before the nonsmoker. It means that all else being the same, smoking increases the probability of death. Refraining from smoking is a type of investment in your future.

Our diet also affects our probability of becoming ill and of dying. As with cigarettes, there are often trade-offs between eating and drinking things we enjoy and the effects of such consumption on our long-term health. Making these types of choices is an economic decision. Each of us makes different choices because we value the taste of particular foods differently, and we value our overall health differently as well. If a thirty-year-old discovers he has elevated cholesterol levels that pose a long-term risk of heart disease, he may decide to adjust his diet, perhaps consuming less red meat. If an eighty-year-old learns the same news, he may not think the long-term benefit is worth giving up his steaks for.

Productivity. Being healthy also means that you can work and earn wages. One of the costs of poor health is lost days at work. This is a cost not only to the individual but also to society as a whole: the economy's population is producing less output. If you are in poor health, then you risk losing wages for the days when you cannot come to work. Many employers provide insurance for these lost wages through the provision of sick days: if you are sick, you are not expected to work but you will still be compensated up to a contracted number of days per year. In addition, there is disability insurance as part of the social security system in the United States. [3] Private employers sometimes also offer disability insurance as part of their compensation packages, and you can also purchase insurance directly from an insurance company.

Information Problems

Health care is an example of a good for which the typical individual is unable to determine the quality of what is being purchased. You can think of other examples, such as legal services and used cars. In such situations, how can we make good decisions? Generally we do so by relying on the advice of experts. In the case of health, these are the doctors, dentists, and other health professionals who are trained to analyze our health situation and make suggestions to us. We listen, try to understand, and, using their advice, make an informed choice.

Suppose you get a phone call from someone telling you they know of a stock, trading on Wall Street, that will double in price the following day. You might be very skeptical, suspecting that they have other reasons for wanting you to buy. Compare this to a conversation with a medical expert. Generally you are going to believe that the expert is acting in your best interests. Although you might get a second (or third) opinion, you do so because health problems are complex and the first expert may have missed something, not because you are afraid the doctor is misleading you in order to profit from your visit. But why do we trust medical experts so much more than the provider of stock tips? We generally do so because we trust that their incentives are aligned with our goals; that is, we hope that they are motivated to act in our best interests.

The Supply of Health Care

We now turn to the supply side of health care. Economists often talk of output being produced using a **production function** that uses labor, capital, and intermediate inputs. What is the production function of a hospital?

- The labor in a hospital includes doctors, surgeons, orderlies, technicians, nurses, administrative staff, janitors, and many others.
- The hospital buildings are part of the hospital's capital stock. In addition, hospitals contain an immense quantity of other capital goods, such as hospital beds and diagnostic tools—everything from stethoscopes to x-ray machines.
- Intermediate inputs in a hospital include dressings for wounds, and pharmaceutical products, such as anesthetics used for operations.

Other sectors of the health-care industry likewise employ labor, capital, and intermediate inputs.

Toolkit: Section 17.17 "Production Function"

You can review the meaning and definition of a production function in the toolkit.

Doctors

If you look at the wall in your doctor's office, you will typically see a large number of framed degrees and other qualifications. To become a doctor, you must first succeed as an undergraduate and then go through multiple years in medical school. After this comes an internship and then you finally graduate and can practice on your own. In most countries, you must have a license to practice medicine. This makes sense: you would not want anyone to advertise as a doctor regardless of their skill level. Most of us would be unable to tell whether a particular individual was a qualified professional or a quack. When buyers cannot easily evaluate the quality of the good or the service they are purchasing, it is useful to have external validations of quality.

Licensing provides more than a guarantee of quality, however. It also limits entry into the profession. Suppose you learned that a small group of lobbyists in your hometown wanted gas station owners to be licensed in the same way as physicians. You would quite rightly suspect that their goal was not to guarantee high-quality gasoline. More likely, they would be trying to limit the number of gas stations to increase their market power. Your suspicions would not be allayed if these lobbyists argued that gas was potentially a very harmful commodity, so by licensing the sellers of gas, they were protecting the community. In the case of doctors, the underlying reason for licensing is not so nefarious. But it still creates a barrier to entry that limits competition and increases market power, just as it would with gas stations.

Doctors differ from gas station owners in many other ways. Typically, we suppose that gas stations and other firms in an economy have profit maximization as a goal. It is this presumption that allows us to develop our theory of supply. Doctors not only think about profits but also take an oath of office, called the Hippocratic Oath, which is as follows:

> *I swear by Apollo, the healer, Asclepius, Hygieia, and Panacea, and I take to witness all the gods, all the goddesses, to keep according to my ability and my judgment, the following Oath and agreement:*
>
> …
>
> *I will prescribe regimens for the good of my patients according to my ability and my judgment and never do harm to anyone…*
>
> *I will not give a lethal drug to anyone if I am asked,…* [4]

This oath is administered to nearly everyone obtaining a medical degree.

Other Health-Care Workers

In addition to doctors and specialists, there are many other kinds of workers in the health care industry, including nurses, dental hygienists, administrative staff, technicians, staff in care facilities such as hospices and nursing homes, and many others. The health-care industry employs almost 10 percent of all civilian workers in the United States.

Table 15.1 "Employment in Health Services (in Millions)" shows the breakdown of employment by health-service site for three years: 2000, 2003, and 2006. We list some (not all) of the types of health sites. From the table, we see that the number of workers in this industry has increased from 12.2 million in 2000 to about 14.4 million in 2006. This increase is typical of many service industries and contrasts with manufacturing, where the number of workers employed is declining.

Hospitals are the most important type of employment site for health-care workers. In 2006, 40 percent of health-care workers were employed in hospitals. About three-fourths of the workers in the health-care sector are women. Women are particularly prevalent in nursing care facilities: of the 1.6 million workers in nursing care facilities in 2000, about 1.4 million were women.

There is a wide variety of occupations within health care: managers, professionals (doctors, dentists, pharmacists, etc.), service occupations (assistants, cooks, cleaners, etc.), and office workers. Professional groups account for about 44 percent of all workers, while about 32 percent of the jobs are in service occupations. There will typically be considerable variation of wages within a sector because of the different occupations of workers in that sector. For example, individuals working in diagnostic laboratories earn, on average, close to twice the wage of workers in nursing homes.

Health-Care Capital

When we look at an industry such as health care, one way of describing it is by counting the number of doctors' offices, clinics, and so on. There are many different kinds of establishments that provide health services. Hospitals are only one example; others include doctors' offices, clinics, nursing homes, and so on. According to the Bureau of Labor Statistics (http://www.bls.gov/oco/cg/cgs035.htm), in 2008, there were 595,800 establishments in the health-care sector in the United States. Of these, doctors' offices are 36 percent. Hospitals are only 1 percent. [5] Another way to describe the industry is by detailing the number of workers employed in different activities, as in Table 15.1 "Employment in Health Services (in Millions)".

Taken together, these statistics paint an interesting picture. Hospitals are a small fraction of the total health-care establishments but employ 35 percent of the workers. This tells us that there are relatively few hospitals (compared to doctors' offices), but they are big. About 70 percent of hospitals employ more than 1,000 workers. Such a pattern is not peculiar to the health-care sector. In US manufacturing, the majority of establishments are small, and a few large establishments employ the majority of workers. [6]

The ownership of hospitals is also complicated. Some are private, while others are public, meaning that federal, state, or even county governments run them. In addition, not all private hospitals are in business for profit; some are classified as not-for-profit institutions. Table 15.2 "Hospital Activity, 2005"provides a breakdown of hospitals by type. From this table, you can see that most admissions are in not-for-profit hospitals that are not federally run. Their goal is largely to provide a public service. These hospitals have a length of patient stay averaging about 5.5 days.

There are other capital goods that enter the production function for the health sector. For example, pharmaceutical production facilities are part of this capital stock. So too is the capital stock of companies that produce the machines, such as MRIs, used in doctors' offices and hospitals.

Technological Progress

Technological advances in health care are truly staggering. Technological progress in this sector, as in other sectors, comprises both product and process innovations. By product innovations, we mean increases in the types of goods and services available to households and doctors. A leading example is the vast array of drugs now available on the market, which is the outgrowth both of research and development at pharmaceutical companies and of publically funded research. Another example is the advanced machinery used in modern health-care facilities. A modern dentist's office is filled with high-speed drills, x-ray machines, and other pieces of technology that would have been unthinkable in your grandparents' day. MRI machines are another example: these are a significant advance over previous imaging techniques such as X-rays, but they are expensive—a new MRI machine will typically cost a hospital more than $1 million. [7]

Process innovations refer to how techniques are implemented. For example, surgeons today can perform operations that surgeons of previous generations could not even imagine. The knowledge for these procedures was created by a few people and then taught to others in medical school and other training programs.

Process and product innovations come together when you compare how certain procedures are performed now relative to years ago. Consider surgery to repair a hernia. The first hernia (hiatal) surgery took place around 1919, and the procedure was risky and painful. Even relatively recently, a procedure like this involved hospital stays, many days lost from work, and a significant risk of medical complications. [8] Today, the leading method for surgery uses a piece of capital called a laparoscope—a tube with light that allows a surgeon to see inside a patient's abdominal cavity. Then, using another instrument, again inserted through a small incision, the surgeon can repair the hernia. Remarkably, this is an outpatient procedure. The patient emerges from the hospital with a few small wounds and can return to work and normal life within a few days.

Price Determination

If this chapter were like most others in this book, we would now turn to a discussion of how supply and demand interact in a competitive market to determine the price. Or, recognizing that firms with market power set prices, we might use the condition that marginal revenue equals marginal cost to talk about price determination. Unfortunately, when it comes to understanding the market for health care, these tools are not as useful.

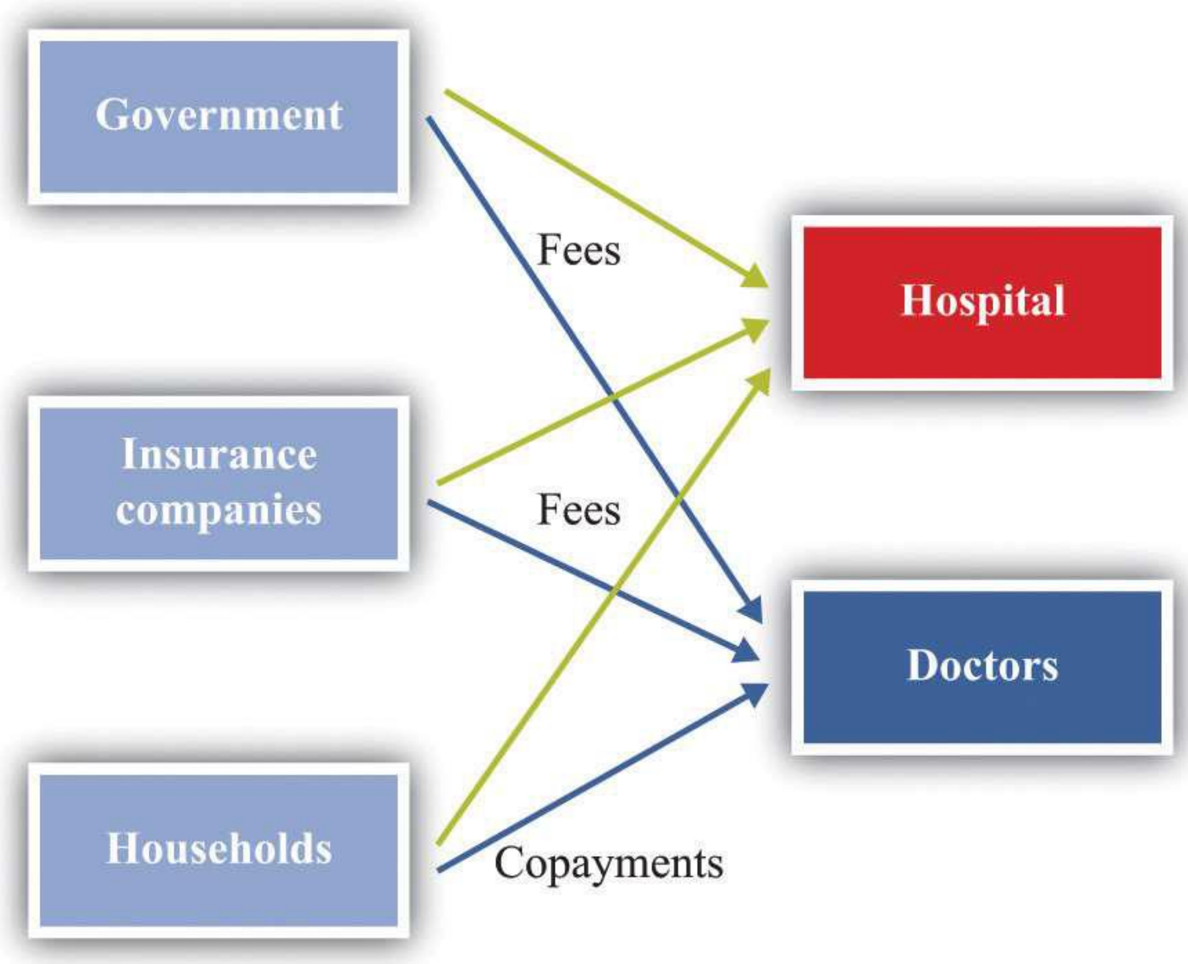
To understand why, imagine you want to book a hotel room in New York City. You can call up any hotel and find the price of a room. Or you can go on the Internet and check prices either at the hotel's website or at any number of other sites that provide booking services. You can find information about the hotel online, read reviews from previous guests, and talk to hotel staff members on the telephone if you need more information. If you are in the city, you can also walk into a hotel and find out the price and the hotel's amenities.

Now compare this to a hospital. It is much harder to get information about prices, and you cannot simply walk in off the street and purchase an operation. You can in fact find out prices for hospital procedures if you look hard enough. For example, there is a website that allows you to find charges for different procedures in Wisconsin (http://www.wipricepoint.org). Here you can "shop" for, say, different types of knee surgery. But these charges do not necessarily reveal the true price to you as a consumer because they may not include all the costs of doctors and other inputs. If you have insurance coverage, meanwhile, you need to find out what portion of any bill will be covered by your insurance. Figuring out the price of a procedure is quite complicated.

How, then, are prices determined? And, importantly, what price are we talking about: the price you pay or the money received by the hospital? Many of the most important prices are determined by the interaction of a few big players, including the government, insurance companies, and pharmaceutical companies. Figure 15.4 "Payments to Hospitals and Doctors" gives a sense of the sources of income for hospitals and doctors. Hospitals and doctors get paid by insurance companies, households, and the government.

- Medicare (http://www.medicare.gov/default.aspx) is a federal program intended to provide health services to elderly (over 65) and disabled people. It covers nearly 40 million people. Under this program, the government sets fees for services provided by physicians. A listing of those fees is available through the Health and Human Services website. [9] These are the fees the government will pay physicians and hospitals for these services.
- *Medicaid* (http://www.cms.gov/MedicaidGenInfo) is run by the US government in conjunction with state governments. [10] This program provides health care to low-income households through payments made directly to a health-care service provider, such as a hospital.

Figure 15.4 *Payments to Hospitals and Doctors*



Because of these programs, the government is a big player in the health-care market. Government decisions determine the demand for health-care services. Governments do not take prices as given. In some cases, the government sets rates for certain procedures, and health-care providers respond. In other cases, the government is involved in negotiations—with pharmaceutical companies, for example.

Insurance companies provide additional sources of revenues to the hospital and a doctor. If you are a policyholder and are admitted to a hospital, your insurance company will reimburse the hospital for part of the cost of your care. It also reimburses your doctor directly. How much of that cost is reimbursed depends on your insurance policy. If you enter a hospital, say, for an operation, the amount of money the insurance company will pay the hospital is set by an existing agreement. As a result, hospital administrators face a complex set of repayment schedules. Reimbursement rates for a given service depend on who is buying the service, as the following quotation illustrates: “Medicaid pays 80 percent of what Medicare pays and about 50 percent of what a commercial insurance carrier like Blue Cross/Blue Shield pays. For example, if Medicaid reimbursed $500 for a gall bladder removal, Medicare would pay $625 and the commercial carrier $1,000 for the same procedure.” [11] Because of these differences in reimbursement rates, doctors and hospitals may sometimes decide not to provide services to certain patients. The same article notes that doctors sometimes turn down Medicaid patients because of these low rates.

And what will you pay if you walk into a hospital without health insurance? An April 28, 2008, article in the *Wall Street Journal* describes the plight of a cancer patient without adequate insurance. The patient was looking for treatment at a not-for-profit hospital in Texas. Her treatment required a payment of $105,000*in advance*. This practice of requiring prepayment is part of a trend in the industry.

> *Hospitals are adopting a policy to improve their finances: making medical care contingent on upfront payments. Typically, hospitals have billed people after they receive care. But now, pointing to their burgeoning bad-debt and charity-care costs, hospitals are asking patients for money before they get treated.*
>
> *Hospitals say they have turned to the practice because of a spike in patients who don't pay their bills. Uncompensated care cost the hospital industry $31.2 billion in 2006, up 44 percent from $21.6 billion in 2000, according to the American Hospital Association.* [12]

KEY TAKEAWAYS

- Due to informational problems for households, market power by suppliers, and government intervention, the market for health care cannot be analyzed by using standard supply-and-demand curves.
- Spending on health care today has an effect on your health status in the future. In that sense, this spending is an investment.
- The demand for health services, like other goods, depends on your income and the price of the services. Unlike your demand for many other goods, your demand for health services is influenced by the costs of health insurance. Also, unlike the case for many other goods, consumers who demand health services are relatively uninformed about the service they are buying.

- The production function for health takes inputs, such as doctors, nurses, and machines, and produces health-care services.

CHECKING YOUR UNDERSTANDING

1. List three reasons why the conventional supply-and-demand model may not fit the market for health services well.
2. How is the demand for health services influenced by age?
3. Give an example of two intermediate inputs into the provision of health-care services.

[1] Chapter 3 "Everyday Decisions" explains this idea in more detail.

[2] Mortality rates can be found at the National Center for Health Statistics, National Vital Statistics System, "Mortality Tables," accessed March 14, 2011,http://www.cdc.gov/nchs/nvss/mortality_tables.htm.

[3] The government program is summarized at http://www.ssa.gov/disability.

[4] *Wikipedia*, s.v. "Hippocratic_Oath," accessed March 14, 2011,http://en.wikipedia.org/wiki/Hippocratic_Oath.

[5] US Department of Labor, Bureau of Labor Statistics, "Career Guide to Industries, 2010–11 Edition: Healthcare," accessed March 14, 2011, http://www.bls.gov/oco/cg/cgs035.htm.

[6] This is described in Steven J. Davis, John C. Haltiwanger, and Scott Schuh, *Job Creation and Destruction* (Boston, MA: MIT Press, 1998). We discuss this phenomenon in more detail in Chapter 8 "Growing Jobs".

[7] An article describes this technique at NOVA, "The Picture Becomes Clear for Magnetic Resonance Imaging," accessed March 15, 2011,http://www.science.org.au/nova/062/062key.htm.

[8] For the history of hiatal hernia surgery, see Nicholas Stylopoulos and David W. Rattner, "The History of Hiatal Hernia Surgery: From Bowditch to Laparoscopy," *Annals of Surgery* 241, no. 1 (2005): 185–193, accessed March 14, 2011,http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1356862.

[9] See Centers for Medicare and Medicaid Services, "Overview," accessed March 14, 2011,http://www.cms.gov/apps/physician-fee-schedule/overview.aspx. A search engine at that site allows you to look for specific fees.

[10] The states design the programs subject to approval by the federal government. Thus there are differences across states. The federal government reimburses states according to a rule that depends on the average income per person in that state.

[11] This quote comes from a hospital in Topeka, Kansas. See Jan Biles, "Medicare: Care versus Cost, *cjonline.com*, October 23, 2007, accessed February 1, 2011,http://www.cjonline.com/stories/102307/sta_211162036.shtml.

[12] Barbara Martinez, "Cash before Chemo: Hospitals Get Tough. *Wall Street Journal*, April 28, 2008, A1.

15.2 Health Insurance

LEARNING OBJECTIVES

1. What are the incentive issues associated with the demand for health insurance?
2. Why is health insurance linked to employment in the United States?
3. How does the law of demand apply to the demand for health services when there is health insurance?

Insurance is something that human beings have developed to help us deal with the risks we face in life. Here are some examples of risks that you might confront.

- Your car or other property will be stolen.
- You will lose some of your possessions due to a fire, flood, storm, or other natural disaster.
- Your car will be damaged in an accident.
- You will lose your job.
- You will be injured in an accident—for example, while working, driving, or playing sports.
- You will become ill.

You can easily add to this list. We always have to worry about bad things happening. One consolation is that, for all the risks listed, you can obtain insurance. This means that we pay a fee (the premium) to an insurer; in return, we receive payment from the insurer if the bad thing happens.

Insurance is based on the idea of the diversification of risk. [1] As an illustration, suppose you face a 1 in 5,000 chance of breaking your leg in a given year. If this happens, it will be very costly to you: between hospital bills and lost earnings, perhaps you would lose $10,000. If you are like most people, you arerisk-averse, meaning that you don't like facing this risk. Suppose, however, you can get together in a group of 5,000 people and agree that if any one of you breaks a leg, you will all share in the bill. The most likely outcome is that only one person will suffer a broken leg, and your share of the costs will be $2. There is still a bit of uncertainty: maybe no one will break a leg; maybe two, three, or four people will. But the likelihood that you will have to pay out more than a few dollars is very small.

Insurance companies are firms that carry out such diversification of risk by bringing together large groups of people. Insurance companies set a premium equal to the **expected value** of the loss (in the example, 15,000 × $10,000 = $2), plus a fee to ensure the insurance company also profits from the deal.

Toolkit: Section 17.7 "Expected Value"

You can review the calculation of expected value in the toolkit.

Insurance, like other services, is traded in a market. You can choose to buy from a variety of sellers at a price that reflects the risk of the type of insurance you purchase. The gains from trade come from the fact that an insurance company is capable of pooling risk. The insurance company assumes your risk at a price you are willing to pay. Because people differ in terms of their attitudes toward risk, some people buy insurance against certain events, while others do not. If you are very cautious (more precisely, very risk-averse), then you are more likely to buy insurance.

What Makes Health Insurance Different?

Health insurance has the same basic structure as any other insurance: you pay a premium to an insurance company that then pays your medical bills if the need arises. Like other types of insurance, there are gains from the sharing of risk. However, health insurance differs from other kinds of insurance in a couple of ways: [2] (1) health insurance is largely provided by employers, and (2) informational problems are particularly acute.

Who Pays for Health Insurance?

In most European countries, health insurance is largely provided by the government. In some cases, the government is also a provider of health services. In the United States, the government provides some health insurance—to the very poor, the old, and military veterans. But for the most part, the provision of health insurance in the United States is very different. Table 15.3 "Sources of Health Insurance in the United States" shows the types of health insurance that households can obtain in the United States. The "Total" column indicates the fraction of households with insurance. Since 1999, this has averaged about 85 percent but has been falling somewhat. On the bright side, this tells us that most people are covered by insurance. It also tells us that about 50 million people in the United States have no health insurance. The table reveals in addition that by far the most important source of health insurance is through employers: about 60 percent of all individuals have insurance provided through a firm. The other forms of insurance are through the government (about 30 percent) and direct purchase (about 9 percent). These numbers add to more than 85 percent because many individuals have insurance from more than one source.

It might seem odd that your health insurance is likely to be linked to your job. After all, your employer doesn't pay for your car insurance or for insuring your bank deposits. Historically, this phenomenon has its roots in the Stabilization Act of 1942, which was signed into law by President Franklin Roosevelt. The idea of the legislation was to stabilize wages and prices during World War II. Although President Harry Truman repealed most of the provisions of the act in 1946, some of the effects of that act remain today. [3]

A key provision of the act established wage and price controls. This meant that wages were no longer determined by market forces but were instead set (in part) by the government. But when the government places restrictions on the way people trade, they will often try to find ways around those restrictions. [4] The loophole in the Stabilization Act was that it exempted pensions and insurance from the calculation of wages. This meant that firms could vary the overall compensation they offered workers through the provision of pensions and health insurance. Even though wage and price controls are no longer in place, the practice of offering health insurance as part of a compensation package persisted.

An employment-based health insurance system was furthered by tax actions, such as the 1954 Internal Revenue Code, which made employer contributions to employee health insurance nontaxable. Individuals were also allowed to deduct medical expenses from taxable income. So if you are paid $1,000 in extra income by your firm and you use these funds to buy health insurance, you are taxed on the $1,000 of income. But if the firm buys the insurance for you, then you do not pay tax on the $1,000 worth of benefits.

Being employed also changes the price you pay for insurance. If you contact an insurance company directly, the rates you will be quoted for health insurance are much higher than the rate (for you and your employer combined) if you buy a health policy through your job. One explanation for this is that it is cheaper to write an insurance policy for many people together than individually. A second explanation, which we explain in more detail later, is that, on average, employed people are likely to be a lower risk than those not working. A third factor is that a group of employees is already partly diversified, so the group is less risky than a single individual.

We saw in Table 15.3 "Sources of Health Insurance in the United States" that about 15 percent of individuals in the United States do not possess health insurance. But who are these individuals, and why are they uninsured? The "who" is easier to answer than the "why" because we have statistics on the uninsured. Table 15.4 "The Uninsured (in Millions)" reveals the following: [5]

- Many of the uninsured are poor. Of the nearly 45 million uninsured in 2005, about 14.5 million had incomes less than $25,000. (As a benchmark, according to the US Census Bureau, the 2006 poverty level for a family of four was $19,350.) Only about 17.6 percent of the uninsured had incomes in excess of $75,000. In 2009, of the 50.7 million uninsured, about 15.5 million had income less than $25,000.
    - Many of the uninsured are young. In 2005, there were 8.0 million uninsured people under the age of 18. This number was 7.5 million in 2009.
    - Many of the uninsured are young *and* poor. About 2.5 million of the 8.5 million have family incomes below the poverty line. According to the US Census Bureau, 65.5 percent of the children in poverty were covered by the government Medicaid program. [6]
- Many of the uninsured are working. There were over 20 million individuals in 2005 who were working full time and yet did not have health insurance. In 2009, the number of full-time workers without health insurance was lower—14.6 million—while the number of part-time workers without health insurance was higher. We do not know from these data whether they were offered health insurance at work and declined or had jobs that did not offer this benefit.

Adverse Selection

One complication of health insurance markets is that those who demand insurance are the ones who are more likely to need insurance. This in itself might not be a problem, except that individuals also know more about their own health than do the companies that are insuring them.

Suppose that half the population carries a gene that gives them a 1 percent risk each year of contracting a particular kind of cancer. The other half does not carry this gene and has only a 0.1 percent risk of this cancer. If an individual becomes sick, suppose that the cost of treatment plus lost work time is $100,000. Table 15.5 "Probabilities and Expected Losses" summarizes the situation. Group A has a 0.1

percent chance of contracting the cancer, so the expected loss for them is $100 (= 0.001 × $100,000). Group B has a 1 percent chance of contracting the cancer, so their expected loss is $1,000 (= 0.01 × $100,000).

Table 15.5 Probabilities and Expected Losses

Group	Probability of Cancer (%)	Expected Loss ($)
A	0.1	100
B	1	1,000

Now let us think about an insurance company that wants to make money by selling insurance policies against this loss. Suppose these policies completely cover all losses in the event that the individual contracts the disease. If the insurance company were to set the price of a policy very high (say, $5,000), then only very risk-averse people would buy the policy. If it were to set the price very low—say, $50—then everyone in the population would want this policy, but the insurance company would make a loss on every individual.

However, suppose that the insurance company were to offer a policy for $550. It might reason as follows: if everyone buys this policy, then we will lose $450 on average from the group B individuals, but we will gain $450 on average from the group A individuals. Because there are equal numbers of both groups in the population, we should expect to make no profits on average. For example, if there are 2,000 typical individuals (1,000 of each type), then on average one group A person will become sick (because their chance is 1 in 1,000) and 10 group B individuals will become sick. In total (reasons the firm), we expect to have to pay out for 11 people, implying payments of $1.1 million (= 11 × $100,000). We will get revenues of $1.1 million (= 2,000 × $550). At this price, our expected revenues and costs are the same. If we were to charge a slightly higher premium, then we could make profits from this contract.

As long as the individuals in the population do not know their own type, this works fine. Risk-averse individuals would find it worthwhile to buy this contract. The problem comes if individuals can make a good guess as to which group they are in—perhaps because they know the history of this cancer in their own families. Then the insurance company might be in for a shock. Group B people would definitely want to buy the contract. The premium is less than their expected loss. But group A people might reason that they are very unlikely to get this disease and might decide that an insurance policy that costs \$550 is much too expensive, given that their expected loss is only \$100. This means that group A people—unless they are very risk-averse—choose not to buy the contract. Now the insurance company will only sell 1,000 contracts, bringing in revenue of \$550,000, but it will have to pay out \$1 million (= 10 × \$100,000).

If the insurance company could distinguish members of group A from members of group B, then it could offer insurance at different rates to the two groups. It could offer insurance to group A at a premium of \$100. They would find it worthwhile to buy this insurance. Likewise, the insurance company could offer insurance to members of group B with a premium of \$1,000, and they would also find it worthwhile to buy insurance. In practice, insurance companies often cannot classify people into such precise risk groups nor offer such targeted policies. In this case, the only kind of contract that is profitable for the insurance company is one that is aimed at the group B people only, with a premium of \$1,000. Group A people are left with a choice of buying no insurance at all or buying a policy that is vastly overpriced given their actual risk of contracting the disease.

This is an example of what economists call adverse selection: a situation in which individuals of different risk types decide whether or not to buy insurance (this is the selection). Lower-risk individuals opt out of the insurance market, leaving only high-risk individuals in the market (this is the sense in which the selection is adverse). Adverse selection is an information problem that is a source of market failure: low-risk individuals also want insurance, but it is unavailable to them at a reasonable price.

