1

CHAPTER 1

Thinking Like an Economist

Scarcity, opportunity cost, and the marginal way of thinking that runs through every later chapter.

CHAPTER 1

What this chapter covers

1.1–1.3 Core ideas

- What economics and microeconomics are
- Scarcity & opportunity cost
- Thinking at the margin

1.4–1.5 The method

- Models & comparative statics
- Reading graphs
- Positive vs. normative

1.6–1.7 Evidence

- Case: the 2020 toilet-paper shortage
- Empirical: returns to college

CHAPTER 1

What this chapter covers

THE THREAD

Two habits of mind unify the whole book: weighing opportunity costs, and deciding at the margin.

1.1

What Are Economics and Microeconomics?

- Economics, a social science
- Micro vs. macro
- Three kinds of decision maker
- The three questions every economy answers

Everyday puzzles micro tries to answer

Prices differ

Why can a gallon of gasoline cost two dollars more in one state than in another?

Prices move

Why did a basic face mask cost a few cents in 2019 and several times as much a year later?

Everyday puzzles micro tries to answer

Budgets differ

Why does a family in India spend ~half its budget on food, a U.S. family about a tenth?

Policy choices

Why tax cigarettes heavily yet subsidize electric cars?

IN ONE SENTENCE

Microeconomics studies how individual decision makers choose when they cannot have everything, and how those choices meet in markets.

Economics is a social science

Economics

The social science that studies how people, firms, and governments use scarce resources (time, money, land, workers, machines) when they cannot have everything they want.

WHY "SOCIAL"

Its subject is people and their choices, so its tools reach far beyond prices: cramming for exams, saving, traffic jams, climate change, why some nations grow rich. Two branches: micro and macro.

Micro vs. macro

Microeconomics

The study of how individual consumers, firms, and governments make choices under scarcity, and how those choices interact in markets.

Macroeconomics

The study of the economy as a whole: total output, unemployment, inflation, and growth.

Micro vs. macro

WHY "MICRO"

"Micro" refers to how we build understanding: from the ground up, from the decision makers whose choices produce the prices and quantities we observe.

Three kinds of decision maker

C

Consumers

How people choose among the things they could buy, given what they can afford. Coffee on the way to class, hesitation over a $1,200 phone, or a front-row concert seat at double the price.

F

Firms

How producers turn inputs into output and set prices. How many stores to open, workers to hire, what to charge for a latte.

Three kinds of decision maker

G

Governments

How regulators predict responses before acting. Cap rents? Tax sugary drinks? Auction the airwaves?

Every economy must answer three questions

1

WHAT?

Which goods get produced: more housing or more highways, more food or more entertainment?

2

HOW?

By what method: many workers with simple tools, or few workers with advanced machines?

3

FOR WHOM?

Who receives the output, and on what basis?

Every economy must answer three questions

TWO WAYS TO ANSWER

In a centrally planned economy, officials decide. In a market economy, the questions are settled by the decentralized choices of millions of buyers and sellers, coordinated by prices.

The rationality assumption

Rational decision maker

Has a clear objective (more satisfaction for a consumer, higher profit for a firm) and, among the options available, chooses the one that serves it best.

Not selfish. Rationality does not mean people never care about others.

Not all-knowing. A rational shopper may still be unsure about a used car's quality (Chapter 16).

Bounded rationality. Limited time and attention lead people to use shortcuts and settle for "good enough" (Simon, 1955). Behavioral sections show when this matters.

1.2

Scarcity, Trade-offs, Opportunity Cost, and Trade

- The starting point: scarcity
- Opportunity cost defined
- The production possibilities frontier
- Comparative advantage and gains from trade

Everything starts with scarcity

Scarcity

We cannot have everything we want, because time, money, and resources are limited: a student has 24 hours, a family one income, a country only so much land, labor, and capital.

THE CONSEQUENCE

Scarcity forces trade-offs: giving up one thing in order to get another.

The most important cost in the book

- **Opportunity cost**

  The value of the next-best alternative you give up when you make a choice. Not just the money you spend: everything you sacrifice.

A HABIT TO BUILD

Whenever you catch yourself asking “what is this really costing me?” you are thinking about opportunity cost.

A free Saturday: one choice, three options

Option	Value to you	
Work a shift	$120	(chosen)
Attend the concert	$90	**next best alternative**
Study for the exam	$70	

Table 1.1 Opportunity cost is the value of the next best forgone option.

A free Saturday: one choice, three options

READ IT RIGHT

Cost of working is not $160 (concert + study). It is $90, the single best forgone option.

Net gain = $120 − $90 = $30.

The idea generalizes far beyond Saturday

College. Its true cost is not only tuition but the salary given up by not working full time.

Highways. A country's true cost of more highways is the schools or hospitals it could have built instead.

The idea generalizes far beyond Saturday

SELF-CHECK

Spend Saturday tutoring ($60), hiking (you value at $45), or sleeping in ($30). Opportunity cost of tutoring? Net gain?

Opportunity cost = $45 (the single best alternative, not $45 + $30). Net gain = $60 − $45 = $15.

An economy's trade-offs: computers vs. wheat

Point	Computers (1000s)	Wheat (1000s of tons)	Opportunity cost (wheat given up for 20,000 more computers)
A	0	100	—
B	20	90	$100 - 90 = 10$
C	40	70	$90 - 70 = 20$
D	60	40	$70 - 40 = 30$
E	80	0	$40 - 0 = 40$

Table 1.2 An economy's production possibilities for computers and wheat.

An economy's trade-offs: computers vs. wheat

THE PPF

A model of scarcity for a whole economy (also called the production possibilities curve): it can make wheat, computers, or any mix, but not unlimited amounts of both.

Four simplifying assumptions

Two goods. Computers and wheat; one good can stand for "everything else."

Fixed resources. Workers, machines, and land do not change during the period.

Fixed technology. The know-how for turning resources into goods does not change.

Full and efficient use. All resources employed, each where it is most productive.

WHAT MOVES THE ECONOMY

Relaxing fixed resources or fixed technology shifts the whole frontier; failing full use puts the economy inside it.

The frontier, plotted

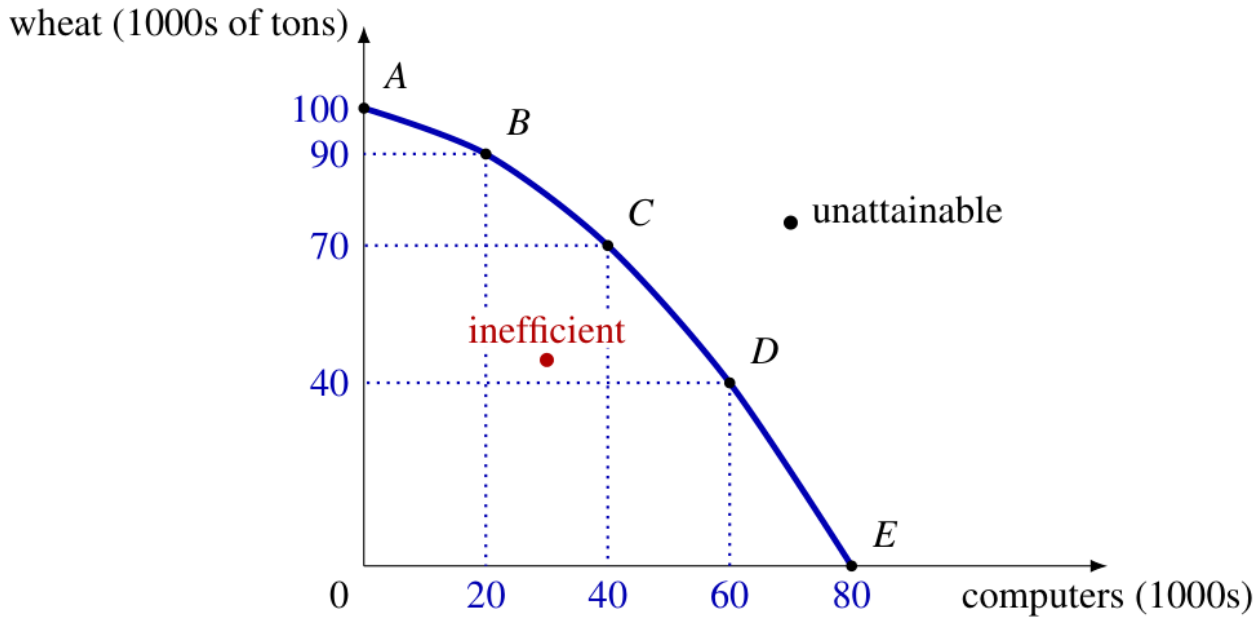


Figure 1.1 The production possibilities frontier.

Slopes down. More computers means less wheat.

Slope = opportunity cost. Of one good in terms of the other.

Points A to E. The rows of Table 1.2.

The frontier, plotted

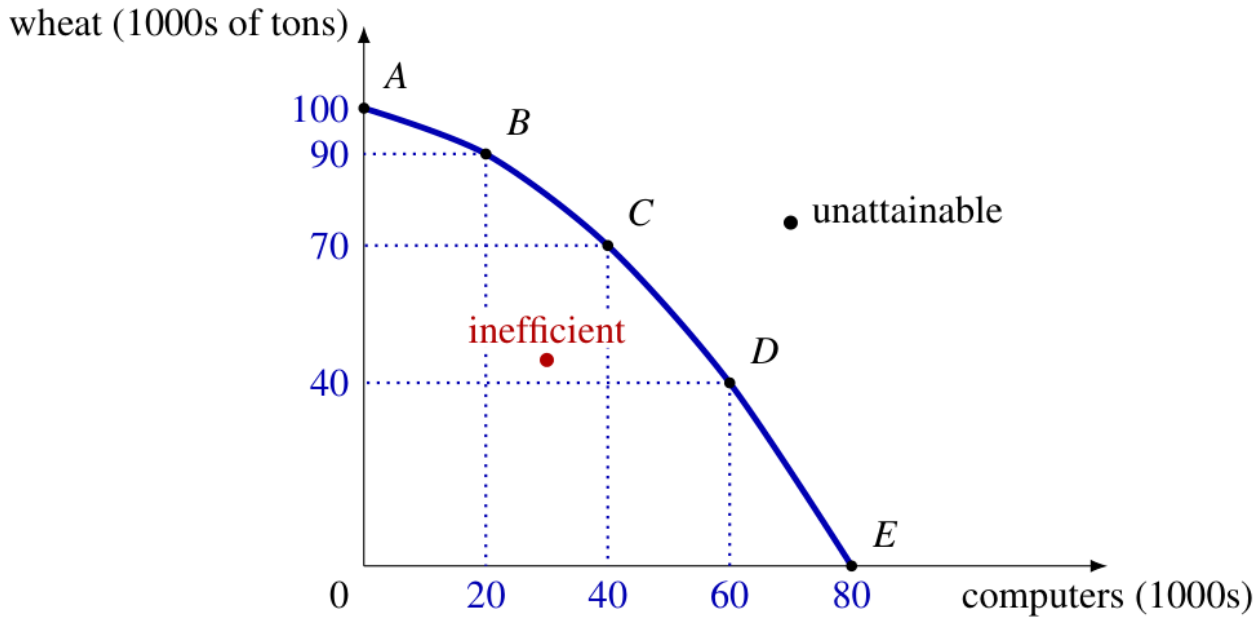


Figure 1.1 The production possibilities frontier.

EXPLORE FIGURE 1.1

Move a slider along an interactive PPF and watch the opportunity cost of one good rise as you produce more of it.

intro-micro-companion-website.netlify.app/ppf

Why the frontier bows outward

INCREASING OPPORTUNITY COST

From Table 1.2, the cost rises: the first 20,000 computers cost only 10,000 tons of wheat, but the last 20,000 cost 40,000.

- **The reason**

  Resources are not equally suited to both tasks. Pushing computer output higher pulls in workers and land ever less suited to computers and ever more valuable for growing wheat.

1.2.1 PRODUCTION POSSIBILITIES FRONTIER

Efficient, inefficient, unattainable, and growth

On the curve

Efficient: all resources used. More of one good only by giving up the other.

Inside the curve

Inefficient: resources wasted (e.g., unemployment). More of both is possible.

Outside the curve

Unattainable: beyond what current resources allow.

Efficient, inefficient, unattainable, and growth

GROWTH

Economic growth shifts the whole frontier outward, as the economy gains resources or better technology (Chapter 8). Growth is the only way to produce the formerly unattainable combinations.

Why trade can make everyone better off

	Bread (if all day on bread)	Fish (if all day on fish)	Opportunity cost
Anna	6	12	12/6 = 2 fish per loaf
Ben	4	4	4/4 = 1 fish per loaf

Table 1.3 Production possibilities of Anna and Ben.

Why trade can make everyone better off

THE PUZZLE

Anna is better at both goods, 6 loaves to Ben's 4, 12 fish to Ben's 4. She has an absolute advantage in each.

So why trade at all?

Look at opportunity cost, not absolute skill

Anna's cost of bread

6 loaves means giving up 12 fish → each loaf costs her 2 fish.

Ben's cost of bread

4 loaves means giving up 4 fish → each loaf costs him just 1 fish.

Comparative advantage

Ben gives up fewer fish per loaf, so he is the lower-cost bread producer, even though he is worse at making it. By the same logic, Anna has the comparative advantage in fish.

The gains from specializing and trading

Specialize. Each does the task with the lower opportunity cost: Ben makes bread, Anna makes fish.

Trade. Swap at a rate between the two opportunity costs, for instance one loaf for 1½ fish.

Both win. Each can end up with more fish, and no less bread, than each would have without trade.

The gains from specializing and trading

THE LESSON

Mutually beneficial trade rests not on who is better in absolute terms, but on comparative advantage: differences in opportunity cost. It is why countries trade with partners far richer or poorer than themselves.

Gains from trade, made numerical

	Autarky		Specialization and trade		
	Bread	Fish	Bread	Fish	Gain
Anna	3	6	3	7	$7 - 6 = 1$ fish
Ben	2	2	2	3	$3 - 2 = 1$ fish
Combined	5	8	5	10	$10 - 8 = 2$ fish

Table 1.4 Daily consumption under autarky and under specialization and trade.

Gains from trade, made numerical

GAINS FROM TRADE

Each ends up with the same bread as under autarky and one more fish.

That extra fish, for both at once, is the gain from trade.

Production, trade, consumption: three steps

Step 1, production. Ben bakes 4 loaves; Anna bakes 1 loaf and catches $12 - 2 = 10$ fish. Together: 5 loaves and 10 fish, against 5 and 8 under autarky.

Step 2, trade. Ben gives 2 loaves for 3 fish: 1.5 fish per loaf, between Ben's cost (1 fish) and Anna's (2 fish).

Step 3, consumption. Anna: 3 loaves + 7 fish. Ben: 2 loaves + 3 fish. Same bread, one more fish each.

WHY ANNA BAKES ONE LOAF

Fishing all day, she would end with 2 loaves and 9 fish: one loaf less than under autarky. One loaf of her own keeps her bread at 3.

The three steps in one table

	Production		Trade		Consumption	
	Bread	Fish	Bread	Fish	Bread	Fish
Anna	1	10	+2	−3	3	7
Ben	4	0	−2	+3	2	3
Combined	5	10	0	0	5	10

Table 1.5 Specialization and trade between Anna and Ben.

Explore Section 1.2.2: intro-micro-companion-website.netlify.app/comparative-advantage

The three steps in one table

THE 2 EXTRA FISH

Specialization: each person makes the good with the lower opportunity cost, so the pair produces 2 more fish.

Trade at 1.5 fish per loaf, between the two opportunity costs, splits the extra fish between them.

1.3

Thinking at the Margin

- Marginal benefit and marginal cost
- The marginal decision rule
- A worked bakery example
- Incentives matter

Good decisions are made at the margin

To decide how much to do, a rational decision maker asks whether one more unit is worthwhile: does its extra benefit exceed its extra cost?

Marginal benefit (MB)

The extra benefit from one more unit.

Marginal cost (MC)

The extra cost of one more unit.

The marginal decision rule

THE RULE

Keep doing more as long as the marginal benefit of the next unit is at least as large as its marginal cost; stop when they are equal.

This rule reappears, in different costumes, in almost every chapter.

A bakery decides how many cakes to bake

Cake number	Marginal benefit (price)	Marginal cost
1	$20	$8
2	$20	$12
3	$20	$16
4	$20	$20
5	$20	$25

Table 1.6 Marginal benefit and marginal cost for the bakery.

- Each cake sells for $20, so MB is flat at $20.
- MC rises as ovens and staff get busier.
- Cake 4 is the break-even case.

Bake until MC meets MB

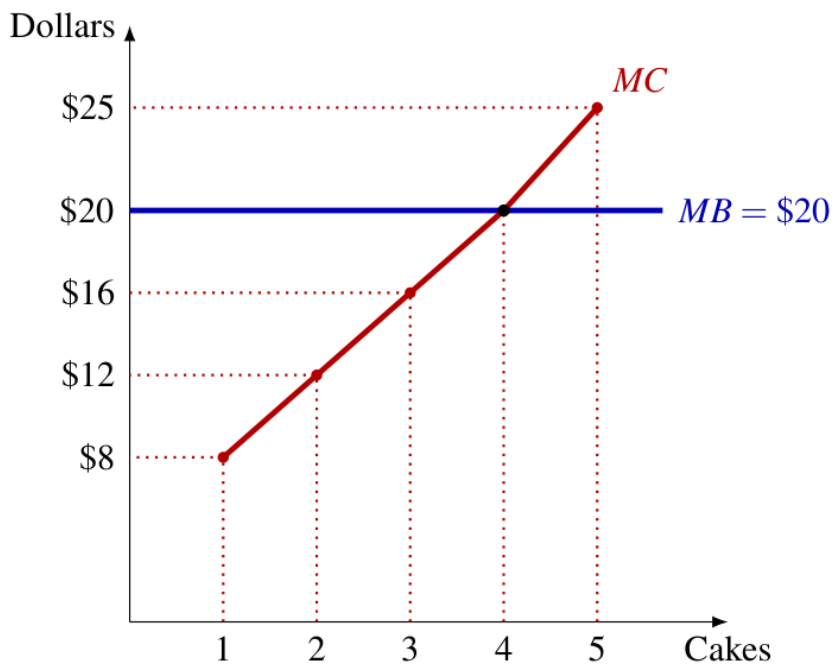


Figure 1.2 Marginal benefit and marginal cost for the bakery.