How do insurance companies deal with their informational disadvantage? One thing they can do is look for other sources of information. For example, firms presumably want to hire healthy, responsible individuals and put some time and effort into making good hires. Insurance companies can use the fact that you work for a firm as a (highly imperfect) signal of your health risk. This is one of the reasons why it is usually cheaper to get insurance through your firm than directly from an insurance company.

A second form of information about you comes from your history. If you have a car accident, your car insurance premium will increase. After an accident, your car insurance company revises its view of your riskiness and resets the price of your insurance. The analogous situation in health care is called preexisting conditions, meaning some disease or disability you already possess when you apply for insurance. For example, someone who has previously suffered a heart attack will find that insurance coverage is more expensive because the insurance company knows that this person is at greater risk of another attack.

If you apply for insurance and have a preexisting condition, then the terms of the insurance will reflect the chance that the condition will recur. This is reasonable enough: insurance is meant to provide protection against things that might happen to you in the future, not those that happened in the past. But it raises a problem with employer-provided health insurance. Suppose an individual has health-care coverage on the job and then suffers a heart attack. His current policy covers him because the heart attack was not a preexisting condition when he obtained insurance. But if he wishes to change jobs, his heart attack becomes a preexisting condition for his new insurer. This can make it very costly to change jobs—in turn making the economy function less efficiently.

Moral Hazard

Another complicating element for insurance is the moral hazard: the idea that, after purchasing insurance, individuals may behave in riskier ways. For example, think about your likelihood of being in a car accident. The probability that you will have an accident depends on many things: road conditions, the actions of other drivers, luck, and many others. It also depends on the actions you take as a driver of the car. There are many things we do that influence our likelihood of having an accident, including (but not limited to) the following:

- Properly maintaining the car
- Paying attention when driving
- Driving when tired
- Driving after consuming alcohol

These items are influenced by decisions that we make. The link back to insurance is that, if we are insured, we may make different choices about the condition of our car, the way we drive, and our physical state when we drive. The analogous idea with health insurance is that we may choose to live a less healthy lifestyle or engage in riskier behavior if we know that we have health insurance to cover our expenses if we become sick or injured.

Insurance companies understand very well that their policies influence the choices that people make. Their response is to design insurance contracts that provide insurance without affecting individuals' incentives too much. In the case of automobile insurance, you will not receive full coverage for your loss in case of an accident. Instead, insurance contracts typically include the following: (1) a deductible, which is the amount of a loss you have to cover before any insurance payment occurs, and (2) a copayment, which is the share of the loss for which you are responsible. The same applies to medical insurance. In the event you are ill, health insurance will typically cover a wide variety of medical costs, but there will usually be a deductible and often a copayment as well. As with property or automobile insurance, the deductible provides an incentive for you to take actions that make you less likely to claim against the policy.

There are two main moral hazard issues with health care. First, health care is an individual investment. Although no one wants to get sick, the more you pay for your own treatment, the more likely you will invest in your own health. Choices pertaining to exercise, diet, and preventive care can all depend on the insurance payments we anticipate if we need health care. The more insurance we have, the less incentive we have to take care of ourselves. And the less we take care, the more likely we are to present the insurance company with a sizable health bill.

Second, the size of the health bill also depends on your choices about treatment. When you are ill, you will meet with your doctor to jointly decide on treatments. Although your doctor will probably talk to you about various treatment options, their price will not be the focus of the discussion. Eventually you will meet with someone else in the office to discuss how your treatment will be paid for and, in particular, how much will be covered by your insurance. In the end, you have a menu of treatments and a menu of prices that you have to pay. You will then make a choice from this menu that is in your best interest.

The insurance company pays some of your bill, so the amount you pay is lower than the actual price of treatment. By the law of demand, you purchase more than you would if you had to pay the full price. For example, you might be much more inclined to get second and third opinions if you don't have to pay the full price for these.

Even if you are not ill but are instead going to see your doctor for a checkup, incentives still come into play. Many insurance policies include funding for an annual checkup with a small copayment. We respond to those incentives by going for the annual checkups covered under the policy. We don't go for checkups every month because such visits are not covered by most policies. The insurance company deliberately designs the incentives so you are likely to find it worthwhile to engage in basic preventive care.

Health Insurance and the Law of Demand

Now that we have a better understanding of health insurance contracts, we can say more about the demand for health care. We start with the cost of health care to us as households. We have just seen that if you have health insurance, the cost to you of a trip to the doctor is determined by your health insurance contract. Many of these contracts have a copayment provision—for example, you must pay $20 for an office visit. Of course, the doctor charges the insurance company much more for the visit, but you don't pay that cost. To you, a trip to the doctor costs $20.

The economic approach to individual choice still applies. Your demand for visits to the doctor comes from comparing the marginal valuation of these visits against this cost of $20. The law of demand works in the usual way: if your insurance company increased its copayment to, say, $50, you would make fewer visits to the doctor. The extent to which the quantity demanded responds to the price depends, of course, on what exactly is wrong with you. If you are seriously ill, your demand is likely to be inelastic. If you have only a sore throat, you might wait a few days to see if you really think you need medical care.

There is another element of the health insurance contract that has a direct effect on your demand. Consider a dental contract. These contracts often provide insurance up to an annual limit. If you need dental work, your dentist may design a treatment plan spread out over several years so that you can obtain maximum insurance coverage for the plan of work. In this case, you and your dentist are responding to the incentives of the dental contract.

KEY TAKEAWAYS

- Incentive problems of adverse selection (the health insurance provider not knowing your risk class) and moral hazard (actions you take to influence your probability of needing health care) are pervasive in the provision of health insurance. These incentive problems are present when the insurance is provided by private companies and the government.
- Health insurance in the United States is linked to your job as a consequence of legislation in 1942 that exempted the provision of insurance from controls on wages.
- When you have health insurance, your demand for health services will reflect the marginal cost to you of the service. This is usually through the copayment.

CHECKING YOUR UNDERSTANDING

1. Suppose the provision of health insurance at your firm induces you to stop exercising. Is this an example of a moral hazard or adverse selection?
2. If the copayment increases on your dental insurance, what will this do to the frequency of your visits to the dentist and the time you devote to taking care of your teeth?
3. What is the difference between a copayment and a deductible?

[1] Chapter 4 "Life Decisions" goes into much more detail about insurance and diversification.

[2] Melissa Thomasson, “The Importance of Group Coverage: How Tax Policy Shaped U.S. Health Insurance,” *American Economic Review* 93, no. 4 (2003): 1373–1384. See also the related discussion of the history of health insurance: Melissa Thomasson, “Health Insurance in the United States,” *EH.net*, February 1, 2010, accessed February 1, 2011,http://eh.net/encyclopedia/article/thomasson.insurance.health.us.

[3] The end of the act by President Harry Truman is documented at the American Presidency Project, “Executive Order 9801: Removing Wage and Salary Controls Adopted Pursuant to the Stabilization Act of 1942,” accessed March 14, 2011,http://www.presidency.ucsb.edu/ws/index.php?pid=60709.

[4] This idea is at the heart of Chapter 11 "Barriers to Trade and the Underground Economy".

[5] This table comes from US Census Bureau, *Income, Poverty and Health Insurance Coverage in the United States: 2009*, table 8, page 23, accessed March 14, 2011,http://www.census.gov/prod/2010pubs/p60-238.pdf.

[6] Medicaid (http://www.cms.hhs.gov/MedicaidGenInfo) is a joint federal-state program providing funding for health care to qualifying low-income households (defined relative to the poverty level).

15.3 Government Policy

LEARNING OBJECTIVES

1. What is the basis for government intervention in the market for health-care services?
2. What forms does this intervention take?

In the United States and around the world, governments are involved in the provision of health care. Some form of national health insurance is commonplace in Europe and Canada. In the United States, bills promoting national health care and universal health insurance have been debated for many years.

Here is a quote from President Dwight Eisenhower's 1954 State of Union Address: "I am flatly opposed to the socialization of medicine. The great need for hospital and medical services can best be met by the initiative of private plans. But it is unfortunately a fact that medical costs are rising and already impose severe hardships on many families. The Federal Government can do many helpful things and still carefully avoid the socialization of medicine." [1]

As we noted earlier, the federal tax code was modified in 1954 to provide incentives for employer-provided health insurance. We structure our discussion of government health policy by answering two questions: (1) why do governments intervene? and (2) how do they intervene?

Why Do Governments Intervene?

As usual, we analyze government involvement in the economy through the lens of market failure. When it comes to health care, there are several market failures to consider.

Externalities. One argument for public involvement in health care is the presence of externalities. If one individual is sick, then the likelihood that others around that person get sick increases. Individuals typically make decisions about their health care without thinking much about the effects of their decisions on the welfare of others. You may decide to go to work even though you are suffering from the flu because you need the money, and you may not think very much about your likelihood of infecting others. This is a classic example of an externality.

Commitment. Hospitals are often unwilling or unable to turn away individuals needing care but lack the resources to pay for that care. Through legislation passed in 1986, hospitals are required to treat patients in emergency situations, no matter what their insurance coverage. [2] In many cases, this is an inefficient way to treat people. For example, one consequence is that the uninsured have an incentive to seek normal care in hospital emergency rooms, even though this is an expensive place to provide care.

If hospitals could commit not to serve people unless they had health insurance, then some of the uninsured might be induced to purchase health insurance, instead of relying on emergency wards. But hospitals are not able to make such a commitment; although this might be more efficient, it is also unacceptably callous and runs counter to the Hippocratic Oath.

Adverse selection and moral hazard. We explained earlier that when insurance companies are unable to observe the probability of illnesses, some individuals will obtain insurance while others do not. Although these choices may be optimal from the standpoint of an individual, the market outcome is not efficient.

Drug quality. The health-care market is filled with gaps in information. Patients and even their doctors cannot fully assess the safety and efficacy of pharmaceutical products. Although the drug companies test their own products, the government has a role in assessing this information and determining the safety and effectiveness of medications.

Doctor quality. Another informational problem in the health-care market is the inability of a patient to properly evaluate the quality of a doctor. You as a patient can look at some indications of your doctor's ability, such as years of practice, school of graduation, and number of people in the waiting room. But it is not possible to make a fully informed judgment about the quality of your doctor. Again, the government plays a role by requiring that doctors obtain specialized training and pass a licensing examination before they are allowed to practice.

Patents. The research and development needed to create a new drug is substantial. [3] For firms to earn a return on this investment, the patent system exists to provide them protection from other firms producing the same product and selling it at a lower price. Although this type of competition may be valued*given that a product exists*, it destroys the initial incentives that a firm has to undertake research and development. Governments provide patent protection to induce firms to undertake the necessary research and development.

Market power. Market outcomes are not efficient when there are relatively few sellers of a product. This may occur in various health-care markets because there may be relatively few doctors and few hospitals in a given location. Furthermore, pharmaceutical companies have market power based on exclusive knowledge of their specific product, as protected through patents. Finally, there are relatively few health insurance providers, and some are very large.

Equity and fairness. Even if health-care markets were efficient (and we have explained many ways in which they are not), they may not be equitable. One argument for government involvement is to provide for a more equitable allocation of goods and services. From this perspective, the fact that many Americans lack health insurance and adequate health care is also a basis for government involvement.

Article 25 of the Universal Declaration of Human Rights includes the right to health care: "Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control" (http://www.un.org/en/documents/udhr/index.shtml). [4] To the extent that basic health care is viewed as a basic human right, then the government ought to guarantee access to at least a minimal level of care.

Indeed, if the aim is to move toward equality of well-being, there is an even stronger equity argument for health care. Imagine for a moment that people could decide how to allocate health care and other resources before they knew anything about their own health or well-being. (Thought experiments of this kind are associated with the philosopher John Rawls. [5]) People might well agree that those who became sick or disabled should be given *extra resources* to compensate them for their ill health.

How Do Governments Intervene?

Now that we have some understanding of the sources of market failure in the health-care market, we turn to a discussion of government policy.

Taxes and Subsidies

We have already mentioned one of the key ways in which the government subsidizes health care—by allowing employees tax-free health insurance benefits provided by an employer. In this way, the government reduces the cost of firm-provided health care. It is now common for employment contracts in the United States to include provision for health care.

One of the main issues surrounding employer-provided health insurance is the possibility of losing insurance when you change jobs (sometimes called the “portability problem”). In our economy, shifts in demand for goods and services and changes in productivity naturally lead to the creation of new jobs by some firms and the destruction of jobs by other (perhaps less profitable) firms. The efficient working of an economy therefore requires that workers leave old jobs for new ones. Unfortunately, insurance can get in the way of worker mobility. If you have a job with health insurance, then quitting your job to look for another may be costly for several reasons. First, you may lose insurance coverage during the period of job search. Second, an ailment that was covered by insurance by your existing firm could be viewed as a preexisting condition when you acquire insurance at a new firm. This can have an adverse effect on your insurance rates and the type of coverage you can obtain. In some cases, people choose not to change jobs purely because of the implications for health insurance.

Health care is also subsidized through income taxes. If you look carefully at your income tax forms, you will see that you can deduct medical expenses. If you itemize deductions on your tax form, and if your medical expenses are substantial enough, you can offset those payments against your taxes.

Regulation

Government regulations are common in the health industry. These regulations influence both demand and supply in this market.

On the demand side, households are required to obtain certain medical services. Some vaccinations are mandatory, for example. At the college level, there is ongoing concern about the spread of meningitis. [6] With concerns like this in mind, it is also common for colleges to require some vaccinations prior to admission. The argument for such interventions is that there are externalities from your health to the health of others.

The government licenses many of the actors on the supply side of the health-care market. This is another form of quality control. Doctors who practice in a state must pass exams called medical boards. Hospitals are certified for the types of activities they offer. Often the certification occurs at the state level. [7]Other providers of health care are also licensed. For example, a nursing home must be certified as a Medicare provider to receive reimbursements. The rationale for such interventions stems from the extensive information problems in the health-care market. As consumers are unable to accurately assess the quality of care provided by doctors and hospitals, the government provides a service to us all by regulating health-care providers.

Provision of Insurance

The government, through its Medicaid and Medicare programs, provides insurance to both low-income and elderly households. There is continuing debate about expanding the availability of health insurance to the general population. On March 23, 2010, President Obama signed a health-care reform bill. [8] The main goal of the bill is to reduce the number of individuals without health insurance in the United States. This bill seeks to achieve this by requiring that everyone purchase health insurance, either through an employer or individually. The legislation provides opportunities for households to obtain insurance on their own through subsidies and a "marketplace for insurance."

The bill regulates insurance policies in several ways. For example, it places limits on the ability of insurance companies to exclude people from coverage due to preexisting health conditions or other health risks. It also restricts the ability of insurance companies to set and change rates on insurance policies.

The new policy will not be fully in force until 2014. It is extremely complex (the bill itself is almost 1,000 pages long), and its impact on health-care outcomes, health-care costs, and the deficit remains an open question. (Even the short summary of the act [9] does not make easy reading.) By the time of its implementation, details on the new law may be clearer. In particular, exactly how insurance markets will be organized and regulated will be made more precise. Further, when the bill was passed, estimates were made of the cost savings from the measure. Over time, we will be better able to forecast the spending and taxation implications of the bill once household and firm responses are observed. Then we can see if this legislation improves the efficiency of the health-care market.

A more fundamental question is whether the government should even be in the business of providing health insurance. One set of arguments for government involvement rests on the various market failures that we have identified in this chapter. Health care is complicated, and there are many ways in which health-care markets depart from the competitive ideal. It is sometimes argued that spending on health services in the United States is very high because the market is very inefficient. From that perspective, having the government in charge of this sector of the economy might reduce inefficiencies. Second, government involvement can be justified on the grounds of equity and fairness.

Provision of Information

One of the primary roles of the government is to provide information to the public about health matters. This comes in a variety of forms. In January 1966, the following warning first appeared on cigarette packs: "Warning: Cigarette Smoking May be Hazardous to Your Health." This initial warning from the Surgeon General's office of the United States was followed by many others concerning the consumption of cigarettes and other potentially harmful products. Such warnings are a good example of government provision of information. Each consumer of these products wants to know the impact on health. Gathering such information is a public good because the information is available to everyone and can be "consumed" by everyone simultaneously.

Another form of information is through drug testing. The US Food and Drug Administration (FDA; http://www.fda.gov) is responsible for testing drugs before they appear on the market. The FDA also supplies public information about a wide range of food items.

KEY TAKEAWAYS

- Government intervention in this market reflects inefficiencies in the market as well as concerns over equity.

- Government intervention takes many forms around the world, including the provision of health insurance, the direct provision of health-care services, and the regulation of drug companies.

CHECKING YOUR UNDERSTANDING

1. What is the commitment problem of a hospital?
2. Recent concerns about the H1N1 virus led the governments in many countries to intervene. How would you explain the basis for this intervention using the list of market failures provided in this section?
3. Give a recent example where the government provided a health warning.

[1] AMDOCS: Documents for the Study of American History, "Dwight Eisenhower, 'State of the Union, 1954,'" accessed March 14, 2011,http://www.vlib.us/amdocs/texts/dde1954.htm.

[2] The regulation is called EMTALA (http://www.emtala.com/history.htm). It literally applies only to those hospitals that accept Medicare, but this is almost universal.

[3] We also discuss this in Chapter 14 "Busting Up Monopolies".

[4] United Nations, "Universal Declaration of Human Rights," accessed March 14, 2011,http://www.un.org/en/documents/udhr/index.shtml.

[5] In Chapter 12 "Superstars", we discuss this kind of thought experiment in more detail.

[6] See "Meningitis on Campus," *American College Health Association*, April 27, 2005, accessed February 1, 2011,http://www.acha.org/projects_programs/meningitis/index.cfm.

[7] For details about accreditation in Texas, for example, see Texas Department of State Health Service, "General Hospitals—Health Facility Program," accessed March 14, 2011,http://www.dshs.state.tx.us/HFP/hospital.shtm.

[8] Ample discussion by the White House along with the final bill is available at "Health Reform Puts American Families and Small Business Owners in Control of Their Own Health Care," The White House, accessed March 14, 2011, http://www.whitehouse.gov/health-care-meeting/proposal.

[9] Library of Congress, "Bill Summary & Status: 111th Congress (2009–2010), H.R.3590, CRS Summary," accessed March 14, 2011, http://thomas.loc.gov/cgi-bin/bdquery/z?d111:HR03590:@@@D&summ2=m&.

15.4 End-of-Chapter Material

In Conclusion

The debates that we have introduced in this chapter are far from settled. The issue of health-care policy is one that is not likely to be quickly resolved in the United States or elsewhere in the world. At the moment, we see many different ways in which health care is provided in different countries.

There is enormous pressure within the United States to deal with the perceived problem of uninsured households. As we have seen, this is a key element of the health-care bill that was signed into law in 2010. In addition, there are unsolved problems associated with Medicare and Medicaid. These government programs are in need of reform to deal with the escalating costs of medical care. According to the General Accounting Office, the current level of Medicare and Medicaid spending is about 4 percent of the gross domestic product (GDP). If there are no changes in current programs, this fraction is expected to increase to about 11 percent by 2050. [1] To put this in perspective, government outlays as a fraction of GDP have typically been about 20 percent of GDP over the past 40 or so years. This number is forecast to increase to nearly 27 percent by 2050, with a significant fraction of this driven by Medicare and Medicaid programs.

There are two overarching lessons to be drawn from this chapter.

1. **Incentives matter.** Whatever programs are deemed to be desirable for dealing with health-care problems, we need to be sure to take into account the incentives that these policies will create.
2. **People are different.** Health-care providers, insurance companies, and the government cannot observe many of these differences. Policies must take into account all these differences and recognize the importance of adverse selection. To the extent that markets fail because of adverse selection, government policy must address this source of the underlying problem of health care.

The nature of adverse selection is heavily influenced by technology. Recently, the US Congress has been considering legislation, HR 493, [2]which limits the ability of insurance companies to use genetic information. If insurance companies have more information about individual health risks, then they can design more targeted insurance contracts. From the perspective of efficiency, this might seem to be a good thing because it eliminates some adverse selection problems. Imagine that technology were to reach a point where all your major health risks could be identified from your genetic code at birth. Almost all the uncertainty over your health would disappear, and there would be almost no role for health insurance. (There could, of course, still be insurance for accident risk.) Although this world might be more efficient, it would also be much more unequal. People with genetic predispositions to certain illnesses would face steep medical bills, while healthy people would not.

In this world people would want insurance before they were born, while there was still uncertainty about their genetic makeup, which takes us back to the Rawlsian thought experiment that we discussed earlier. It might be that, paradoxically, one of the most compelling arguments for government-provided universal health care will turn out to be the gradual *elimination* of market failures from adverse selection.

Key Links

- World Health Organization: http://www.who.int/nha/en
- US Department of Health and Human Services: http://www.hhs.gov
- US Census Bureau, "Income, Poverty and Health Insurance Coverage in the United States: 2009" http://www.census.gov/prod/2010pubs/p60-238.pdf
- Bureau of Labor Statistics, Consumer Price Index in medical care:http://www.bls.gov/cpi/cpifact4.htm

EXERCISES

1. Is health insurance a complement or a substitute for the demand for health-care services?
2. If doctors no longer needed a license to practice so that there was free entry into the provision of that service, what would happen to the price and quantity of health-care services? What would happen to the quality?
3. Give an example of technological innovation in the health-care industry.
4. If adverse selection is a problem, does allowing an insurance company to know your entire health history improve market efficiency?
5. What type of inefficiencies does the commitment problem of a hospital create? Why don't restaurants have this same problem?
6. One government intervention in health care is compulsory vaccinations for children against various infectious diseases. Can you explain why governments might enact such policies? (Hint: are there any externalities involved?)

7. In some countries and some regions, there are shortages of doctors. Why is this problem not quickly resolved by the normal workings of supply and demand in the labor market?
8. As the United States has become richer, an increasing proportion of GDP is spent on health care. Does this fact, in and of itself, indicate inefficiencies in health care? (Hint: to what extent are aspects of health care luxury goods?)
9. One complicated part of the demand for health care is that consumers are not quite sure of the quality of the product they are buying. Can you think of other goods or services that have this same property? Are there measures to protect consumers? Why don't sellers sell only low-quality goods to consumers who are not able to judge quality?
10. In the health-care market, private and public hospitals coexist. Can you think of another market in which both public and private providers exist? What are the differences in that market among the public and private firms?
11. What are the incentives for parents to provide health care for their children?
12. The next time you visit a doctor, ask for a price list. Discuss what happens.
13. The government provides deposit insurance so that funds deposited at a bank are insured even if the bank goes out of business. What are the moral hazard implications of providing deposit insurance?
14. What does the link of health-care coverage to employment do to the incentives of someone to quit one job and look for another?

Economics Detective

1. Table 15.4 "The Uninsured (in Millions)" came from a 2006 and 2010 census. What is the current number of uninsured? What fraction is under 18 and in the 18–24 age group?
2. Find out about the health-care system in France, Sweden, or Canada. How does it compare with the US system? How do health outcomes in the country you have chosen compare with those in the United States?
3. The election results in November 2010 reflected, in part, concerns over the health-care bill signed into law in March 2010. What were the main concerns discussed in the election campaign with regard to the bill?

[1] This draws the discussion of the US budget through 2075: "A 125-Year Picture of the Federal Government's Share of the Economy, 1950 to 2075," *Congressional Budget Office*, July 3, 2002, accessed February 1,

2011, http://www.cbo.gov/ftpdoc.cfm?index=3521. Also see David M. Walker, "U.S. Financial Condition and Fiscal Future Briefing," US Government Accountability Office, January 2, 2008, accessed February 1, 2011,http://www.gao.gov/cghome/d08395cg.pdf.

[2] Govtrack.us, "Text of H.R. 493 [110th]: Genetic Information Nondiscrimination Act of 2008," accessed March 14, 2011, http://www.govtrack.us/congress/billtext.xpd?bill=h110-493&show-changes=0&page-command=print.

Chapter 16
Cars

Looking into Cars

Cars are so common that we rarely give them a second thought. If you live in the United States, then you either already have a car or will most likely have one in a few years' time; over 90 percent of households in the United States either own or lease a car. Although other countries do not have quite the same levels of car ownership as the United States, there are more than half a billion automobiles in the world.

The familiarity of the car is so great that it is easy to forget how the automobile transformed the world. The automobile made modern cities and suburbs possible because people were no longer obliged to live close to where they worked. The automobile made it easier to transport goods from place to place, dramatically altering patterns of trade in the global economy. At the same time, automobile emissions have degraded the air we breathe to the point where they sometimes seriously damage people's health. Indeed, because emissions also contribute to the accumulation of greenhouse gases in the atmosphere, the automobile may be changing the very climate of the planet.

Although you may own a car, it is likely that you are unfamiliar with how it works. Half a century ago, car owners were typically very knowledgeable about how their vehicles operated. They needed to be because cars broke down frequently. People knew how to adjust spark plugs, clean distributor caps, and so on. But the modern automobile is a remarkably sophisticated and complex piece of engineering. Today, it is unlikely that an owner of a modern-day car knows how to do much more than very basic maintenance. Even car mechanics rely on computer diagnostics to perform repairs.

Just as the product itself has become increasingly complex, so too has its method of manufacture. In the early years of the 20th century, cars were produced in small numbers and largely by hand. In 1913, however, Henry Ford introduced mass production of cars at Ford's Highland Park plant. [1] By the middle of the 20th century, cars were typically produced on assembly lines. In contrast to the early years of car production, there were far fewer workers at this stage of the production process. [2]

If you were to visit the production line at a modern automobile manufacturing plant, you would hardly see any people at all. Modern production uses a great deal of capital and relatively little labor. Computerized robots perform manufacturing roles. Yet despite the relative absence of workers on the production line, over one million workers are employed in this sector of the economy in the United States. [3] In 2006, about 360,000 jobs were associated with the production of automobile parts alone. [4] In the 21st century, though, there have been significant job losses in this part of the economy.

Road Map

This chapter is different from others in this book because it is a capstone discussion. We use the automobile industry to illustrate the different ideas we have explained in the book. We also use this industry to provide further examples of how to use the different tools we introduced in previous chapters. Whereas other chapters were largely self-contained, here we will repeatedly remind you of ideas that we have already studied.

We begin our look at cars in a familiar way, using the supply-and-demand framework. All of us who own cars reside on the demand side of the market. We make choices about the type of car we want, whether to buy a new car or a used one, and when to replace it. We also make decisions about related products like gasoline and insurance. The supply side of the car market illustrates technological progress, enormous growth in product diversity, the impact of trade on domestic markets, and the social costs of automobiles. We examine some decisions of automobile producers, including where to locate their operations, why they introduce new models, and what price they should set.

We then study the equilibrium of the car industry. The US car industry began around the start of the 20th century, survived the Great Depression of the early 1930s, and has been transformed by international competition. Understanding these dynamics provides a perspective on other industries. After understanding industry equilibrium, we look at the variety of government policies that impact this industry, including trade and environmental policies.

[1] The Library of Congress has an extensive discussion of Henry Ford, including photos of production in 1923 at http://memory.loc.gov/ammem/today/jul30.html.

[2] You can find more details at How Stuff Works, "1957–1959 Ford Fairlane 500 Skyliner," accessed March 14, 2011, http://auto.howstuffworks.com/1957-1959-ford-fairlane-skyliner3.htm.

[3] Basic information about firms and workers in this sector comes from the Bureau of Labor Statistics, "Career Guide to Industries, 2010–11 Edition,"http://www.bls.gov/oco/cg/cgs012.htm.

[4] This number is from a Bureau of Labor Statistics study of employment in the automobile parts industry: Benjamin Collins, Thomas McDonald, and Jay A. Mousa, "The Rise and Decline of Auto Parts Manufacturing in the Midwest," *Monthly Labor Review Online* 130, no. 10 (2007): 14–20, accessed March 14, 2011,http://www.bls.gov/opub/mlr/2007/10/art2full.pdf.

16.1 The Demand for Automobiles

LEARNING OBJECTIVES

1. What determines a household's demand for a car?
2. In what sense is a car an asset?
3. What are some of the complementary goods and services for cars?

People don't buy cars only because they want to look at a piece of fine engineering and enjoy a luxurious ride (although this sometimes plays a role). They buy cars because they want to be able to travel from one place to another. The demand for automobiles is a piece of a larger market: the demand for transportation in general. As the price of a particular car increases, the law of demand tells us that the quantity demanded of that car will decrease. There are three kinds of substitution at work here. In response to a price increase, households can

- delay the purchase of a new car,
- choose to purchase another type of car, or
- choose not to buy a car and use another mode of transportation instead.

Suppose you are thinking of buying a car, but the price of your favorite model increases. One possible response would be to delay your purchase until a later time. With this form of substitution, you decide not to buy a new car right now. This does not mean you will never buy a new car. Instead, you are keeping your options open for the future: you will drive your old car for perhaps another year and then search again next year for a replacement. Next year, of course, you might decide to defer your purchase still further. A second possible response to a price increase in your preferred model is to purchase another type of car. There is a substitution effect at work again, but now it applies across cars rather than over time. Perhaps you are indifferent between buying a Ferrari racing car and a Mini Cooper. If the price of the Mini Cooper increases, you would be induced to buy the Ferrari. Finally, if the price of your preferred car increases, you might substitute another form of transportation for your car, such as a bicycle or the bus. From this perspective, your demand for a car is really a demand for transportation.