THINKING AT THE MARGIN

Bake cakes 1–3 (benefit exceeds cost). Cake 4 breaks even. Cake 5 would lose money ($25 > $20).

Answer: 4 cakes, and we never computed total profit.

Change the incentives, and behavior changes

- **Incentives**

  Rewards and penalties that change the marginal benefit or marginal cost of an action. Because people decide at the margin, their choices respond to incentives.

1.3.1 INCENTIVES MATTER

Change the incentives, and behavior changes

Costlier parking

Drivers circle less, take transit more.

Solar tax credit

More households install panels.

A store sale

Shoppers who would have waited buy now.

Incentives can surprise us

Safer cars. Seatbelts and airbags made crashes survivable; some drivers, feeling safer, drove faster, offsetting part of the benefit.

Plastic-bag tax. Shoppers respond by bringing their own.

Daycare late fine. Can increase lateness if parents treat the penalty as a price they will pay.

THE LESSON

To predict what a policy will do, ask how it changes the incentives facing the people it touches, then follow those incentives to their conclusion.

A closer look: the Israeli daycare study

Setup. Gneezy and Rustichini (2000) followed 10 Israeli daycare centers. At 6 of them, a small fine of NIS 10 (about $3) was introduced for parents arriving late to pick up their child.

What happened. Late pickups roughly doubled within a few weeks and stayed high. When the fine was withdrawn 12 weeks later, lateness did not revert to its pre-fine level.

The lesson. The fine converted an implicit moral obligation (do not inconvenience the staff) into an explicit price parents were willing to pay. A monetary incentive can crowd out a social norm, and removing the price does not restore it.

A closer look: the Israeli daycare study

WHY IT MATTERS

To predict a policy, ask not only what marginal cost or benefit it changes, but how it reframes the decision, from a moral obligation into a market transaction.

1.4

Models, Assumptions, and Comparative Statics

- What an economic model is
- Why assumptions are fine
- Comparative statics & ceteris paribus
- Reading graphs

A model is a deliberately simplified map

Model

A deliberately simplified description of reality that keeps only the features needed to answer a question.

THE SUBWAY-MAP ANALOGY

A subway map is a model. It distorts distances and leaves out streets, parks, and rivers, yet it is far more useful for crossing a city than a perfectly accurate satellite photo. By leaving out detail, models make the essential forces visible.

Why unrealistic assumptions are fine

■ When we assume a firm maximizes profit, or buyers prefer lower prices, we only claim that these statements hold often enough that a model built on them predicts well.

THE TEST OF A MODEL

A model is judged not by whether its assumptions are perfectly realistic, but by whether its predictions are borne out by evidence, which is why this book pairs every chapter's theory with real cases and data.

Comparative statics & ceteris paribus

Comparative statics

Change one thing, hold everything else fixed, and ask how the outcome changes. If the gas price rises, what happens to truck sales? If cigarette taxes rise, what happens to smoking?

CETERIS PARIBUS

Latin for “other things equal”: hold everything else fixed to isolate the effect of one cause. A higher gas price cuts truck purchases with incomes, other vehicle prices, and tastes held fixed.

First feature of a curve: its slope

Slope

How steeply one variable responds to the other.

Steep curve

Quantity barely responds to price.

Flat curve

Quantity responds a lot to price.

We make this precise in Chapter 3 (elasticity).

Second feature: movement along vs. shift of a curve

Movement along

Happens when the variable on the axis changes. Coffee's own price rises → we slide up the demand curve to a smaller quantity.

Shift of the curve

Happens when something not on the axis changes. Incomes rise → the entire curve moves to a new position.

Second feature: movement along vs. shift of a curve

WATCH OUT

Confusing the two is the single most common error students make. Always ask: what is on each axis, what does the slope mean, and does this event move us along the curve or shift it?

1.5

Positive versus Normative Analysis

- Statements that can be checked against evidence
- Statements that rely on values

Two kinds of statement

IS

Positive

Describes how the world is and can be checked against evidence. “A higher minimum wage reduces the number of low-wage jobs.”

⊕

Normative

Describes how the world ought to be and relies on values. “The government should raise the minimum wage.”

Two kinds of statement

BOTH MATTER

Most of this book is positive: models and evidence predict what will happen. But good policy needs both: economics can say how a rent cap affects apartment supply; whether a city should adopt it depends on values.

1.6

Case Study: The Great Toilet-Paper Shortage of 2020

Every idea in the chapter, at work in one episode.

1.6 CASE STUDY

Stripped shelves, a normally steady product

+845%

U.S. online and in-store sales, March 11 and 12, 2020

April

in early April, sales were still above their level before the pandemic

1.6 CASE STUDY

Stripped shelves, a normally steady product

- **Thinking at the margin, under a frightening but unlikely risk**

  Each family judged that one more package carried a large marginal benefit (avoiding the small but alarming chance of running out) against a marginal cost of a few dollars. Each choice was reasonable; together they produced a surge.

Why the shelves emptied

Opportunity cost. Low to each household but high for society: every extra package in a closet was one unavailable to a nurse or elderly neighbor.

Supply could not keep up. Mills already ran near full capacity around the clock, so production could not rise quickly to meet the extra demand.

A demand shock. The shortage came from a sudden rightward jump in demand colliding with a supply that could not move in the short run.

NEXT CHAPTER

We formalize this collision with supply and demand in Chapter 2.

1.7

Empirical Evidence: The Returns to a College Degree

Opportunity cost, measured.

The largest cost of college is invisible

OPPORTUNITY COST IN ACTION

Tuition is the visible price. For most students the largest cost is the earnings they give up by studying instead of working full time, usually more than half of the true cost of a four-year degree.

Does it pay off?

Two large World Bank studies, Montenegro and Patrinos (2014) and Psacharopoulos and Patrinos (2018), compare earnings across many countries.

The evidence

~9%

rise in earnings per extra year of schooling, worldwide average

>15%

typical return to a year of higher education

among highest

returns a young person can earn on any investment

The evidence

NET OF OPPORTUNITY COST

These returns remain positive even after subtracting the earnings forgone during the study years.

The opportunity cost varies sharply across countries

Free-tuition systems

In Germany, Norway, and Mexico's public universities, forgone earnings make up nearly the entire cost of attending.

High-tuition systems

A U.S. private university can exceed $50,000 a year, so tuition and forgone earnings are closer in size.

The opportunity cost varies sharply across countries

TAKEAWAY

The lesson is not that college is right for everyone, but that judging its value requires counting the opportunity cost, not just the tuition bill.

EXPLORE THE DATA

Compare returns to education, how much an extra year of schooling raises earnings, across countries.

ourworldindata.org/global-education

CHAPTER 1

Key takeaways

1. **Scarcity forces trade-offs**
   Limited time, money, and resources mean every choice has a cost.

2. **Opportunity cost**
   The value of the next-best alternative forgone, not just money spent.

3. **The PPF**
   Pictures efficiency, waste, the unattainable, and growth; its slope is opportunity cost.

4. **Comparative advantage**
   Gains from trade rely on differences in opportunity cost, not absolute skill.

CHAPTER 1

Key takeaways

5

Think at the margin

Compare one more unit's benefit with its cost; stop where MB = MC.

6

Incentives & method

Behavior follows incentives; models, ceteris paribus, and positive vs. normative guide analysis.

2

CHAPTER 2

Supply and Demand

The most widely used tool in all of economics, built one piece at a time, then put to work.

CHAPTER 2

What this chapter covers

2.2–2.3 The pieces

- Demand: buyers at each price
- Supply: sellers at each price
- What shifts each curve

2.4–2.5 The model

- Market equilibrium
- Shortages & surpluses
- Comparative statics: shifts

2.6–2.8 Evidence

- Case: surge pricing
- Empirical: coffee 2021–22
- Behavioral economics

CHAPTER 2

What this chapter covers

THE BIG IDEA

In a competitive market the price emerges from the interaction of all buyers and sellers. Our task is to understand that interaction.

2.1 Introduction

- Prices that move on their own
- What a market is
- Why we focus on competitive markets

Four prices that moved sharply, and no one decreed them

Coffee, 2020–21

The world price of Brazilian coffee roughly doubled in a year.

Surge pricing

A rainy Friday-night ride that normally costs $15 can cost two or three times as much.

Four prices that moved sharply, and no one decreed them

Bikes, 2020

The average price of a bicycle sold in the UK rose by about a quarter.

Oil, Apr 2020

U.S. crude briefly fell below zero: sellers paid buyers to take it.

WHAT THEY SHARE

In each case the price moved because the forces of demand and supply shifted, not because any government or company set the new figure.

What is a market?

Market

Any arrangement that brings buyers and sellers of a good or service into contact: a farmers' market, a stock exchange, or a phone app that matches riders with drivers.

Competitive market

One with so many buyers and sellers that no single one can move the price on its own. The price emerges from everyone's interaction.

2.2

Demand

- The demand schedule and curve
- The law of demand
- Market demand
- What shifts demand

The demand schedule for coffee

Price (per lb)	Quantity demanded (million lbs/week)
$2	100
$3	90
$4	80
$5	70
$6	60
$7	50

Table 2.1 A demand schedule for coffee.

The demand schedule for coffee

Demand

How much of a good buyers are willing and able to purchase at each possible price, holding everything else fixed.

Demand schedule

A table listing the quantity buyers would purchase at each price.

The law of demand

THE LAW OF DEMAND

Holding other things constant, a higher price reduces the quantity demanded, and a lower price raises it.

- When coffee becomes more expensive, some buyers brew at home, some switch to tea, and some simply drink less.
- When it becomes less expensive, they do the reverse.
- The pattern is so common that economists call it a law.

Why demand slopes down: two forces

Substitution effect

Coffee is now pricier relative to other goods, so buyers switch to tea or coffee brewed at home.

Income effect

The same budget buys less, so buyers are effectively poorer and cut back, coffee included.

Why demand slopes down: two forces

For most goods

Both forces reduce the quantity demanded when the price rises. Chapter 6 studies each of them, including the rare cases in which they point in different directions.

The demand curve slopes downward

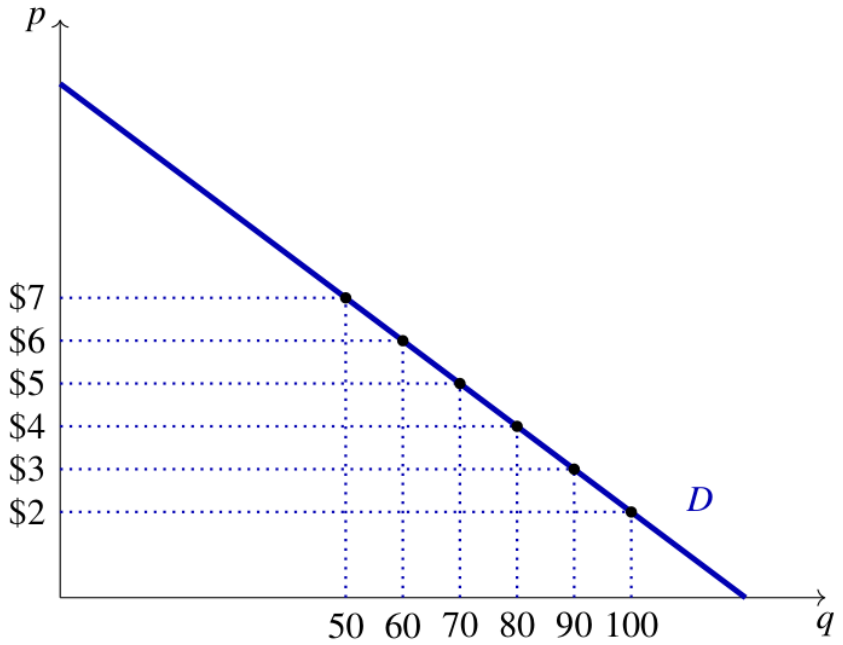


Figure 2.1 The demand curve for coffee, plotted from Table 2.1.

- Plot each price–quantity pair and connect the points.
- The curve slopes downward from left to right: the law of demand, drawn.

Market demand is the sum of individual demands

Market demand schedule

It adds up the amounts demanded by every individual buyer in the market.

HOW TO BUILD IT

To get the market quantity at $4, simply add the amount each buyer would purchase at $4. Market demand is the horizontal sum of individual demands.

Movement along vs. shift of the curve

Change in own price

Moves us along a fixed curve, from one row of the schedule to another.

Change in anything else

Shifts the whole curve: buyers want a different quantity at every price.

How to tell them apart

Only a change in the good's own price moves us along the curve; everything else shifts the curve. A rightward shift is an increase in demand; a leftward shift, a decrease.

Five shifters of demand (1–2)

1. Income

Normal goods: higher income raises demand (more restaurant meals, more coffee). Inferior goods: higher income lowers demand (fewer instant-noodle packets).

2. Prices of related goods

Substitutes: a rise in one's price raises demand for the other (coffee & tea). Complements: a rise in one's price lowers demand for the other (coffee & sugar).

Five shifters of demand (3–5)

3

3. Tastes

Anything that makes the good more appealing (a study praising coffee, a new café culture) raises demand.

4

4. Expectations

If buyers expect coffee to be costlier next month, some buy extra today, raising demand now.

Five shifters of demand (3–5)

5

5. Number and composition of buyers

More buyers mean greater demand at every price. Composition matters too: an aging population buys more health care and fewer school supplies.

What shifts demand: a summary

Factor	Increases demand (shifts the curve right)	Decreases demand (shifts the curve left)
Income	Higher income (normal good); lower income (inferior good)	Lower income (normal good); higher income (inferior good)
Prices of related goods	A higher price of a substitute; a lower price of a complement	A lower price of a substitute; a higher price of a complement
Tastes	The good becomes more appealing	The good becomes less appealing
Expectations	Buyers expect a higher price in the future	Buyers expect a lower price in the future
Number and composition of buyers	More buyers, or more of the buyers who want the good	Fewer buyers, or fewer of the buyers who want the good

What shifts demand: a summary.

2.3

Supply

- The supply schedule and curve
- The law of supply
- What shifts supply

The supply schedule for coffee

Price (per lb)	Quantity supplied (million lbs/week)
$2	40
$3	50
$4	60
$5	70
$6	80
$7	90

Table 2.2 A supply schedule for coffee.

The supply schedule for coffee

Supply

How much of a good sellers are willing and able to offer for sale at each possible price.

WHY IT SLOPES UP

A higher price makes it worthwhile to harvest more carefully, bring marginal land into production, and draws new sellers in.

The law of supply

THE LAW OF SUPPLY

Holding other things constant, a higher price raises the quantity sellers want to sell.

Plotting the pairs gives the upward-sloping supply curve.

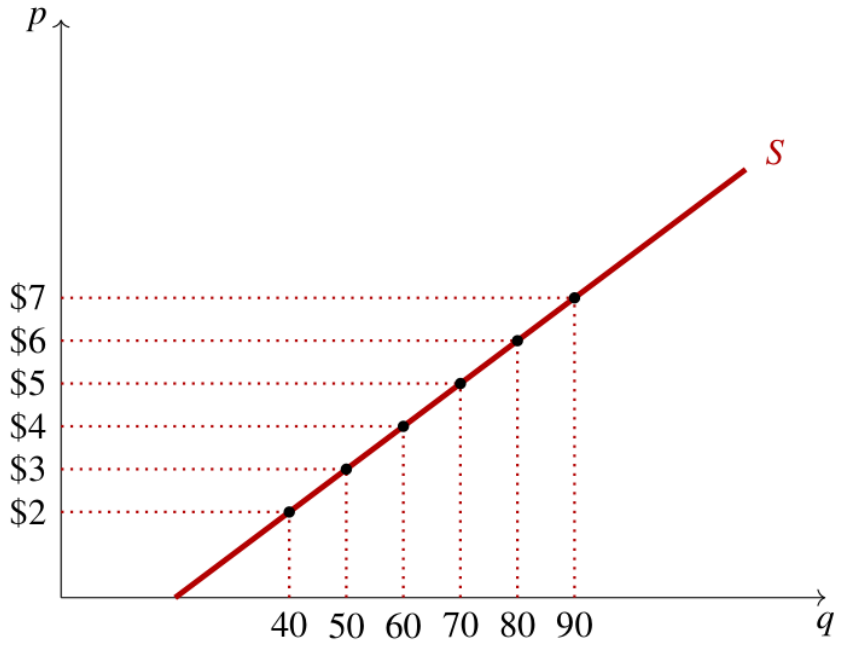


Figure 2.2 The supply curve for coffee, plotted from Table 2.2.

Six shifters of supply (1–4)

1

1. Input prices

Less expensive fertilizer, fuel, or labor increases supply; costlier inputs decrease it.

2

2. Technology

A better drying or roasting method that lowers cost per pound increases supply.

Six shifters of supply (1–4)

3

3. Prices of related goods

Land that grows coffee can grow cocoa; a higher cocoa price pulls land away and decreases coffee supply.

4

4. Expectations

Growers expecting much higher prices next season store part of the harvest, reducing supply today.

Six shifters of supply (5–6)

5

5. Natural conditions

Weather, pests, and disease: a frost or drought reduces supply; a good season increases it.

6

6. Number of sellers

More growers or roasters raise supply at every price.

And one more: taxes and subsidies

A PREVIEW OF CHAPTER 4

Taxes and subsidies also shift supply, since they change the net price sellers receive.

We study their effects in detail in Chapter 4.

What shifts supply: a summary

Factor	Increases supply (shifts the curve right)	Decreases supply (shifts the curve left)
Input prices	Inputs become less expensive	Inputs become more expensive
Technology	A better method lowers the cost of producing each unit	Rarely, a change that raises the cost of producing each unit
Prices of related goods	A lower price of another good the same land and labor can produce	A higher price of that other good (cocoa for coffee growers)
Expectations	Sellers expect a lower price in the future	Sellers expect a higher price in the future
Natural conditions	An unusually good growing season	A frost, drought, pests, or plant disease
Number of sellers	More sellers	Fewer sellers
Taxes and subsidies (Chapter 4)	A subsidy	A tax

What shifts supply: a summary.

2.4

Market Equilibrium

- Bringing buyers and sellers together
- Shortages and surpluses
- Price as signal and rationing device

Where the two schedules meet

Price	Quantity demanded	Quantity supplied	Pressure on price
$2	100	40	shortage of 100 − 40 = 60, price rises
$3	90	50	shortage of 90 − 50 = 40, price rises
$4	80	60	shortage of 80 − 60 = 20, price rises
$5	70	70	**no shortage or surplus (70 − 70 = 0), equilibrium**
$6	60	80	surplus of 80 − 60 = 20, price falls
$7	50	90	surplus of 90 − 50 = 40, price falls

Table 2.3 Finding the equilibrium price for coffee.

Supply meets demand at $5 and 70 million lbs

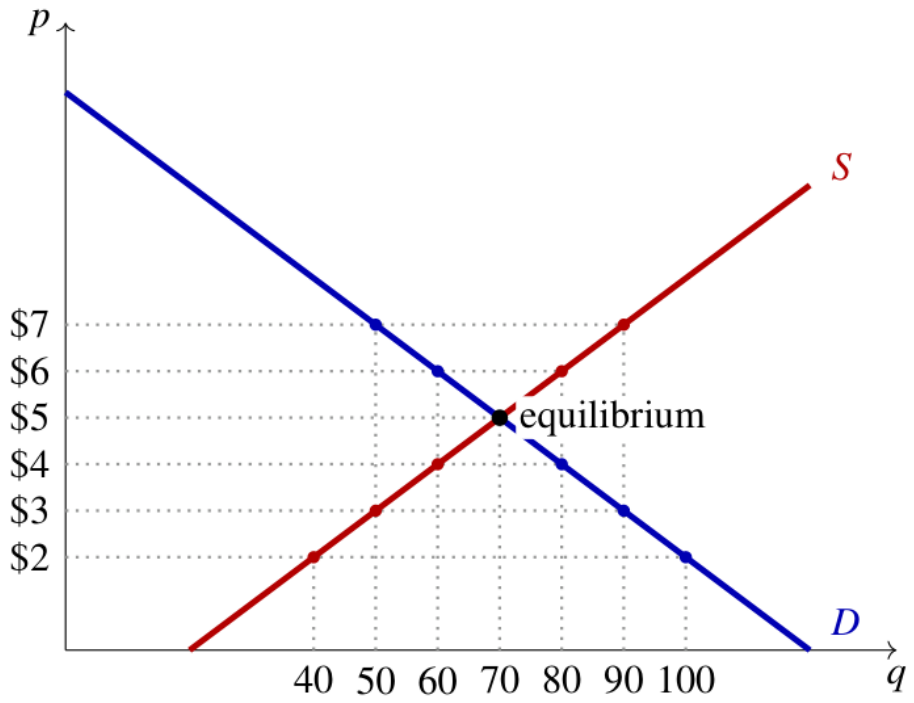


Figure 2.3 Market equilibrium for coffee, where supply meets demand.

Equilibrium

At $5, quantity demanded (70) equals quantity supplied (70). The market clears: every buyer willing to pay $5 finds coffee, every seller willing to sell at $5 finds a buyer.

Explore Figure 2.3: intro-micro-companion-website.netlify.app/supply-demand

Too low a price: a shortage

- **Shortage, price below $5 (for instance, $4)**

  Buyers want 80 million lbs, sellers offer only 60. A shortage of 20 million: more buyers want coffee than there is to go around.

WHAT HAPPENS NEXT

Frustrated buyers offer to pay more; sellers, seeing eager customers, raise prices. The price rises toward $5.

Too high a price: a surplus

- **Surplus, price above $5 (for instance, $6)**

  Sellers offer 80 million lbs, buyers want only 60. A surplus of 20 million: unsold coffee piles up.

THE SELF-CORRECTING MARKET

Sellers cut prices to move inventory; the price falls toward $5. From either direction the price is pushed to equilibrium, the only price with no pressure to change. That self-correction makes equilibrium a useful prediction.

A price does two jobs at once (1)

- **Rationing**

  When coffee is scarce relative to demand, the rising price gently pushes the least eager buyers out, so the limited supply goes to those willing to pay the most. No official decides who goes without; the price quietly settles it.

First job: it allocates the scarce good among buyers.

A price does two jobs at once (1)

WHEN PRICE CANNOT RATION

When the price is not allowed to ration, some other rule must: rent-controlled apartments in New York and San Francisco are rationed by waiting lists and personal connections, university dorm rooms by lottery, and wartime or Soviet-era systems used paper coupons for bread, gasoline, and meat.

A price does two jobs at once (2)

- **Signal**

  A high price tells producers that buyers value the good and rewards making more; a low price says cut back. A shortage that bids the price up says “make more of this”; a surplus says “make less.”

A price does two jobs at once (2)

THE INVISIBLE HAND

No one sends these messages deliberately. Economists call this coordination the “invisible hand,” after Adam Smith, who used the phrase once in The Wealth of Nations (1776), about a merchant who invests at home rather than abroad. We return to it, and where it fails, in Chapters 10 and 18.

2.5

Changes in Equilibrium: Shifts in Demand and Supply

- A supply shift: the Brazil frost
- A demand shift, worked out
- When both curves shift, two dominance cases
- GLP-1 drugs; the same model, many markets
- A second worked case: oil in 2020

A supply shock: the 2021 Brazil frost

Price	Quantity demanded	Old quantity supplied	New quantity supplied	Pressure on price
$4	80	60	40	shortage of 80 − 40 = 40, price rises
$5	70	70	50	shortage of 70 − 50 = 20, price rises
$6	60	80	60	**no shortage or surplus (60 − 60 = 0), new equilibrium**
$7	50	90	70	surplus of 70 − 50 = 20, price falls

Table 2.4 A leftward shift in supply raises the equilibrium price.

A supply shock: the 2021 Brazil frost

THE SHOCK

A severe frost in Brazil, the world's largest producer, destroyed part of the harvest. A decrease in supply: at every price, sellers can offer ~20 million lbs less. Demand is unchanged.

The frost shifts supply left: price up, quantity down

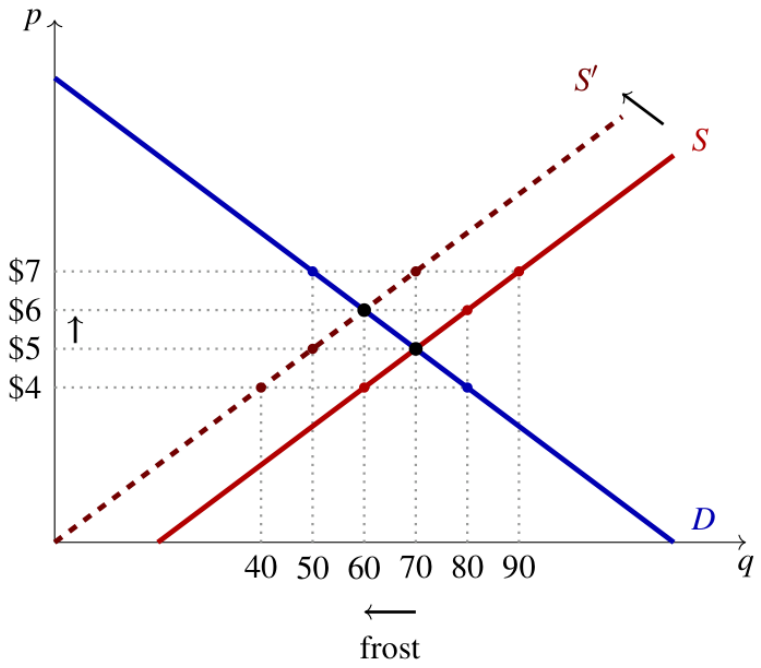


Figure 2.4 A decrease in supply: the frost.

$5 → $6

equilibrium price rises

70 → 60

quantity traded falls

The frost shifts supply left: price up, quantity down

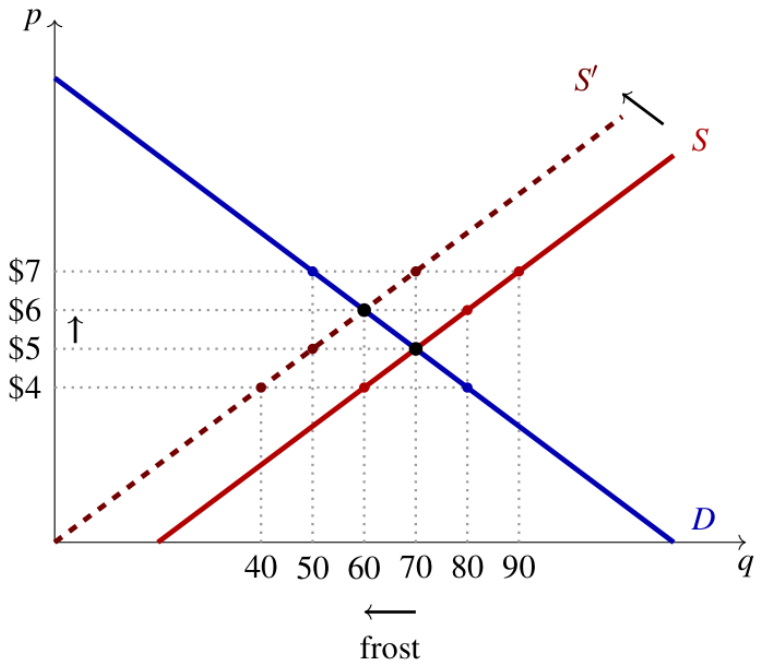


Figure 2.4 A decrease in supply: the frost.

INTUITIVELY

A smaller harvest makes coffee scarcer; buyers compete for the reduced supply, bidding the price up until enough drop out to match what is available.

How equilibrium responds to a single shift

Shift	Equilibrium price	Equilibrium quantity
Demand increases (shifts right)	rises	rises
Demand decreases (shifts left)	falls	falls
Supply increases (shifts right)	falls	rises
Supply decreases (shifts left)	rises	falls

Table 2.5 How the equilibrium responds to a single shift.

A three-step habit for any “what happens?” question

1

Step 1

Does the event shift demand, supply, or neither?

2

Step 2

In which direction does the curve shift?

3

Step 3

Find the new equilibrium in the schedule.

A three-step habit for any “what happens?” question

SELF-CHECK

A new process lowers the cost of making each flat-screen TV. Predict price and quantity.
(1) It shifts supply: it changes the cost of producing, not buyers’ willingness to pay. (2) Supply shifts right. (3) Price falls, quantity rises.

A health study raises demand by 20 million

Price	Old qty demanded	New qty demanded	Qty supplied
$4	80	100	60
$5	70	90	70
$6	60	80	80
$7	50	70	90

Table 2.6 A health study shifts coffee demand right by 20 million pounds.

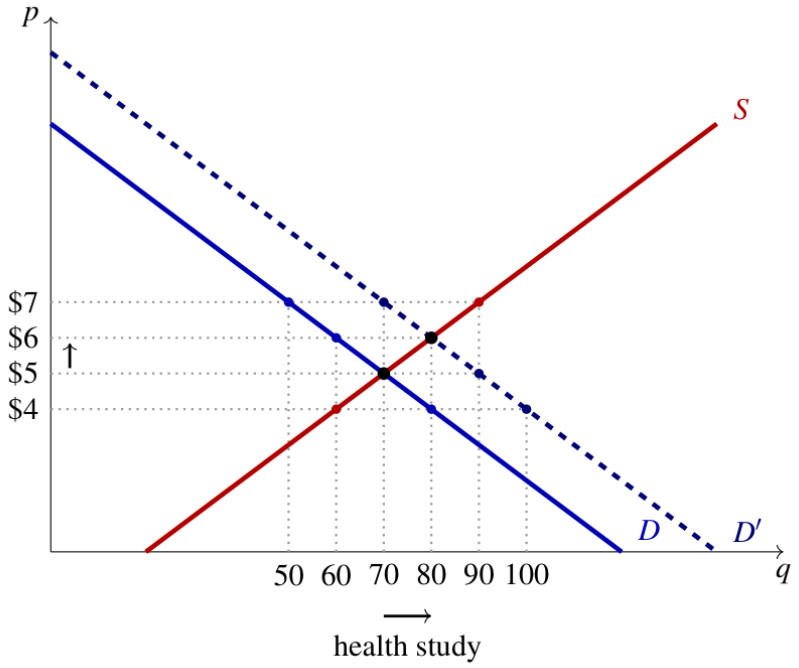
A health study raises demand by 20 million

FIND THE NEW EQUILIBRIUM

New equilibrium where the new quantity demanded = quantity supplied: at $6, both are 80 million.

Price: $5 → $6. Quantity: 70 → 80. Both rise, as Table 2.5 predicts.

The health study shifts demand right



$5 → $6

equilibrium price rises

70 → 80

equilibrium quantity rises

Figure 2.5 An increase in demand: the health study.

The health study shifts demand right

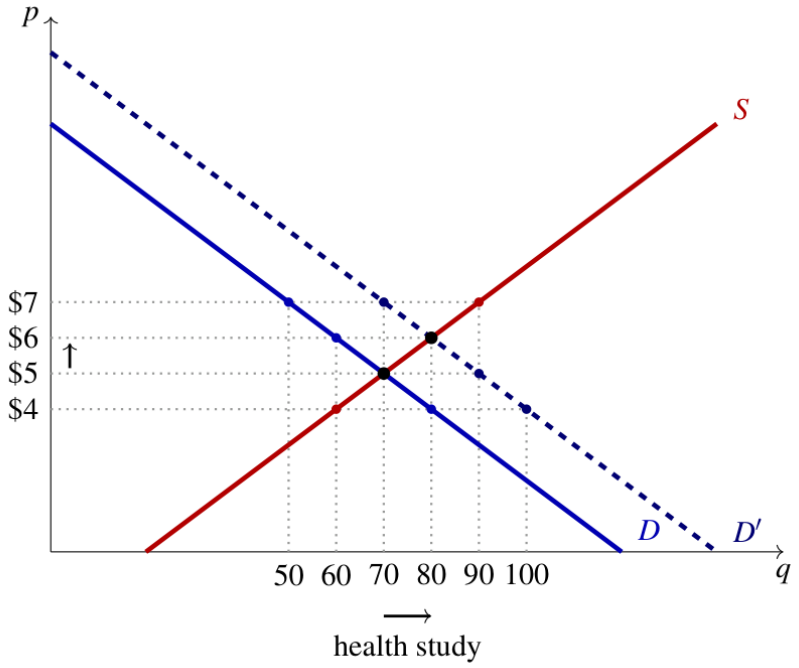


Figure 2.5 An increase in demand: the health study.

INTUITIVELY

The favorable news makes coffee more desirable; buyers compete for it and bid the price up, and the higher price brings out the extra coffee they now want.

Health study and frost in the same season

Price	Old qty demanded	New qty demanded	Old qty supplied	New qty supplied
$5	70	90	70	50
$6	60	80	80	60
$7	50	70	90	70

Table 2.7 Demand rises and supply falls at the same time.

Health study and frost in the same season

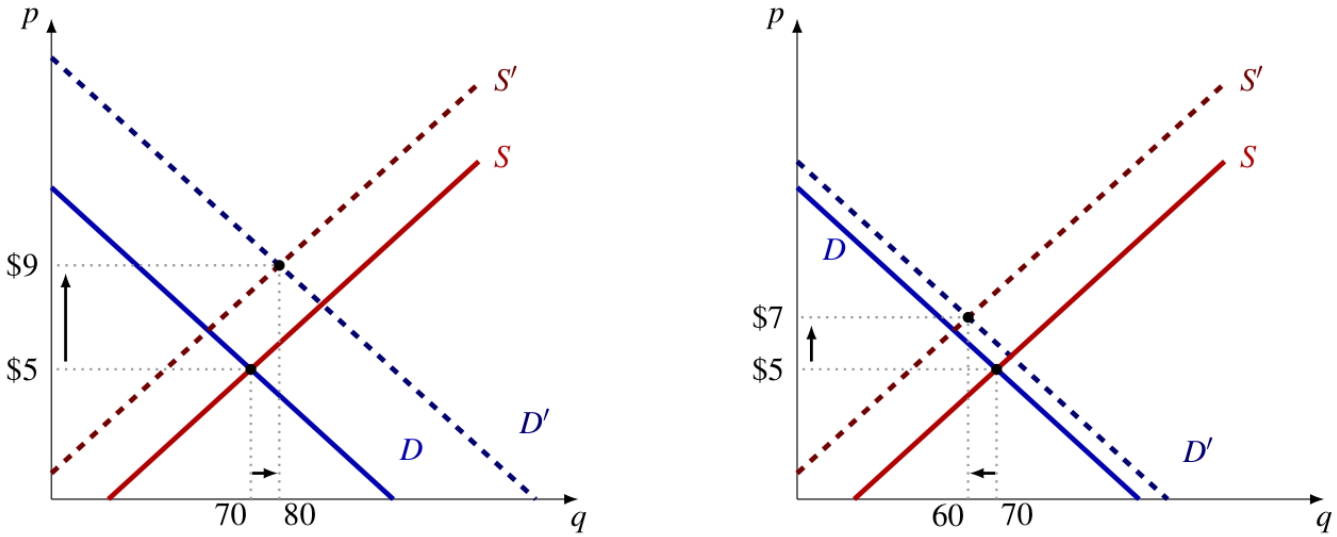
THE OUTCOME

Old equilibrium: $5, 70 million (columns 2 and 4).

New equilibrium: $7, with both new quantities at 70 million.

Price rises clearly ($5 → $7); quantity is unchanged (70 → 70).

Two dominance cases: which shift is larger?



Case 1: demand shift dominates

Case 2: supply shift dominates

Figure 2.6 Which shift dominates determines whether quantity increases or decreases.

Two dominance cases: which shift is larger?

Case 1, demand dominates

Supply −30, demand +50: price and quantity rise, to $9 and 80 million.

Case 2, supply dominates

Same supply −30, demand +10: price rises, quantity falls, to $7 and 60 million.

Explore Figure 2.6: intro-micro-companion-website.netlify.app/comparative-statics

Sign what you can; the rest depends on which shift is larger

THE GENERAL LESSON

When demand rises and supply falls, the price must rise: both forces push it up. The quantity moves in the direction of the dominant shift: up when the demand shift dominates, down when the supply shift dominates.

- In the equal-shift case (Table 2.7), the two forces on quantity cancel and quantity is unchanged.
- Whenever both curves move, one change has a definite direction and the other depends on which shift is larger; when both curves shift to the right, the quantity must rise.

A huge demand shift meets a vertical supply curve

- GLP-1 drugs, first approved for type-2 diabetes, sharply reduce appetite. Approval for weight loss (Wegovy 2021, Zepbound 2023) shifted demand far right.
- Injectable biologics need purpose-built plants that take years to license: in the short run, supply was nearly vertical.
- The shift pushed toward a much higher price, but manufacturers kept list prices near $900 to $1,350 a month before insurance in the U.S., so the excess demand showed up as shortages and waiting lists (Section 2.4.1).