Household Demand for Cars

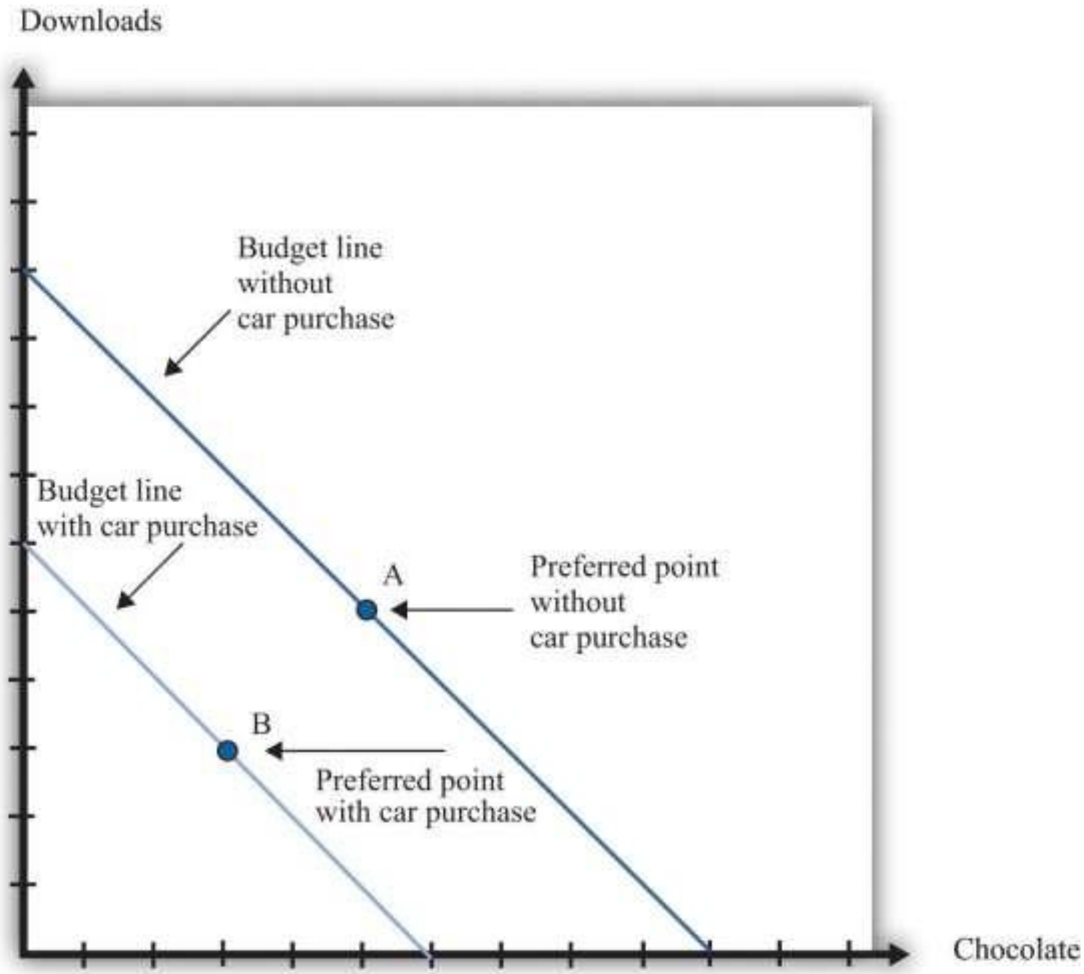
The decision to buy a car is best understood as an example of *unit demand*. Most households—even if they own more than one car—do not buy a large number of cars at a time. Instead, they buy a single car. The decision about whether or not to purchase a car thus involves comparing the **valuation** a household places on the car with the price of the car. One way to illustrate this is to look at the household's **budget line** when it does or does not purchase a car.

Toolkit: Section 17.1 "Individual Demand"

You can review unit demand, valuation, and budget lines in the toolkit.

The household's choice is shown in Figure 16.1 "The Household's Budget Line When It Does or Does Not Buy a Car". (This draws on our presentation of unit demand in Chapter 3 "Everyday Decisions".) The household can spend income on three items: chocolate bars, downloads, and a new car. If the household chooses not to buy a car, then it consumes the combination of downloads and chocolate bars indicated by point A in the graph. This is the household's most preferred point in the budget line, given that it does not buy a car. If the household buys a car, then the combination of downloads and chocolate bars it consumes is given by point B. The budget line is shifted inward by the amount of income the household spends on the car—that is, by an amount equal to the cost of the car. The household's decision about whether to purchase the car involves comparing bundle A to (bundle B plus a car).

Figure 16.1 The Household's Budget Line When It Does or Does Not Buy a Car



Here we illustrate a household's choice about whether or not to purchase a car.

Remembering the idea of **buyer surplus**, this is the same as saying that the household would buy the car if the purchase gave it some surplus. In other words, the household's decision rule is

purchase car if valuation of car − price of car = buyer surplus > 0.

If the price of the car is greater than the household's valuation, the household prefers point A to point B and does not buy the car. If the price of the car is less than the household's valuation, it prefers point B to point A and buys the car.

Toolkit: Section 17.10 "Buyer Surplus and Seller Surplus"

Buyer surplus was introduced in Chapter 5 "eBay and craigslist". You can also review the various kinds of surplus in the toolkit.

There is another way of looking at the same decision that gives us a way to measure the household's valuation of the car. Remember that the household's valuation of the car is the *maximum* amount that it would be willing to pay for it. Look again at Figure 16.1 "The Household's Budget Line When It Does or Does Not Buy a Car" and begin at point A. Now move the budget line inward until we find that the household is just as happy at point A or point B. We have now found the point where the household is *indifferent* between the combination of chocolate bars and downloads it buys without a car and the bundle it buys along with the car. The amount by which we have moved the budget line is the household's valuation of the car.

If there were only a single model of car for the household to choose from, we could stop here. The household would compare the valuation of the car against the price and buy the car as long as the valuation is greater than the price. Today, however, cars differ in numerous ways. Like many goods, a car consists of many different features all bundled together. These include the car's performance features, styling, color, and sound system; whether it has leather seats, a sunroof, and air conditioning; and hundreds of other attributes that we could list. The household's valuation of a car embodies a valuation of each attribute of a car.

This complexity makes the decision to buy a car a challenge. How can this decision be made? For every car available on the market, the household can calculate the buyer surplus attainable from that car. After considering all these alternatives, the household should then buy the car that gives the most surplus. Of course, households do not literally sit down with a list of cars and try to calculate the exact surplus from each one. But this is a useful, if stylized, representation of how such choices are made. In effect, the household is making a unit demand decision—buy or not buy—about every single car. For almost all cars, the household chooses to purchase zero.

(There is a subtlety you may be wondering about here. Hundreds of different cars might yield positive surplus, but obviously the household does not buy hundreds of cars. The trick is that, once the household has bought the car that gives the highest surplus, the valuation of all other cars it might consider buying decreases substantially. If you don't own a car, then a Ford Focus might be very valuable to you. If you already own a Mazda 5, then the value of a Ford Focus would be much smaller.)

Deciding what car to buy is only one part of the household's decision. As we already noted, the household must decide *when* it wants to buy a car. A car is adurable good; it lasts for several years. If a household already has a car, it can decide to defer purchase of a new car until later. A household is likely to do this if (1) there is substantial uncertainty about future income (perhaps members of the household fear losing their jobs) or (2) cars are likely to be relatively cheaper in the future. To understand this choice, we turn to some of the tools introduced in Chapter 4 "Life Decisions", and Chapter 9 "Making and Losing Money on Wall Street".

A car is an asset: it yields a flow of services. As a consequence, buying a car is both an act of consumption and an act of saving. This means that the decision to buy a car is an example of decision making over time. The household looks at both current and expected future income when deciding about the purchase, and it knows that the car will yield benefits for many years. Furthermore, because a car is an asset, its valuation today depends on its value in the future. You might buy a car this year and then discover your transportation needs have changed. In that event, you can sell your car in the used car market. The more you expect to get for your car whenever you might sell it, the more you are willing to pay for it today.

The demand for a particular car also depends on factors other than the price of the car itself. The prices of other goods—most importantly, other cars and other forms of transportation—also matter. Household income, both now and in future years, is another determinant of demand. Finally, because you often purchase a car with a car loan, the interest rates charged on loans may matter for your car purchase. A decrease in the interest rate for car loans will increase the demand for cars.

In sum, buying a car is a very complex decision. There are rich substitution possibilities involving the choice of different models, the timing of the purchase, and the possibility of using public transportation rather than owning a car at all. The law of demand applies to cars, just as it does to other goods and services. But as we move along the demand curve in response to a change in the price of cars, the substitution possibilities are complex. Understanding these substitution possibilities is critical when firms are choosing the prices to set for the cars that they produce.

Complementary Products

There are several products that are complementary to the purchase of a car. Here we look at three: gasoline, insurance, and infrastructure. A complementary product is one for which the cross-price elasticity of demand is negative. In other words, we expect that if the price of gasoline increases, the quantity of cars purchased will decrease.

Toolkit: Section 17.2 "Elasticity"

The cross-price elasticity of demand measures the response of the quantity demanded of a good to a change in the price of another good. Formally, it is the percentage change in the quantity demanded of one good divided by the percentage change in the price of another good:

$$\text{cross-price elasticity of demand} = \frac{\text{percentage change in quantity of one good}}{\text{percentage change in price of another good}}.$$

- If the cross-price elasticity of demand is a negative number, then the quantity demanded of one good decreases when the price of the other good increases. In this case, we say that the two goods are**complements**.
- If the cross-price elasticity of demand is a positive number, then the quantity demanded of one good increases when the price of the other good increases. In this case, we say that the two goods are **substitutes**.

Gasoline

During late spring of 2008, the price of oil and gasoline rose considerably. By the end of May, the price of unleaded gas averaged over $4.00 per gallon in major US cities. Figure 16.2 "Gasoline Prices: Pennies per Gallon" shows the price of regular gasoline in the United States. The first series is just the price of a gallon of gas in pennies. The second shows prices after *correcting for inflation*; all prices are quoted relative to the price level in the base year of 1990. You can see from this figure that the dollar price of gas has increased steadily since August 1990, but once we correct for changes in the overall price level, the real price of gas was actually decreasing until around 1998. Since then, it has increased to about twice its 1990 level.

Toolkit: Section 17.8 "Correcting for Inflation"

You can review how to correct for inflation in the toolkit.

Figure 16.2 Gasoline Prices: Pennies per Gallon

What is the impact of an increase in the price of gasoline on the demand for cars?

- As the price of gas increases, buying and operating a car becomes more expensive. Thus we expect the demand for cars to decrease.
- An increase in the price of gasoline induces a move away from cars with low mileage per gallon to cars with higher mileage per gallon. This type of substitution is often seen after rapid increases in the price of gasoline.
- As the price of gasoline increases, individuals substitute competing means of transportation for cars. For example, an increase in the price of gasoline might induce commuters to use bicycles, take buses or trains, or walk to work.

All of these channels were in the news in 2008 as consumers responded to higher gasoline prices by driving less and buying more fuel-efficient vehicles.

If we think of cars in general, then the cross-price elasticity of demand with the price of gasoline is negative. But the second channel reminds us that, for fuel-efficient models, the cross-price elasticity of demand might be positive. Higher gas prices mean that fewer cars will be purchased, but households that do purchase favor cars that are more fuel-efficient. Remembering that cars typically last for several years, households think about not only the current price of gas but also what they expect gas prices to do in the next few years. If gas prices increase but consumers think that this increase is likely to be temporary, then people will drive less, but the demand for cars will be little affected. Conversely, if gas prices increase and consumers expect them to stay high for some years, we see a much bigger effect on the demand for automobiles.

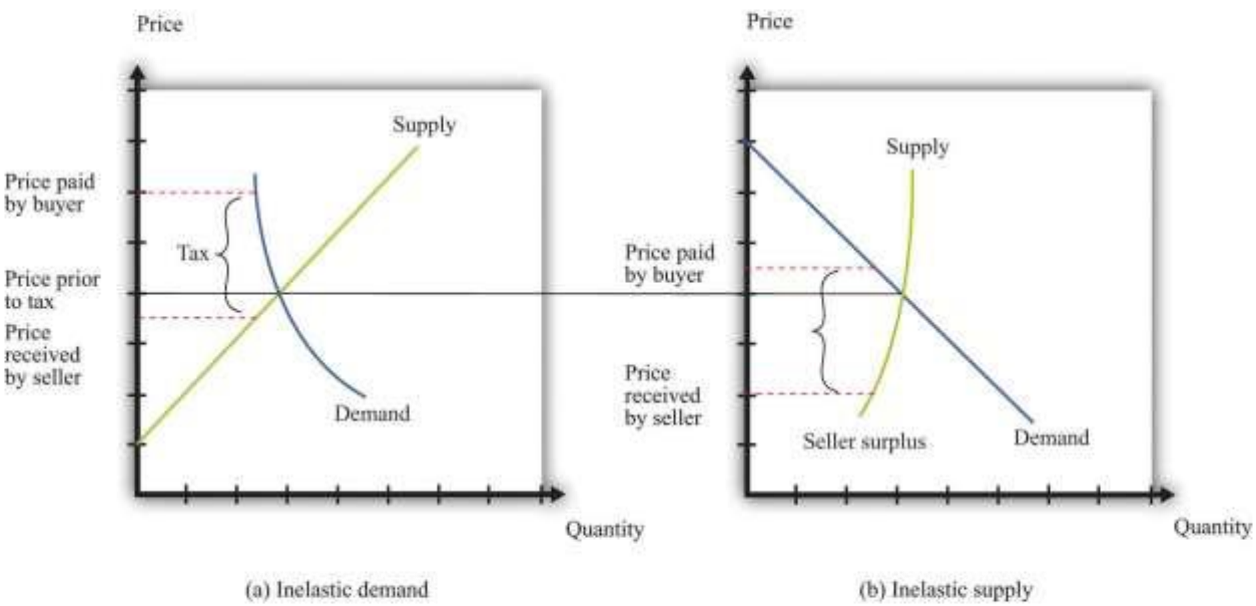
The gas price increase in 2008 coincided with an intense battle between Senator Hilary Clinton and Senator Barack Obama for the Democratic presidential nomination. By this time, Senator John McCain had effectively won the Republican nomination. Senator Clinton and Senator McCain proposed a temporary reduction in the tax on gasoline in an effort to shield households from the high price of gasoline during the summer season. To see the effects of such a tax, we can use the tool of **tax incidence**, which we introduced inChapter 11 "Barriers to Trade and the Underground Economy".

Toolkit: Section 17.11 "Efficiency and Deadweight Loss"

You can review tax incidence in the toolkit.

Figure 16.3 "The Incidence of a Tax on Gasoline" shows the effect of a tax on the price paid by a buyer and the price received by a seller. In part (a) of Figure 16.3 "The Incidence of a Tax on Gasoline", the demand for the product is very inelastic. As a consequence, when a tax is imposed, the price paid by the buyer increases a lot compared to the price in the absence of a tax. This means that when demand is inelastic, buyers bear the tax burden. Consequently, if a tax is removed, even temporarily, the price buyers pay will decrease considerably. As shown in part (b) of Figure 16.3 "The Incidence of a Tax on Gasoline", we reach the opposite conclusion if supply rather than demand is inelastic. When supply is inelastic, the price received by the seller varies with the tax, while the price paid by the buyer is almost independent of the tax.

Figure 16.3 The Incidence of a Tax on Gasoline



When demand is inelastic (a), the price paid by buyers increases a lot, and the price received by sellers decreases only a little bit, so most of the tax burden is borne by buyers. The opposite is true when supply is inelastic (b).

Whether a temporary repeal of the gas tax would reduce the price of gas depends on the elasticities of supply and demand in this market. The more inelastic demand is relative to supply, the more the tax reduction will lower the price paid by households. In fact, the supply of gas tends to be very inelastic in the short run because refining capacity is limited. Part (a) of Figure 16.3 "The Incidence of a Tax on Gasoline" is the one relevant for a temporary change in the gas tax.

Insurance

A second complementary product is car insurance. In most countries, having car insurance is mandatory. Typically, drivers must at the very least purchase some liability coverage, meaning that your insurance company will pay out if you are responsible for an accident that injures or kills another person or causes damage to a car or property. You may also choose to buy collision coverage to cover damage to your own car. You can also purchase "uninsured motorist coverage," which protects you in the event you are in an accident with someone who is uninsured. Exactly what type of coverage you are able (or required) to purchase varies from country to country and from state to state in the United States.

You may be aware that your insurance rates depend on your age, gender, and driving record. Insurance companies work very hard to determine the probability you will have an accident and make a claim for funds from them. But they do not know exactly what that probability is for each person individually. Thus they rely on information about us, such as our age and sex, and also look at driving records for indicators of the likelihood that we will file a claim. If you have an accident and file a claim, the insurance company will often revise its assessment of how safe a driver you are and increase your premium. Such adjustments can be so severe that sometimes people prefer to fix their own car after a minor incident than have the insurance company handle a claim. It is even possible to obtain insurance against the costs of obtaining a traffic ticket. Getting a ticket can be expensive, both directly in terms of a fine and then through an increased insurance premium. Recognizing the desire of a household to shed this risk, insurance policies that compensate ticketed drivers for these costs are available. [1]

Government restrictions on trades often lead people to avoid these restrictions. This was one of our themes in Chapter 11 "Barriers to Trade and the Underground Economy". In many states, insurance is required by law, yet drivers sometimes flout this law, choosing to drive without insurance instead. The law of demand tells us that the higher the cost of insurance, the more likely people are to drive without insurance.

If high insurance rates lead to a large number of uninsured motorists, then it is more likely that, if you have an accident, the other driver will be uninsured. In this case, your insurance company will be obliged to cover your damages, even if the other driver was at fault. Think for a moment about what this means from the perspective of an individual insurance company. If there are more uninsured motorists, then the insurance company expects to make higher payouts per accident, on average. This means that when there are more uninsured motorists, insurance premiums will be higher.

Economists Eric Smith and Randy Wright noticed that insurance premiums are very different in different places. In a paper titled "Why Is Automobile Insurance in Philadelphia So Damn Expensive?" they speculated that this was because of a **coordination game**. They argued that the decision to purchase insurance could lead to payoffs like those in Table 16.1 "A Coordination Game for Automobile Insurance". The rows show your decision about whether to buy insurance or not. The columns show other people's decisions. The numbers in the table refer to your payoff from every combination of what you choose to do and what everyone else chooses. (To keep things simple, we state what your payoffs are when everyone else does the same thing. We also suppose that everyone else faces the same payoffs you do.)

Toolkit: Section 17.18 "Nash Equilibrium"

You can review the coordination game in the toolkit.

Table 16.1 A Coordination Game for Automobile Insurance

	Everyone Else Buys Insurance	**Everyone Else Does Not Buy Insurance**
You Buy Insurance	10	2
You Do Not Buy Insurance	4	8

Look at the first column. This says that if everyone else purchases insurance, then you have an incentive to do so as well. We see this in the table because 10 is greater than 4. Because everyone else buys insurance, the price of insurance will be relatively low, and you will be induced to buy insurance as well. Thus one Nash equilibrium of this game is for everyone to buy insurance. Now look at the second column. Here, everyone else is an uninsured motorist. If no one else buys insurance, your insurance will be very costly. This will induce you not to buy insurance either (8 is greater than 2). Thus there is a second Nash equilibrium of this game in which no one buys insurance.

If, as Smith and Wright suggested, Philadelphia was an example of a city in which no one bought insurance, you can see from Table 16.1 "A Coordination Game for Automobile Insurance" that everyone there *wished* that other people bought insurance. Everybody is better off in the Nash equilibrium where everyone buys insurance. But starting from the equilibrium in which no one buys insurance, no single individual, acting alone, can coordinate everyone else's choices to reach the preferred outcome.

Infrastructure

Gasoline and insurance are products that are complementary to automobiles. There is another significant complementary product—the roads on which you can drive your car. Without roads, cars have limited value. The same argument applies to bridges and highways and even to the police who enforce the laws of the road. These various kinds of infrastructure serve to increase the value of a car.

This cartoon, which is taken from an article on the history of automobiles in the early 20th century, illustrates the link between car demand and roads. [2] The value of a car is much higher in the setting labeled "good roads" compared to that labeled "bad roads." In the developed countries of the world, we now take good roads as a given, but that was not the case at the start of the automobile industry. The evolution of roads was directly linked to the spread of automobiles as a form of transportation. There was a seasonal aspect to this as well. After a long hard winter, the roads were not ready for use, and additional maintenance was needed to put them back into shape for drivers.

The infrastructure of an economy is a special type of good, called apublic good. A public good has two characteristics:

1. **Nonrivalry.** Public goods are nonrival. A good is nonrival if one person's consumption of that good does not prevent others from consuming the good as well. In other words, if you supply a nonrival good to one person, you can supply it to everyone.
2. **Nonexcludability.** Public goods are nonexcludable. We discussed this idea in Chapter 13 "Cleaning Up the Air and Using Up the Oil". A good is nonexcludable if it is not possible to selectively exclude people from access to the good. In other words, if you supply a nonexcludable good to one person, you *must* supply it to everyone.

There are many examples of public goods, such as roads, bridges, highways, police services, national defense, and lighthouses. Because public goods are nonexcludable, it is difficult for a private firm to produce them. After all, it is hard to expect someone to pay a positive price for a good if she can always get it for free. Instead, governments generally provide these goods. In the United States, local or state governments may provide roads and bridges, while the federal government is the provider of the highway system. Because these goods are not privately produced and traded in the economy, we cannot rely on supply and demand to determine the quantity of these goods produced in an economy. The quantity of public goods produced is an outcome of a political process.

Suppose the government is thinking about building a bridge. The cost of building the bridge depends on the design and the cost of materials and labor. We take this cost as given. Because everyone in the community can use the bridge, the benefits flow to everyone, not to any particular individual. Suppose the government knows how much each citizen in the community values the bridge: call these the individual valuations. Consider then the following procedure.

1. The government adds together the individual valuations of the bridge. Call the total the *social benefit* from the bridge.
2. The government builds the bridge if the social benefit of the bridge exceeds the cost of the bridge and does not build it otherwise.
3. The government uses its power of taxation to raise revenues to build the bridge.

This is a rule that determines which public goods should be provided. It is a rule that leads to the efficient provision of public goods: those with a large enough social benefit are provided, those with a lower benefit are not. More precisely, by following this rule, it is possible for the government to make sure that public goods are provided only when their provision makes everybody better off. Whenever the rule justifies the building of the bridge, the government can tax each individual an amount that is less than that person's individual valuation of the bridge and still raise enough money to finance the building of it.

There are two major problems with this scenario. First, we supposed the government knew everybody's valuation. Obviously, this will not be true in practice. The government could ask people to provide their valuations, but the problem here is that people have no incentive to tell the truth. In particular, if people thought that the amount they would be taxed was related to their valuation, then they would have an incentive to understate their valuations. Second, even if people truthfully revealed their valuations, the tax scheme might be perceived as very unfair because different people would be taxed different amounts.

To see in more detail how this incentive problem arises, imagine a different rule. Individuals in the community individually decide how much to contribute to the construction of the bridge. [3] The bridge is built if the sum of everyone's contributions exceeds the cost. Because the bridge is a public good, each resident enjoys the benefit of the total bridge, not only the segment built by their individual contribution. You benefit from the contributions of others, and they benefit from your contribution. If everyone independently decided how much to contribute to the construction of the road, they would be unlikely to contribute at all. If others are not contributing, then there is no reason to contribute because the bridge will not be built anyway. And if others are contributing enough to finance the bridge, then you can benefit without having to pay. [4] We expect that the contributions will, from a social perspective, be small or zero because each member of the community ignores the benefit of his or her contribution to others.

More generally, there will be under provision of public goods because individuals do not take into account the effects of their contributions on others' well-being. This is sometimes called a "free-rider problem." The term comes from the fact that if everyone else pays for the good, you can travel for free.

KEY TAKEAWAYS

- The demand for cars is an example of unit demand.
- A car is an asset because it is a durable good that can be resold.
- Gasoline and insurance are two important complementary products to cars. Infrastructure, such as roads and bridges, is also a complementary product to cars. Because such infrastructure often has the characteristics of a public good, the government often provides it.

CHECKING YOUR UNDERSTANDING

1. Suppose there is a temporary shortfall in refining capacity, so the supply of gasoline decreases for a three-month period. Use a supply-and-demand diagram to show what happens to the price of gasoline and to the quantity traded. What implications will this have for (a) how much people drive and (b) the demand for automobiles?
2. If everyone is required by law to have car insurance, what will happen to the demand for cars?

[1] One such program can be found at the TraffiCare International home page, accessed March 14, 2011, http://trafficare.net.

[2] The cartoon comes from Peter Hugill, "Good Roads and the Automobile in the United States 1880–1929," *Geographical Review*, July 1982, 327–49.

[3] Marco Haan and Peter Kooreman, "Free Riding and the Provision of Candy Bars," *Journal of Public Economics*, February 2002, 277–92.

[4] To be more precise, you would contribute only in the unlikely event that the amount you are willing to pay would make the difference between the bridge being built or not.

16.2 Supply of Cars

LEARNING OBJECTIVES

1. What factors determine the price of a car?
2. How do car manufacturers compete beyond their choice of price?
3. What factors influence the choices that automobile producers make about the location of their production?

If you walk around the streets of your town, you could conduct a survey of the cars you see. For each car, you could make your best guess as to the answers to the following questions:

- What was the sticker price of the car when it was first sold?
- How old is the car?
- What is the car's estimated value now?
- What are the car's most important features?
- Which company produced the car?
- Is the car manufacturer a US company or a foreign company?
- In which country was the car assembled?
- At which manufacturing plant was the car assembled?

We start with the price of a car. We then look at other aspects of the production decision, such as the key attributes of the car and the choice of production location.

Car Prices

The basic rule for pricing is as follows: set the price so that

marginal cost = marginal revenue.

This rule was explained and developed in Chapter 6 "Where Do Prices Come From?". **Marginal cost** is the extra cost incurred by producing an additional unit, and **marginal revenue** is the extra revenue earned by producing an additional unit. To understand how this rule applies to cars, we need to look more carefully at both the costs of production and the demand for cars.

Marginal Cost

Cars are produced in automobile assembly plants using a variety of inputs, such as steel, rubber, glass, and labor. Lying behind the assembly of the car is an organization that engineered the car and designed the production process. At one level, there is nothing special about the cost structure for car production. We can decompose costs into three components: **entry costs**, **fixed operating costs**, and **variable costs**. We explained these notions of cost inChapter 8 "Growing Jobs".

- Entry costs are incurred prior to production. For cars, the most significant entry costs are design costs and the costs of establishing the production line. Many of these expenses are incurred in the research and development stage of the product. Once the car has been designed, the production line must be organized to manufacture the car efficiently. Finally, specialized machinery must be ordered or custom built for the production line. All of these expenses are incurred before a single car can be produced.
- Fixed operating costs include the costs of managing the automobile plant. Think of these as the costs of the various divisions of the plant not directly engaged in production: the operations department, the human resources department, and so on. These costs are the same no matter how many automobiles are being produced.
- Variable costs depend on the number of cars being produced. To build more cars, the firm must hire more labor. If the firm wants to produce fewer cars, it needs to buy less steel. These and other variable costs fluctuate according to the number of cars rolling off the production line.

By definition, entry costs and fixed operating costs are the same no matter how many cars are produced. The only costs that matter for the pricing decision are the firm's variable costs. Managers in auto plants must do their best to determine how much these variable costs change when they produce one extra vehicle. In other words, they need an estimate of the marginal cost of production.

Toolkit: Section 17.14 "Costs of Production", and Section 17.15 "Pricing with Market Power"

You can review different cost definitions and the definition of marginal revenue in the toolkit.

The history of automobile manufacture reveals that costs of production change over time. Technological progress is visible as we compare production processes at different dates. Ford's move to mass production was key to its success in the early 1900s because this new production method reduced costs substantially. Meanwhile, modern, highly automated, capital-intensive production facilities make those Ford production techniques seem primitive.

Even today, however, the labor input into the production process differs across producers. A recent report compared the labor hours required to produce a car at different manufacturing facilities. [1] For 2006, a Nissan plant in Smyrna, Tennessee, required 28.32 labor hours to produce one vehicle. A Toyota plant was next at 29.54 hours. In contrast, a General Motors (GM) car required 44.59 hours of labor input. Thus GM is using a much more labor-intensive method of production than Toyota or Nissan, whose facilities are more automated. These are not exactly measures of marginal cost because they measure average labor hours rather than the labor hours required to produce one extra car. Still, it is very likely that the GM plant has a higher marginal cost than the Nissan plant.

Marginal Revenue

We can take marginal revenue = marginal cost and rewrite it as a markup pricing formula:

price = (1 + markup) × marginal cost.

For example, if the marginal cost of producing the last unit is $30,000 and the markup is 0.50 (50 percent), then the firm sets a price of $45,000. For a given value of marginal cost, a higher markup translates into a higher price. And for a given markup, higher marginal cost translates into a higher price. The **markup** depends on the **own-price elasticity of demand**.

$$\text{markup}=1-(\text{elasticity of demand})-1.$$

Suppose a firm has a lot of market power. This means it can increase its price with relatively small changes in the quantity demanded: that is, demand is inelastic so −(elasticity of demand) is small. In this case, a firm will choose a large markup. If demand is more elastic, a firm will choose a smaller markup.

Toolkit: Section 17.2 "Elasticity", and Section 17.15 "Pricing with Market Power"

You can review the definition and measurement of own-price elasticity of demand and markup pricing in the toolkit.

The markup pricing equations seem easy to implement, at least in principle. For an automobile producer, pricing is actually quite complex. There are several reasons for this:

- With something as complicated as an automobile, the calculation of marginal cost is not straightforward. The operations department must determine how much additional labor and raw materials are needed to produce exactly one more car.
- Extensive market research and a certain amount of careful experimentation may be required to find the elasticity of demand. Remember that we saw that a household's decision about whether to buy a car depends on many different factors, including income, interest rates, the current and expected price of gasoline, the price of public transportation, and so on.
- The price elasticity of demand depends on the decisions of other producers. We discussed this idea in Chapter 14 "Busting Up Monopolies". When there is a relatively small number of suppliers, firms have to keep a close eye on the strategies of their competitors. For example, if other producers increase the prices of their cars, then you can expect households to substitute toward purchasing your car. The demand curve for your car will shift to the right. This is good news. But to determine whether to change your own price in response, you have to determine the elasticity of the demand curve for your product, given the new prices set by the other producers. You also have to worry about whether changes in your price will in turn lead other producers to change their prices again.
- Automobile producers manufacture many different models of vehicles. In effect, they compete with themselves as well as with other producers. If Ford cuts the price of, say, a Thunderbird, then the demand curve for other Ford vehicles will shift to the left as at least some potential buyers now choose the Thunderbird ahead of other Ford products. Ford must take into account how its various pricing decisions interact.
- As we explained earlier, an automobile can be thought of as a "bundle of attributes," such as performance, style, color, and so on. The valuation that a potential buyer places on a car depends on the buyer's valuations of these various attributes. Thus when manufacturers want to assess how much a car is worth to potential buyers, they really need to determine how much each attribute might be worth. Sophisticated statistical techniques are used to develop these numbers,

and this information is used in both the pricing of vehicles and the decisions about which attributes to include in new models, which to exclude, and which to have as available options.