2.5.3 A RECENT CASE: THE GLP-1 BOOM

A huge demand shift meets a vertical supply curve

~$50 bn

worldwide sales of the drug class by 2024

$11 bn

Novo Nordisk’s 2024 purchase of three Catalent fill-and-finish plants

A huge demand shift meets a vertical supply curve

THE MECHANISM

A large demand shock against inelastic supply pushes toward a much higher price; with list prices held, the excess demand shows up as shortages until new capacity restores balance.

One model, with the axes relabeled

Rental housing

Price = monthly rent. New residents shift demand right, raising rents until construction catches up.

Labor

Price = the wage. A shortage of nurses bids their wages up.

One model, with the axes relabeled

Foreign currency

Price = the exchange rate.

Crude oil

A war in a producing region shifts supply left and raises the world price, as in 2022 after Russia’s invasion of Ukraine.

One model, with the axes relabeled

WHY THE MODEL MATTERS

Master the model once with coffee, and you hold a tool for reading the front page of business newspapers and understanding many economic situations.

The oil market in 2020: two shocks at once

- Demand fell: lockdowns grounded flights and emptied roads (IEA: April demand ~29 million barrels a day below a year earlier).
- Supply rose: OPEC+ talks on output cuts collapsed on March 6, 2020; Saudi Arabia cut prices and pledged ~12 million barrels a day for April.
- On April 12, OPEC+ agreed to cut 9.7 million barrels a day from May 1, helping prices recover.

$64→$18

Brent, average price per barrel, Jan. vs. Apr. 2020 (EIA)

2.5.5 A SECOND WORKED CASE: OIL IN 2020

The oil market in 2020: two shocks at once

−$37.63

WTI May futures settlement, April 20, 2020, as storage ran out

THE MECHANISM

Demand fell and supply rose, so the price had to fall. Which effect dominates the quantity? The demand collapse was far larger, so the quantity traded fell.

Both shifts push the price down

Price	Old qty demanded	New qty demanded	Old qty supplied	New qty supplied
$20	116	96	84	88
$30	112	92	88	92
$40	108	88	92	96
$50	104	84	96	100
$60	100	80	100	104

Table 2.8 Stylized demand and supply schedules for crude oil in 2020 (million barrels per day).

Both shifts push the price down

THE OUTCOME

Old equilibrium: $60, 100 million barrels a day. New: $30, 92 million.

The price must fall (both shifts push it down); the quantity falls because the demand shift is far larger.

2.6

Case Study: Surge Pricing at Uber and Lyft

The model of Section 2.4, operating minute by minute.

2.6 CASE STUDY

Surge pricing is the equilibrium model in real time

- On an ordinary afternoon, riders requesting trips roughly match drivers on the road at the standard fare; the market clears.
- At 12:30 a.m. on New Year's Day, or the moment a rainstorm begins, requests jump: a sharp rightward shift in demand.
- At the old fare, riders far exceed drivers: a shortage. The app raises the fare, sometimes to two or three times normal.

2.6 CASE STUDY

The higher fare does two things, both predicted

Demand side

Riders who value the trip least (those who could wait, walk, or share) drop out, rationing cars to those who value them most.

Supply side

Driving during the surge is more lucrative, so more drivers come onto the road; quantity supplied rises.

The higher fare does two things, both predicted

EVIDENCE

Hall et al. (2015), researchers at Uber and the University of Chicago, studied demand spikes (a sold-out concert, New Year's Eve): letting the fare surge sharply reduced the number of riders who could not get a car and shortened wait times, versus a glitch that held the fare fixed. (Why riders still call it unfair: Section 2.8.1.)

2.7

Empirical Evidence: The 2021–22 Global Coffee Price Spike

A clean, large-scale example of a supply shift moving a price.

A supply contraction, and the price it produced

- Brazil grows ~a third of the world's coffee.
- 2021: its worst drought in nearly a century, then a July frost damaged trees across main regions.
- Both reduced the harvest, a leftward shift in supply, as in Table 2.4.

+44%

ICO composite price, Oct 2020 – Jul 2021 (highest since 2014)

A supply contraction, and the price it produced

~2×

price of Brazilian Naturals by Oct 2021

THE MECHANISM

Demand did not surge in 2021; the supply contraction moved the price. The quantity traded fell only slightly: world coffee exports dipped about 1% in 2021/22 (International Coffee Organization, 2022).

2.8

A Look into Behavioral Economics

- 2.8.1 Fairness and price gouging
- 2.8.2 Do markets find the equilibrium?

Why firms often refuse to raise prices

A WELL-KNOWN SURVEY

Kahneman et al. (1986): a hardware store sells snow shovels for $15; the morning after a storm it raises the price to $20. Though demand clearly rose, 82% of respondents called the increase “unfair.”

- People hold a norm that sellers should not exploit a sudden hardship, even when the model says the price “should” rise.
- Firms that violate it risk losing customer loyalty; many keep prices fixed in emergencies, and some governments pass price-gouging laws.
- The same tension surrounds surge pricing: efficient, yet often seen as unfair.

2.8.2

Do Markets Find the Equilibrium?

Evidence from market experiments.

Testing a bold claim in the laboratory

THE CLAIM

The model claims a market converges to the price where supply meets demand, even though no single trader knows the overall supply or demand curve.

Testing a bold claim in the laboratory

Chamberlin (1948)

Students struck private bilateral deals around a room; prices did not settle cleanly on the predicted equilibrium.

Smith (1962)

Replaced haggling with an oral double auction (open bids and asks, repeated periods). Prices converged rapidly to equilibrium.

Prices converge, period by period

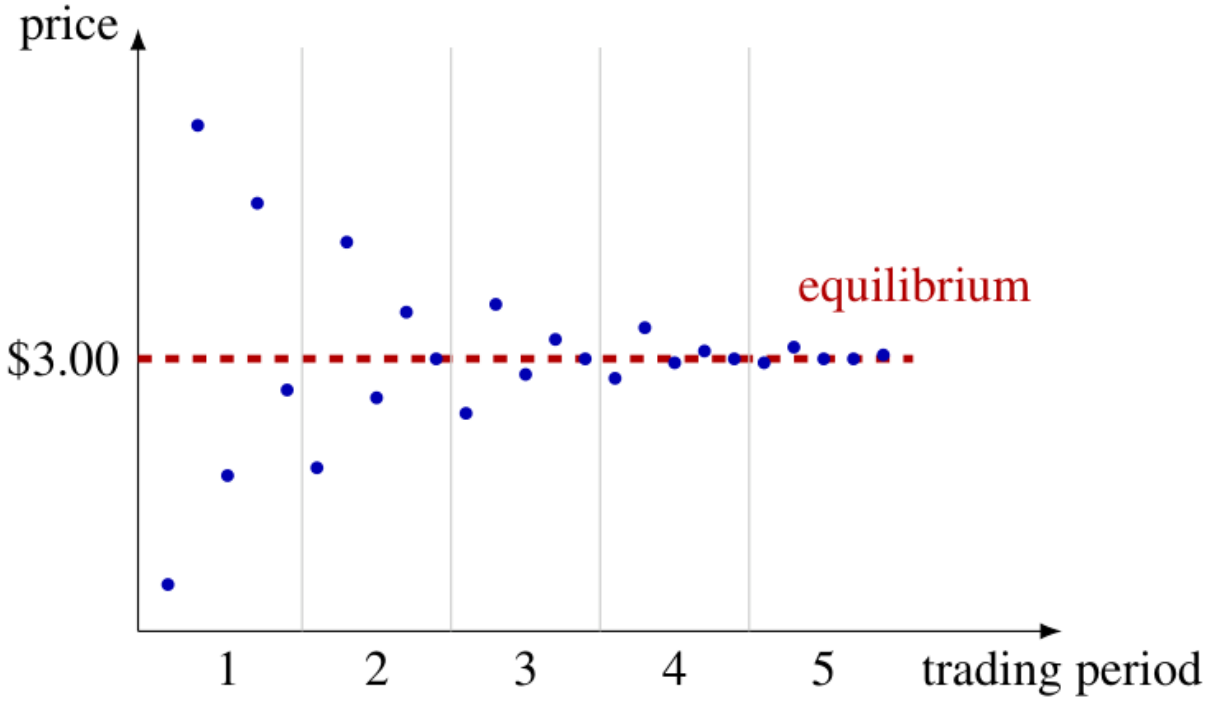


Figure 2.7 Stylized transaction prices in a double auction experiment.

The invisible hand is real. Self-interested traders, each knowing only their own valuation, find the market-clearing price.

How they trade matters. The double auction converged where free-form bargaining did not.

Nobel 2002. Smith's experiments earned the prize.

CHAPTER 2

Key takeaways

1 **Demand & supply**

Demand slopes down (law of demand); supply slopes up (law of supply).

2 **Along vs. shift**

Own price moves us along a curve; everything else shifts it.

3 **Equilibrium**

Where quantity demanded = quantity supplied; the market clears and self-corrects.

4 **Comparative statics**

Table 2.5 signs the four single-shift cases; both-shift cases leave one variable ambiguous.

CHAPTER 2

Key takeaways

5

Price's two jobs

It rations the good and signals producers: Adam Smith's invisible hand.

6

Evidence

Coffee 2021–22, surge pricing, fairness norms, and Smith's converging double auctions.

3

CHAPTER 3

Elasticity and Its Applications

One number that measures how much quantity responds to price, and turns the law of demand into a practical tool.

CHAPTER 3

What this chapter covers

3.2–3.3 The core idea

- Price elasticity of demand
- Elastic vs. inelastic
- Elasticity and total revenue

3.4–3.5 More elasticities

- Income & cross-price elasticity
- Price elasticity of supply
- Corrective (sin) taxes

3.6–3.7 Evidence

- Case: cigarette taxes
- Empirical: gasoline demand
- Short run vs. long run

CHAPTER 3

What this chapter covers

THE BIG IDEA

Chapter 2 gave the direction of the response; this chapter measures its size. How much quantity responds to a change in price is what almost every practical decision turns on.

3.1 Introduction

- Gasoline, cigarettes, and a streaming price hike
- Direction vs. size of the response
- Why size is what matters in practice

Three questions, one underlying issue

Gasoline

When the price rises 50 cents a gallon, do drivers cut back sharply, or hardly at all?

Cigarettes

Why can a tax on cigarettes raise large sums, but a similar tax on a luxury cruise raise far less?

Streaming

If a service raises its monthly price by $2, will it collect more revenue or less?

Three questions, one underlying issue

WHAT THEY SHARE

Each is really the same question: how much does the quantity people buy respond to a change in price? Economists answer it with a number called elasticity.

3.2

Price Elasticity of Demand

- Measuring responsiveness
- The midpoint method
- Elastic, inelastic, unit elastic
- What makes demand elastic

What price elasticity of demand measures

Price elasticity of demand

The percentage change in quantity demanded divided by the percentage change in price. It measures how much quantity demanded responds to a change in price.

What price elasticity of demand measures

THE FORMULA

ε = (% change in quantity demanded) ÷ (% change in price).

Because a higher price lowers quantity, the ratio is negative for almost every good (Section 6.6 looks at the rare exceptions). By convention we drop the minus sign and report its size (absolute value). A larger number means more responsive, or more elastic, demand.

The midpoint method, worked on gasoline

- Price rises from $3.00 to $3.60; quantity falls from 100,000 to 96,000 gallons/week.
- Divide each change by the average of start and end, so the answer is the same whether price rises or falls.
- Price: +$0.60 on an average of $3.30: 0.60 ÷ 3.30 ≈ 0.18, about 18%.
- Quantity: −4,000 on an average of 98,000: 4,000 ÷ 98,000 ≈ 0.04, about 4%.

≈ 0.22

ε = 4 ÷ 18

The midpoint method, worked on gasoline

WHAT IT TELLS US

Gasoline buyers barely cut back: for every 1% increase in price, quantity demanded falls by only 0.22%, less than proportionally.

Why divide by the average?

- Dividing by the starting value gives two answers for the same $0.60 change.
- A rise from $3.00: 0.60 ÷ 3.00 = 0.20, or 20%.
- A fall from $3.60: 0.60 ÷ 3.60 ≈ 0.17, or about 17%.
- Dividing by the average, $3.30, gives 0.60 ÷ 3.30 ≈ 0.18 either way.

≈ 18%

rise or fall: 0.60 ÷ 3.30

Why divide by the average?

THE MIDPOINT METHOD

Dividing each change by the average of the starting and ending values makes the elasticity the same whether the price rises or falls.

The value of ε sorts goods into three cases

Elasticity ε	Name	Meaning
Greater than 1	Elastic	Quantity responds more than price
Less than 1	Inelastic	Quantity responds less than price
Equal to 1	Unit elastic	Quantity responds in equal proportion

Table 3.1 Three cases of price elasticity of demand.

Through one point, steeper means less elastic

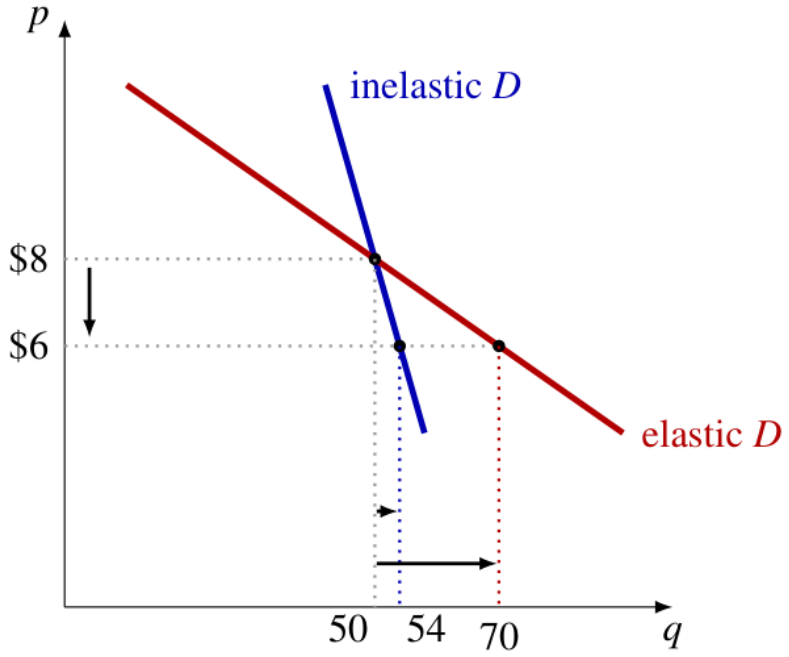


Figure 3.1 A price fall along an inelastic (steep) and an elastic (flat) demand curve.

- From \$8 to \$6, quantity rises from 50 to 54 on the steep curve ($\varepsilon \approx 0.27$) but from 50 to 70 on the flat one ($\varepsilon \approx 1.17$).
- A single brand of bottled water beside a dozen rivals is highly elastic: a small price rise sends nearly all buyers away.

The two elasticities in Figure 3.1, step by step

- Price: $8 → $6, a change of $2 from an average of $7: 2 ÷ 7 ≈ 0.286, or 28.6%, on both curves.
- Steep curve: quantity 50 → 54, a change of 4 from an average of 52: 4 ÷ 52 ≈ 0.077, or 7.7%.
- Flat curve: quantity 50 → 70, a change of 20 from an average of 60: 20 ÷ 60 ≈ 0.333, or 33.3%.

0.27 vs. 1.17

ε = 7.7 ÷ 28.6 (steep) and 33.3 ÷ 28.6 (flat)

The two elasticities in Figure 3.1, step by step

THE SAME PRICE CHANGE

Through the same point, the steep curve is inelastic and the flat curve is elastic.

Four forces make demand more elastic (1–2)

1. Close substitutes

The more easily buyers can switch to another good, the more elastic demand is. Demand for one brand of cola is far more elastic than demand for cola as a whole.

2. Necessities versus luxuries

Necessities (insulin, electricity) have inelastic demand; luxuries (cruises, second homes) have elastic demand, since buyers can more easily do without them.

Four forces make demand more elastic (3–4)

3. Share of the budget

A good that takes a large share of spending (rent) has more elastic demand than one that takes a tiny share (salt), because buyers watch its price more closely.

4. Time horizon

Demand is more elastic over longer periods. When gas prices rise, drivers can do little this week, but over years they buy efficient cars, move closer to work, or switch to transit.

Representative estimates, inelastic to elastic

Good	Approx. ε of demand
Salt	0.1
Gasoline (short run)	0.25
Eggs	0.3
Coffee	0.3
Cigarettes	0.4
Gasoline (long run)	0.6
Automobiles (short run)	1.2
Restaurant meals	2.3
Foreign travel (long run)	4.0

Table 3.2 Representative estimates of price elasticity of demand.

Sources: Espey (1998) for gasoline; Chaloupka and Warner (2000) for cigarettes; Andreyeva et al. (2010) for eggs; Gwartney and Stroup (1997) for the other goods.

Representative estimates, inelastic to elastic

READING THE TABLE

Table 3.2 illustrates the four determinants of Section 3.2.3.

Salt and eggs: inexpensive necessities with few close substitutes and a tiny budget share; quantity barely responds.

Automobiles, meals, foreign travel: costly luxuries with many substitutes; buyers respond sharply.

Eggs and bird flu: inelastic demand at work

- Bird flu: between mid-October 2024 and early March 2025, outbreaks of avian influenza forced U.S. farms to destroy 50.7 million egg-laying hens.
- Elasticity near 0.3: shoppers kept buying eggs, so the price absorbed the shock: a dozen Grade A large eggs climbed from $3.37 in October 2024 to a record $6.23 in March 2025.
- Recovery: as flocks were rebuilt, the price fell to $2.27 by August 2026.

3.3

Elasticity and Total Revenue

- The total-revenue rule
- Revenue along a linear demand curve
- Should the firm raise its price?

Revenue rises, peaks, then falls

Price	Quantity	Total revenue	Elasticity	Demand is
\$9	2	\$9 × 2 = \$18	200/22.2 ≈ 9.00	elastic
\$8	4	\$8 × 4 = \$32	100/25 = 4.00	elastic
\$7	6	\$7 × 6 = \$42	66.7/28.6 ≈ 2.33	elastic
\$6	8	\$6 × 8 = \$48	50/33.3 ≈ 1.50	elastic
\$5	10	\$5 × 10 = \$50	**40/40 = 1.00**	**unit elastic**
\$4	12	\$4 × 12 = \$48	33.3/50 ≈ 0.67	inelastic
\$3	14	\$3 × 14 = \$42	28.6/66.7 ≈ 0.43	inelastic
\$2	16	\$2 × 16 = \$32	25/100 = 0.25	inelastic
\$1	18	\$1 × 18 = \$18	22.2/200 ≈ 0.11	inelastic

Table 3.3 Total revenue rises, peaks, and falls along a linear demand curve.

Revenue rises, peaks, then falls

FOLLOWING THE COLUMN

Raising price from $1 toward $5, revenue climbs $18 → $50: in the inelastic range the higher price outweighs the few lost sales.

Beyond $5, revenue falls: in the elastic range the lost sales dominate.

Revenue peaks at $5, where demand is unit elastic.

Where the elasticity column comes from

- Each row uses prices $1 above and $1 below its own price (midpoint method).
- At $8: price $9 → $7, a change of $2 from an average of $8: 2 ÷ 8 = 0.25, or 25%.
- Quantity 2 → 6, a change of 4 from an average of 4: 4 ÷ 4 = 1, or 100%.
- For $9 and $1, the line continues to $10 (no units sold) and to $0 (20 units).

= 4.00

ε = 100 ÷ 25 at a price of $8

Where the elasticity column comes from

READING DOWN THE COLUMN

The elasticity falls as the price falls: above 1 above $5, equal to 1 at $5, below 1 below $5.

The rule every manager should know

THE TOTAL-REVENUE RULE

When demand is inelastic, raising the price raises total revenue. When demand is elastic, raising the price lowers total revenue.

Inelastic demand

A streaming service with loyal subscribers can raise its price and collect more.

Elastic demand

A roadside fruit stand facing a dozen rivals would only drive customers away.

Elastic above the midpoint, inelastic below

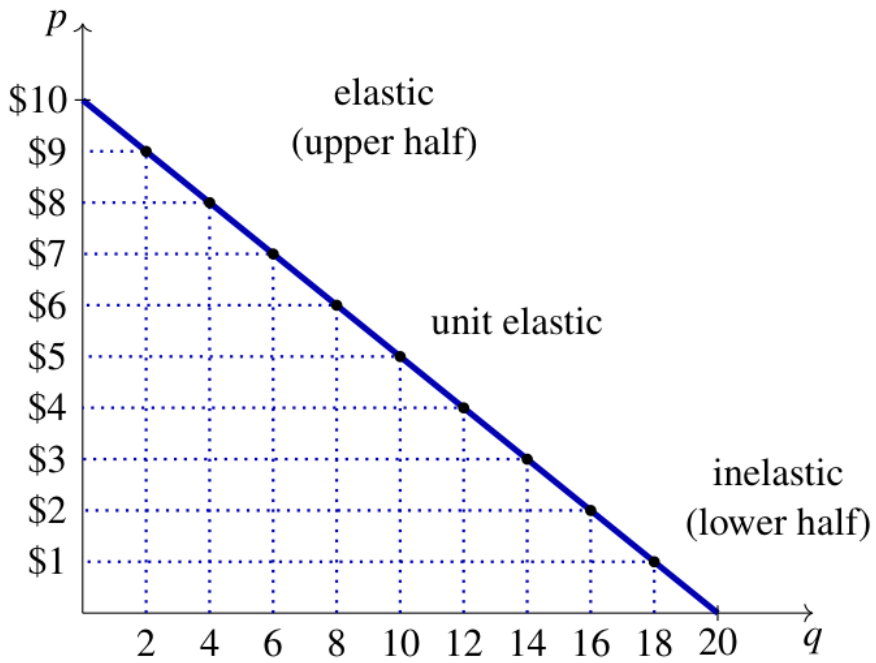


Figure 3.2 Elasticity and total revenue along a straight-line demand curve.

- Upper portion of the curve (above the midpoint): elastic.
- Lower portion (below the midpoint): inelastic.
- Total revenue is largest right at the midpoint.

Explore Figure 3.2: intro-micro-companion-website.netlify.app/elasticity

3.3.1 A WORKED EXAMPLE

Should the streaming firm raise its price?

THE SETUP

A service charges $10/month to 200,000 subscribers and considers $12. The price rise is about 18% by the midpoint method. The answer depends entirely on elasticity.

Should the streaming firm raise its price?

Inelastic, ε = 0.5

- Subscribers fall ~9% → about 182,600
- Revenue $2.0M → ~$2.19M
- Raise the price

Elastic, ε = 2.0

- Subscribers fall ~36% → about 138,500
- Revenue → ~$1.66M (falls)
- Do not raise; consider lowering it

Checking the subscriber numbers

- Price: \$10 → \$12, a change of \$2 from an average of \$11: 2 ÷ 11 ≈ 0.18, about 18%.
- $\varepsilon = 0.5$: subscribers fall about 0.5 × 18 = 9%. A drop of about 17,400, to 182,600, does it: 17,400 ÷ 191,300 ≈ 0.09 (191,300 is the average of 200,000 and 182,600).
- $\varepsilon = 2.0$: about 2.0 × 18 = 36%. A drop of about 61,500, to 138,500: 61,500 ÷ 169,250 ≈ 0.36.

\$2.19M vs. \$1.66M

\$12 × 182,600 and \$12 × 138,500

Checking the subscriber numbers

THE ANSWER

With inelastic demand the price rise raises revenue; with elastic demand it lowers revenue.

Self-check: is the latte elastic or inelastic?

SELF-CHECK

A coffee shop raises a latte from $5 to $6, and daily sales fall from 100 to 80. Using the midpoint method, is demand elastic or inelastic?

Price: +$1 on an average of $5.50: $1 \div 5.50 \approx 0.18$, about 18%. Quantity: −20 on an average of 90: $20 \div 90 \approx 0.22$, about 22%.

$\varepsilon \approx 22 \div 18 \approx 1.2$, greater than 1, so demand is elastic.

3.4

Income and Cross-Price Elasticity

- Income elasticity: normal vs. inferior
- Necessities, luxuries, and Engel's law
- Cross-price elasticity: substitutes vs. complements

Income elasticity of demand

Income elasticity of demand

The percentage change in quantity demanded divided by the percentage change in income: with an income elasticity of 0.5, a 1% rise in income raises quantity demanded by 0.5%. Positive for normal goods, negative for inferior goods.

Income elasticity of demand

Necessities

Normal goods with income elasticity between 0 and 1: demand grows more slowly than income.

Luxuries

Income elasticity above 1: demand grows faster than income (restaurant meals, air travel).

Engel's law: as incomes rise, spending on food as a whole rises more slowly than income, so the share of income spent on food falls (Engel, 1857).

Cross-price elasticity of demand

- **Cross-price elasticity of demand**

  The percentage change in the quantity demanded of one good divided by the percentage change in the price of another. The sign alone tells a firm whether another product is its rival or its partner.

Cross-price elasticity of demand

Substitutes (+)

A higher price for tea raises the demand for coffee: positive cross-price elasticity.

Complements (−)

A higher price for game consoles lowers demand for the games: negative cross-price elasticity.

Sorting goods by income elasticity

Category	Income elasticity	Examples
Necessities	small positive (~0.2–0.5)	bread, electricity
Luxuries	above 1	restaurant meals, vacations, jewelry
Inferior goods	negative	intercity bus travel, store-brand noodles

Income elasticity: luxuries boom in good times and slump in recessions.

The sign reveals the relationship

Pair	Relationship	Cross-price elasticity
Butter and margarine	substitutes	positive (+)
Cars and gasoline	complements	negative (−)

Firms watch these signs: they reveal whether a product is a rival or a partner.

3.5

Price Elasticity of Supply

- How readily sellers change output
- Time, spare capacity, storage
- Corrective (sin) taxes and elasticity

Supply has an elasticity too

Price elasticity of supply

The percentage change in quantity supplied divided by the percentage change in price. Its main determinant is how easily sellers can change output.

Inelastic supply

A parking lot near a stadium: however high the price on game day, the number of spaces is fixed.

Elastic supply

Manufactured goods with spare factory capacity: firms can quickly produce more.

Measuring supply elasticity between two points

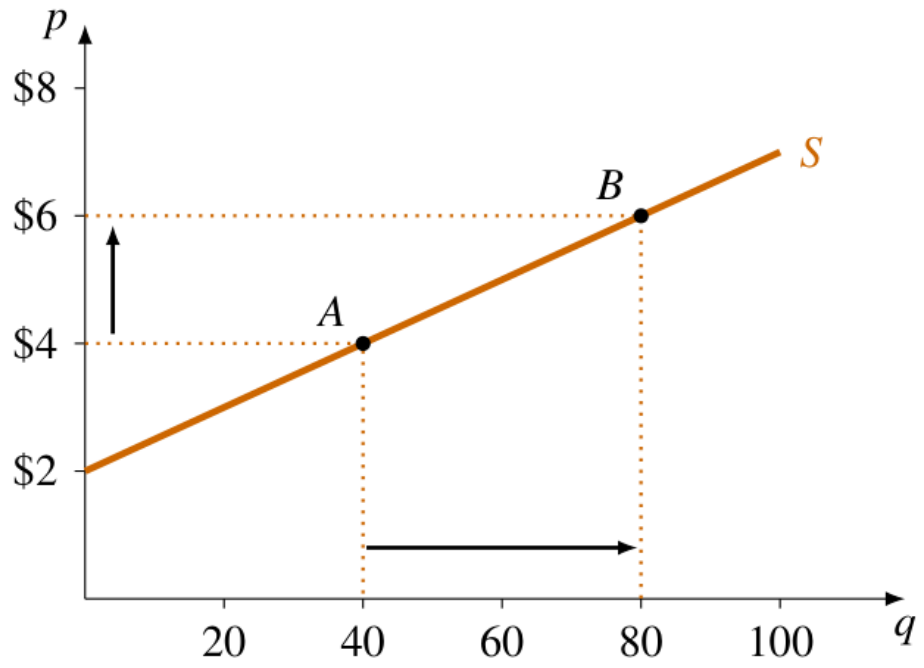


Figure 3.3 Measuring the price elasticity of supply between two points.

- A to B: price \$4 → \$6, 2 ÷ 5 = 0.40 (40%); quantity 40 → 80, 40 ÷ 60 ≈ 0.667 (66.7%).
- Elasticity of supply ≈ 66.7 ÷ 40 ≈ 1.67: a 1% price rise raises quantity supplied by about 1.67%, so supply is elastic here.
- A steeper curve through A (40 → 44 units, 4 ÷ 42 ≈ 9.5%) gives 9.5 ÷ 40 ≈ 0.24: inelastic.

Three factors determine supply elasticity

1

1. Time horizon

Given time, firms hire, build plants, and plant fields, so supply is almost always more elastic in the long run.

2

2. Spare capacity

A factory below its limit can raise output quickly and at low cost; one at full capacity cannot.

3

3. Ability to store

Storable goods can be released when prices rise; fresh fish, hotel rooms, and electricity cannot, so their supply is very inelastic.

Why a sin tax's effect depends on elasticity

Corrective taxes (sin taxes)

Taxes meant to discourage harmful consumption. What such a tax accomplishes depends on the elasticity of demand for the good.

Inelastic good

Cigarettes: a tax changes behavior only modestly but raises a great deal of revenue: people keep buying.

Elastic good

Sugary drinks: a tax produces a larger fall in consumption: the very goal of a health-motivated tax.

The same 10% price rise, five different effects

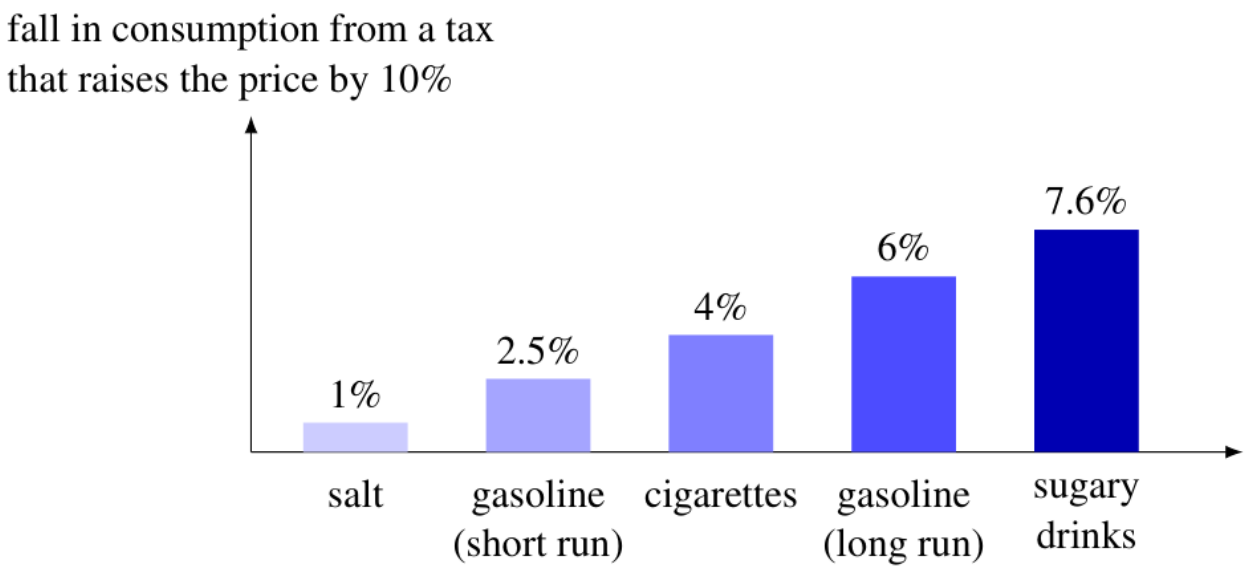


Figure 3.4 Fall in consumption from a 10 percent price increase.

- A tax that raises the price by 10% cuts salt consumption by only about 1% and cigarette consumption by about 4%.
- It cuts sugary-drink consumption by about 7.6% (Mexico's tax, next slide); the other bars use Table 3.2.

Explore Figure 3.4: intro-micro-companion-website.netlify.app/consumption-response

Sugar taxes in practice

UK, 2018

The Soft Drinks Industry Levy (announced in 2016) taxes drinks by sugar content; drinks below a threshold pay nothing. Makers reformulated before it started: by 2019, only 15% of drinks studied exceeded the lowest threshold, against an expected 49% (Scarborough et al., 2020).

Sugar taxes in practice

Mexico, 2014

A 1-peso/liter tax on sugary drinks (~10% price rise). Colchero et al. (2017): purchases fell ~5.5% in year one, ~10% in year two, average ~7.6%, largest drops among the poorest.

To predict what a tax will do (raise revenue, change behavior, or both), a regulator must first know the elasticity of demand. (Explore Figure 3.4: raise a good's price by 10 percent and watch its consumption fall.)

3.6

Case Study: Why Cigarette Taxes Reduce Smoking Only So Much

One inelastic number explains both revenue and limits.

3.6 CASE STUDY

Inelastic demand: revenue yes, behavior only so much

≈ 0.4

adult cigarette price elasticity (Chaloupka and Warner, 2000)

−4%

quantity smoked per +10% price

- Because demand is inelastic, quantity falls proportionally less than price rises.
- Because the quantity smoked falls only modestly, a higher tax rate applies to almost as many packs, so tax revenue rises.
- The same inelasticity that makes cigarettes a reliable revenue source also limits how much a tax can cut smoking.

One group for whom the tax works better

YOUTH DEMAND IS ELASTIC

Young people are roughly twice as responsive to price as adults: a 10% price increase is estimated to lower youth smoking by roughly 7 to 10%.

- Teenagers have lower incomes, are less addicted, and are more swayed by price.
- This is why public-health economists favor large tobacco taxes specifically to keep young people from starting: the population whose demand is elastic enough for the tax to make a real difference.

3.7

Empirical Evidence: How Elastic Is the Demand for Gasoline?

The time-horizon effect, made concrete.

Espey (1998): a deep meta-analysis

- Pooled hundreds of short-run and long-run estimates from three decades of studies.
- Short run: drivers are locked into the cars they own and the distances they commute.
- Long run: they buy efficient cars, change where they live and work, shift to transit.

0.23

median short-run ε (mean 0.26)

Espey (1998): a deep meta-analysis

≈ 0.58

mean long-run ε

THE LESSON

The long-run response is more than twice as large, yet still below 1. Gasoline is inelastic over any horizon, which is why fuel taxes raise revenue and oil shocks pass so painfully to budgets.

CHAPTER 3

Key takeaways

1. **Elasticity**
   $\varepsilon = \%\Delta Q \div \%\Delta P$, reported as an absolute value via the midpoint method.

2. **Three cases**
   $\varepsilon > 1$ elastic, $\varepsilon < 1$ inelastic, $\varepsilon = 1$ unit elastic; driven by substitutes, luxury, budget share, time.

3. **Total-revenue rule**
   Inelastic: raising price raises revenue. Elastic: raising price lowers it. Revenue peaks at unit elasticity.

4. **Income & cross-price**
   The income elasticity signs normal vs. inferior goods; the cross-price elasticity, substitutes vs. complements.

Key takeaways

5 **Supply elasticity**

Determined by time horizon, spare capacity, and storability.

6 **Evidence**

Cigarette taxes ($\varepsilon \approx 0.4$), gasoline short vs. long run (0.25 vs. 0.58), and sugar taxes.

4

CHAPTER 4

Government in the Market: Price Controls, Taxes, and Subsidies

The same supply-and-demand diagrams, now with a policy added, and the new outcome found on the graph.

CHAPTER 4

What this chapter covers

4.2–4.3 Controls

- Price ceilings & shortages
- Price floors & surpluses
- The minimum wage as a floor

4.4–4.5 Taxes & subsidies

- Tax incidence: who really pays
- Elasticity & the burden
- Subsidies: the mirror image

4.6–4.7 Evidence

- Case: does rent control help?
- Empirical: the minimum wage & jobs

CHAPTER 4

What this chapter covers

THE BIG IDEA

When a government caps a price, sets a floor, taxes, or subsidizes, we add the policy to the supply-and-demand diagram and find the new outcome.

4.1

Introduction

- Berlin's rent freeze, 2020
- A federal minimum wage fixed since 2009
- Taxes on every pack, gallon, and bottle

Three ways government touches a price

Berlin, 2020

The city froze rents on ~1.5 million apartments; listings fell sharply, and a court struck the freeze down 14 months later.