Pricing is only one of many decisions made by car producers. They make other key choices as well. Two of the most significant are design changes when they introduce new models and the decision about where to locate their production facilities. We turn to these next.

Model Introductions

A century has passed since Henry Ford introduced one of the most famous automobiles ever: the Ford Model T. [2] This car remained in production for almost two decades, with 15 million automobiles produced. There were two versions of the Model T: a car and a truck. Otherwise, there were very few changes made to the vehicle design throughout its years of production. Famously, Henry Ford is claimed to have said, "You can paint it any color, so long as it's black."

In July 2008, 59 different vehicles were listed on the Ford website, including an entire family of brands: Ford, Land Rover, Lincoln, Mercury, Mazda, and Jaguar. In other words, Ford produced an immense variety of vehicles—available in more than one color. The same is true of other automobile producers. And, of course, such product variety means more than just a large number of models: any particular model may be available with all sorts of different styling, performance, and features. Interestingly, a visit to the Ford website in 2011 yields a different picture. There are Ford vehicles available, but the other brands are gone. Both Jaguar and Land Rover were sold by Ford in 2008, partly in response to the financial crisis. Over time, companies decide both to introduce and to remove models from their range of offerings.

Cars are not the only products that display such diversity. You can buy many different kinds of laptop computers, breakfast cereals, or mobile phones, for example. As economies grow and develop, we typically see an increasing variety of goods available. But product variety is particularly noticeable with cars because automobile producers come out with new models each year.

New model introductions began early in the history of the automobile. In the 1920s, Ford faced stiff competition from other producers, particularly GM. In the mid-1920s, under the leadership of company president Alfred Sloan, GM had adopted a strategy of introducing new models. [3] In part, the strategy came from recognizing that automobiles were durable goods that households kept for many years. The introduction of new models was a strategy to motivate the exchange of old for new cars. This strategy worked. Ford's sales of the Model T fell off and, at the end of the decade, Ford also adopted the strategy of model turnover.

The tactic remains in place today. Each year, car companies introduce new models. In some years, they make radical changes, while in other years new cars do not deviate much from previous models. The design and production of new models is one element of the competition among automobile producers. Although we often emphasize price competition, producers also compete in terms of the attributes of their models. Thus competition is very complex.

Plant Location

You have probably given little thought to why firms build factories in one location rather than another. But imagine for a moment that you must decide where to construct a new automobile plant. What kinds of factors might influence your decision?

You would certainly think about the cost of your inputs—that is, the items you need to manufacture new vehicles. Cars require substantial amounts of raw materials, such as steel, that have to be brought to your factory. If those inputs have to be brought in from a long way away, then your inputs will be more expensive. These costs depend also on the local infrastructure: are there good road and rail links to your prospective site? Another input, of course, is labor. Ideally, you want to locate your factory where labor is cheap but also sufficiently skilled for the positions you need to fill.

Once you have manufactured the cars, you have to get them to their final destinations: dealers throughout the country or even throughout the world. Because cars are large and heavy, they are expensive to ship to other locations. Thus, other things being equal, you would also like to locate your manufacturing site near your final demand. Of course, producers must usually serve many markets from a single plant.

Where you ultimately choose to locate the plant will depend on the costs of transporting both inputs and output. If your inputs are very costly to transport, then you will produce near the source of inputs and ship your finished goods to your markets. Alternatively, if your inputs are easy to transport but your output is costly to ship, then you might locate your production near some of your markets. You might even consider multiple production plants to lower the costs of transporting the final good.

You also care about local policies, such as the level of taxes. Countries, states, regions, and cities often compete to attract factories. They do so because a factory brings with it jobs and greater prosperity for a region. In some places in the world, you also have to worry about whether your property rights are well protected. If you set up a factory in the United States, you can be reasonably confident that the government will not try to confiscate either your capital or profits. In some other countries, however, you may justifiably be concerned for the safety of your assets.

The automobile industry in the United States was initially located in and around Detroit. This was partly due to the fact that access to the Great Lakes provided low-cost transportation of the necessary inputs into the production process. As time passed, plants began to appear outside the Detroit area, particularly in the southern part of the United States.

One of the factors motivating these location decisions was labor costs. The automobile plants in and around Detroit were dominated by a union, the United Auto Workers (http://www.uaw.org/node/39), which was formed near the end of the Great Depression. In the short run, firms must negotiate with the unions that represent its workers. In the longer run, though, firms have other options. One of them is to locate plants in areas with cheaper labor costs. Over time, firms have indeed shifted some of their production facilities to other parts of the United States and other countries around the world where labor is cheaper.

KEY TAKEAWAYS

- The price of a car is a markup over the marginal cost of production. The markup depends on the elasticity of demand.
- Car producers compete by introducing new models.
- Plant location choices are made in an effort to reduce the costs of production as well as the costs of transporting intermediate goods to a plant and finished goods to the market.

CHECKING YOUR UNDERSTANDING

1. How might you explain the differences in labor input per car across automobile assembly plants?
2. Under what conditions would a car producer locate a production plant in Alaska?
3. What other goods have new models introduced into the market? Does this happen every year? Why do the producers of these goods change models?

[1] See Gary S. Vasilash, “Assembly Plants: How They Compare,” *Automotive Manufacturing & Production*, August 1997, accessed March 14, 2011,http://findarticles.com/p/articles/mi_m0FWH/is_n8_v109/ai_20855370/pg_2.

[2] For details on the history of the Model T, see The Henry Ford Museum website, “The Model T,” accessed March 2011,http://www.thehenryford.org/exhibits/showroom/1908/model.t.html.

[3] This discussion draws on the history of automobiles at David Gartman, “Tough Guys and Pretty Boys The Cultural Antagonisms of Engineering and Aesthetics in Automotive History,” Automobile in American Life and Society, accessed March 14, 2011,http://www.autolife.umd.umich.edu/Design/Gartman/D_Casestudy/D_Casestudy3.htm.

16.3 Market Outcomes in the Automobile Industry

LEARNING OBJECTIVES

1. What kind of competition is there in the automobile industry?
2. How do market outcomes differ in the short run compared to the long run?

The interactions among buyers and sellers in the car market ultimately lead to prices and quantities of all the different cars that are produced. But what is the right way to think about that interaction? Automobile markets are not examples of competitive markets—many firms each producing an identical product. Nor is there a single car producer acting as a monopolist. To study markets such as the car market, we have an intermediate situation where firms

- sell goods that are imperfect substitutes for other goods in the markets in the short run, and
- enter and exit in response to profit opportunities in the long run.

Competition among Producers in the Short Run

When we think about market outcomes for automobiles, there are two different markets to consider. There are business-to-business markets in which manufacturing firms sell cars to dealerships, and there is the business-to-consumer market in which dealerships sell cars to the final consumer. This pattern of trade is quite normal: most firms do not sell directly to the final consumer but instead sell their goods through retailers.

So far we have said that automobile producers determine prices for their cars. But the companies do not actually set the price you will ultimately pay for a new vehicle. That price is determined through a bargain between you and a dealer. The price that the company sets is the price at which it sells to the dealer. Given the numerous dealers, you would not expect them to be able to make much profit. Competition will force the price to be close to the cost of the car to the dealer. But the producer retains market power and can dictate a price for selling the car to the dealer.

This might make it tempting to think about the final market for cars as being roughly competitive. After all, one of the conditions for a competitive market is that there should be a large number of buyers and sellers. Another condition, though, is that sellers should be selling identical goods. In the case of cars, this is evidently not the case. We have already pointed out that firms produce many different models of vehicles with various options available. On top of that, dealerships may differ in terms of the quality of service they offer both before and after the sale of a vehicle. Consumers, when choosing which car to buy and where to buy it, are choosing from a large set of different, imperfectly substitutable products. We call these differentiated products.

Each dealer therefore possesses a degree of market power. Some of this market power comes from the fact that there will be only a small number of sellers of a particular model in a given region. Some of the dealer's market power stems from specific features of the dealership, such as location and after-sales service. The key point is that each dealership faces a downward-sloping demand curve for the cars that it sells. The seller chooses a point on the demand curve. Because there are competing cars available from other dealerships in the market, the position of the demand curve depends on the prices set by other firms for other models.

Although dealerships possess some market power, the retail market for automobiles is still quite competitive. Demand is relatively elastic because consumers have different dealerships and cars to choose from. In addition, information about the price at which dealers obtain vehicles from manufacturers is readily available. Under most circumstances, therefore, dealers are able to enjoy only a small markup over this price. (The exception is when a particular model of vehicle is in particularly high demand for some reason.)

From a dealer's perspective, marginal cost is determined largely by the price at which it obtains the car from the manufacturer. The producer sets the price to the dealer to maximize its own profit. Producers understand that the demand for their products is affected by the prices of competing vehicles. This strategic interaction means that the elasticity of demand (and hence the markup) for a particular car depends on the prices set by other manufacturers. We explained this in detail in Chapter 14 "Busting Up Monopolies". Likewise, dealers set their prices based in part on the prices at other dealerships. On the demand side, households take the set of products offered in the market and their prices (subject to a little bargaining with dealers) as given as well. Their decisions about which cars to purchase and when to purchase them generate the market demand curves faced by dealerships.

Market Dynamics in the Long Run

So far we have taken as given the types of cars produced, the location of plants, and the identity of the automobile producers. Over a short period of time, such as a year, this is a good way to think about the market for cars. But over longer periods of time, the market is much more dynamic. There are changes in the models of vehicles; there are changes in the location of manufacturing plants; and there is entry and exit of manufacturers. One way to see this is to look at the evolution of the automobile market since the early part of the 20th century.

The beautiful car shown in this picture is called a Marmon. [1] The photo is of a 1932 model. A Marmon won the first Indianapolis 500, and nearly 22,000 models were sold in 1929. But by 1934, the company was gone, a casualty of the Great Depression. Small fringe producers like Marmon disappeared from the automobile industry. Left behind were the large producers who were to dominate the US automobile industry from that time onward. By the mid-1930s, the US market was largely ruled by three manufacturers.

Economists Tim Bresnahan and Daniel Raff looked at data on automobile plants during this time period. [2] They found that the number of plants (remember that one firm may have multiple plants) that were producing cars fell from 211 in 1929 to 121 in 1935. There is no single explanation of exactly why these producers failed and had to close their plants. The Great Depression evidently led to a large decrease in the demand for automobiles. But on top of that, surviving firms were marked by advances in product and process development. In the early stages of the automobile industry, small producers operated at a small scale. Such producers simply could not compete with Ford's lower-cost production process. This competition from Ford led to the exit of producers of cars like the Marmon. In the end, the industry was left with a small number of powerful firms.

In this market, firms were selling differentiated products, so they had market power. Over the long run, there was entry and exit of competing products (that is, firms introduced new products and retired old ones). There was also entry and exit of entire firms. The conditions governing entry and exit are the same as those that we explained in Chapter 8 "Growing Jobs". A firm will introduce a new product if it expects to make sufficient profits (in terms of discounted present value) to justify the fixed entry costs. A firm will discontinue a product if the discounted present value of profits that it expects from that product is less than the value of the firm's recoverable assets. Similar conditions apply to entire firms in the market.

Over the past 70 or so years, after the shakeout in the 1920s and 1930s, the big three producers have remained the dominant sellers. From that perspective, you might think that there was little entry and exit. However, the story is more complicated. First, the market share of the three main producers declined due to foreign competition. American consumers started buying cars made in Europe, Japan, and elsewhere. Second, the products produced by the firms have evolved considerably over time. This is a very dynamic market in terms of product innovation. Although there may not have been very much entry and exit of firms, there was considerable entry and exit of products. Sometimes, manufacturers retire entire brands, such as the Hummers that General Motors (GM) stopped producing in 2010.

The Used Car Market

When households choose a car, one option is not to purchase a new car but instead to buy a (as the dealers like to put it) "preowned" vehicle. From the perspective of the buyer, there is one critical difference between a new car and a used car. With a new car, it is relatively easy to make a reasonably good judgment about the attributes of a product, partly from reviews in magazines and on the Internet. With a used car, it is much harder to judge the quality of the product and thus place an accurate valuation on it. We explained a similar problem in terms of health care in Chapter 15 "A Healthy Economy".

With new cars, you bear only a small risk that the car will not perform properly when you buy them. This is not the case in the market for used cars. Imagine (or perhaps you have actually experienced this) going to a used car lot to look for a car. Here is what you might hear from a member of the sales force: "This is the best used car I have ever seen. No lie—it was purchased new by an elderly woman a few years back, and she treated it like one of her kids. It is only here on our lot because she has decided to stop driving. At this price, it is a steal." You are much less likely to hear this: "Yeah, that car is a lemon. Some guy bought it from the dealer a few months back, and it never was right. One problem after another; it was back in the shop every week. Sure there is low mileage, but my guess is that there are no more miles from that car anyway. Go ahead, buy it if you like. But don't say I didn't warn you."

When you see a used car for sale, ask yourself: why is that car here? The true answer could be one of these two stories. If it is the first situation, then the car is probably a good buy. But if it is the second, then you could be getting ripped off. And the problem is that the seller may give you the first story even when the second is the truth.

The fundamental difficulty here is that you and the seller have very different information. The seller of the product knows its quality (is the car good or bad?) while you, as a buyer, do not know its quality. This does not mean you should never buy a used car. But it does mean that your willingness to pay for a used car should reflect the uncertainty you face with regard to the quality of that car. Because all buyers face the same problem, the end result is that the market valuation of used cars will be low. Accordingly, the price of a used car is lower than it might be if the quality of cars was known. And this can also mean that there are fewer good used cars on the market. This is the problem that economists call adverse selection.

You can also perhaps spare some sympathy for the used car dealer as well. We have described this problem from the perspective of a buyer. Even if a dealer really does have a car that is of high quality, it is hard for him to convince prospective buyers of that fact. If you want to sell a car you own, you will probably encounter this problem: you may know that your car is high quality, but you cannot convince buyers.

KEY TAKEAWAYS

- Car companies compete in markets where they sell differentiated products.
- In the long run, the entry of competitors (in the form of either new firms or new products) continues until profits are equal to zero.

CHECKING YOUR UNDERSTANDING

1. Use the condition that marginal revenue = marginal cost (consult the toolkit if needed) to explain the difference in the price of two cars of your choice.
2. Some used car sellers include a warranty with your purchase. Would that help overcome the lemons problem?
3. If two cars are close substitutes, what do you predict about their prices?

[1] This discussion draws a history of the Marmon that can be found at Bill Vance, “Motoring Memories: Marmon,” accessed March 14, 2011,http://www.canadiandriver.com/2000/03/16/motoring-memories-marmon.htm.

[2] See Timothy Bresnahan and Daniel Raff, “Intra-industry Heterogeneity and the Great Depression: The American Automobile Industry, 1929-1935,” *The Journal of Economic History*, June 1991, 317–31.

16.4 Policy Issues

LEARNING OBJECTIVES

1. What are some policies to control the emissions from cars?
2. How do trade policy and international investment affect the car market?
3. What is the problem of congestion, and what can governments do about it?

There are many ways in which government policies impinge on automobiles. Here we highlight a few such issues.

Environmental and Resource Concerns

In Chapter 2 "Microeconomics in Action", we showed a photograph of smog in Mexico City. At the same time that cars have transformed the economic world, they have also transformed our natural environment. The exhaust from cars contributes to air pollution, which is hazardous to health. Car exhausts are a source of greenhouse gas emissions and thus contribute to climate change.

Pollution from cars is a classic example of an **externality**. (We discussed externalities in detail in Chapter 13 "Cleaning Up the Air and Using Up the Oil".) An individual's decision to purchase and drive a car does not take into account the effects on third parties. In this case, some of the affected third parties are those in the immediate vicinity who suffer from a reduction in air quality. To the extent that emissions contribute to climate change, however, the third parties potentially include everyone in the world.

Toolkit: Section 17.19 "Externalities and Public Goods"

You can review the definition and use of externalities in the toolkit.

Governments in the United States and elsewhere have enacted various policies that are motivated, at least in part, by the desire to take into account such environmental externalities and resource use. First, there are taxes on gasoline. These are relatively low in the United States but are much higher in Europe. Second, there are technological restrictions, such as the requirement that automobiles be fitted with catalytic converters and designed to run on unleaded fuel. In the United States, the government has taken action to improve the fuel consumption of cars produced within US borders. These are called Corporate Average Fuel Economy (CAFE) standards. [1] You can notice two things from this term: (1) the restrictions are in terms of fuel economy (miles per gallon), and (2) the restriction does not apply to individual automobiles but rather to the set of cars sold by a corporation. For example, the standard is applied to the entire set of models produced by General Motors (GM), not model by model, so GM makes some cars that are below the standard and others that are above. Corporations that do not meet the standard are fined.

The CAFE standard comes from legislation passed in 1975 in response to the embargo by oil-producing countries in 1973. The initial motivation was to reduce energy consumption and, in part, make the United States less dependent on imported oil. The arguments today for these standards also include the effect of car emissions on global warming.

Trade and Investment Policies

A second government policy that has had a huge impact on the automobile industry is the opening of the world economy to trade and international investment. The current automobile market is no longer just a US market. The United States is part of the world market. US producers interact with the rest of the world by

- selling cars in many countries,
- buying parts from suppliers throughout the world,
- producing in many countries,
- being financed by debt and equity held in foreign countries.

  Meanwhile, US citizens
- own cars produced in other countries and imported into the United States,
- consume imported oil,
- work for foreign companies that produce cars in the United States,
- work at car production facilities in other countries.

For example, let us look at Ford Motor Company. In 2007, Ford had 95 plants worldwide and employed about 246,000 people. The Ford operations in North America (United States, Canada, and Mexico) had 94,000 employees. In other words, 62 percent of the workforce was employed outside North America. There are Ford plants all over the world. Ford’s 2009 annual report tells us that Ford sold 4.82 million cars in 2009. Of these, 2.0 million were sold in North America, 1.6 million in Europe, and the remainder in South America and Asia. [2]

This international structure permits diversification. Ford produces and sells cars in China, South America, and elsewhere around the world. If you browse Ford's global activities, [3] you will get a sense of its worldwide sales and production operations. [4]

Each producer of cars has its own story of expansion across international borders, both through trade and through production. Honda began operations in the United States by creating a motorcycle sales division in the late 1950s. This eventually led to the production of motorcycles in the United States in 1978 and ultimately the production and sales of Honda cars in the United States.[5]

None of this would be possible without governments permitting the movement of goods and capital.

The first trade policy action directly impacting car production was the Canadian-US Automotive Products Trade Agreement of 1965. The goal of this agreement was to create an integrated market for cars between the United States and Canada by eliminating tariffs. Concerns that US companies would sell but not produce cars in Canada were met by some restrictions on production, including requirements that cars built in Canada had to have a certain domestic content.

The second trade policy action was called the North American Free Trade Agreement (NAFTA). [6] NAFTA was a controversial trade agreement. One of the big issues was whether the reduction in trade barriers would lead to job destruction in the United States. (We discussed this in Chapter 8 "Growing Jobs".) A 2001 study looking back at the effects of NAFTA directly on the production of cars did not find large effects at all.

> *Most fears about the ill effects of NAFTA on the U.S. auto industry, whether in term of employment, wages, or investment, have been proven wrong. The U.S. auto industry did experience rationalization of production and hence job displacements. But overall, NAFTA appears to have helped the U.S. auto sector (U.S. Trade Representative, 1997). Employment in the American automotive industry grew by 14.1 percent overall, with an increase of 16.1 percent in the auto parts sector and 10.1 percent in the motor vehicle assembly sector from 1994 to 1996.*

> *Hourly earnings for production workers in the U.S. automotive sector grew by 5.6 percent between 1993 and 1996. The Big Three U.S. automobile manufacturers invested $39.1 billion from 1993 to 1996 in new manufacturing plants and equipment in the United States, while investing only $3 billion in Mexico over the same period.* [7]

These statistics, of course, refer to what actually happened in the auto sector over this time period. What *would* have happened had NAFTA not been implemented requires a more sophisticated analysis.

Congestion

If you travel to Mexico City or Manchester, Beijing or Buenos Aires, Jakarta or Johannesburg, Los Angeles or Lagos, you will see that these cities all have something in common: traffic jams. Such road congestion is another example of an externality. The decision of one person to drive has an effect on other drivers.

One way of solving externality problems is to create new markets. In most cases, there is no market for the use of roads. However, if we charge people to use roads, then market incentives come into play. Toll roads are an example of the introduction of a market mechanism to combat congestion problems.

Congestion fees and tolls are in use in some cities around the world, such as London and Singapore. [8] The system in London, started in February 2003, charges drivers for entering the central city area between certain hours. Details of the system are available from Transport for London. [9] The cost in May 2008 was £8 (about $15.60) for access to the charging zone in Central London between 7 a.m. and 6 p.m., Monday through Friday. The system is enforced by a series of cameras that record license numbers and then check them against a record of who has paid for access to the zone. According to the Transport for London, the traffic flow into the zone has been reduced by 21 percent, and there is now less pollution and more cycling in the area.

The Electronic Road Pricing (ERP) system in Singapore, although older, is much more sophisticated. It was introduced in April 1998 along Singapore's expressways and in the city's central business district. All vehicles contain a transponder, mounted on the windscreen, into which the driver inserts a prepaid cash card. There are gantries located at various points around the city, and whenever a car passes under a gantry, a toll is automatically deducted. The rates differ for different categories of vehicle: motorcycles and light goods vehicles pay less than cars; heavy goods vehicles pay more than cars.

The most striking feature of the Singapore system is that the charges vary by time of day. Charges are imposed only at the peak hours, and the charges vary within those hours. Thus, for example, a driver passing a typical gantry might pay SGD 0.80 (about $0.58) from 08:00 to 08:05, SGD 1.50 from 8:05 to 8:30, SGD 2.00 from 08:30 to 09:00, SGD 1.50 from 09:00 to 09:25, SGD 1.00 from 09:25 to 09:30, and so on. You can see that these rates are quite finely tuned, with some rates being in effect for only a five-minute period.

The rates just quoted were in effect in mid-2008. By now, they may be quite different because a second feature of the system is that these rates are revised frequently. The Singapore Land Transport Authority has targets for the desired average speed of traffic on Singaporean roads: the target speed for expressways is 45–65 kilometers per hour (28–41 miles per hour), and the target speed for arterial roads is 20–30 kilometers per hour (13–19 miles per hour). Thus if they observe that traffic is flowing below these speeds, they consider raising the rates; if traffic is flowing smoothly, they consider reducing rates. They also adjust rates on a seasonal basis—for example, ERP charges are lower during school vacations.

KEY TAKEAWAYS

- Gas taxes and actions to improve fuel efficiency of cars are policies that reduce pollution from cars.
- In the United States, households benefit from the importation of foreign-produced cars and also from the ability to work at automobile factories owned by foreign companies.
- The opening of the car market to imports creates some job displacements.
- In some countries, governments tax the use of roads when they are congested.

CHECKING YOUR UNDERSTANDING

1. Give two reasons why the government taxes gasoline.
2. Why might a government choose to limit car access to a city center? What policies are available to a government for doing that?

[1] CAFE standards are described in detail at the National Highway Traffic Safety Administration site, "CAFE Overview—Frequently Asked Questions," accessed March 14, 2011, http://www.nhtsa.gov/CARS/rules/CAFE/overview.htm.

[2] Ford Motor Company, "Annual Reports," accessed March 14, 2011,http://corporate.ford.com/microsites/annual-reports.

[3] Ford Motor Company, "About Ford: Global Vehicles Sites," accessed March 14, 2011,http://corporate.ford.com/about-ford/global-vehicles-sites.

[4] Ford Motor Company, "About Ford: Global Operations," accessed March 14, 2011,http://corporate.ford.com/about-ford/global-operations.

[5] Honda Worldwide, "Establishing American Honda Motor Co. (1959)," accessed March 14, 2011,http://world.honda.com/history/challenge/1959establishingamericanhonda/index.html.

[6] US Department of Agriculture, Foreign Agricultural Service, "North American Free Trade Agreement (NAFTA)," accessed March 14, 2011,http://www.fas.usda.gov/itp/Policy/NAFTA/nafta.asp.

[7] Mary E. Burfisher, Sherman Robinson, and Karen Thierfelder, "The Impact of NAFTA on the United States," *Journal of Economic Perspectives* 15 (Winter 2001): 125–144.

[8] This June 21, 2006, press release from the UK Commission for Integrated Transport provides some discussion of the London system and others around the world: "New study shows road pricing progress," accessed March 14, 2011,http://cfit.independent.gov.uk/pn/060621/index.htm.

[9] Transport for London, "Congestion Charging," accessed March 14, 2011,http://www.tfl.gov.uk/roadusers/congestioncharging/default.aspx.

16.5 End-of-Chapter Material

In Conclusion

When you study economics from a textbook such as this, you learn different economic theories. In this book, we have looked at the theory of the consumer, which helps explain how individuals make their choices about what goods to buy. We have looked at the theory of the firm, which explains how firms make decisions about which goods to produce and what price to sell them at.

Our goal in this book is to help you see that economics is not only a matter of graphs and definitions but also the study of the world around you. Economists see economic decisions and economic forces everywhere they look. This chapter gave you many examples linked to one particular and very familiar product. Yet we have only scratched the surface in terms of the ways in which we could apply economic analysis to cars and the car industry.

Perhaps you can think of other ways in which you could apply the things you learned from your study of economics to the market for automobiles. In any case, we hope that, now, every time you see a car, you will remember that you are also seeing economics in action.

Key Links

- GM: http://www.gm.com/corporate/about/history
- Rouge River Plant: http://www.hfmgv.org/rouge/historyofrouge.aspx
- History of automobiles from the Smithsonian Institution:http://www.si.edu/Encyclopedia_SI/nmah/autohist.htm
- Bureau of Labor Statistics, motor vehicle and parts manufacturing:http://www.bls.gov/oco/cg/cgs012.htm
- BMW plant in South Carolina: http://www.bmwusfactory.com
- Electronic Road Pricing in Singapore: http://www.lta.gov.sg/motoring_matters/index_motoring_erp.htm

EXERCISES

1. (Advanced) A car is an asset, like a house. How would you use the principles of asset valuation to ascertain the value of a car? (Hint: see Chapter 9 "Making and Losing Money on Wall Street".)
2. List the products and/or services that are substitutes for cars.
3. Because a household can delay the purchase of a new car when the price increases, what does this do to the price elasticity of demand for cars?
4. List the car characteristics and household characteristics that would increase a household's valuation of a car.
5. Would an increase in the price of gasoline decrease the demand for all cars or only some? What are the effects of an increase in gas prices on alternative forms of transportation?
6. In the coordination game for automobile insurance, is it desirable for the government to require that everyone buy automobile insurance?
7. There are three people in a community. The government is proposing to build a bridge in that community. The bridge costs $120 to construct. Everyone values the bridge at $60. Is it efficient to build the bridge? Suppose the government asks people what the bridge is worth to them and plans to tax them the amount that they say. If two people truthfully reveal to the government that the bridge is worth $60 to them, will the third person give his true valuation as well?
8. Using the equation linking the price of a car to markup and marginal cost, if the markup is 60 percent and marginal cost is $20,000, what is the price? Show with this example that a firm facing a more elastic demand curve will set a lower price.
9. Is a car dealership more valuable in a small city with little competition or in a big city with more competition but also more potential buyers?
10. Car dealers make profits from goods and services that are complementary to the cars they sell. List and discuss some of these complementary products.
11. Why do automobile companies offer warranties with new cars? Why do these warranties expire after a few years?
12. What factors determine the choice of an automobile producer to create a new model?
13. If you were a seller of a used car and you knew it was high quality, would you have an incentive to offer the buyer a warranty on the car? Would you have that same incentive if you knew the car was low

quality? What then should you as a buyer infer if you see a seller willing to offer a used car with a warranty?

14. One concern with opening markets in the United States is that domestic car producers are forced to compete with foreign producers. In some cases, this competition leads to job losses in the United States. Who gains from this competition?
15. (Advanced) If you were designing a policy of charging for road use, would you include or exempt taxis?
16. In what sense is buying a car like buying health care? What do car companies do to provide buyers with assurance over the functioning of the car? Why doesn't this happen in the health-care market?
17. If you look at the market for cars, high-income households are more likely to buy new cars, and lower income households are more likely to buy used cars. Why is that the case?
18. Following a recession, it is likely that the average age of cars will be higher than prior to the recession. Why?
19. Would you say that the market for higher education is a market with monopolistic competition or a monopoly? How does competition occur in that market?

Economics Detective

1. In addition to car production being shifted to the southern part of the United States to reduce labor costs, car production has also moved overseas. What information can you find on the wages of workers at automobile plants outside the United States?
2. What is the current penalty on companies that do not meet the CAFE standard?
3. In June 2009, the US government became a majority owner of GM. What prompted that action? Why did the US government not buy Ford Motor Company? What has happened to US government ownership of GM? What has happened to the sales and profitability of Ford and GM in recent years?
4. What products other than cars have a "model year"? Are new models introduced simultaneously or at different times of the year? What factors might determine the timing of new model introductions?