Minimum wage

The U.S. federal floor has been $7.25/hour since July 2009, while several states set $16 or more (Washington: $17.13 in 2026) and some cities exceed $21.

Three ways government touches a price

Taxes everywhere

Every pack of cigarettes, gallon of gas, and bottle of soda carries a tax that changes its price.

THE PLAN

In Chapters 2 and 3 prices were set by supply and demand alone. Here a government intervenes, and the tools are the diagrams we already built.

4.2

Price Ceilings

- A legal maximum price
- When a ceiling binds
- The shortage it creates
- How the good gets rationed instead

A legal maximum price

Price ceiling

A legal maximum price. It matters only if it is set below the equilibrium price, in which case we call it binding.

THE SETUP

Return to the coffee market of Chapter 2: equilibrium price $5, equilibrium quantity 70 million pounds. Consider that the government, hoping to make coffee affordable, caps the price at $3.

A binding ceiling creates a shortage

- **At a $3 ceiling**

  Buyers want 90 million pounds, but sellers offer only 50 million. The ceiling creates a shortage of 40 million pounds that cannot disappear: the price is forbidden to rise.

WHAT ACTUALLY TRADES

With the price held below equilibrium, the quantity actually traded is the smaller of the two: the 50 million pounds sellers are willing to supply.

The ceiling, drawn

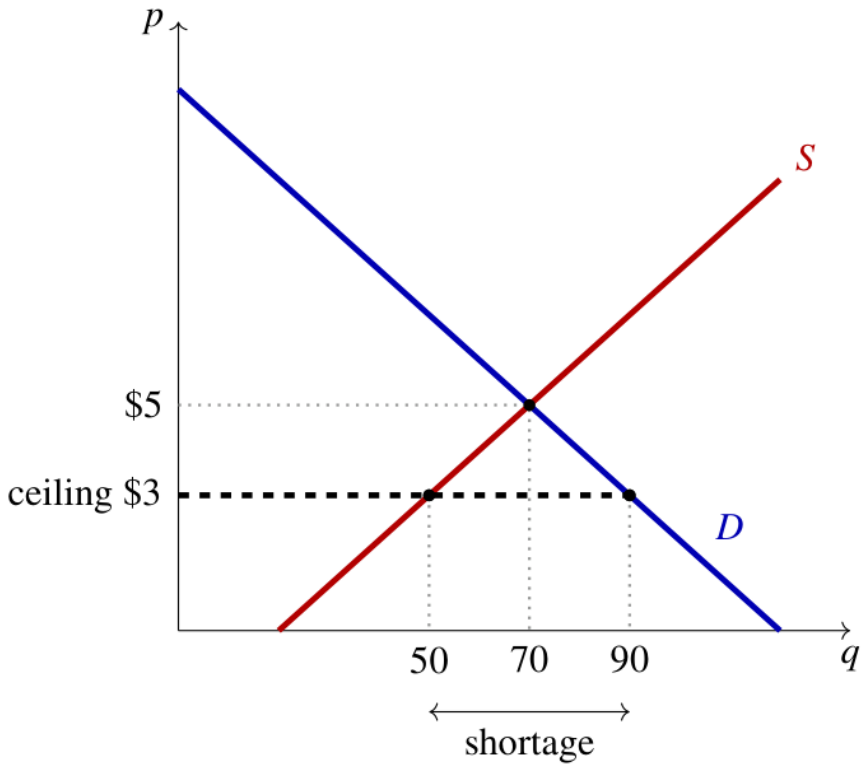


Figure 4.1 A binding price ceiling below the equilibrium price creates a persistent shortage.

- The ceiling sits below the $5 equilibrium, so it binds.
- Quantity demanded (90) exceeds quantity supplied (50): a gap of 40 million.
- Only 50 million pounds change hands.

Explore Figure 4.1: intro-micro-companion-website.netlify.app/price-controls

When price can't ration, something else must

Lines

Long queues, as at U.S. gas stations under the 1970s price controls.

Waiting lists

Stockholm's rent-controlled apartments carry waits measured in years.

When price can't ration, something else must

Favoritism & side payments

Sellers choose whom to serve; illegal side payments appear.

Lower quality

Facing excess demand, sellers have little reason to maintain the good.

We examine the most studied price ceiling, rent control, in Section 4.6.

4.3

Price Floors

- A legal minimum price
- When a floor binds
- The surplus it creates
- The minimum wage as a price floor

A legal minimum price

Price floor

A legal minimum price. It binds only when set above the equilibrium price.

At a $7 floor on coffee

Sellers offer 90 million pounds but buyers want only 50 million: a surplus of 40 million pounds (the mirror image of the ceiling).

The floor, drawn

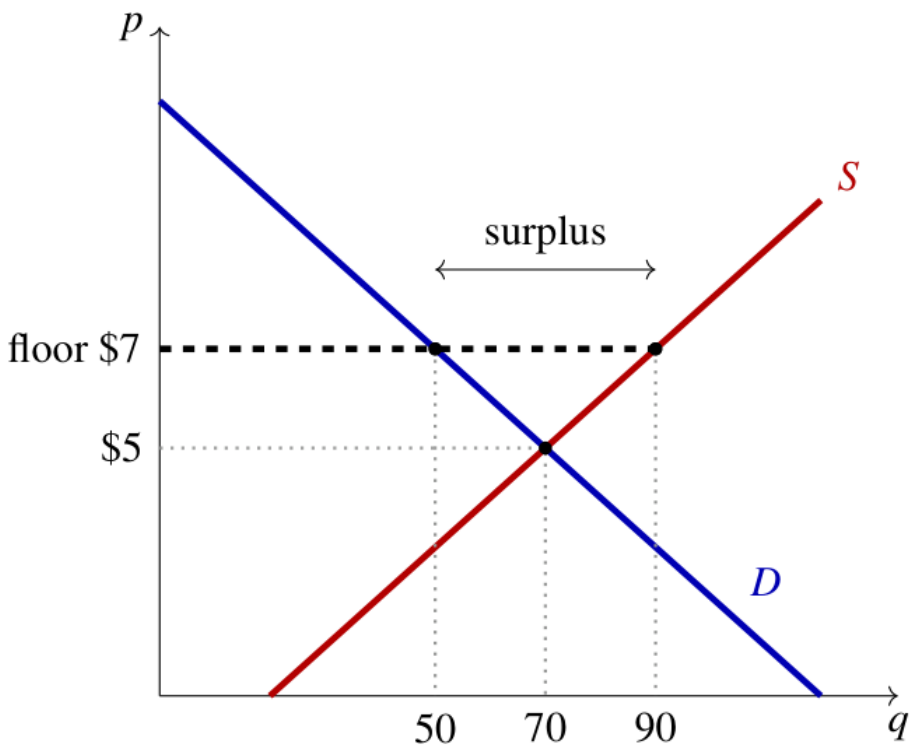


Figure 4.2 A binding price floor above the equilibrium price creates a persistent surplus.

- The floor sits above the $5 equilibrium, so it binds.
- Quantity supplied (90) exceeds quantity demanded (50): a surplus of 40 million.
- The mirror image of Figure 4.1.

The two most important floors

Agricultural supports

Guaranteed crop prices produced the surplus the model predicts: the European Community's 1980s "butter mountains" and "wine lakes," which governments had to buy and store.

The minimum wage

A floor in the market for low-wage labor. If set above the market-clearing wage, the model predicts a surplus of labor, that is, unemployment.

The two most important floors

Whether that prediction holds in practice is one of the most studied questions in economics (see Section 4.7).

A market for low-wage labor

Wage	Workers demanded	Workers supplied	Status
$6	60,000	40,000	shortage of 60,000 − 40,000 = 20,000
$8	50,000	50,000	**no shortage or surplus (50,000 − 50,000 = 0), equilibrium**
$10	40,000	60,000	surplus of 60,000 − 40,000 = 20,000 (unemployment)
$12	30,000	70,000	surplus of 70,000 − 30,000 = 40,000

Table 4.1 A minimum wage above the equilibrium wage creates a surplus of labor.

A $10 floor binds above the $8 equilibrium

THE PREDICTION

At $10, firms want only 40,000 workers but 60,000 wish to work: a surplus of 20,000. The floor helps the workers who keep their jobs (they earn more) but harms those who cannot find one.

A $10 floor binds above the $8 equilibrium

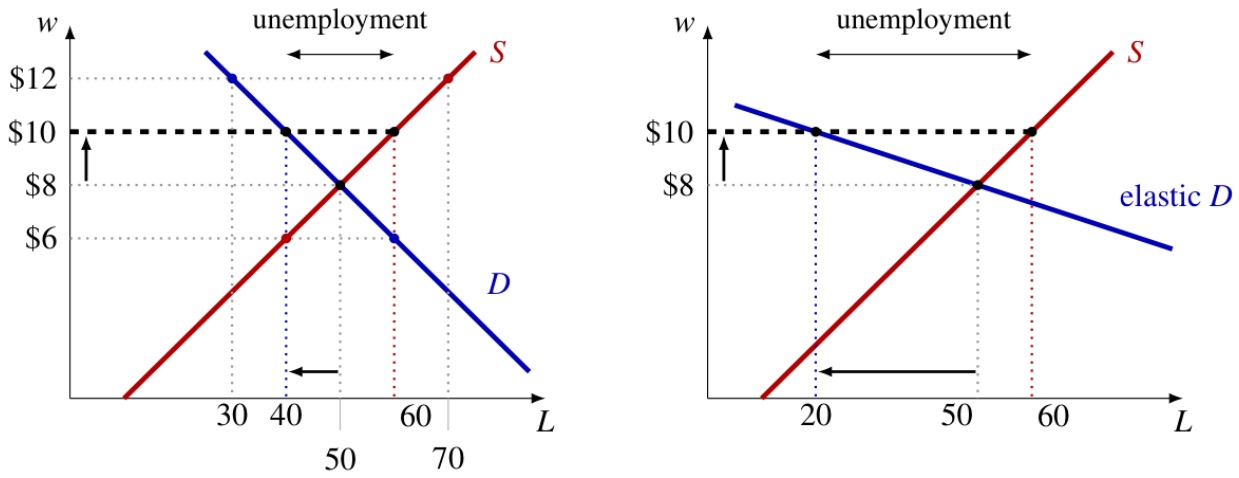
Caution 1: how far above?

The effect depends on how far the floor is above equilibrium and on the elasticity of labor demand. A floor only slightly above, or inelastic demand, produces little unemployment (Figure 4.3).

Caution 2: monopsony

A single dominant employer can hold wages below the competitive level (Chapter 11). Then a minimum wage can raise both wages and employment.

The same floor, two labor markets



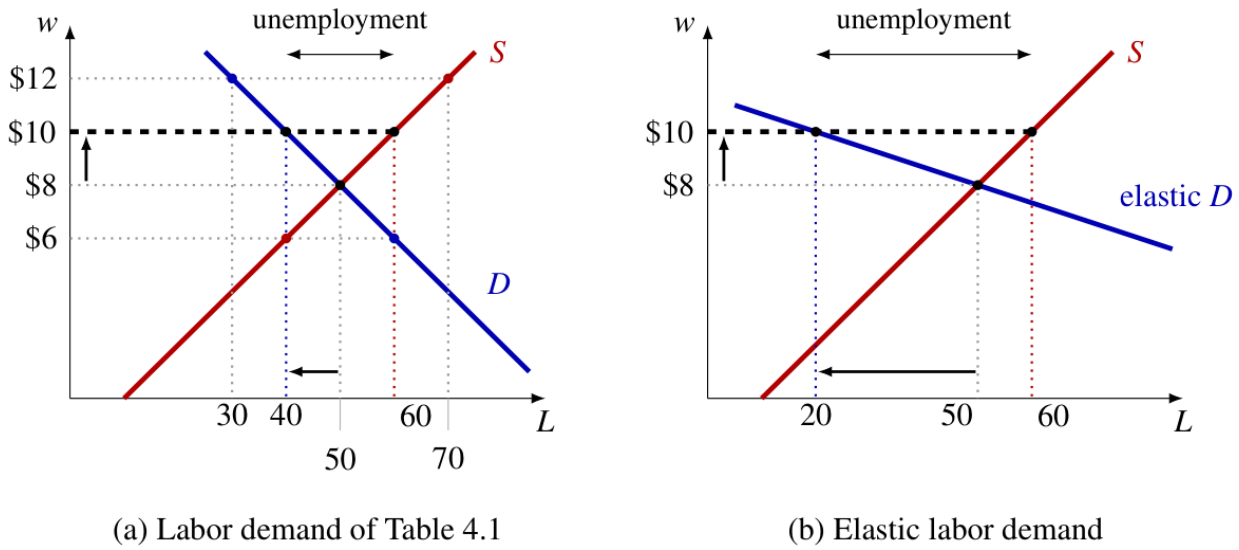
(a) Labor demand of Table 4.1

(b) Elastic labor demand

Figure 4.3 A $10 minimum wage with the labor demand of Table 4.1 (a) and elastic labor demand (b), workers in thousands.

- Same supply, same $8 equilibrium, same $10 minimum wage.
- (a) The labor demand of Table 4.1, unit elastic between $8 and $10: unemployment is 20,000 workers.

The same floor, two labor markets



(a) Labor demand of Table 4.1

(b) Elastic labor demand

Figure 4.3 A $10 minimum wage with the labor demand of Table 4.1 (a) and elastic labor demand (b), workers in thousands.

- (b) Elastic labor demand: hiring falls to 20,000, and unemployment doubles, to 40,000 workers.

The two panels of Figure 4.3, in numbers

- Wage \$8 → \$10: a change of 2 from an average of 9, 2 ÷ 9 ≈ 22% (midpoint method), in both panels.
- (a) Hiring 50,000 → 40,000: 10 ÷ 45 ≈ 22%, the same percentage, so labor demand is unit elastic between \$8 and \$10. Unemployment: 60,000 − 40,000 = 20,000.
- (b) Hiring 50,000 → 20,000: 30 ÷ 35 ≈ 86%, an elasticity of about 3.9. Unemployment: 60,000 − 20,000 = 40,000.

The two panels of Figure 4.3, in numbers

THE LESSON

The more elastic the demand for labor, the more jobs the same minimum wage costs.

Self-check: does the policy bind?

SELF-CHECK

The competitive price of a good is $5. State whether each policy binds and what it does: (a) a price ceiling of $3; (b) a price floor of $4.

(a) A $3 ceiling lies below $5, so it binds and creates a shortage. (b) A $4 floor also lies below $5, so it does not bind: the price stays at $5 and nothing changes.

4.4

Taxes and Tax Incidence

- The tax as a wedge
- Who really bears the burden
- Deadweight loss
- Elasticity and the split

A tax drives a wedge between two prices

The wedge

A $2 tax on each pound of coffee makes the price buyers pay and the price sellers keep differ by the amount of the tax. We find the quantity at which the buyer's price exceeds the seller's price by $2.

THE NEW OUTCOME

Working through the coffee schedules, a $2 tax moves the market to 60 million pounds: buyers pay $6, sellers keep $4, a difference equal to the $2 tax.

The tax, with and without

	No tax	With $2 tax
Price buyers pay	$5	$6
Price sellers keep	$5	$4
Quantity traded (million lbs)	70	60

Table 4.2 A $2 tax on coffee raises the buyer's price by $1 and lowers the seller's price by $1.

READ IT

The split is even here because supply and demand are equally responsive. Buyers and sellers each absorb $1 of the $2 tax.

Two lessons about who pays

1. Legal vs. economic burden

It does not matter whether the law collects the tax from buyers or sellers: either way buyers pay $6, sellers keep $4, government collects $2. Who hands over the check is a legal detail; who bears the burden (the tax incidence) is an economic outcome the market decides.

2. In our coffee example, an even split

Supply and demand are equally responsive, so buyers and sellers split the $2 evenly; Section 4.4.1 explains what happens when they are not.

The wedge and the deadweight loss

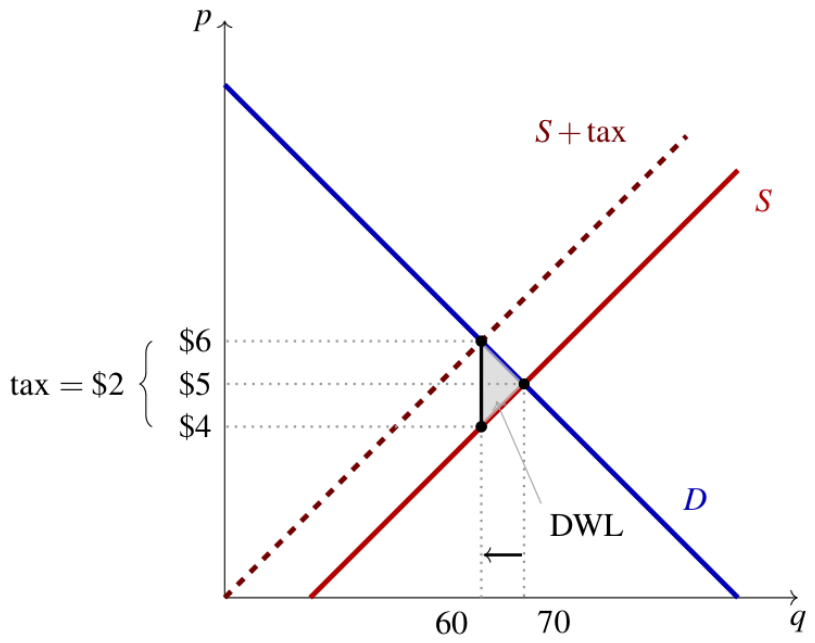


Figure 4.4 A $2 tax on coffee.

Deadweight loss

Quantity falls from 70 to 60: the tax blocks 10 million pounds of trades that buyers valued above what sellers would accept: $10 million of lost, mutually beneficial trades.

This is why taxes have a cost beyond the revenue they raise.

The wedge and the deadweight loss

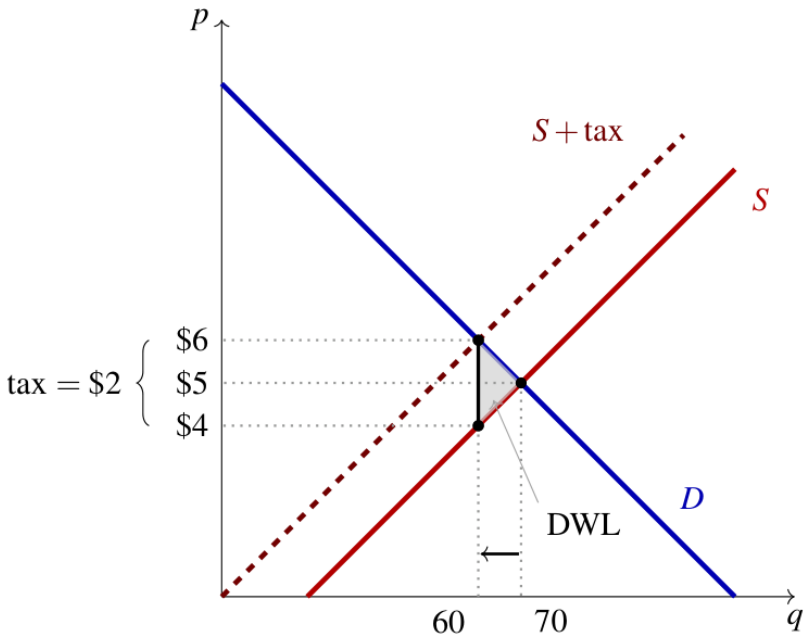


Figure 4.4 A $2 tax on coffee.

Explore Figure 4.4: intro-micro-companion-website.netlify.app/supply-demand

The less elastic side bears more

THE RULE

The side of the market that is less elastic (less able to walk away) bears the larger share of the tax.

The less elastic side bears more

Inelastic demand → buyers pay

Smokers' demand is highly inelastic, so a tax barely cuts their purchases. Sellers pass almost the whole tax through; consumers bear most of it. The same holds for gasoline and electricity.

Elastic demand → sellers pay

A good with many substitutes, made by producers who cannot switch products, taxes mainly the sellers, who absorb it to keep buyers from fleeing.

The lesson reaches beyond taxes

THE GENERAL PRINCIPLE

Whenever a cost is imposed on a market (a tax, a fee, a tariff, a regulation), the party that bears it is determined not by who legally pays, but by who is less able to change their behavior in response.

Explore Figure 4.4: slide the tax up and down and watch the wedge widen, quantity fall, and the deadweight-loss triangle grow.

4.5

Subsidies

- A subsidy is a negative tax
- Lower buyer price, higher seller price
- More quantity traded
- Subsidy: a worked-out example

A subsidy is a negative tax

Subsidy

The government pays sellers (or buyers) for each unit traded. It lowers the price buyers pay, raises the price sellers receive, and increases the quantity traded beyond the free-market level.

- Governments subsidize goods they want more of: solar panels, electric cars, childhood vaccines, university education.
- Like a tax, a subsidy has costs, which the worked example of Section 4.5.1 measures.

A $2 subsidy on coffee

	No subsidy	With $2 subsidy
Price buyers pay	$5	$4
Price sellers receive	$5	$6
Quantity traded (million lbs)	70	80

Table 4.3 A $2 subsidy on coffee lowers the buyer's price by $1 and raises the seller's price by $1.

A $2 subsidy on coffee

THE SETUP

Sellers receive the price buyers pay plus $2, so they supply more at any buyer price. The market moves to 80 million pounds: buyers pay $4, sellers receive $6: a difference equal to the $2 subsidy.

The mirror image of a tax, but not free

The mirror

Where the tax raised the buyer's price and lowered the seller's, the subsidy does the reverse; where the tax shrank quantity 70→60, the subsidy expands it 70→80.

The bill

The government pays $2 on each of 80 million pounds ($160 million), raised through taxes elsewhere (such as taxes on other goods, income taxes, or corporate taxes).

The mirror image of a tax, but not free

THE COST

Pushing output beyond the efficient level creates a deadweight loss of its own, now from overproduction. How the $2 is split between buyers and sellers again depends on the relative elasticities of supply and demand.

4.6

Case Study: Does Rent Control Help Tenants?

A price ceiling on housing, and a natural experiment in San Francisco.

A natural experiment in San Francisco

THE STUDY

Diamond et al. (2019) studied a 1994 ballot measure that suddenly extended rent control to a set of older, smaller buildings while leaving similar buildings uncovered, then compared the two groups for over two decades.

- For tenants who stayed put, rent control worked as intended: it held rents down and reduced the chance of displacement: large benefits to those specific renters.
- But landlords responded as the supply-and-demand model predicts.

4.6 CASE STUDY

It helps sitting tenants, and worsens the shortage

-15%

rental housing supply (landlords converted, redeveloped, or sold)

↑ rents

on the city's uncontrolled apartments

THE VERDICT

By shrinking the rental stock, the policy pushed up rents on uncontrolled apartments, so renters as a whole (especially future and newly arriving ones) faced higher rents. Rent control transferred wealth to sitting tenants while making the broader shortage worse.

4.6 CASE STUDY

The same pattern elsewhere

Stockholm

Controlled rents come with waiting lists years long.

Berlin, 2020

After the rent freeze, the number of apartments advertised for rent fell sharply.

THE BOTTOM LINE

The policy helps the tenants it covers and harms those it does not.

4.7

Empirical Evidence: The Minimum Wage and Jobs

Which case, competitive or monopsony, better describes actual labor markets?

A landmark natural experiment

THE STUDY

Card and Krueger (1994): in April 1992 New Jersey raised its minimum wage from $4.25 to $5.05, while neighboring Pennsylvania left its unchanged. Surveying 410 fast-food restaurants on both sides of the border, before and after.

THE FINDING

Employment in New Jersey did not fall relative to Pennsylvania; if anything, it rose slightly. The simplest model's prediction did not appear.

A generation of research

Study	Setting	Key finding
Card and Krueger (1994)	NJ vs. PA, 410 fast-food restaurants, 1992	Employment did not fall; if anything rose slightly
Cengiz et al. (2019)	138 U.S. state increases, 1979–2016	Low-wage jobs essentially unchanged; earnings rose
Large-increase studies	Very large minimum-wage hikes	Some employment losses appear

Studies of the minimum wage and employment.

Theory gives the prediction; measurement tests it

- The evidence does not say minimum wages never cost jobs; the effect depends on how high the floor is relative to local wages.
- But for the moderate increases studied, the large job losses the simplest model predicts have generally not appeared.

THE TAKEAWAY

A recurring lesson of the book: a model gives a clear prediction, and careful measurement tells us when that prediction holds and when reality is more complicated.

Why the job losses did not appear

- The increases studied were moderate: the floor was set only slightly above the market wage.
- Labor demand is inelastic (restaurants and stores cannot easily do without cooks and cashiers), so firms cut hiring only a little, unlike panel (b) of Figure 4.3.
- If labor supply is also inelastic, few more people look for work, so unemployment barely rises and the deadweight loss is small.

A SECOND EXPLANATION

Monopsony, which we study in Chapter 11.

CHAPTER 4

Key takeaways

1. **Price ceilings**
   A binding ceiling (below equilibrium) creates a persistent shortage; price can no longer ration the good.

2. **Price floors**
   A binding floor (above equilibrium) creates a surplus; for the minimum wage, that surplus is unemployment.

3. **Tax incidence**
   It does not matter who legally pays; the burden falls on whichever side is less elastic.

4. **Deadweight loss**
   Taxes block mutually beneficial trades; subsidies cause them in excess. Both have a cost beyond the budget.

Key takeaways

5

Subsidies

A negative tax: lower buyer price, higher seller price, more quantity, but a bill and overproduction.

6

Evidence

Rent control helps sitting tenants but shrinks supply; moderate minimum-wage hikes show little job loss.

5

CHAPTER 5

Consumer Preferences

What a consumer wants (preferences, utility, indifference curves) and the rate at which she trades one good for another.

CHAPTER 5

What this chapter covers

5.2–5.3 The basics

- Ranking bundles
- Total & marginal utility
- Indifference curves

5.4–5.5 Trading goods

- The MRS and why it falls
- Link to marginal utility
- Substitutes & complements

5.6–5.8 Evidence

- Case: razors & printer ink
- Empirical: too much choice
- Behavioral: fairness

CHAPTER 5

What this chapter covers

THE BIG IDEA

Preferences (the likes and dislikes that guide every purchase) are the most basic ingredient of microeconomics. We keep to words, rankings, and simple numbers throughout.

5.1

Introduction

- Why a second cup brings less satisfaction
- What a consumer wants vs. can afford
- Preferences, utility, and indifference curves

Four puzzles that preferences will solve

The second cup

Why does a second cup of coffee bring less extra satisfaction than the first?

The same aisle

Why do supermarkets place hot dogs and buns together?

Four puzzles that preferences will solve

Costly ink

Why does a $60 printer use ink pricier, ounce for ounce, than fine champagne?

Less is more

Why did shoppers buy more jam when offered fewer flavors?

THE PLAN

This chapter describes what a consumer wants; Chapter 6 adds what she can afford, and puts the two together to explain what she actually buys.

5.2

Preferences and the Ranking of Bundles

- What a bundle is
- Three rules for ranking
- Utility as a ranking number

A bundle, and three reasonable rules

Bundle

A list of how much of each good a person consumes: for example, 2 coffees and 3 donuts. We assume a consumer can look at any two bundles and rank them.

A bundle, and three reasonable rules

1

1. Completeness

Faced with any two bundles, she can say which she prefers, or that she likes them equally. She is never unable to compare, which assumes she has enough information and time.

2

2. Transitivity

If she prefers A to B, and B to C, then she prefers A to C. Her rankings do not run in circles.

A bundle, and three reasonable rules

3

3. More is better

Other things equal, more of a good is preferred to less.

Utility is a ranking, not an amount

Utility

A number attached to each bundle, where a higher number marks a more preferred bundle. It measures order, not amount.

Why the Number Is Only Ordinal

If one bundle has utility 10 and another 5, the first is preferred, but it is not "twice as good"; utilities of 45 and 30 would tell us the same. Utility is like the finishing places in a race: they tell us who came first and second, but not by how many seconds.

Total vs. marginal utility

Total utility

The overall satisfaction a good provides.

LOOKING CLOSER

It helps to look closely at how utility grows as a person consumes more of a single good. Table 5.1 shows both measures for cups of coffee over a day.

Marginal utility

The extra satisfaction from one more unit.

Total and marginal utility from coffee

Cups	Total utility	Marginal utility
0	0	—
1	10	10 − 0 = 10
2	18	18 − 10 = 8
3	24	24 − 18 = 6
4	28	28 − 24 = 4
5	30	30 − 28 = 2
6	30	30 − 30 = 0

Table 5.1 Total and marginal utility from cups of coffee.

Total and marginal utility from coffee

READING THE TABLE

Total utility rises, but by ever smaller amounts: the first cup adds 10, the second only 8, the third 6. At cup 6 marginal utility hits zero: the point of satiation, where an extra cup adds nothing; beyond it, marginal utility would be negative.

The law of diminishing marginal utility

THE LAW

As a person consumes more of a good, the extra satisfaction from each additional unit tends to fall.

- One of the most dependable regularities in economics: the first slice of pizza is a delight, the fifth a chore.
- It underlies the downward-sloping demand curves of Chapter 2: a buyer pays less for a unit that brings less extra satisfaction.
- It also bows the indifference curves we meet next.

Self-check: the third slice of pizza

SELF-CHECK

Q. After two slices of pizza, you find each delicious and consider a third. If the extra utility of the third slice is smaller than that of the second, what pattern of preferences does this describe?

A. Diminishing marginal utility: the extra satisfaction from each additional unit falls as consumption rises. Total utility can still rise (you may still enjoy the third slice); only the additional enjoyment from each new slice is smaller than the last.

5.3

Indifference Curves

- Bundles that are equally liked
- The indifference map
- Three properties of the curves

Equally-liked bundles trace a curve

Indifference curve

A curve connecting bundles that give a consumer equal satisfaction: she is indifferent among them. One good on each axis.

HIGHER IS BETTER

Every bundle on a single curve is equally preferred. Bundles on a higher curve (farther from the origin) are preferred to those on a lower one.

An indifference map

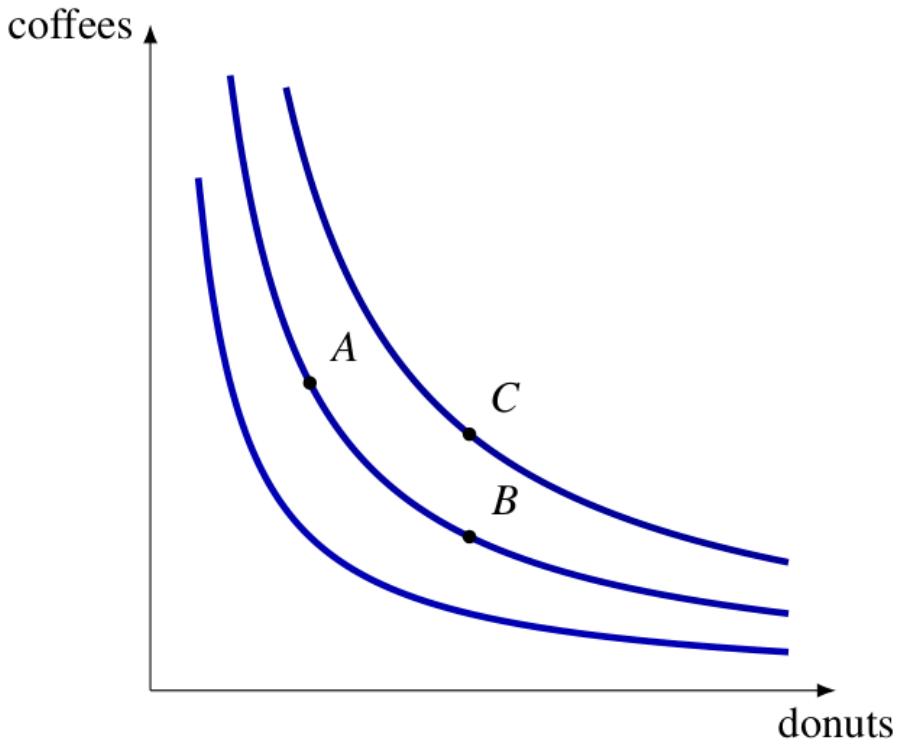


Figure 5.1 An indifference map.

A and B. On the same curve, equally preferred.

C. On a higher curve, preferred to both, because it contains more.

Three properties of indifference curves

Slope downward

Giving up coffee for nothing leaves her worse off, so she must receive more donuts to stay equally satisfied.

Never cross

If they did, transitivity and "more is better" would contradict each other.

Bowed to the origin

For the reason we examine next: the diminishing MRS.

Why indifference curves never cross

CONSIDER THAT TWO CURVES CROSSED

Let two indifference curves cross at bundle X. Take bundle A on one curve, and bundle C on the other with the same donuts as A but more coffee.

- A and X lie on the same curve, and so do X and C: by transitivity, she is indifferent between A and C.
- But C has more coffee and the same donuts, so she prefers C to A (more is better).
- The two conclusions contradict each other, so indifference curves cannot cross.

5.4

The Marginal Rate of Substitution

- The slope of an indifference curve
- A worked schedule (Table 5.2)
- Diminishing MRS bows the curve

The slope has a name

- **Marginal rate of substitution (MRS)**
  The number of units of one good a consumer will give up to gain one more unit of the other, while staying equally satisfied. It measures her willingness to trade.

A WORKED SCHEDULE

Table 5.2 lists four equally-preferred bundles for a consumer choosing between donuts and cups of coffee, all on the same indifference curve, which Figure 5.2 depicts.

The MRS falls as donuts grow plentiful

Bundle	Donuts	Coffees	Coffees given up for one more donut
A	1	10	—
B	2	7	$10 - 7 = 3$
C	3	5	$7 - 5 = 2$
D	4	4	$5 - 4 = 1$

Table 5.2 The marginal rate of substitution falls as donuts become more plentiful.

The MRS falls as donuts grow plentiful

READING THE TABLE

From A to B she gives up 3 coffees for her second donut; from B to C only 2; from C to D only 1. Her willingness to trade (her MRS) falls as she gets more donuts.

The principle of diminishing MRS

DIMINISHING MRS

The more a person already has of a good, the less of the other good she will sacrifice for still more units of it.

- When donuts are scarce and coffee plentiful, an extra donut is worth a lot of coffee.
- Once donuts are abundant, an extra one is worth little.
- Diminishing MRS is what makes indifference curves bow toward the origin.

A bowed indifference curve

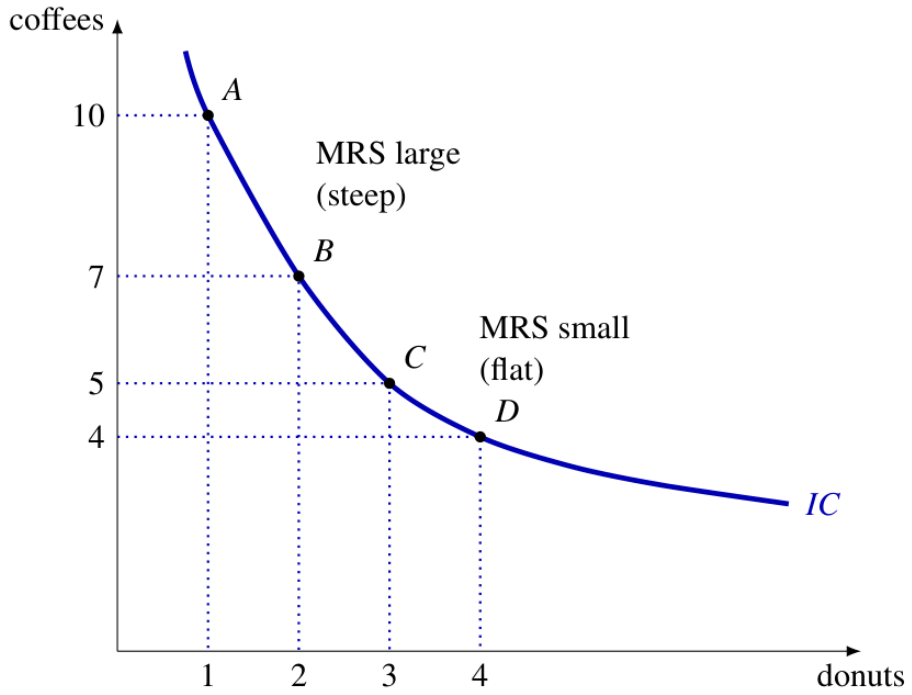


Figure 5.2 A bowed indifference curve.

Explore Figure 5.2: intro-micro-companion-website.netlify.app/indifference-curves

Steep where coffee is plentiful and donuts scarce: a large MRS.

Flat where donuts are plentiful: a small MRS.