Spreadsheet Exercise

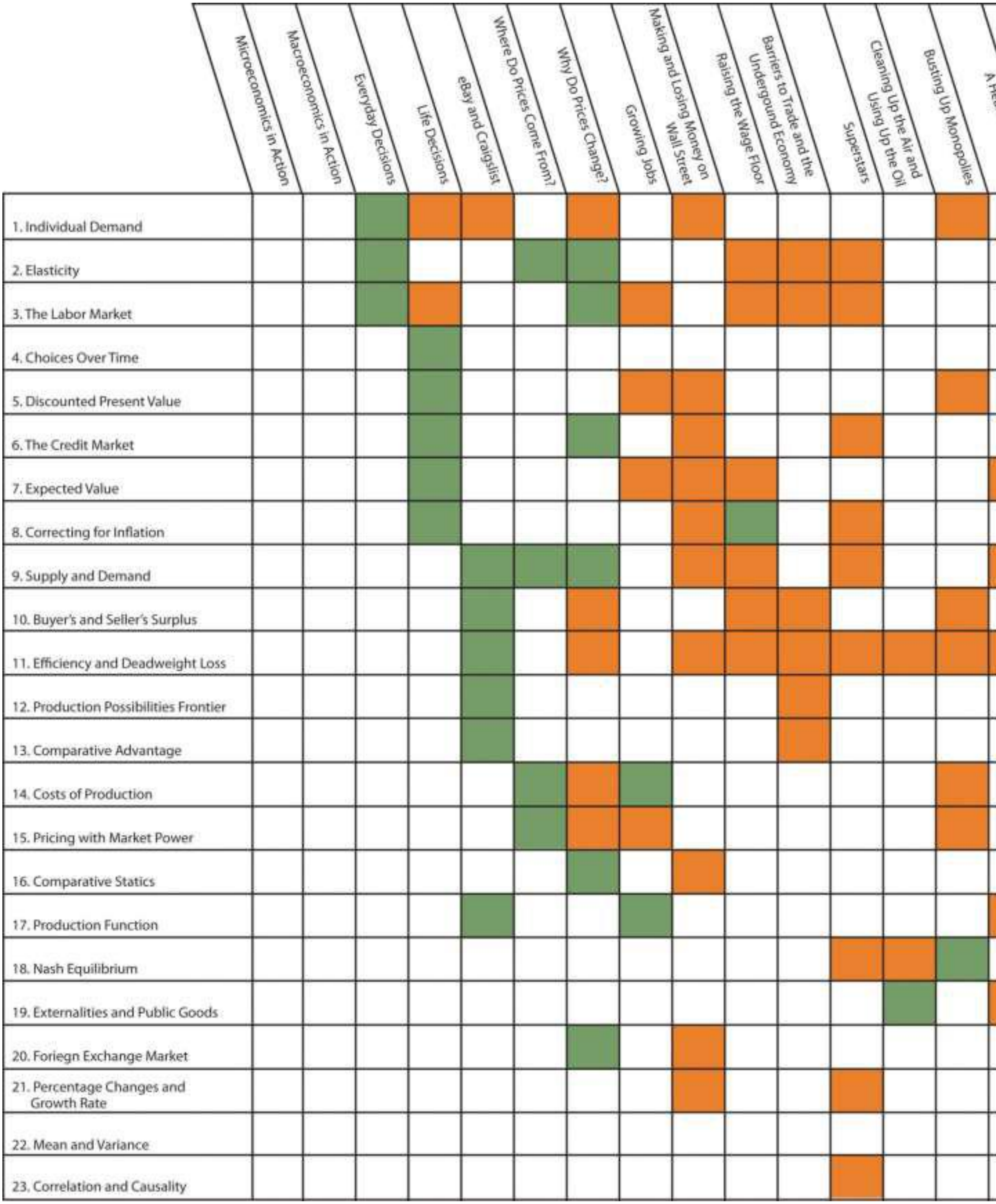
1. Based on the equations for markup pricing, create a spreadsheet to calculate the price you would set as a car producer. To do this, input the elasticity of demand and marginal cost. Use your program to calculate

the markup and product price. Produce graphs to show how the product price will change as the elasticity of demand changes.

Chapter 17 Microeconomics Toolkit

In this chapter we present the key tools used in the microeconomics part of this textbook. This toolkit serves two main functions:

1. Because these tools appear in multiple chapters, the toolkit serves as a reference. When using a tool in one chapter, you can refer back to the toolkit to find a more concise description of the tool as well as links to other parts of the book where the tool is used.
2. You can use the toolkit as a study guide. Once you have worked through the material in the chapters, you can review the tools using this toolkit.


Microeconomics in Action
Macroeconomics in Action
Everyday Decisions
Life Decisions
eBay and Craigslist
Where Do Prices Come From?
Why Do Prices Change?
Growing Jobs
Making and Losing Money on Wall Street
Raising the Wage Floor
Barriers to Trade and the Undergound Economy
Superstars
Cleaning Up the Air and Using Up the Oil
Busting Up Monopolies
1. Individual Demand
2. Elasticity
3. The Labor Market
4. Choices Over Time
5. Discounted Present Value
6. The Credit Market
7. Expected Value
8. Correcting for Inflation
9. Supply and Demand
10. Buyer's and Seller's Surplus
11. Efficiency and Deadweight Loss
12. Production Possibilities Frontier
13. Comparative Advantage
14. Costs of Production
15. Pricing with Market Power
16. Comparative Statics
17. Production Function
18. Nash Equilibrium
19. Externalities and Public Goods
20. Foriegn Exchange Market
21. Percentage Changes and Growth Rate
22. Mean and Variance
23. Correlation and Causality

17.1 Individual Demand

Individual demand refers to the demand for a good or a service by an individual (or a household). Individual demand comes from the interaction of an individual's desires with the quantities of goods and services that he or she is able to afford. By desires, we mean the likes and dislikes of an individual. We assume that the individual is able to compare two goods (or collections of goods) and say which is preferred. We assume two things: (1) an individual prefers more to less and (2) likes and dislikes are consistent.

An example is shown in part (a) of Figure 17.1 "Individual Demand". (This example is taken from Chapter 3 "Everyday Decisions".) In this example, there are two goods: music downloads ($1 each) and chocolate bars ($5 each). The individual has income of $100. The budget line is the combination of goods and services that this person can afford if he spends all of his income. In this example, it is the solid line connecting 100 downloads and 20 chocolate bars. The horizontal intercept is the number of chocolate bars the individual could buy if all income was spent on chocolate bars and is income divided by the price of a chocolate bar. The vertical intercept is the number of downloads the individual could buy if all income was spent on downloads and is income divided by the price of a download. The budget set is all the combinations of goods and services that the individual can afford, given the prices he faces and his available income. In the diagram, the budget set is the triangle defined by the budget line and the horizontal and vertical axes.

An individual's preferred point is the combination of downloads and chocolate bars that is the best among all of those that are affordable. Because an individual prefers more to less, all income will be spent. This means the preferred point lies on the budget line. The most that an individual would be willing to pay for a certain quantity of a good (say, five chocolate bars) is his valuation for that quantity. The marginal valuation is the most he would be willing to pay to obtain one *extra* unit of the good. The decision rule of the individual is to buy an amount of each good such that

marginal valuation = price.

The individual demand curve for chocolate bars is shown in part (b) of Figure 17.1 "Individual Demand". On the horizontal axis is the quantity of chocolate bars. On the vertical axis is the price. The demand curve is downward sloping: this is the law of demand. As the price of a chocolate bar increases, the individual substitutes away from chocolate bars to other goods. Thus the quantity demanded decreases as the price increases.

In some circumstances, the individual's choice is a zero-one decision: either purchase a single unit of the good or purchase nothing. The unit demand curve shows us the price at which a buyer is willing to buy the good. This price is the same as the buyer's valuation of the good. At any price above the buyer's valuation, the individual will not want to buy the good. At any price below the buyer's valuation, the individual wants to buy the good. If this price is exactly equal to the buyer's valuation, then the buyer is indifferent between purchasing the good or not.

In other words, the individual's decision rule is to purchase the good if the valuation of the good exceeds its price. This is consistent with the earlier condition because the marginal valuation of the first unit is the same as the valuation of that unit. The difference between the valuation and the price is the buyer surplus. (See Section 17.10 "Buyer Surplus and Seller Surplus" for more discussion.)

Key Insights

- The individual demand for a good or a service comes from the interactions of desires with the budget set.
- The individual purchases each good to the point where marginal valuation equals price.
- As the price of a good or a service increases, the quantity demanded will decrease.

More Formally

Let p_d be the price of a download, p_c the price of a chocolate bar, and I the income of an individual. Think of prices and income in terms of dollars. Then the budget set is the combinations of downloads (d) and chocolate bars (c) such that

$$I \geq p_d \times d + p_c \times c.$$

The budget line is the combinations of d and c such that

$$I = p_d \times d + p_c \times c.$$

In the graph, with downloads on the vertical axis, the equation for the budget line is

$$d=\frac{I}{p_d}-\frac{p_c}{p_d}\times c$$

You can use this equation to understand how changes in income and prices change the position of the budget line. You can also use this equation to find the vertical and horizontal intercepts of the budget line, along with the slope of $-(p_c/p_d)$.

The individual purchases downloads up to the point where

$MV_d = p_d$

(where *MV* represents marginal valuation) and purchases chocolate bars up to the point where

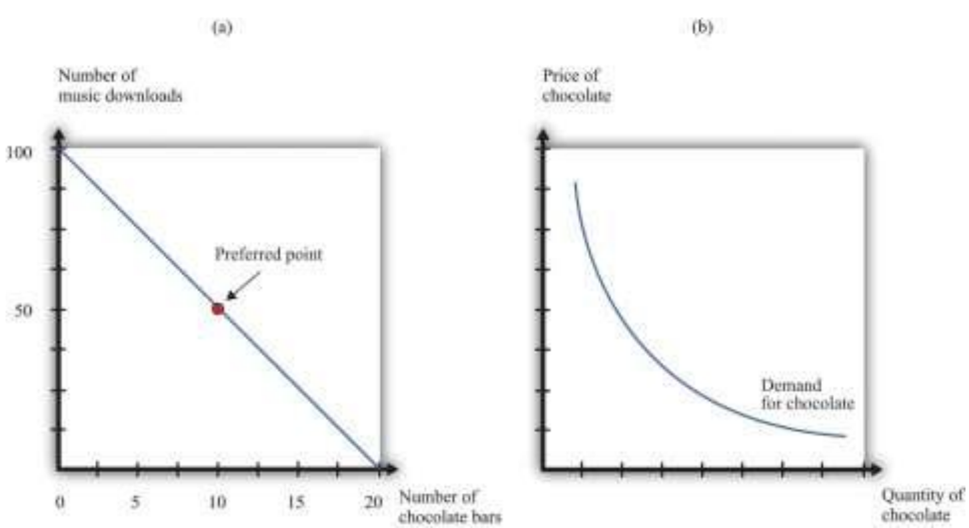
$MV_c = p_c$.

Combining these expressions we get

$$-\frac{MV_c}{MV_d}=\frac{p_c}{p_d},$$

which tells us that (minus) the ratio of marginal valuations equals the slope of the budget line. The ratio of marginal valuations is the rate at which an individual would like to trade one good for the other. The ratio of prices (the slope of the budget line) is the rate at which the market allows an individual to make these trades.

Figure 17.1 Individual Demand



The Main Uses of This Tool

- Chapter 3 "Everyday Decisions"
- Chapter 4 "Life Decisions"
- Chapter 5 "eBay and craigslist"
- Chapter 7 "Why Do Prices Change?"
- Chapter 9 "Making and Losing Money on Wall Street"
- Chapter 12 "Superstars"
- Chapter 13 "Cleaning Up the Air and Using Up the Oil"
- Chapter 14 "Busting Up Monopolies"
- Chapter 16 "Cars"

17.2 Elasticity

Elasticity measures the proportionate change in one variable relative to the change in another variable. Consider, for example, the response of the quantity demanded to a change in the price. The price elasticity of demand is the percentage change in the quantity demanded divided by the percentage change in the price:

$$\text{price elasticity of demand} = \frac{\text{percentage change in quantity}}{\text{percentage change in price}}.$$

When the price increases (the percentage change in the price is positive), the quantity decreases, meaning that the percentage change in the quantity is negative. In other words, the law of demand tells us that the elasticity of demand is a negative number. For this reason we often use −(elasticity of demand) because we know this will always be a positive number.

- If −(elasticity of demand) > 1, demand is relatively elastic.
- If −(elasticity of demand) < 1, demand is relatively inelastic.

We can use the idea of the elasticity of demand whether we are thinking about the demand curve faced by a firm or the market demand curve. The definition is the same in either case.

If we are analyzing the demand curve faced by a firm, then we sometimes refer to the elasticity of demand as the own-price elasticity of demand. It tells us how much the quantity demanded changes when the firm changes its price. If we are analyzing a market demand curve, then the price elasticity of demand tells us how the quantity demanded in the market changes when the price changes. Similarly, the price elasticity of supply tells us how the quantity supplied in a market changes when the price changes. The price elasticity of supply is generally positive because the supply curve slopes upward.

The income elasticity of demand is the percentage change in the quantity demanded divided by the percentage change in income. The income elasticity of demand for a good can be positive or negative.

- If the income elasticity of demand is negative, it is an inferior good.
- If the income elasticity of demand is positive, it is a normal good.
- If the income elasticity of demand is greater than one, it is a luxury good.

The cross-price elasticity of demand tells us how the quantity demanded of one good changes when the price of another good changes.

- If the cross-price elasticity of demand is positive, the goods are substitutes.
- If the cross-price elasticity of demand is negative, the goods are complements.

In general, we can use elasticity whenever we want to show how one variable responds to changes in another variable.

Key Insights

- Elasticity measures the responsiveness of one variable to changes in another variable.
- Elasticities are *unitless*: you can measure the underlying variables in any units (for example, dollars or thousands of dollars), and the elasticity will not change.
- Elasticity is not the same as slope. For example, the price elasticity of demand depends on both the slope of the demand curve *and* the place on the demand curve where you are measuring the elasticity.

The Main Uses of This Tool

- Chapter 3 "Everyday Decisions"
- Chapter 6 "Where Do Prices Come From?"
- Chapter 7 "Why Do Prices Change?"
- Chapter 10 "Raising the Wage Floor"
- Chapter 11 "Barriers to Trade and the Underground Economy"
- Chapter 12 "Superstars"
- Chapter 16 "Cars"

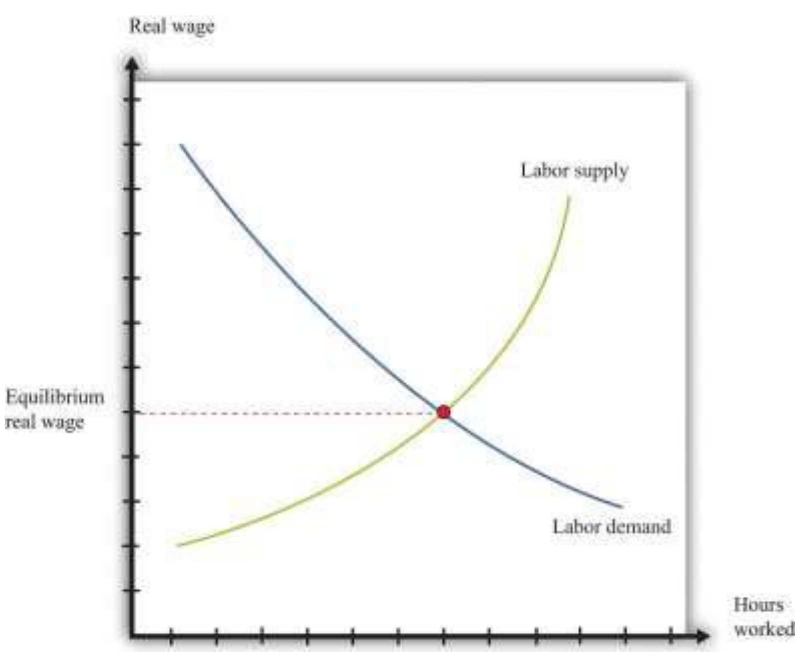
17.3 The Labor Market

The labor market is the market in which labor services are traded. *Individual labor supply* comes from the choices of individuals or households about how to allocate their time. As the real wage (the nominal wage divided by the price level) increases, households supply more hours to the market, and more households decide to participate in the labor market. For both of these reasons, the quantity of labor supplied increases. The labor supply curve of a household is shifted by changes in wealth. A wealthier household supplies less labor at a given real wage.

Labor demand comes from firms. As the real wage increases, the marginal cost of hiring more labor increases, so each firm demands fewer hours of labor input—that is, a firm's labor demand curve is downward sloping. The labor demand curve of a firm is shifted by changes in productivity. If labor becomes more productive, then the labor demand curve of a firm shifts to the right: the quantity of labor demanded is higher at a given real wage.

The labor market equilibrium is shown in Figure 17.2 "Labor Market Equilibrium". The real wage and the equilibrium quantity of labor traded are determined by the intersection of labor supply and labor demand. At the equilibrium real wage, the quantity of labor supplied equals the quantity of labor demanded.

Figure 17.2 *Labor Market Equilibrium*



Key Insights

- Labor supply and labor demand depend on the real wage.
- Labor supply is upward sloping: as the real wage increases, households supply more hours to the market.
- Labor demand is downward sloping: as the real wage increases, firms demand fewer hours of work.
- A market equilibrium is a real wage and a quantity of hours such that the quantity demanded equals the quantity supplied.

The Main Uses of This Tool

- Chapter 3 "Everyday Decisions"
- Chapter 4 "Life Decisions"
- Chapter 8 "Growing Jobs"
- Chapter 10 "Raising the Wage Floor"
- Chapter 12 "Superstars"

17.4 Choices over Time

Individuals make decisions that unfold over time. Chapter 4 "Life Decisions"gives several examples. Because individuals choose how to spend income earned over many periods on consumption goods over many periods, they sometimes wish to save or borrow rather than spend all of their income in every period.

Figure 17.3 "Choices over Time" shows examples of these choices over a two-year horizon. The individual earns income this year and next. The combinations of consumption that are affordable and that exhaust all of an individual's income are shown on the budget line, which in this case is called an intertemporal budget constraint. The slope of the budget line is equal to (1 + real interest rate), which, for convenience, we sometimes call the *real interest factor* (that is, real interest factor = 1 + real interest rate). The slope is the amount of consumption that can be obtained tomorrow by giving up a unit of consumption today.

The preferred point is indicated in the figure as well. This is the combination of consumption this year and consumption next year that the individual prefers to all the points on the budget line. The individual in part (a) of Figure 17.3 "Choices over Time" is consuming less this year than she is earning: she is saving. Next year she can use her savings and can consume more than her income. The individual in part (b) of Figure 17.3 "Choices over Time" is consuming more this year than he is earning: he is borrowing. Next year, his consumption will be less than his income because he must repay the amount borrowed in the first period.

When the real interest rate increases, individuals will borrow less and (usually) will save more (the effect of interest rate changes on saving is unclear as a matter of theory because income effects and substitution effects act in opposite directions). Thus individual loan supply slopes upward.

Of course, individuals live for many periods and make frequent decisions on consumption and saving. The lifetime budget constraint is obtained using the idea of discounted present value:

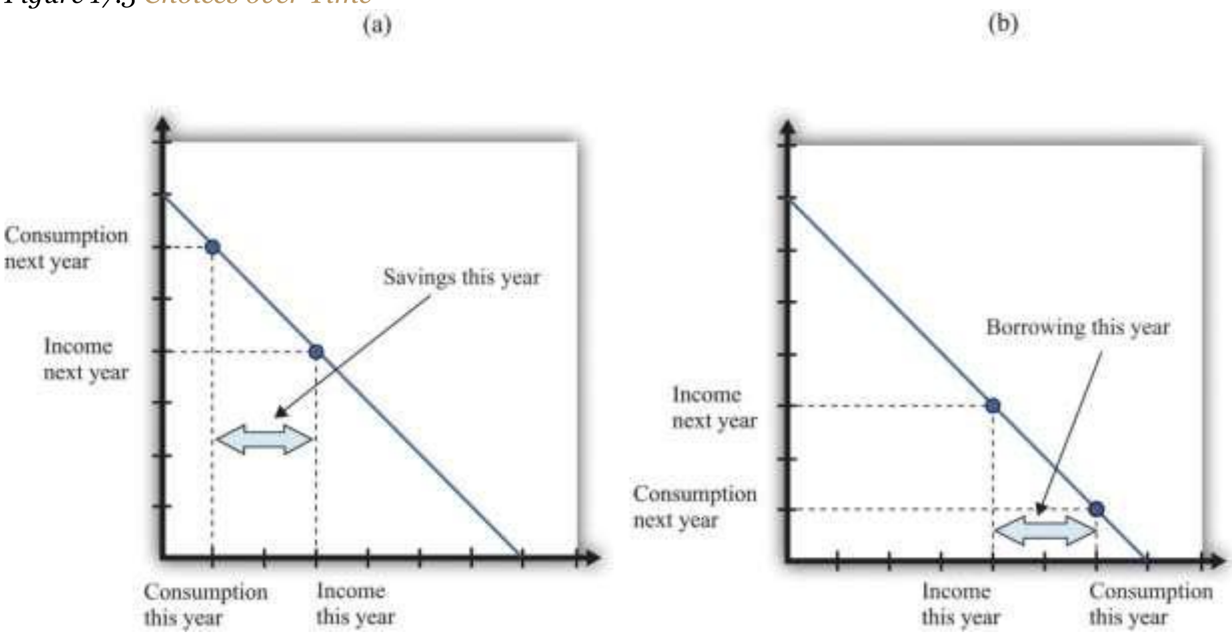
discounted present value of lifetime income = discounted present value of lifetime consumption.

The left side is a measure of all the disposable income the individual will receive over his lifetime (disposable means after taking into account taxes paid to the government and transfers received from the government). The right side calculates the value of consumption of all goods and services over an individual's lifetime.

Key Insights

- Over a lifetime, an individual's discounted present value of consumption will equal the discounted present value of income.
- Individuals can borrow or lend to obtain their preferred consumption bundle over their lifetimes.
- The price of borrowing is the real interest rate.

Figure 17.3 Choices over Time



The Main Uses of This Tool

- Chapter 4 "Life Decisions"

17.5 Discounted Present Value

Discounted present value is a technique used to add dollar amounts over time. We need this technique because a dollar today has a different value from a dollar in the future.

The discounted present value this year of $1.00 that you will receive next year is

$1normal interest factor=$11+nominal interest rate.

If the nominal interest rate is 10 percent, then the nominal interest factor is 1.1, so $1 next year is worth $1/1.1 = $0.91 this year. As the interest rate increases, the discounted present value decreases.

More generally, we can compute the value of an asset this year from this formula:

value of asset this year=flow benefit from asset this year+price of asset next yearnominal interest factor.

The flow benefit depends on the asset. For a bond, the flow benefit is a coupon payment. For a stock, the flow benefit is a dividend payment. For a fruit tree, the flow benefit is the yield of a crop.

If an asset (such as a bond) yields a payment next year of $10 and has a price next year of $90, then the "flow benefit from asset + price of the asset next year" is $100. The value of the asset this year is then $100nominal interest factor. If the nominal interest rate is 20 percent, then the value of the asset is $100/1.2 = 83.33.

We discount nominal flows using a nominal interest factor. We discount real flows (that is, flows already corrected for inflation) using a real interest factor, which is equal to (1 + real interest rate).

Key Insights

- If the interest rate is positive, then the discounted present value is less than the direct sum of flows.
- If the interest rate increases, the discounted present value will decrease.

More Formally

Denote the dividend on an asset in period t as D_t. Define R_t as the *cumulative*effect of interest rates up to period t. For example, $R_2 = (1 + r_1)(1 + r_2)$. Then the value of an asset that yields D_t dollars in year t is given by

$$\text{asset value}=D_1R_1+D_2R_2+D_3R_3+\ldots+D_TR_T.$$

If the interest rate is constant (equal to r), then the one period interest factor is$R = 1 + r$, and $R_t = Rt$.

The discounted present value tool is illustrated in Table 17.1 "Discounted Present Value with Different Interest Rates". The number of years (T) is set equal to 5. The table gives the value of the dividends in each year and computes the discounted present values for two different interest rates. For this example, the annual interest rates are constant over time.

Table 17.1 Discounted Present Value with Different Interest Rates

Year	**Dividend ($)**	**Discounted Present Value with *R* = 1.05 ($)**	**Discounted Present Value with *R* = 1.10 ($)**
1	100	100	100
2	100	95.24	90.91
3	90	81.63	74.38
4	120	103.66	90.16
5	400	329.08	273.20
Discounted Present Value		709.61	628.65

The Main Uses of This Tool

- Chapter 4 "Life Decisions"
- Chapter 7 "Why Do Prices Change?"
- Chapter 8 "Growing Jobs"
- Chapter 9 "Making and Losing Money on Wall Street"

- Chapter 14 "Busting Up Monopolies"
- Chapter 16 "Cars"

17.6 The Credit Market

An individual's choices over time determine how much he or she will borrow or lend. In particular, individual loan supply is upward sloping: when the real interest rate increases, a typical household will supply a greater quantity of funds to the credit market. Market loan supply is obtained by adding together the individual loan supplies of everyone in an economy. We use the terms "credit" and "loans" interchangably.

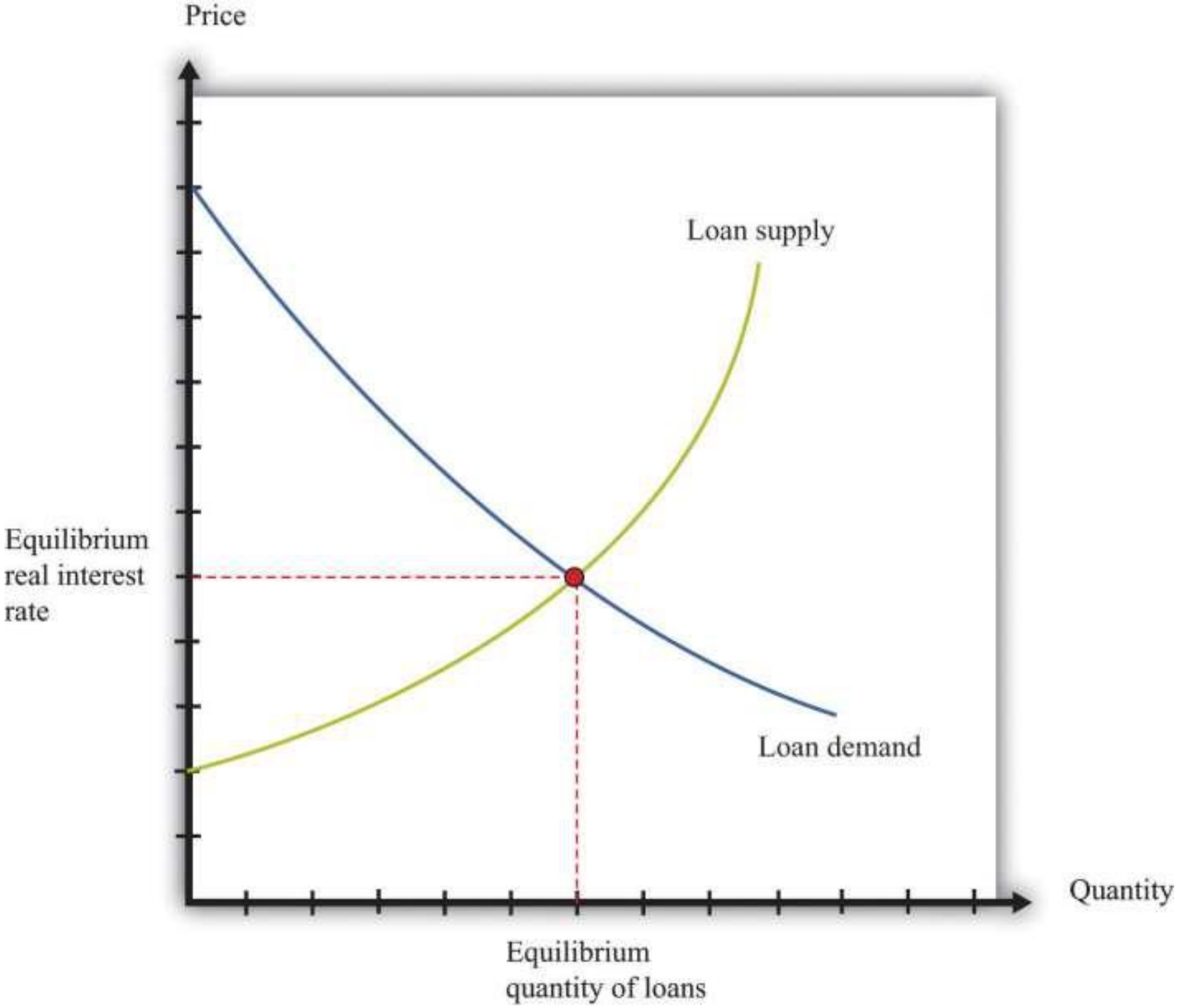
The demand for credit comes from households and firms that are borrowing. Market loan demand is obtained by adding together all the individual demands for loans. When real interest rates increase, borrowing is more expensive, so the quantity of loans demanded decreases. That is, loan demand obeys the law of demand.

Borrowers and lenders interact in the credit market (or loan market), which is illustrated in Figure 17.4 "The Credit Market (or Loan Market)". Credit market equilibrium occurs at the real interest rate where the quantity of loans supplied equals the quantity of loans demanded. At this equilibrium real interest rate, lenders lend as much as they wish, and borrowers can borrow as much as they wish. All gains from trade through loans are exhausted in equilibrium.

Key Insight

- As the real interest rate increases, more loans are supplied, and fewer loans are demanded.

Figure 17.4 *The Credit Market (or Loan Market)*



The Main Uses of This Tool

- Chapter 4 "Life Decisions"
- Chapter 7 "Why Do Prices Change?"
- Chapter 9 "Making and Losing Money on Wall Street"
- Chapter 12 "Superstars"

17.7 Expected Value

Probability is the percentage chance that something will occur. For example, there is a 50 percent chance that a tossed coin will come up heads. We say that the probability of getting the *outcome* "*heads*" is 1/2. There are five things you need to know about probability:

1. The list of possible outcomes must be complete.
2. The list of possible outcomes must not overlap.
3. If an outcome is certain to occur, it has probability 1.
4. If an outcome is certain not to occur, it has probability 0.
5. If we add together the probabilities for all the possible outcomes, the total must equal 1.

The expected value of a situation with financial risk is a measure of how much you would expect to win (or lose) on average if the situation were to be replayed a large number of times. You can calculate expected value as follows:

- For each outcome, multiply the probability of that outcome by the amount you will receive.
- Add together these amounts over all the possible outcomes.

For example, suppose you are offered the following proposal. Roll a six-sided die. If it comes up with 1 or 2, you get \$90. If it comes up 3, 4, 5, or 6, you get \$30. The expected value is

$(1/3) \times \$90 + (2/3) \times \$30 = \$50.$

Most people dislike risk. They prefer a fixed sum of money to a gamble that has the same expected value. *Risk aversion* is a measure of how much people want to avoid risk. In the example we just gave, most people would prefer a sure \$50 to the uncertain proposal with the expected value of \$50.

Suppose we present an individual with the following gamble:

- With 99 percent probability, you lose nothing.
- With 1 percent probability, you lose \$1,000.

The expected value of this gamble is −\$10. Now ask the individual how much she would pay to avoid this gamble. Someone who is risk-neutral would be willing to pay only \$10. Someone who is risk-averse would be willing to pay more than $10. The more risk-averse an individual, the more the person would be willing to pay.

The fact that risk-averse people will pay to shed risk is the basis of insurance. If people have different attitudes toward risky gambles, then the less risk-averse individual can provide insurance to the more risk-averse individual. There are gains from trade. Insurance is also based on diversification, which is the idea that people can share their risks so it is much less likely that any individual will face a large loss.

Key Insights

- Expected value is the sum of the probability of an event times the gain/loss if that event occurs.
- Risk-averse people will pay to avoid risk. This is the basis of insurance.

More Formally

Consider a gamble where there are three and only three possible outcomes (x, y, z) that occur with probabilities $Pr(x)$, $Pr(y)$, and $Pr(z)$. Think of these outcomes as the number of dollars you get in each case. First, we know that

$$Pr(x) + Pr(y) + Pr(z) = 1.$$

Second, the expected value of this gamble is

$$EV = (Pr(x)*x) + (Pr(y)*y) + (Pr(z)*z).$$

The Main Uses of This Tool

- Chapter 4 "Life Decisions"
- Chapter 8 "Growing Jobs"
- Chapter 9 "Making and Losing Money on Wall Street"
- Chapter 10 "Raising the Wage Floor"
- Chapter 15 "A Healthy Economy"

17.8 Correcting for Inflation

If you have some data expressed in nominal terms (for example, in dollars), and you want to convert them to real terms, you should use the following four steps.

1. Select your deflator. In most cases, the Consumer Price Index (CPI) is the best deflator to use. You can find data on the CPI (for the United States) at the Bureau of Labor Statistics website (http://www.bls.gov).
2. Select your base year. Find the value of the index in that base year.
3. For all years (including the base year), divide the value of the index in that year by the value in the base year. Notice that the value for the base year is 1.
4. For each year, divide the value in the nominal data series by the number you calculated in step 3. This gives you the value in "base year dollars."

Table 17.2 "Correcting Nominal Sales for Inflation" shows an example. We have data on the CPI for three years, as listed in the second column. The price index is created using the year 2000 as a base year, following steps 1–3. Sales measured in millions of dollars are given in the fourth column. To correct for inflation, we divide sales in each year by the value of the price index for that year. The results are shown in the fifth column. Because there was inflation each year (the price index is increasing over time), real sales do not increase as rapidly as nominal sales.

Table 17.2 Correcting Nominal Sales for Inflation

Year	CPI	Price Index ($2000 Mil.)	Sales (Millions)	Real Sales ($2,000 Mil.)
2000	172.2	1.0	21.0	21.0
2001	177.1	1.03	22.3	21.7
2002	179.9	1.04	22.9	21.9

This calculation uses the CPI, which is an example of a price index. To see how a price index like the CPI is constructed, consider Table 17.3 "Constructing a Price Index", which shows a very simple economy with three goods: T-shirts, music downloads, and meals. The prices and quantities purchased in the economy in 2007 and 2008 are summarized in the table.

Table 17.3 Constructing a Price Index

Year	**T-Shirts**		**Music Downloads**		**Meals**		**Cost of 2008 Basket**	**Price Index**
	Price ($)	Quantity	Price ($)	Quantity	Price ($)	Quantity	($)	
2007	20	10	1	50	25	6	425	1.0
2008	22	12	0.80	60	26	5	442	1.04

To construct a price index, you must decide on a fixed basket of goods. For example, we could use the goods purchased in 2008 (12 T-shirts, 60 downloads, and 5 meals). This fixed basket is then priced in different years. To construct the cost of the 2008 basket at 2008 prices, the product of the price and quantity purchased for each of the three goods in 2008 is added together. The basket costs $442. Then we calculate the cost of the 2008 basket at 2007 prices: that is, we use the prices of each good in 2007 and the quantities purchased in 2008. The sum is $425. The price index is constructed using 2007 as a base year. The value of the price index for 2008 is the cost of the basket in 2008 divided by its cost in the base year, 2007.

When the price index is based on a bundle of goods that represents total output in an economy, it is called the price level. The CPI and the *gross domestic product (GDP) deflator* are examples of measures of the price level (they differ in terms of exactly which goods are included in the bundle). The growth rate of the price level (its percentage change from one year to the next) is called the inflation rate.

We also correct interest rates for inflation. The interest rates you typically see quoted are in nominal terms: they tell you how many dollars you will have to repay for each dollar you borrow. This is called a nominal interest rate. The real interest rate tells you how much you will get next year, in terms of goods and services, if you give up a unit of goods and services this year. To correct interest rates for inflation, we use the Fisher equation:

real interest rate ≈ nominal interest rate − inflation rate.

Key Insights

- Divide nominal values by the price index to create real values.
- Create the price index by calculating the cost of buying a fixed basket in different years.

The Main Uses of This Tool

- Chapter 4 "Life Decisions"
- Chapter 9 "Making and Losing Money on Wall Street"
- Chapter 10 "Raising the Wage Floor"
- Chapter 12 "Superstars"
- Chapter 16 "Cars"

17.9 Supply and Demand

The supply-and-demand framework is the most fundamental framework in economics. It explains both the price of a good or a service and the quantity produced and purchased.

The market supply curve comes from adding together the individual supply curves of firms in a particular market. A competitive firm, taking prices as given, will produce at a level such that

price = marginal cost.

Marginal cost usually increases as a firm produces more output. Consequently, an increase in the price of a product creates an incentive for firms to produce more—that is, the supply curve of a firm is upward sloping. The market supply curve slopes upward as well: if the price increases, all firms in a market will produce more output, and some new firms may also enter the market.

A firm's supply curve shifts if there are changes in input prices or the state of technology. The market supply curve is shifted by changes in input prices and changes in technology that affect a significant number of the firms in a market.

The market demand curve comes from adding together the individual demand curves of all households in a particular market. Households, taking the prices of all goods and services as given, distribute their income in a manner that makes them as well off as possible. This means that they choose a combination of goods and services preferred to any other combination of goods and services they can afford. They choose each good or service such that

price = marginal valuation.

Marginal valuation usually decreases as the household consumes more of a product. If the price of a good or a service decreases, a household will substitute away from other goods and services and toward the product that has become cheaper—that is, the demand curve of a household is downward sloping. The market demand curve slopes downward as well: if the price decreases, all households will demand more.

The household demand curve shifts if there are changes in income, prices of other goods and services, or tastes. The market demand curve is shifted by changes in these factors that are common across a significant number of households.

A market equilibrium is a price and a quantity such that the quantity supplied equals the quantity demanded at the equilibrium price (Figure 17.5 "Market Equilibrium"). Because market supply is upward sloping and market demand is downward sloping, there is a unique equilibrium price. We say we have a competitive market if the following are true:

- The product being sold is homogeneous.
- There are many households, each taking the price as given.
- There are many firms, each taking the price as given.

A competitive market is typically characterized by an absence of barriers to entry, so new firms can readily enter the market if it is profitable, and existing firms can easily leave the market if it is not profitable.

Key Insights

- Market supply is upward sloping: as the price increases, all firms will supply more.
- Market demand is downward sloping: as the price increases, all households will demand less.
- A market equilibrium is a price and a quantity such that the quantity demanded equals the quantity supplied.

Figure 17.5 Market Equilibrium

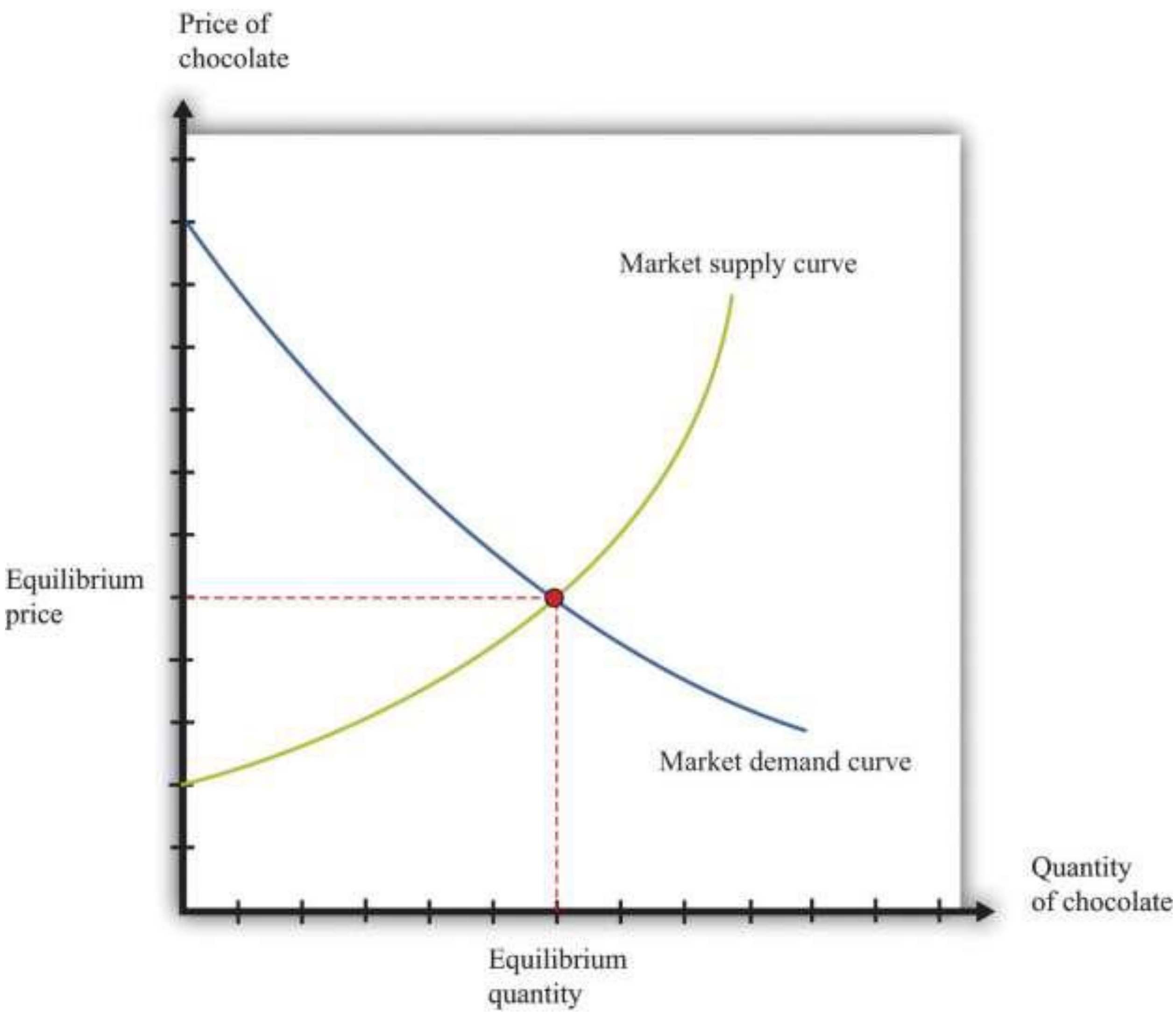


Figure 17.5 "Market Equilibrium" shows equilibrium in the market for chocolate bars. The equilibrium price is determined at the intersection of the market supply and market demand curves.

More Formally

If we let *p* denote the price, *qd* the quantity demanded, and *I* the level of income, then the market demand curve is given by

$qd = a - bp + cI,$

where a, b, and c are constants. By the law of demand, $b > 0$. For a normal good, the quantity demanded increases with income: $c > 0$.

If we let qs denote the quantity supplied, and t the level of technology, the market supply curve is given by

$qs = d + ep + ft$,

where d, e, and f are constants. Because the supply curve slopes upwards, $e > 0$. Because the quantity supplied increases when the technology improves, $f > 0$.

In equilibrium, the quantity supplied equals the quantity demanded. Set $qs = qd = q^*$ and set $p = p^*$ in both equations. The market clearing price (p^*) and quantity (q^*) are as follows:

$$p^*=(a+cI)-(d+ft)a+cI$$

and

$q^* = d + ep^* + ft$.

The Main Uses of This Tool

- Chapter 5 "eBay and craigslist"
- Chapter 6 "Where Do Prices Come From?"
- Chapter 7 "Why Do Prices Change?"
- Chapter 9 "Making and Losing Money on Wall Street"
- Chapter 10 "Raising the Wage Floor"
- Chapter 11 "Barriers to Trade and the Underground Economy"
- Chapter 12 "Superstars"
- Chapter 15 "A Healthy Economy"

17.10 Buyer Surplus and Seller Surplus

If you buy a good, then you obtain buyer surplus. If you did not expect to obtain any surplus, then you would not choose to buy the good.

- Suppose you buy a single unit of the good. Your surplus is the difference between your valuation of the good and the price you pay. This is a measure of how much you gain from the exchange.
- If you purchase many units of a good, then your surplus is the sum of the surplus you get from each unit. To calculate the surplus from each unit, you subtract the price paid from your marginal valuation of that unit.

  If you sell a good, then you obtain seller surplus. If you did not expect to obtain any surplus, you would not sell the good.
- Suppose you sell a single unit of a good. Your surplus is equal to the difference between the price you receive from selling the good and your valuation of the good. This valuation may be a measure of how much you enjoy the good or what you think you could sell it for in some other market.
- If you sell many units of a good, then the surplus you receive is the sum of the surplus for each unit you sell. To calculate the surplus from selling each unit, you take the difference between the price you get for each unit sold and your marginal valuation of that extra unit.

Buyer surplus and seller surplus are created by trade in a competitive market (Figure 17.6 "A Competitive Market"). The equilibrium price and the equilibrium quantity are determined by the intersection of the supply and demand curves. The area below the demand curve and above the price is the buyer surplus; the area above the supply curve and below the price is the seller surplus. The sum of the buyer surplus and the seller surplus is called total surplus or the gains from trade.

Buyer surplus and seller surplus can also arise from individual bargaining (Figure 17.7 "Individual Bargaining"). When a single unit is traded (the case of unit demand and unit supply), the total surplus is the difference between the buyer's valuation and the seller's valuation. Bargaining determines how they share the gains from trade. The quantity of trades, indicated on the horizontal axis, is either zero or one. The valuations of the buyer and the seller are shown on the vertical axis. In this case, the valuation of the buyer ($3,000) exceeds the valuation of the seller ($2,000), indicating that there are gains from trade equal to $1,000. How these gains are shared between the buyer and seller depends on the price they agree on. In part (a) of Figure 17.7 "Individual Bargaining", the buyer gets most of the surplus; in part (b) of Figure 17.7 "Individual Bargaining", the seller gets most of the surplus.

Key Insights

- Buyer surplus and seller surplus are created by trade.
- Buyer surplus is the difference between the marginal value of a good and the price paid.
- Seller surplus is the difference between the price received and the marginal value of a good.

Figure 17.6 *A Competitive Market*

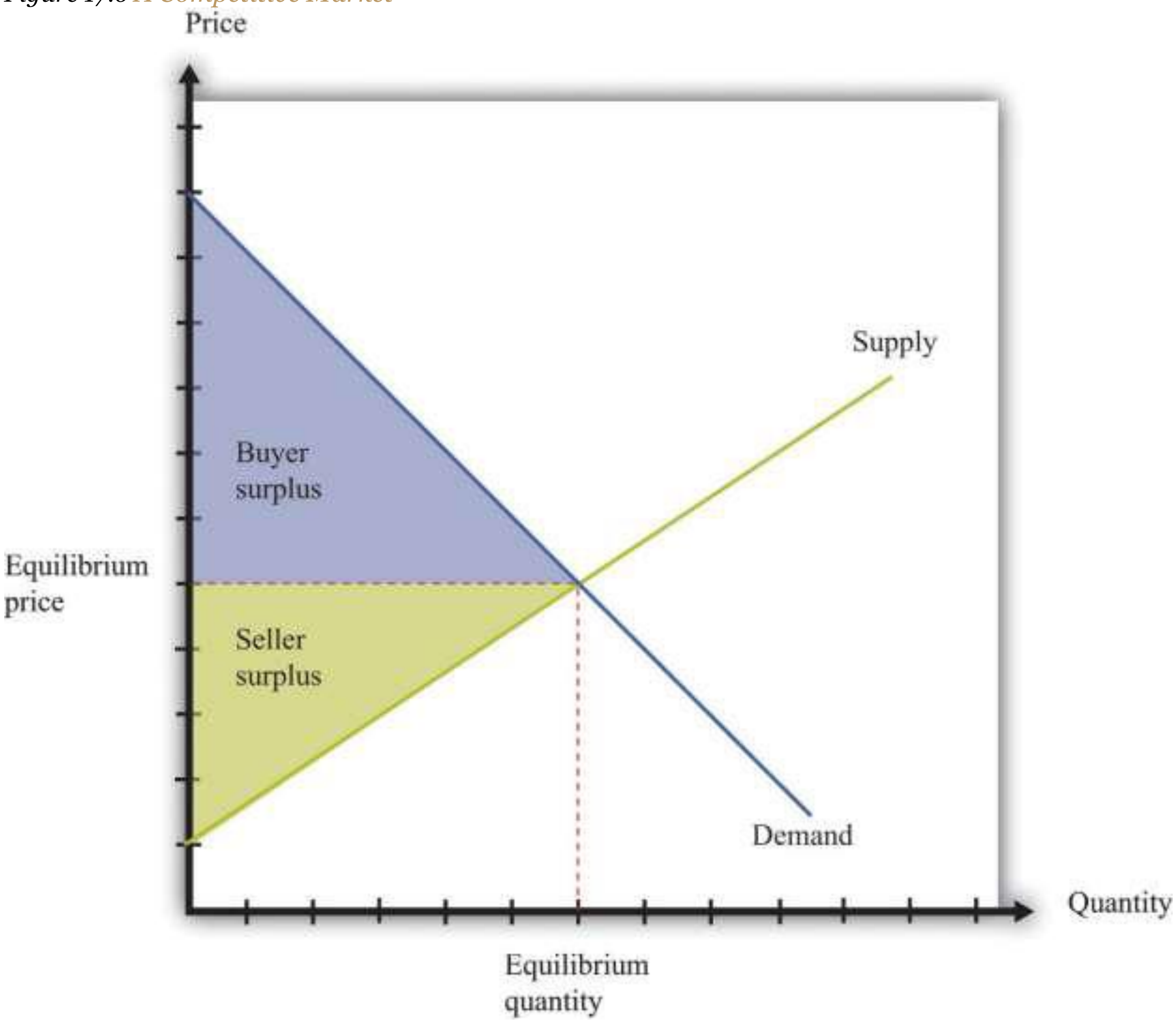
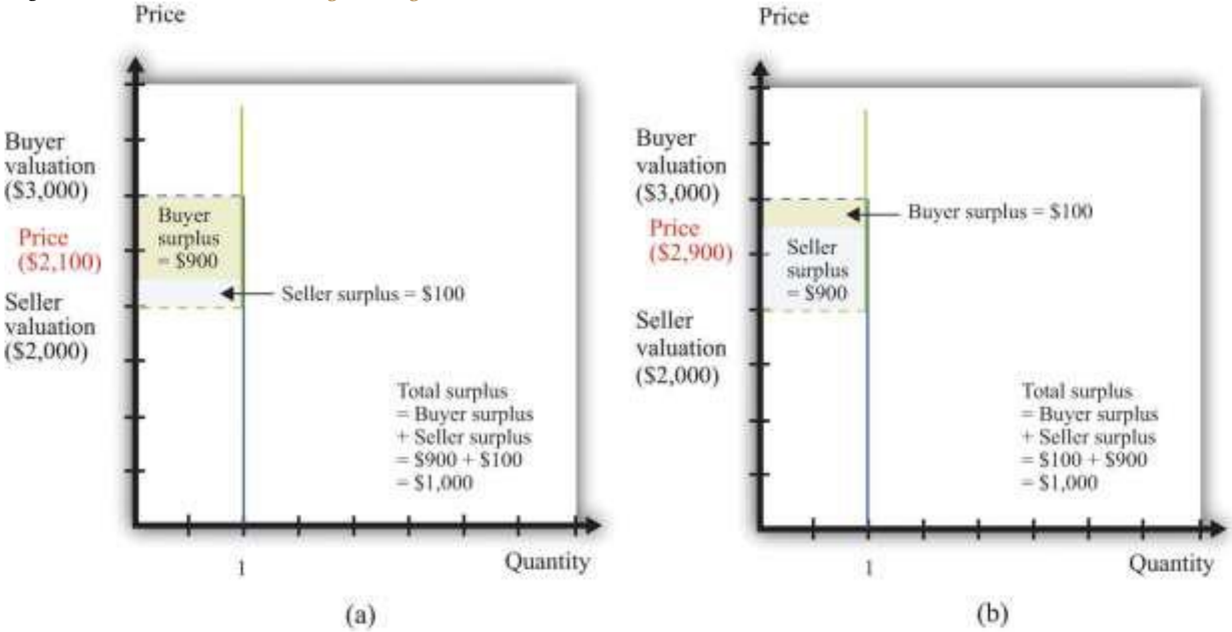


Figure 17.7 Individual Bargaining



The Main Uses of This Tool

- Chapter 5 "eBay and craigslist"
- Chapter 7 "Why Do Prices Change?"
- Chapter 10 "Raising the Wage Floor"
- Chapter 11 "Barriers to Trade and the Underground Economy"
- Chapter 14 "Busting Up Monopolies"
- Chapter 16 "Cars"

17.11 Efficiency and Deadweight Loss

The outcome of a competitive market has a very important property. In equilibrium, all gains from trade are realized. This means that there is no additional surplus to obtain from further trades between buyers and sellers. In this situation, we say that the allocation of goods and services in the economy is efficient. However, markets sometimes fail to operate properly and not all gains from trade are exhausted. In this case, some buyer surplus, seller surplus, or both are lost. Economists call this a deadweight loss.

The deadweight loss from a monopoly is illustrated in Figure 17.8 "Deadweight Loss". The monopolist produces a quantity such that marginal revenue equals marginal cost. The price is determined by the demand curve at this quantity. A monopoly makes a profit equal to total revenue minus total cost. When the total output is less than socially optimal, there is a deadweight loss, which is indicated by the red area in Figure 17.8 "Deadweight Loss".

Deadweight loss arises in other situations, such as when there are quantity or price restrictions. It also arises when taxes or subsidies are imposed in a market. Tax incidence is the way in which the burden of a tax falls on buyers and sellers—that is, who suffers most of the deadweight loss. In general, the incidence of a tax depends on the elasticities of supply and demand.

A tax creates a difference between the price paid by the buyer and the price received by the seller (Figure 17.9 "Tax Burdens"). The burden of the tax and the deadweight loss are defined relative to the tax-free competitive equilibrium. The tax burden borne by the buyer is the difference between the price paid under the tax and the price paid in the competitive equilibrium. Similarly, the burden of the seller is the difference between the price in the competitive equilibrium and the price received under the equilibrium with taxes. The burden borne by the buyer is higher—all else being the same—if demand is less elastic. The burden borne by the seller is higher—all else being the same—if supply is less elastic.

The deadweight loss from the tax measures the sum of the buyer's lost surplus and the seller's lost surplus in the equilibrium with the tax. The total amount of the deadweight loss therefore also depends on the elasticities of demand and supply. The smaller these elasticities, the closer the equilibrium quantity traded with a tax will be to the equilibrium quantity traded without a tax, and the smaller is the deadweight loss.

Key Insights

- In a competitive market, all the gains from trade are realized.
- If sellers have market power, some gains from trade are lost because the quantity traded is below the competitive level.
- Other market distortions, such as taxes, subsidies, price floors, or price ceilings, similarly cause the amount to be traded to differ from the competitive level and cause deadweight loss.

Figure 17.8 Deadweight Loss

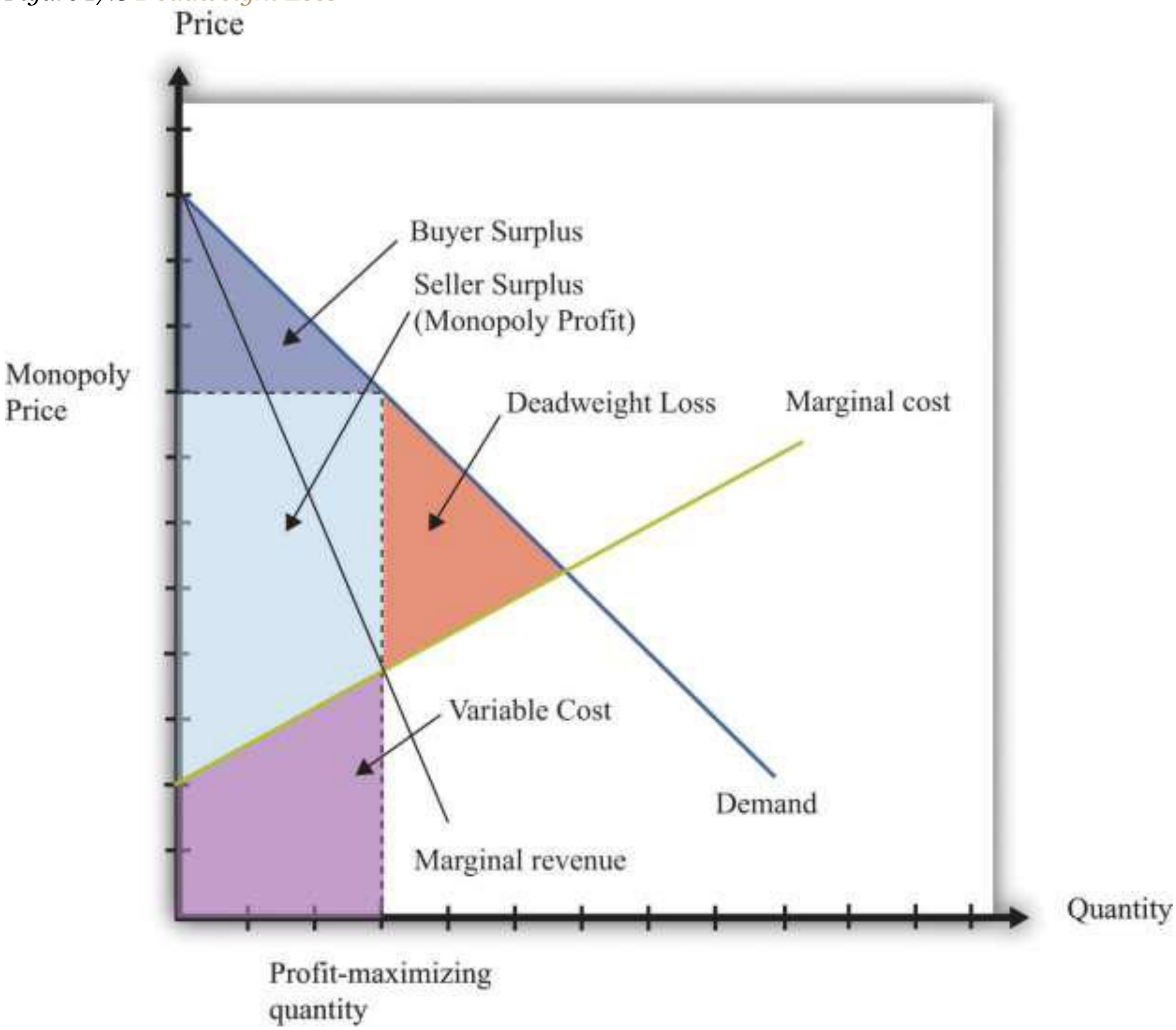
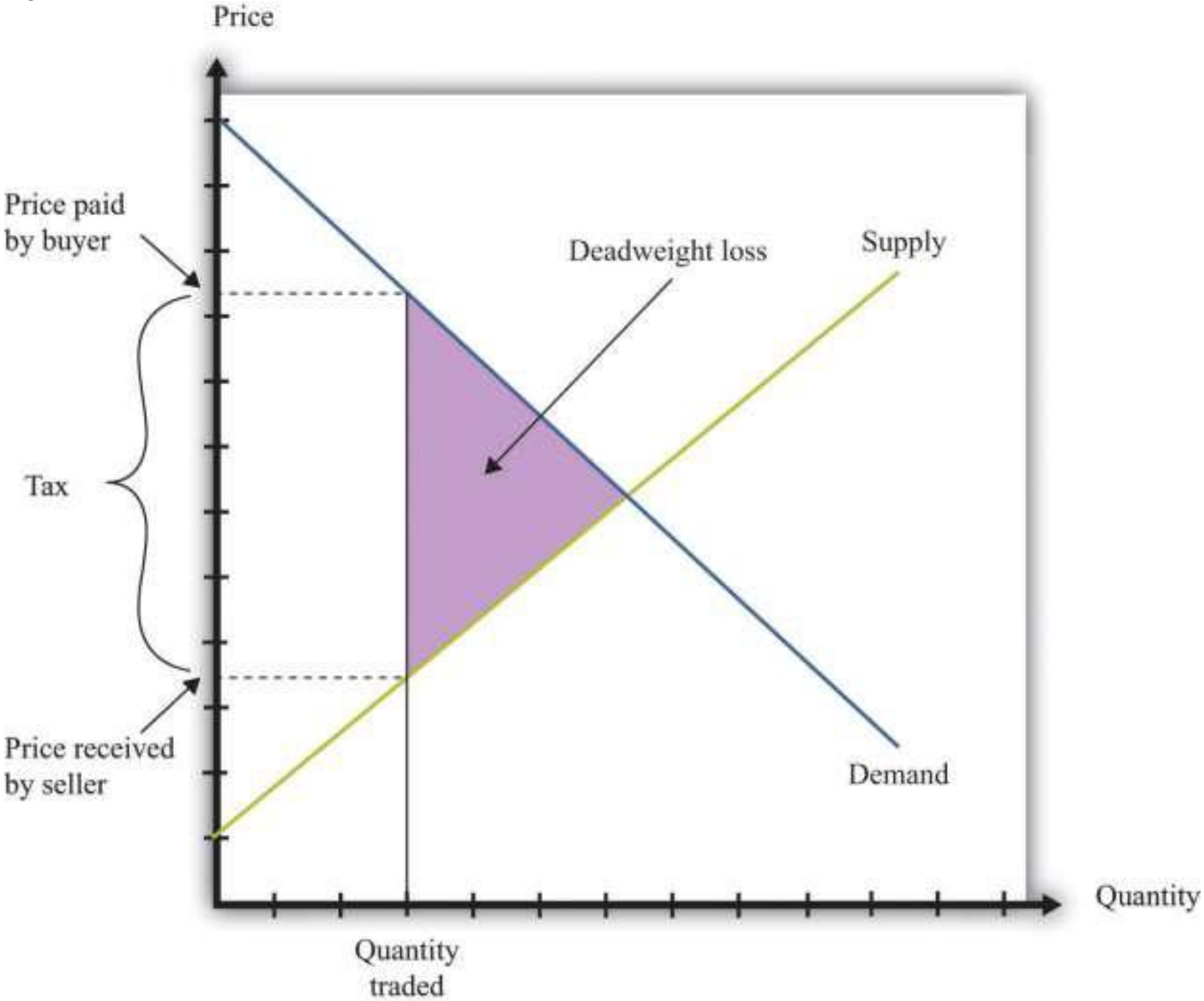


Figure 17.9 Tax Burdens



The Main Uses of This Tool

- Chapter 5 "eBay and craigslist"
- Chapter 7 "Why Do Prices Change?"
- Chapter 9 "Making and Losing Money on Wall Street"
- Chapter 10 "Raising the Wage Floor"
- Chapter 11 "Barriers to Trade and the Underground Economy"
- Chapter 12 "Superstars"
- Chapter 13 "Cleaning Up the Air and Using Up the Oil"
- Chapter 14 "Busting Up Monopolies"
- Chapter 15 "A Healthy Economy"

- Chapter 16 "Cars"

17.12 Production Possibilities Frontier

The production possibilities frontier shows the combinations of goods and services that an economy can produce if it is efficiently using every available input. A key component in understanding the production possibilities frontier is the term *efficiently*. If an economy is using its inputs in an efficient way, then it is not possible to produce more of one good without producing less of another.

Figure 17.10 "The Production Possibilities Frontier" shows the production possibilities frontier for an economy producing web pages and meals. It is downward sloping: to produce more web pages, the production of meals must decrease. Combinations of web pages and meals given by points inside the production possibilities frontier are possible for the economy to produce but are not efficient: at points inside the production possibilities frontier, it is possible for the economy to produce more of both goods. Points outside the production possibilities frontier are not feasible given the current levels of inputs in the economy and current technology.

The negative slope of the production possibilities frontier reflects opportunity cost. The opportunity cost of producing more meals is that fewer web pages can be created. Likewise, the opportunity cost of creating more web pages means that fewer meals can be produced.

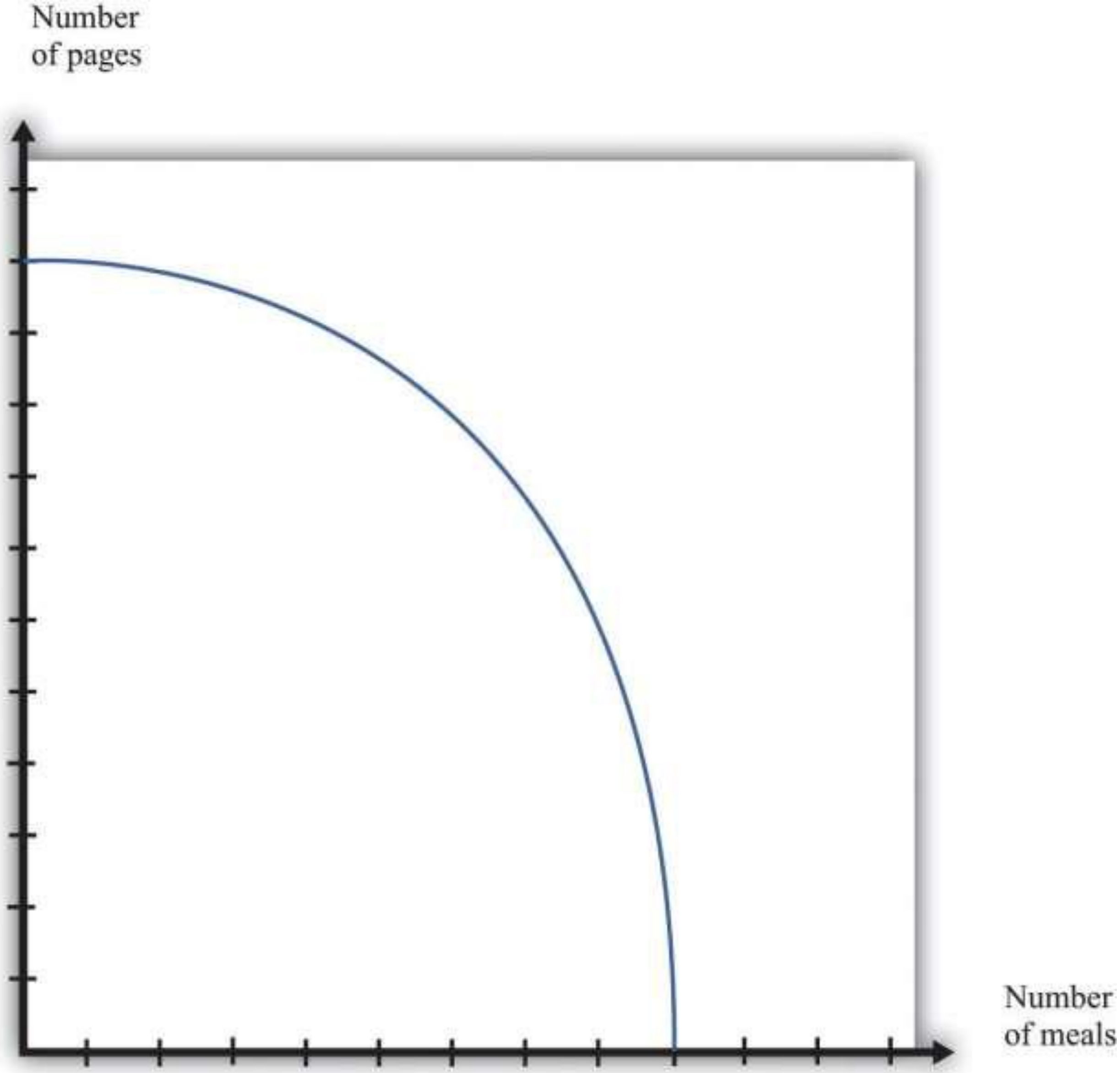
The production possibilities frontier shifts over time. If an economy accumulates more physical capital or has a larger workforce, then it will be able to produce more of all the goods in an economy. Further, it will be able to produce new goods. Another factor shifting the production possibilities frontier outward over time is technology. As an economy creates new ideas (or receives them from other countries) on how to produce goods more cheaply, then it can produce more goods.

Key Insights

- The production possibilities frontier shows the combinations of goods and services that can be produced efficiently in an economy at a point in time.
- The production possibilities frontier is downward sloping: producing more of one good requires producing less of others. The production of a good has an opportunity cost.

- As time passes, the production possibilities frontier shifts outward due to the accumulation of inputs and technological progress.

Figure 17.10 The Production Possibilities Frontier



The Main Uses of This Tool

- Chapter 5 "eBay and craigslist"
- Chapter 11 "Barriers to Trade and the Underground Economy"

17.13 Comparative Advantage

Comparative advantage is an idea that helps explain why individuals and countries trade with each other. Trade is at the heart of modern economies: individuals specialize in production and generalize in consumption. To consume many goods while producing relatively few, individuals must sell what they produce in exchange for the output of others. Countries likewise specialize in certain goods and services and import others. By so doing, they obtain gains from trade.

Table 17.4 shows the productivity of two different countries in the production of two different goods. It shows the number of labor hours required to produce two goods—tomatoes and beer—in two countries: Guatemala and Mexico. From these data, Mexico has an absolute advantage in the production of both goods. Workers in Mexico are more productive at producing both tomatoes and beer in comparison to workers in Guatemala.

Table 17.4

	Hours of Labor Required	
	Tomatoes (1 kilogram)	Beer (1 liter)
Guatemala	6	3
Mexico	2	2

In Guatemala, the opportunity cost of 1 kilogram of tomatoes is 2 liters of beer. To produce an extra kilogram of tomatoes in Guatemala, 6 hours of labor time must be taken away from beer production; 6 hours of labor time is the equivalent of 2 liters of beer. In Mexico, the opportunity cost of 1 kilogram of tomatoes is 1 liter of beer. Thus the opportunity cost of producing tomatoes is lower in Mexico than in Guatemala. This means that Mexico has a comparative advantage in the production tomatoes. By a similar logic, Guatemala has a comparative advantage in the production of beer.

Guatemala and Mexico can have higher levels of consumption of both beer and tomatoes if they trade rather than produce in isolation; each country should specialize (either partially or completely) in the good in which it has a comparative advantage. It is never efficient to have both countries produce both goods.

Key Insights

- Comparative advantage helps predict the patterns of trade between individuals and/or countries.
- A country has a comparative advantage in the production of a good if the opportunity cost of producing that good is lower in that country.
- Even if one country has an absolute advantage in all goods, it will still gain from trading with another country.
- Although the example is cast in terms of countries, the same logic is also used to explain production patterns between two individuals.

The Main Uses of This Tool

- Chapter 3 "Everyday Decisions"
- Chapter 5 "eBay and craigslist"
- Chapter 11 "Barriers to Trade and the Underground Economy"

17.14 Costs of Production

The costs of production for a firm are split into two categories. One type of cost, fixed costs, is independent of a firm's output level. A second type of cost, variable costs, depends on a firm's level of output. Total costs are the sum of the fixed costs and the variable costs.

The change in costs as output changes by a small amount is called marginal cost. It is calculated as follows:

$$\text{marginal cost} = \frac{\text{change in total cost}}{\text{change in quantity}} = \frac{\text{change in variable cost}}{\text{change in quantity}}.$$

Because fixed costs do not depend on the quantity, if we produce one more unit, then the change in total cost and the change in the variable cost are the same. Marginal cost is positive because variable costs increase with output. Marginal cost is usually increasing in the level of output, reflecting the diminishing marginal product of factors of production.

For example, suppose that total costs are given by

$\text{total costs} = 50 + 10 \times \text{quantity}.$

Here the fixed cost is 50, and the variable cost is 10 times the level of output. In this example, marginal cost equals 10. These costs are shown in Table 17.5.

Table 17.5

Output	Fixed Cost	Variable Cost	Total Cost
0	50	0	50
10	50	100	150
20	50	200	250
50	50	500	550

We sometimes divide fixed costs into two components: entry costs, which are the one-time fixed costs required to open a new business or set up a new plant, and fixed operating costs, which are the fixed costs incurred regularly during the normal operation of a business.

Some costs are sunk costs; once incurred, these costs cannot be recovered. Such costs should be ignored in forward-looking business decisions. Other costs are partially or fully recoverable costs. For example, if a firm purchases an asset that can be resold, then the cost of that asset is recoverable.

Figure 17.11 Cost Measures

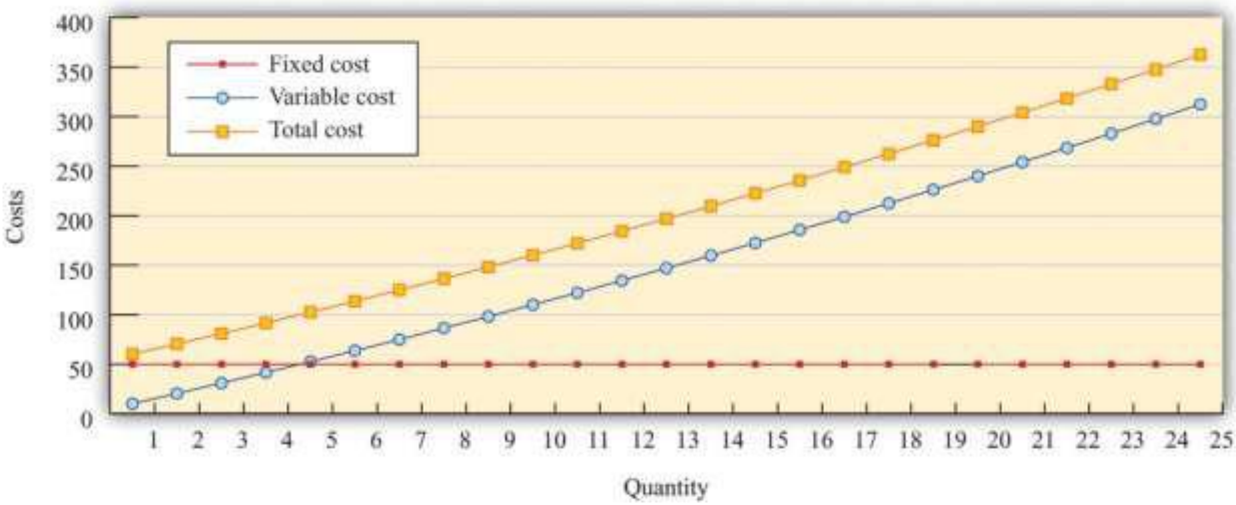


Figure 17.11 "Cost Measures" shows these various measures of costs. It is drawn assuming a fixed cost of 50 and variable costs given by

variable costs = 10 × quantity + 0.1 × quantity^2.

For this example, marginal cost is positive and increasing.

Key Insights

- Fixed costs are independent of the level of output, whereas variable costs depend on the output level of a firm.
- Pricing decisions depend on marginal costs.
- Decisions to enter and/or exit an industry depend on both fixed and variable costs.

The Main Uses of This Tool

- Chapter 6 "Where Do Prices Come From?"

- Chapter 7 "Why Do Prices Change?"
- Chapter 8 "Growing Jobs"
- Chapter 14 "Busting Up Monopolies"

17.15 Pricing with Market Power

The goal of the managers of a firm is to maximize the firm's profit.

profit = revenues − costs.

We can think of a firm as choosing either the price to set or the quantity that it sells. Either way, the firm faces a demand curve and chooses a point on that curve that maximizes its profits. In reality, most firms choose the price of the good that they sell. However, it is often simpler to analyze a firm's behavior by looking at the quantity that it chooses.

Profits are maximized (Figure 17.12 "Markup Pricing") when the extra revenue from selling one more unit of output (marginal revenue) is equal to the extra cost of producing one more unit (marginal cost). The firm's decision rule is to select a point on the demand curve such that

marginal revenue = marginal cost.

We can rearrange this condition to obtain a firm's pricing rule:

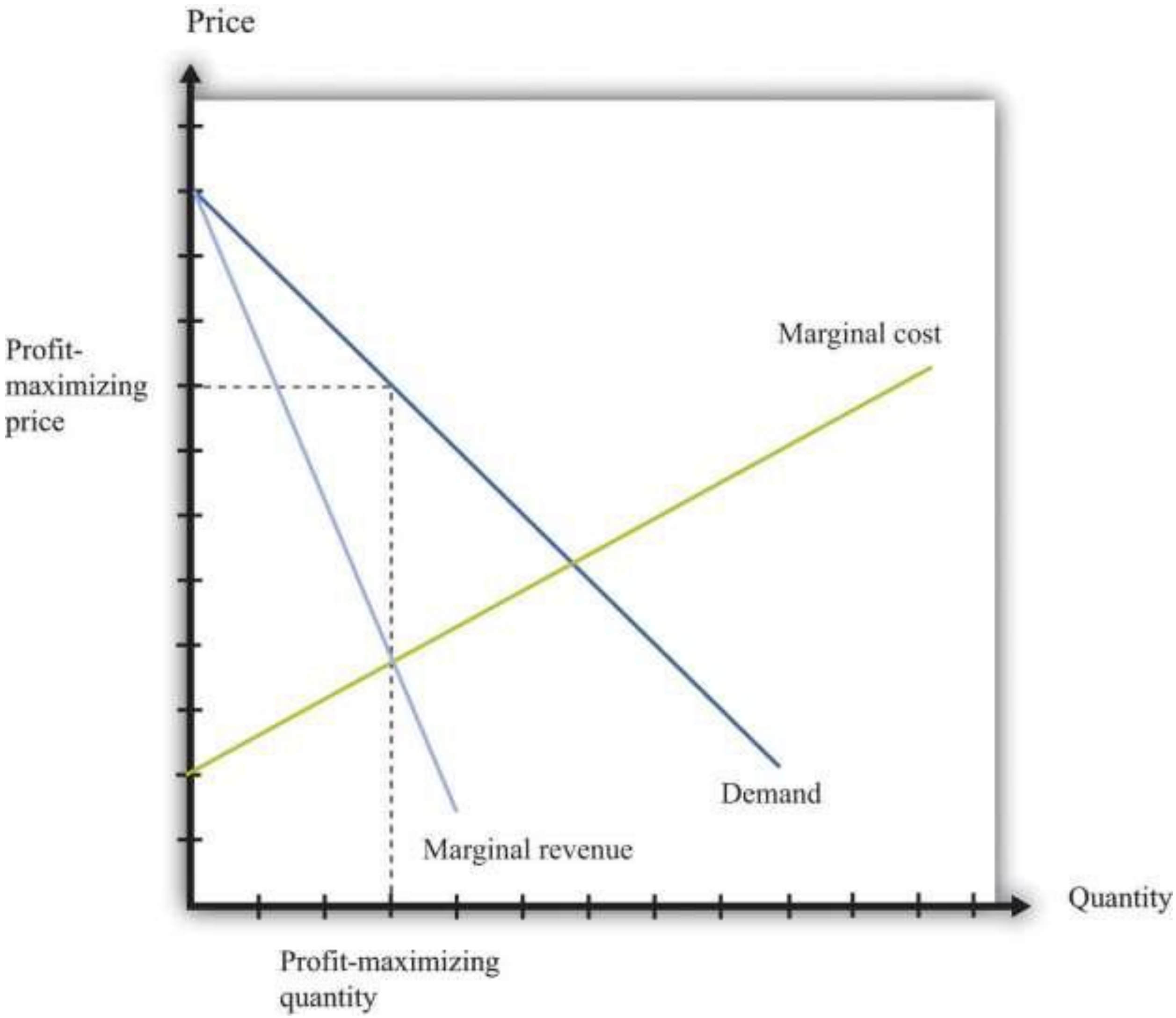
price = markup × marginal cost.

Figure 17.12 "Markup Pricing" illustrates this pricing decision. The markup depends on the price elasticity of demand. When demand is relatively inelastic, firms have a lot of market power and set a high markup. This is *not* a "plug-and-play" formula because both the markup and marginal cost depend, in general, on the price that a firm chooses. However, it does provide a useful description of a firm's decision.

Key Insights

- When marginal cost is higher, a firm sets a higher price.
- When demand is more inelastic (so a firm has more market power), the markup is higher, so a firm sets a higher price.
- When demand is perfectly elastic, the markup is 1, and the firm sets its price equal to marginal cost. This is the case of a competitive market.
- Any price you see has two components: the marginal cost and the markup. When a price changes, one or both of these must have changed.

Figure 17.12 Markup Pricing



More Formally

We can derive the markup pricing formula as follows, where π = profit, *R* = revenues, *C* = costs, *MR* = marginal revenue, *MC* = marginal cost, *P* = price, *Q* = output, $\varepsilon = (\Delta Q/Q)/(\Delta P/P)$ = elasticity of demand, and μ = markup.

First we note that

$$MR=\Delta R\Delta Q=P(1-1-\varepsilon).$$

The firm sets marginal revenue equal to marginal cost:

$$MR = P\left(1 - \frac{1}{-\varepsilon}\right) = MC.$$

Rearranging, we obtain

$P = \mu \times MC$,

where the markup is given by

$$\mu = \frac{1}{1 - \frac{1}{-\varepsilon}}.$$

The Main Uses of This Tool

- Chapter 6 "Where Do Prices Come From?"
- Chapter 7 "Why Do Prices Change?"
- Chapter 8 "Growing Jobs"
- Chapter 14 "Busting Up Monopolies"
- Chapter 16 "Cars"

17.16 Comparative Statics

Comparative statics is a tool used to predict the effects of exogenous variables on market outcomes. By exogenous variables, we mean anything that shifts either the market demand curve (for example, news about the health effects of consuming a product) or the market supply curve (for example, weather effects on a crop). By market outcomes, we mean the equilibrium price and the equilibrium quantity in a market. Comparative statics is a comparison of the market equilibrium before and after a change in an exogenous variable.

A comparative statics exercise consists of a sequence of five steps:

1. Begin at an equilibrium point where the quantity supplied equals the quantity demanded.
2. Based on the description of the event, determine whether the change in the exogenous factor shifts the market supply curve or the market demand curve.
3. Determine the direction of this shift.
4. After shifting the curve, find the new equilibrium point.
5. Compare the new and old equilibrium points to predict how the exogenous event affects the market.

Figure 17.13 "Shifts in the Demand Curve" and Figure 17.14 "Shifts in the Supply Curve" show comparative statics in action. In Figure 17.13 "Shifts in the Demand Curve", the market demand curve has shifted to the left. The consequence is that the equilibrium price and the equilibrium quantity both decrease. Notice that the demand curve shifts along a fixed supply curve. In Figure 17.14 "Shifts in the Supply Curve", an increase in the price of an input (hops) has shifted the market supply curve of beer to the left. The consequence is that the equilibrium price increases and the equilibrium quantity decreases. Notice that the supply curve shifts along a fixed demand curve.

Key Insights

- Comparative statics is used to determine the market outcome when the market supply and demand curves are shifting.
- Comparative statics is a comparison of equilibrium points.

- If the market demand curve shifts, then the new and old equilibrium points lie on a fixed market supply curve.
- If the market supply curve shifts, then the new and old equilibrium points lie on a fixed market demand curve.

Figure 17.13 Shifts in the Demand Curve

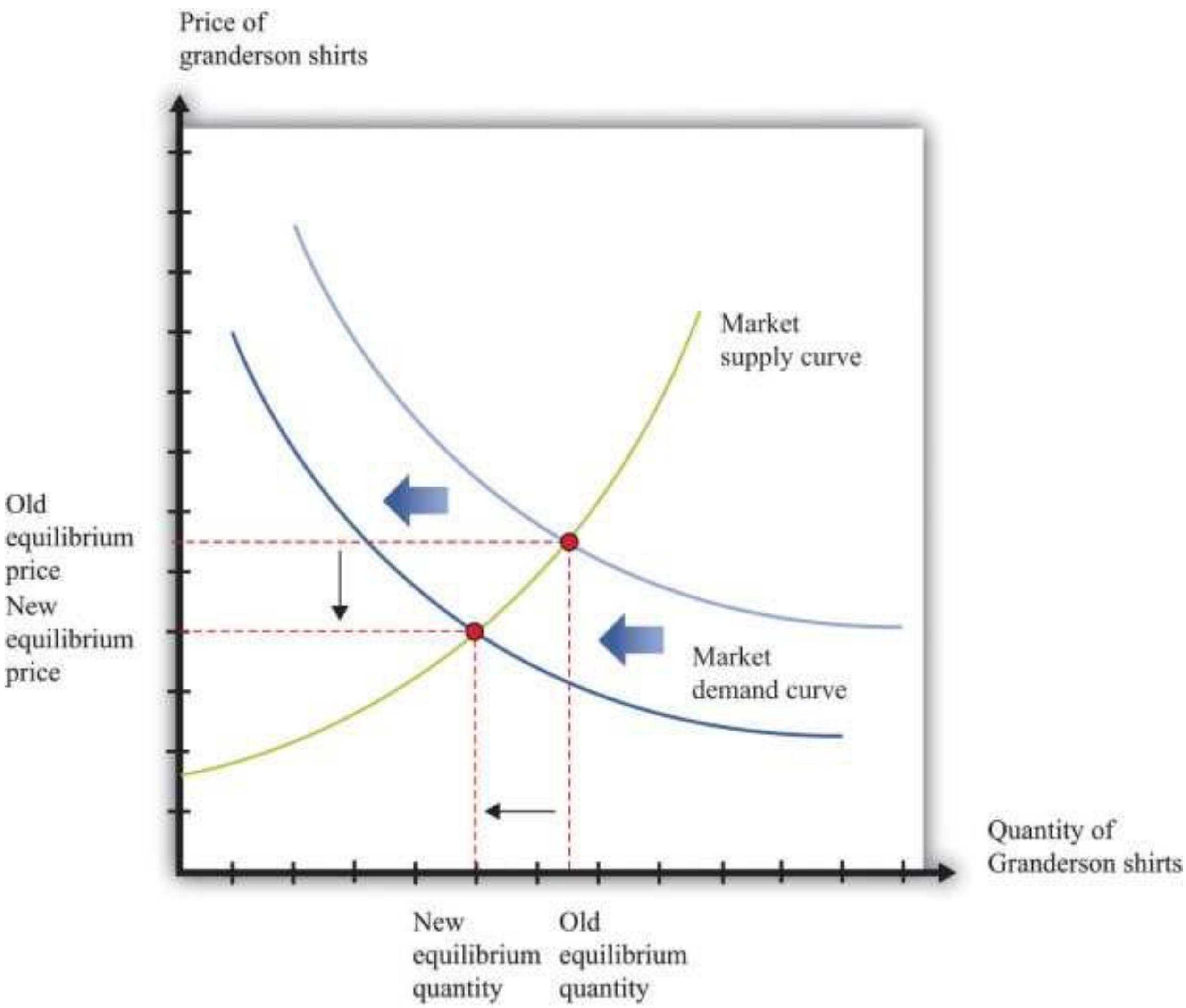
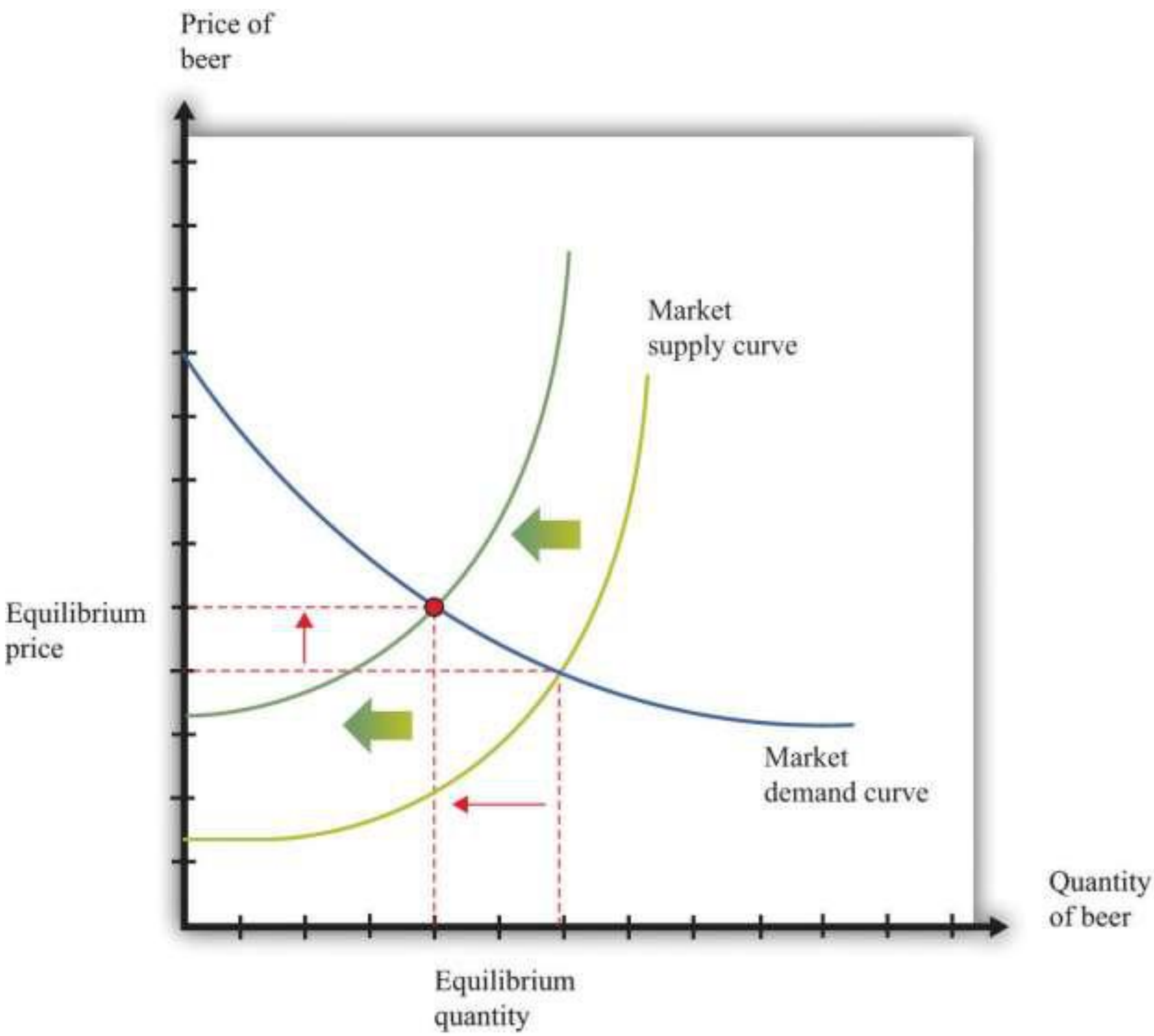


Figure 17.14 *Shifts in the Supply Curve*



The Main Uses of This Tool

- Chapter 7 "Why Do Prices Change?"
- Chapter 9 "Making and Losing Money on Wall Street"

17.17 Production Function

The production function characterizes the output of a firm given the inputs it uses. The link between inputs and output is shown Figure 17.15 "The Production Function". The production function combines a firm's physical capital stock, labor, raw materials (or intermediate inputs), and technology to produce output. Technology is the knowledge (the "blueprints") that the firm possesses, together with managerial skills.

Production functions generally have two important properties:

1. Positive marginal product of an input
2. Diminishing marginal product of an input

By input, we mean any of the factors of production, such as physical capital, labor, or raw materials. The marginal product of an input is the extra output obtained if extra input is used. In this conceptual exercise, all other inputs are held fixed so that we change only one input at a time.

The first property asserts that additional output will be obtained from additional units of an input. Adding another machine, another worker, some more fuel, and so on, increases the output of a firm. A positive marginal product does not necessarily mean that the extra output is profitable: it might be that the cost of the extra input is high relative to the value of the additional output obtained.

The second property explains how the marginal product of an input changes as we increase the amount of that input, keeping the quantities of other inputs fixed. An additional unit of an input will (usually) increase output more when there is a small (rather than a large) amount of that input being used. For example, the extra output obtained from adding the first machine is greater than the additional output obtained from adding the 50th machine.

A simple production function relating output to labor input is shown in Figure 17.16 "Labor Input in the Production Function". This figure illustrates the two properties of positive and diminishing marginal product of labor. As more labor is added, output increases: there is a positive marginal product of labor (that is, the slope of the relationship is positive). But the extra output obtained from adding labor is greater when the labor input is low: there is diminishing marginal product of labor. From the graph, the slope of the production function (which is the marginal product of labor) is greater at low levels of the labor input.

Key Insights

- The production function shows the output produced by a firm given its inputs.
- The production function displays two important properties: positive marginal product and diminishing marginal product.

Figure 17.15 *The Production Function*

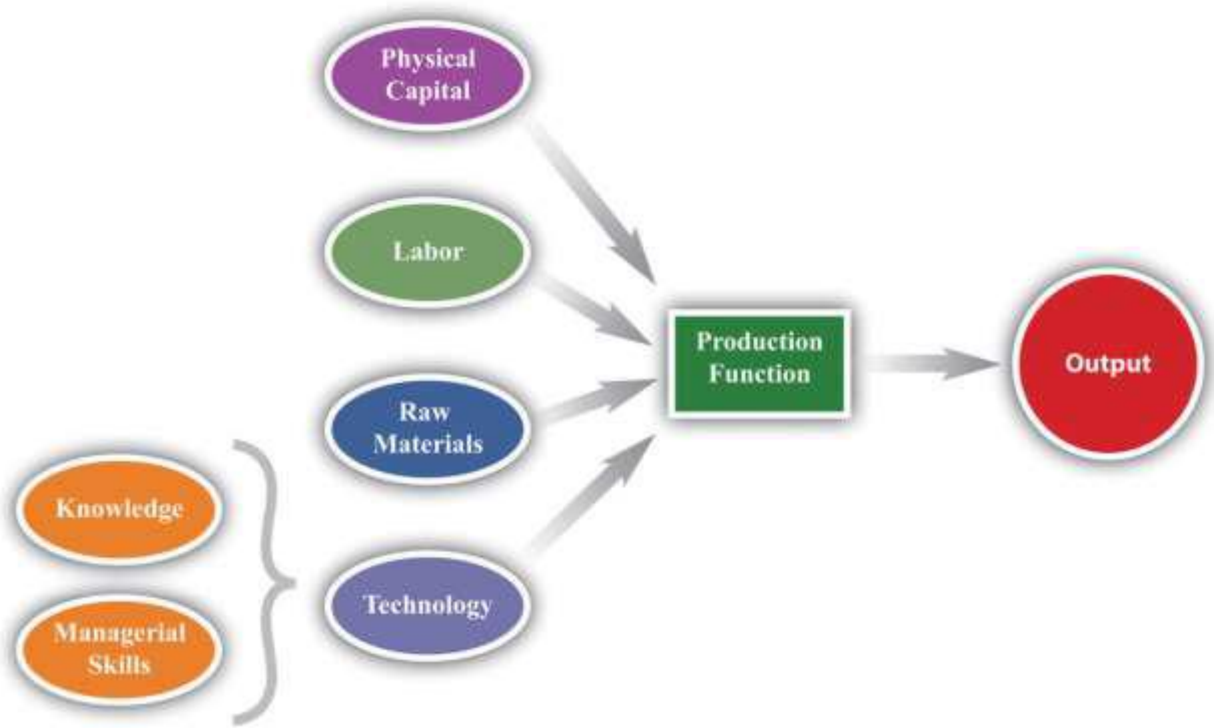
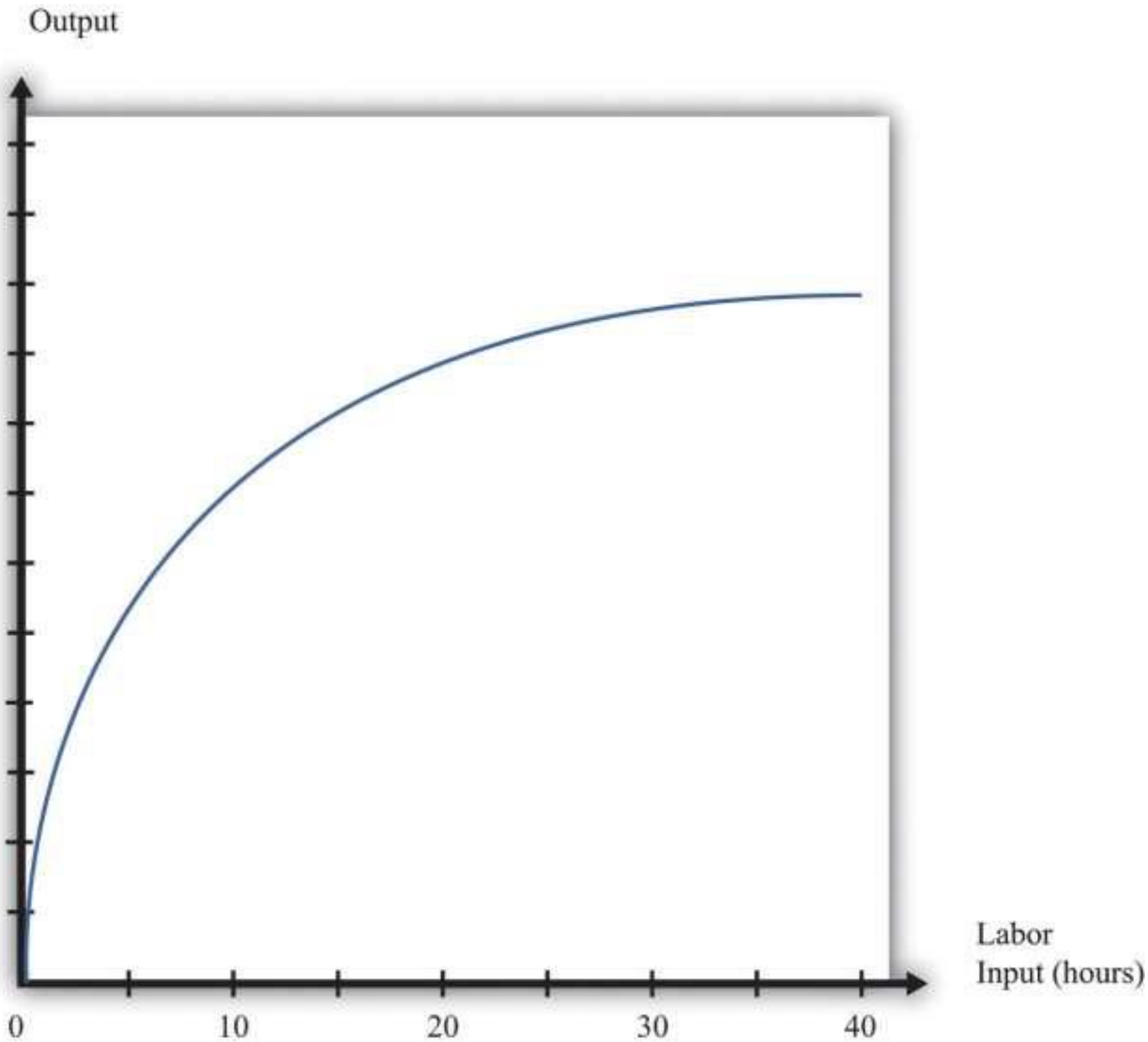


Figure 17.16 Labor Input in the Production Function



The Main Uses of This Tool

- Chapter 5 "eBay and craigslist"
- Chapter 8 "Growing Jobs"
- Chapter 15 "A Healthy Economy"

17.18 Nash Equilibrium

A Nash equilibrium is used to predict the outcome of a game. By a game, we mean the interaction of a few individuals, called *players*. Each player chooses an*action* and receives a *payoff* that depends on the actions chosen by everyone in the game.

A Nash equilibrium is an action for each player that satisfies two conditions:

1. The action yields the highest payoff for that player given her predictions about the other players' actions.
2. The player's predictions of others' actions are correct.

Thus a Nash equilibrium has two dimensions. Players make decisions that are in their own self-interest, and players make accurate predictions about the actions of others.

Consider the games in Table 17.6 "Prisoners' Dilemma", Table 17.7 "Dictator Game", Table 17.8 "Ultimatum Game", and Table 17.9 "Coordination Game". The numbers in the tables give the payoff to each player from the actions that can be taken, with the payoff of the row player listed first.

Table 17.6 Prisoners' Dilemma

	Left	**Right**
Up	5, 5	0, 10
Down	10, 0	2, 2

Table 17.7 Dictator Game

Number of Dollars (x)	$100 - x, x$

Table 17.8 Ultimatum Game

	Accept	**Reject**
Number of Dollars (x)	$100 - x, x$	0, 0

Table 17.9 Coordination Game

	Left	**Right**

	Left	Right
Up	5, 5	0, 1
Down	1, 0	4, 4

- **Prisoners' dilemma**. The row player chooses between the action labeled "Up" and the one labeled "Down." The column player chooses between the action labeled "Left" and the one labeled "Right." For example, if row chooses "Up" and column chooses "Right," then the row player has a payoff of 0, and the column player has a payoff of 10. If the row player predicts that the column player will choose "Left," then the row player should choose "Down" (that is, down for the row player is her *best response* to left by the column player). From the column player's perspective, if he predicts that the row player will choose "Up," then the column player should choose "Right." The Nash equilibrium occurs when the row player chooses "Down" and the column player chooses "Right." Our two conditions for a Nash equilibrium of making optimal choices and predictions being right both hold.
- **Social dilemma**. This is a version of the prisoners' dilemma in which there are a large number of players, all of whom face the same payoffs.
- **Dictator game**. The row player is called the dictator. She is given $100 and is asked to choose how many dollars (x) to give to the column player. Then the game ends. Because the column player does not move in this game, the dictator game is simple to analyze: if the dictator is interested in maximizing her payoff, she should offer nothing ($x = 0$).
- **Ultimatum game**. This is like the dictator game except there is a second stage. In the first stage, the row player is given $100 and told to choose how much to give to the column player. In the second stage, the column player accepts or rejects the offer. If the column player rejects the offer, neither player receives any money. The best choice of the row player is then to offer a penny (the smallest amount of money there is). The best choice of the column player is to accept. This is the Nash equilibrium.
- **Coordination game**. The coordination game has two Nash equilibria. If the column player plays "Left," then the row player plays "Up"; if the row player plays "Up," then the column player

plays “Left.” This is an equilibrium. But “Down/Right” is also a Nash equilibrium. Both players prefer “Up/Left,” but it is possible to get stuck in a bad equilibrium.

Key Insights

- A Nash equilibrium is used to predict the outcome of games.
- In real life, payoffs may be more complicated than these games suggest. Players may be motivated by fairness or spite.

The Main Uses of This Tool

- Chapter 12 "Superstars"
- Chapter 13 "Cleaning Up the Air and Using Up the Oil"
- Chapter 14 "Busting Up Monopolies"
- Chapter 16 "Cars"

17.19 Externalities and Public Goods

Some economic transactions have effects on individuals not directly involved in that transaction. When this happens, we say there is an externality present. An externality is generated by a decision maker who disregards the effects of his actions on others. In the case of a positive externality, the individual's actions increase the welfare of others (for example, research and development by firms). In the case of a negative externality, an individual's actions decrease the welfare of others (for example, pollution).

Economic outcomes are not efficient when externalities are present. So the government may be able to improve on the private outcome. The possible remedies are as follows:

- Subsidies (in the case of positive externalities) and taxes (in the case of negative externalities)
- The creation of markets by the government

If people are altruistic, then they may instead take into account others' welfare and may internalize some of the effects of their actions.

We typically see externalities associated with nonexcludable goods (or resources)—goods for which it is impossible to selectively deny access. In other words, it is not possible to let some people consume the good while preventing others from consuming it. An excludable good (or resource) is one to which we can selectively allow or deny access. If a good is nonexcludable or partially excludable, there are positive externalities associated with its production and negative externalities associated with its consumption.

We say that a good is a rival if one person's consumption of the good prevents others from consuming the good. Most of the goods we deal with in economics are rival goods. A good is nonrival if one person can consume the good without preventing others from consuming the same good. Knowledge is a nonrival good.

If a good is both nonexcludable and nonrival, it is a public good.

Key Insights

- When externalities are present, the outcome is inefficient.
- The market will typically not provide public goods.

The Main Uses of This Tool

- Chapter 11 "Barriers to Trade and the Underground Economy"
- Chapter 13 "Cleaning Up the Air and Using Up the Oil"
- Chapter 15 "A Healthy Economy"
- Chapter 16 "Cars"

17.20 Foreign Exchange Market

A foreign exchange market is where one currency is traded for another. There is a demand for each currency and a supply of each currency. In these markets, one currency is bought using another. The price of one currency in terms of another (for example, how many dollars it costs to buy one Mexican peso) is called the exchange rate.

Foreign currencies are demanded by domestic households, firms, and governments who wish to purchase goods, services, or financial assets that are denominated in the currency of another economy. For example, if a US auto importer wants to buy a German car, it must buy euros. The law of demand holds: as the price of a foreign currency increases, the quantity of that currency demanded will decrease.

Foreign currencies are supplied by foreign households, firms, and governments that wish to purchase goods, services, or financial assets denominated in the domestic currency. For example, if a Canadian bank wants to buy a US government bond, it must sell Canadian dollars. As the price of a foreign currency increases, the quantity supplied of that currency increases.

Exchange rates are determined just like other prices: by the interaction of supply and demand. At the equilibrium exchange rate, the supply and demand for a currency are equal. Shifts in the supply or demand for a currency lead to changes in the exchange rate. Because one currency is exchanged for another in a foreign exchange market, the demand for one currency entails the supply of another. Thus the dollar market for euros (where the price is dollars per euro and the quantity is euros) is the mirror image of the euro market for dollars (where the price is euros per dollar and the quantity is dollars).

To be concrete, consider the demand for and supply of euros. The supply of euros comes from the following:

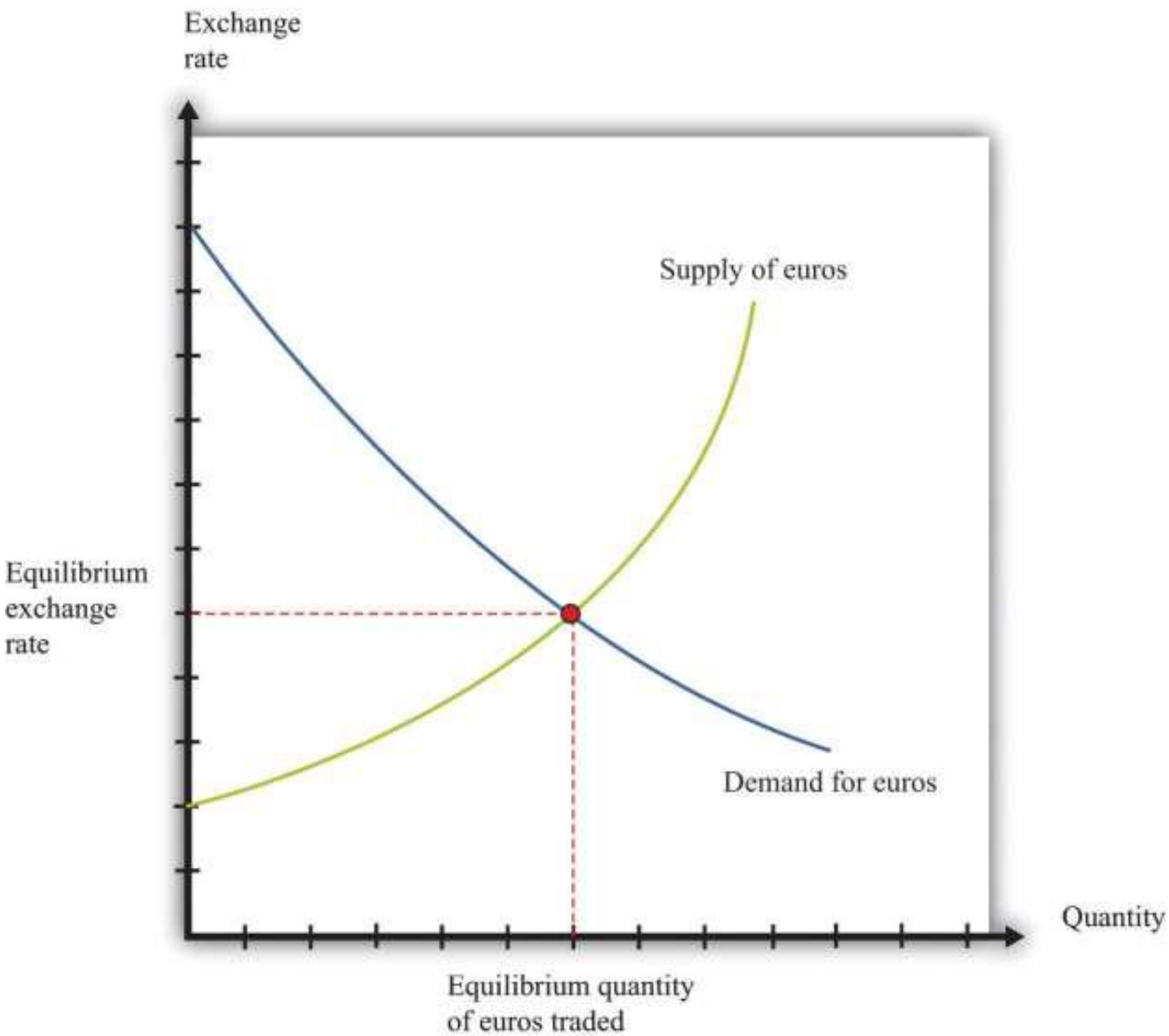
- European households and firms that wish to buy goods and services from non-euro countries
- European investors who wish to buy assets (government debt, stocks, bonds, etc.) that are denominated in currencies other than the euro

  The demand for euros comes from the following:
- Households and firms in non-euro countries that wish to buy goods and services from Europe

- Investors in non-euro countries that wish to buy assets (government debt, stocks, bonds, etc.) that are denominated in euros

Figure 17.17 "The Foreign Exchange Market" shows the dollar market for euros. On the horizontal axis is the quantity of euros traded. On the vertical axis is the price in terms of dollars. The intersection of the supply and demand curves determines the equilibrium exchange rate.

Figure 17.17 The Foreign Exchange Market



The foreign exchange market can be used as a basis for comparative statics exercises. We can study how changes in an economy affect the exchange rate. For example, suppose there is an increase in the level of economic activity in the United States. This will lead to an increase in the demand for European goods and services. To make these purchases, US households and firms will demand more euros. This will cause an outward shift in the demand curve and an increase in the dollar price of euros.

When the dollar price of a euro increases, we say that the dollar has depreciated relative to the euro. From the perspective of the euro, the depreciation of the dollar represents an appreciation of the euro.

Key Insights

- As the exchange rate increases (so a currency becomes more valuable), a greater quantity of the currency is supplied to the market and a smaller quantity is demanded.

The Main Uses of This Tool

- Chapter 7 "Why Do Prices Change?"
- Chapter 9 "Making and Losing Money on Wall Street"

17.21 Percentage Changes and Growth Rates

If some variable x (say, the number of gallons of gasoline sold in a week) changes from x_1 to x_2, then we can simply define the change in that variable as $\Delta x = x_2 - x_1$. But there are problems with this simple definition. The number that we calculate will change depending on the units in which we measure x. If we measure in millions of gallons, x will be a much smaller number than if we measure in gallons. If we measured x in liters rather than gallons (as it is indeed measured in most countries), it would be a bigger number. So the number we calculate depends on the units we choose. To avoid these problems, we look at percentage changes and express the change as a fraction of the individual value. In what follows, we use the notation $\%\Delta x$ to mean the percentage change in x, and we define it as follows: $\%\Delta x = (x_2 - x_1)/x_1$. A percentage change equal to 0.1 means that gasoline consumption increased by 10 percent. Why? Because 10 percent means 10 "per hundred," so 10 percent 10/100 = 0.1.

Very often in economics, we are interested in changes that take place over time. Thus we might want to compare gross domestic product (a measure of how much our economy has produced) between 2012 and 2013. Suppose we know that gross domestic product in the United States in 2012 was \$14 trillion and that gross domestic product in 2013 was \$14.7 trillion. Using the letter Y to denote gross domestic product measured in trillions, we write: $Y_{2012} = 14.0$ and $Y_{2013} = 14.7$. If we want to talk about gross domestic product at different points in time without specifying a particular year, we use the notation Yt. We express the change in a variable over time in the form of a growth rate, which is just an example of a percentage change. Thus the growth rate of gross domestic product in 2013 is calculated as

$\%\Delta\ Y_{2013} = (Y_{2013} - Y_{2012})/\ Y_{20126} = (14.7 - 14)/14 = 0.05.$

The growth rate equals 5 percent. In general, we write $\%\Delta\ Yt_{+1} = (Yt_{+1} - Yt)/\ Yt$.

The Main Uses of This Tool

- Chapter 9 "Making and Losing Money on Wall Street"
- Chapter 12 "Superstars"

17.22 Mean and Variance

To start our presentation of descriptive statistics, we construct a data set using a spreadsheet program. The idea is to simulate the flipping of a two-sided coin. While you might think it would be easier just to flip a coin, doing this on a spreadsheet gives you a full range of tools embedded in that program. To generate the data set, we drew 10 random numbers using the spreadsheet program. In the program we used, the function was called RAND, and this generated the choice of a number between zero and one. Those choices are listed in the second column of the table.

The third column creates the two events of heads and tails that we normally associate with a coin flip. To generate this last column, we adopted a rule: if the random number was less than 0.5, we termed this a "tail" and assigned a 0 to the draw; otherwise, we termed it a "head" and assigned a 1 to the draw. The choice of 0.5 as the cut-off for heads reflects the fact that we are considering the flips of a fair coin in which each side has the same probability of 0.5.

Table 17.10

Draw	Random Number	Heads (1) or Tails (0)
1	0.94	1
2	0.84	1
3	0.26	0
4	0.04	0
5	0.01	0
6	0.57	1
7	0.74	1
8	0.81	1
9	0.64	1

Draw	Random Number	Heads (1) or Tails (0)
10	0.25	0

Keep in mind that the realization of the random number in draw *i* is independent of the realizations of the random numbers in both past and future draws. Whether a coin comes up heads or tails on any particular flip does not depend on other outcomes.

There are many ways to summarize the information contained in a sample of data. Even before you start to compute some complicated statistics, having a way to present the data is important. One possibility is a bar graph in which the fraction of observations of each outcome is easily shown. Alternatively, a pie chart is often used to display this fraction. Both the pie chart and bar diagram are commonly found in spreadsheet programs.

Economists and statisticians often want to describe data in terms of numbers rather than figures. We use the data from Table 17.10 to define and illustrate two statistics that are commonly used in economics discussions. The first is the mean (or average) and is a measure of *central tendency*. Before you read any further, ask yourself what you think the average ought to be from the coin-flipping exercise. It is natural to say 0.5, since half of the time the outcome will be a head and thus have a value of zero, while the remainder of the time the outcome will be a tail and thus have a value of one.

Whether or not that guess holds can be checked by looking at Table 17.10 and calculating the mean of the outcome. We let ki be the outcome of draw i. For example, from the table, $k_1 = 1$ and $k_5 = 0$. Then the formula for the mean if there are N draws is $\mu = \Sigma i\ \mathrm{k}i/N$. Here $\Sigma i\ ki$ means the sum of the ki outcomes. In words, the mean, denoted by μ, is calculated by summing the draws and dividing by the number of draws, N. In the table, $N = 10$ and the sum of the draws of random numbers is about 51.0. Thus the mean of the 10 draws is about 0.51.

We can also calculate the mean of the heads/tails column, and this is 0.6, since heads came up 6 times in our experiment. This calculation of the mean differs from the mean of the draws because the numbers in the two columns differ, with the third column being a very discrete way to represent the information in the second column.

A second commonly used statistic is a measure of dispersion of the data called the variance. The variance, denoted σ^2, is calculated as $\sigma^2 = \Sigma i\ (ki - \mu)^2/(N - 1)$. From this formula, if all the draws were the same (thus equal to the mean), then the variance would be zero. As the draws spread out from the mean (both above and below), the variance increases. Since some observations are above the mean and others below, we square the difference between a single observation (ki) and the mean (μ) when calculating the variance. This means that values above and below the mean both contribute a positive amount to the variance. Squaring also means that values that are a long way away from the mean have a big effect on the variance.

For the data given in Table 17.10, the mean of the 10 draws was given: $\mu = 0.51$. So to calculate the variance, we would subtract the mean from each draw, square the difference, and then sum up the squared differences. This yields a variance of 0.118 for this draw. A closely related concept is that of the standard deviation, which is just the square root of the variance. For our example, the standard deviation is 0.34. The standard deviation is bigger than the variance because the variance is less than 1.

17.23 Correlation and Causality

Correlation is a statistical measure describing how two variables move together. In contrast, causality (or causation) goes deeper into the relationship between two variables by looking for cause and effect.

Correlation is a statistical property that summarizes the way in which two variables move either over time or across people (firms, governments, etc.). The concept of correlation is quite natural to us, as we often take note of how two variables interrelate. If you think back to high school, you probably have a sense of how your classmates did in terms of two measures of performance: grade point average (GPA) and results on a standardized college entrance exam (SAT). It is likely that classmates with high GPAs also had high scores on the SAT exam. In this instance, we would say that the GPA and SAT scores were positively correlated: looking across your classmates, when a person's GPA score is higher than average, that person's SAT score is likely to be higher than average as well.

As another example, consider the relationship between a household's income and its expenditures on housing. If you conducted a survey across households, it is likely that you would find that richer households spend more on most goods and services, including housing. In this case, we would conclude that income and expenditures on housing are *positively correlated*.

When economists look at data for the whole economy, they often focus on a measure of how much is produced, which we call real gross domestic product (GDP), and the fraction of workers without jobs, called the unemployment rate. Over long periods of time, when GDP is above average (so that the economy is doing well), the unemployment rate is below average. In this case, GDP and the unemployment rate are *negatively correlated*, as they tend to move in opposite directions.

The fact that one variable is correlated with another does not inform us about whether one variable *causes* the other. Imagine yourself on an airplane in a relaxed mood, reading or listening to music. Suddenly, the pilot comes on the public address system and requests that you buckle your seat belt. Usually, such a request is followed by turbulence. This is a correlation: the announcement by the pilot is positively correlated with air turbulence. The correlation is, of course, not perfect, because sometime you hit some bumps without warning and sometimes the pilot's announcement is not followed by turbulence.

But—obviously!—this doesn't mean that we could solve the turbulence problem by turning off the public address system. The pilot's announcement does not*cause* the turbulence. The turbulence is there whether the pilot announces it or not. In fact, the causality runs the other way. The turbulence causes the pilot's announcement.

We noted earlier that real GDP and unemployment are negatively correlated. When real GDP is below average, as it is during a recession, the unemployment rate is typically above average. But what is the causality here? If unemployment caused recessions, we might be tempted to adopt a policy that makes unemployment illegal. For example, the government could fine firms if they lay off workers. This is not a good policy because we don't think that low unemployment causes high real GDP. But nor do we necessarily think that high real GDP causes low unemployment. Instead, based on economic theory, we think there are other influences that affect both real GDP and unemployment.

More Formally

Suppose that you have N observations of two variables, x and y, where x_i and y_i are the values of these variables in observation I = 1,2,...,N. The mean of x, μ_x, is the sum over the values of x in the sample divided by N, and likewise for y.

$$\mu_x = x_1 + x_2 + \ldots + x_N; \mu_y = y_1 + y_2 + \ldots + y_N / N.$$

We can also calculate the variance and standard deviations of x and y. The calculation for the variance of x, denoted σ_x^2 is

$$\sigma_x^2 = (x_1 - \mu_x) + (x_2 - \mu_x) + \ldots (x_N - \mu_x) / N$$

The standard deviation of x is the square root of σ_x^2,

$$\sigma_x = \sqrt{(x_1 - \mu_x)^2 + (x_2 - \mu_x)^2 + \ldots\ (x_N - \mu_x)^2 / N}$$

With these ingredients, the correlation of (x,y), denoted corr(x,y), is given by

$$\text{corr}(x,y) = (x_1 - \mu_x)(y_1 - \mu_y) + (x_2 - \mu_x)(y_2 - \mu_y) + \ldots (x_N - \mu_x)(y_N - \mu_y) / N\sigma_x\sigma_y.$$

The Main Uses of This Tool

- Chapter 12 "Superstars"