MRS and marginal utility: one story, twice

THE LINK

The MRS at any bundle is the ratio of the two goods' marginal utilities: MRS = (marginal utility of donuts) ÷ (marginal utility of coffee). So diminishing marginal utility for each good helps produce the diminishing MRS that bows the indifference curve toward the origin.

MRS and marginal utility: one story, twice

- Many donuts, little coffee: the marginal utility of a donut is low and of coffee high, so she parts with little coffee for another donut: a small MRS.
- Donuts scarce, coffee plentiful: the reverse holds, and the MRS is large.
- The graphical story of 5.4 and the numerical story of 5.2.1 are the same story told twice.

5.5

Two Special Cases: Perfect Substitutes and Perfect Complements

- Perfect substitutes: straight lines
- Perfect complements: L-shapes
- The goods in between

Two extremes worth knowing

Perfect substitutes

Goods a consumer trades at a fixed rate, no matter how many she has: a one-dollar coin and two fifty-cent coins, or two brands of bottled water she considers identical. Indifference curves are straight lines: the MRS never changes.

Two extremes worth knowing

Perfect complements

Goods a consumer wants in a fixed proportion: left and right shoes, or consoles and the controllers they require. Indifference curves are L-shaped: a seventh left shoe is worthless without a seventh right shoe.

Straight lines and L-shapes

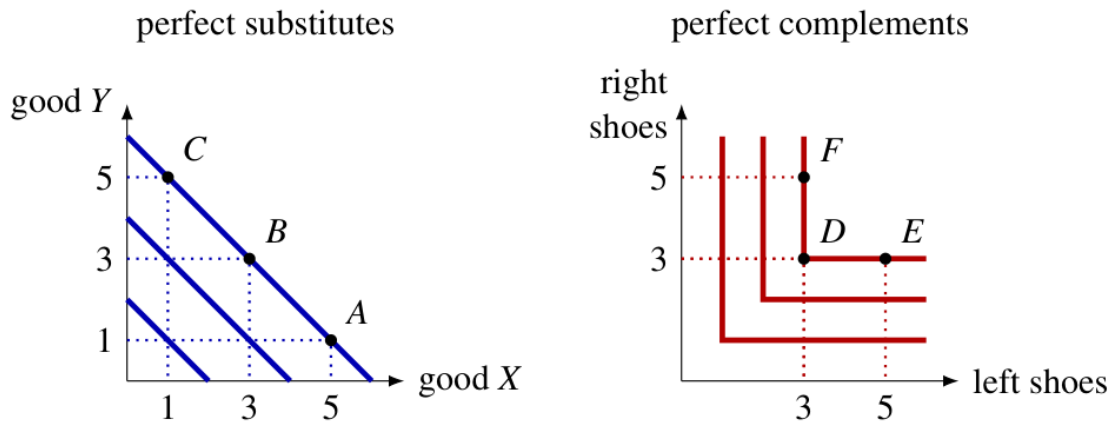


Figure 5.3 Perfect substitutes (left) and perfect complements (right).

Substitutes. Straight lines: a constant MRS.

Complements. Right angles: extra of one good alone is worthless.

Perfect substitutes: two brands of water

LEFT PANEL OF FIGURE 5.3

Goods X and Y are two brands of bottled water that the consumer considers identical.

- A has 5 bottles of X and 1 of Y; B has 3 of each; C has 1 of X and 5 of Y.
- All three contain 5 + 1 = 3 + 3 = 1 + 5 = 6 bottles, so they give the same utility and lie on the same indifference curve.
- Whatever her bundle, she gives up one bottle of Y for one more of X: the MRS never changes, so the curves are straight lines.

Perfect complements: left and right shoes

RIGHT PANEL OF FIGURE 5.3

D has 3 left and 3 right shoes, E has 5 left and 3 right, and F has 3 left and 5 right.

- Each bundle makes only 3 pairs: an extra left shoe is worthless without a matching right shoe, and vice versa.
- So D, E, and F give the same utility and lie on the same indifference curve.
- Only more of both goods, such as 4 left and 4 right shoes, makes her better off: the curve is L-shaped, with its corner where the goods come in the right proportion.

Most goods lie between the extremes

THE BOWED MIDDLE

A consumer choosing between two similar coffee brands, or between pizza and tacos, has some willingness to substitute, but at a rate that changes as the mix changes. These "well-behaved" preferences give the smoothly bowed curves of Section 5.4.

- Neither the straight lines of perfect substitutes nor the sharp corners of perfect complements.
- Placing a pair of goods on this spectrum is often the first step in predicting how a consumer responds when a price changes (the subject of Chapter 6).

Four curves through the same bundle

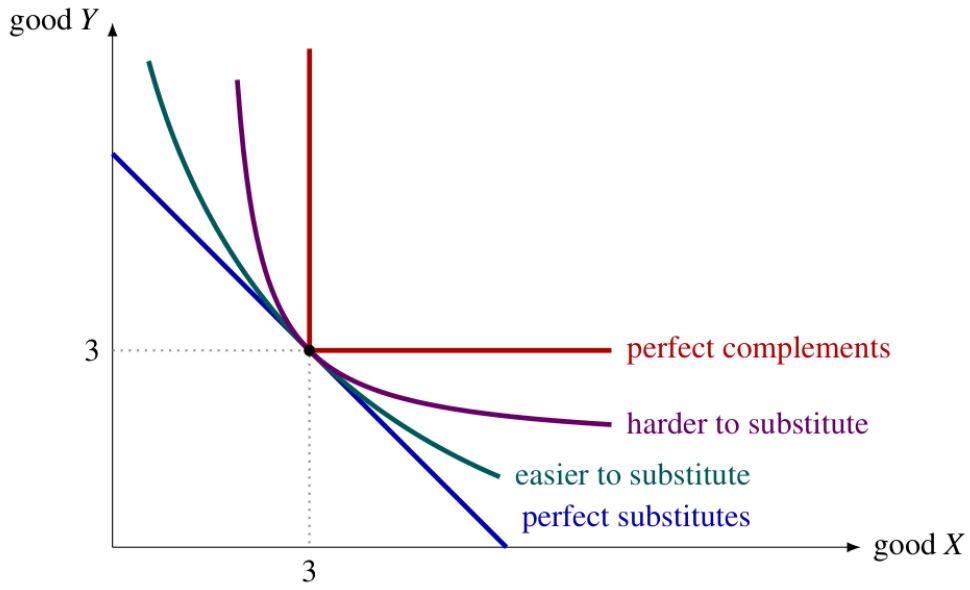


Figure 5.4 Four indifference curves through the same bundle.

Same bundle. All four pass through 3 units of each good.

The extremes. The straight line is perfect substitutes; the L-shape, perfect complements.

In between. Slightly bowed: goods she substitutes easily (two similar coffee brands). More bowed: goods she substitutes less easily.

Four curves through the same bundle

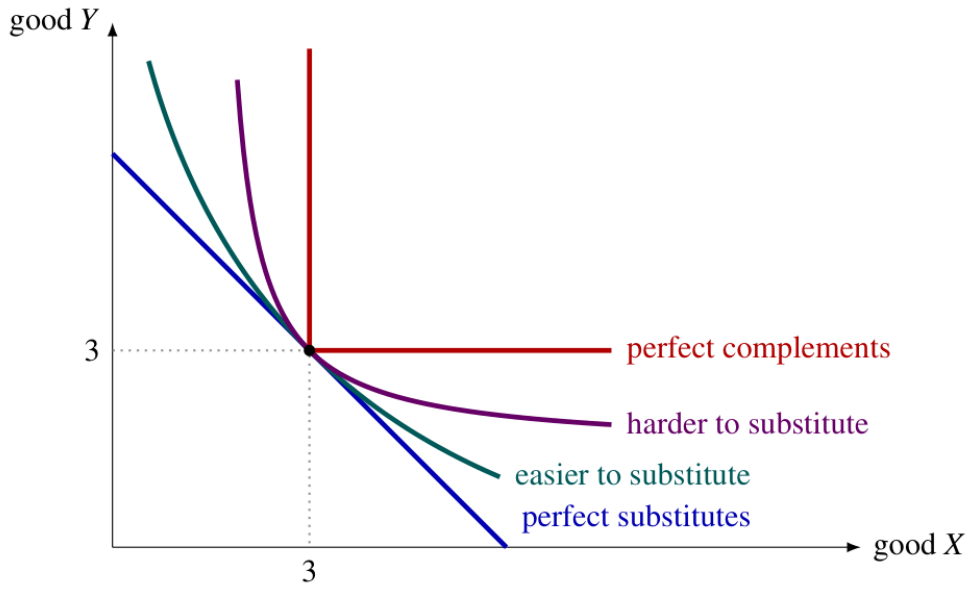


Figure 5.4 Four indifference curves through the same bundle.

The lesson. The more bowed the curve, the faster her MRS falls, and the closer the goods are to perfect complements.

5.6

Case Study: Razors, Blades, and the Price of Printer Ink

Complements explain one of the most common pricing strategies in business.

5.6 CASE STUDY

Sell the device for little, charge for the supplies

THE RAZOR MODEL

The classic example is the shaving razor. Selling razors and blades is profitable precisely because they are complements (a razor is useless without blades), so a customer who buys an inexpensive razor is locked into buying the maker's blades for years: a firm may sell a razor for $8 or less but charge $12 or more for a pack of replacement blades.

- The strategy works only when the seller can tie the two goods together.
- When refill makers or generic cartridges break the link, the high markup on supplies erodes.

5.6 CASE STUDY

Printers and ink: the same logic, made vivid

< $60

what an inkjet printer can retail for, at or below cost

$1,000s

per liter for branded ink: more, ounce for ounce, than fine champagne

WHY IT WORKS

Consumers tolerate this because, having bought the printer, they treat it and ink as perfect complements. The manufacturer expects to profit from replacement cartridges (one black cartridge from the maker can cost $20 to $50). We return to this kind of pricing in Chapter 12, on price discrimination and bundling.

5.7

Empirical Evidence: Can There Be Too Much Choice?

“More is better” applies to quantities, but does it apply to options?

The Iyengar and Lepper (2000) jam experiment

Display	Stopped	Bought
6 jams	40%	30%
24 jams	60%	3%

- A tasting booth in an upscale grocery store displayed, at different times, either 6 or 24 varieties of jam.
- The larger display drew more interest, but the purchases ran the other way.

Among those who stopped at the 6-jam table, the purchase rate was 10 times as high.

The Iyengar and Lepper (2000) jam experiment

THE NUMBERS

60% stopped at the 24 jams vs. 40% at the 6, yet only 3% of the first group bought, against 30% of the second.

Choice overload

Choice overload

A vast menu is attractive but can leave shoppers unable to decide and less satisfied with the option they pick.

WHAT IT MEANS

The result does not overturn the theory of preferences, but it shows that more options do not always leave a consumer better off. The effect is less general than one study suggests: pooling 50 experiments, Scheibehenne et al. (2010) found a mean effect near zero.

5.8

A Look into Behavioral Economics: Fairness and Social Preferences

The standard model assumes a consumer cares only about her own bundle.

The ultimatum game

- **Ultimatum game (Güth et al., 1982)**

  One player is given a sum, for instance $10, and proposes how to split it with a second player, who can accept (both are paid as proposed) or reject (both get nothing).

The ultimatum game

THE PUZZLE

A purely self-interested responder should accept any positive offer. Yet across hundreds of experiments, proposers most often offer close to half, and responders routinely reject low offers, so that both players walk away empty-handed.

Inequity aversion and social preferences

- **Inequity aversion (Fehr and Schmidt, 1999)**
  People dislike unequal outcomes, both when the inequality favors others and, to a lesser degree, when it favors themselves.

WHY IT MATTERS

Such social preferences explain behavior the narrow model misses: why people tip in restaurants they will never revisit, contribute to public broadcasting, and punish firms they regard as having priced unfairly, as with price gouging in Chapter 2.

The endowment effect

THE MUG EXPERIMENT

Kahneman et al. (1990) gave half a group of students a coffee mug, then opened a market. Standard theory predicts brisk trading; instead, far fewer trades occurred: owners demanded about twice as much to give up their mug as nonowners would pay for one.

The endowment effect

- **Endowment effect**

  Valuing a thing more simply because you own it: a form of loss aversion (Chapter 7). Giving up the mug feels like a loss, which looms larger than the equivalent gain. It helps explain why people hold onto possessions, stocks, and even opinions longer than the standard model of preferences would suggest.

CHAPTER 5

Key takeaways

1. **Ranking bundles**
   Completeness, transitivity, and more-is-better; utility is a ranking, not an amount.

2. **Marginal utility**
   Extra satisfaction from one more unit falls: the law of diminishing marginal utility.

3. **Indifference curves**
   Equally-liked bundles; downward-sloping, never crossing, bowed to the origin.

4. **The MRS**
   Willingness to trade; it diminishes because it equals the ratio of marginal utilities.

Key takeaways

5

Special cases

Perfect substitutes (straight lines), perfect complements (L-shapes), and the bowed middle.

6

Evidence

Razors & ink, choice overload (jam), the ultimatum game, and the endowment effect.

6

CHAPTER 6

Consumer Choice, Demand, and Welfare

What consumers want meets what they can afford, and out of it comes the demand curve and a measure of the gains from buying.

CHAPTER 6

What this chapter covers

6.2–6.3 The choice

- The budget constraint
- The best affordable bundle
- The “bang for the buck” rule

6.4–6.5 The payoff

- Income & price changes
- From choices to demand
- Consumer surplus & welfare

6.6–6.8 Evidence

- Case: Giffen goods
- Empirical: free digital goods
- Experimental tests of choice

CHAPTER 6

What this chapter covers

THE MOTIVATING QUESTION

$40 to spend, coffee at $4 a cup and donuts at $2 each: what is the best a consumer can do? The model in this chapter answers that, and explains where the demand curve comes from.

6.1 Introduction

- Three questions about choice
- Combining wants with means
- The road through the chapter

Three questions this chapter answers

Best affordable

With $40, coffee at $4 and donuts at $2, what is the best a consumer can do?

A price falls

Why does she usually buy more of a good, and could she ever buy less?

A free service

How much is a service worth when its price is zero, like a search engine?

Three questions this chapter answers

THE PLAN

We combine what consumers want, from Chapter 5, with what they can afford. Budget line (6.2), best bundle (6.3), responses to income and prices (6.4), and consumer surplus (6.5).

6.2

The Budget Constraint

- What income and prices allow
- Intercepts and slope
- Shifts vs. pivots

What a consumer can afford

Budget line

With $40 to spend, coffee at $4 a cup and donuts at $2 each: spend it all on coffee and buy $40 \div 4 = 10$ cups; spend it all on donuts and buy $40 \div 2 = 20$. Every combination that uses all of the $40 lies on the straight line between these two points.

What a consumer can afford

- Below the line: 2 coffees and 3 donuts cost $\$4 \times 2 + \$2 \times 3 = \$14$, less than \$40: affordable, but money is left unspent.
- Above the line: 8 coffees and 16 donuts cost $\$4 \times 8 + \$2 \times 16 = \$64$, more than \$40: out of reach.

The budget line for $40 of income

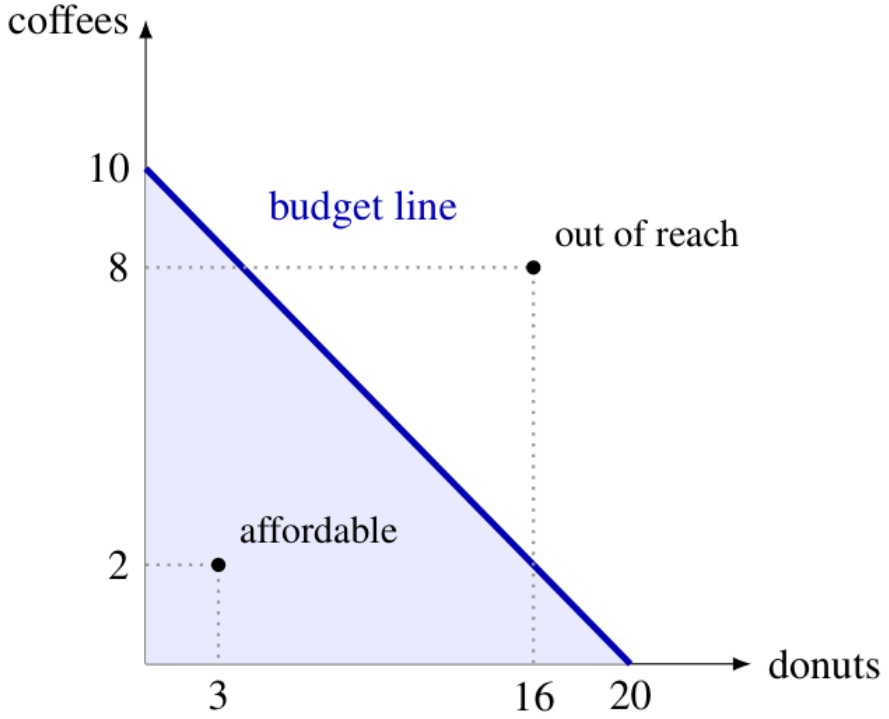


Figure 6.1 The budget line for $40 of income, with coffee at $4 and donuts at $2.

- Coffee intercept: 40 ÷ 4 = 10 cups. Donut intercept: 40 ÷ 2 = 20 donuts.
- Slope: 10 ÷ 20 = 1/2 cup of coffee per donut, the price ratio $2 ÷ $4.

The slope is the price ratio

SLOPE = PRICE RATIO

The slope of the budget line is the ratio of the prices. Since a donut costs $2 and a coffee $4, buying one more donut means giving up half a cup of coffee: 10 cups ÷ 20 donuts = 1/2 = $2 ÷ $4. Read the other way, a cup of coffee costs $4 ÷ $2 = 2 donuts.

- The slope is the opportunity cost, in coffee, of one more donut.
- It is set entirely by relative prices, not by income.

Two kinds of change: shift vs. pivot

Income change → shift

With $50 instead of $40, she could buy 50 ÷ 4 = 12.5 coffees or 50 ÷ 2 = 25 donuts: a parallel outward shift. The slope does not change.

Price change → pivot

If the donut price falls to $1, she can now buy 40 donuts rather than 20. The line swings out along the donut axis; the coffee intercept stays at 10, and one more donut now costs only 1/4 cup of coffee.

Two kinds of change: shift vs. pivot

THE DIFFERENCE

A change in income moves the line in or out without changing its slope; a change in a price pivots it.

A pivot and a shift

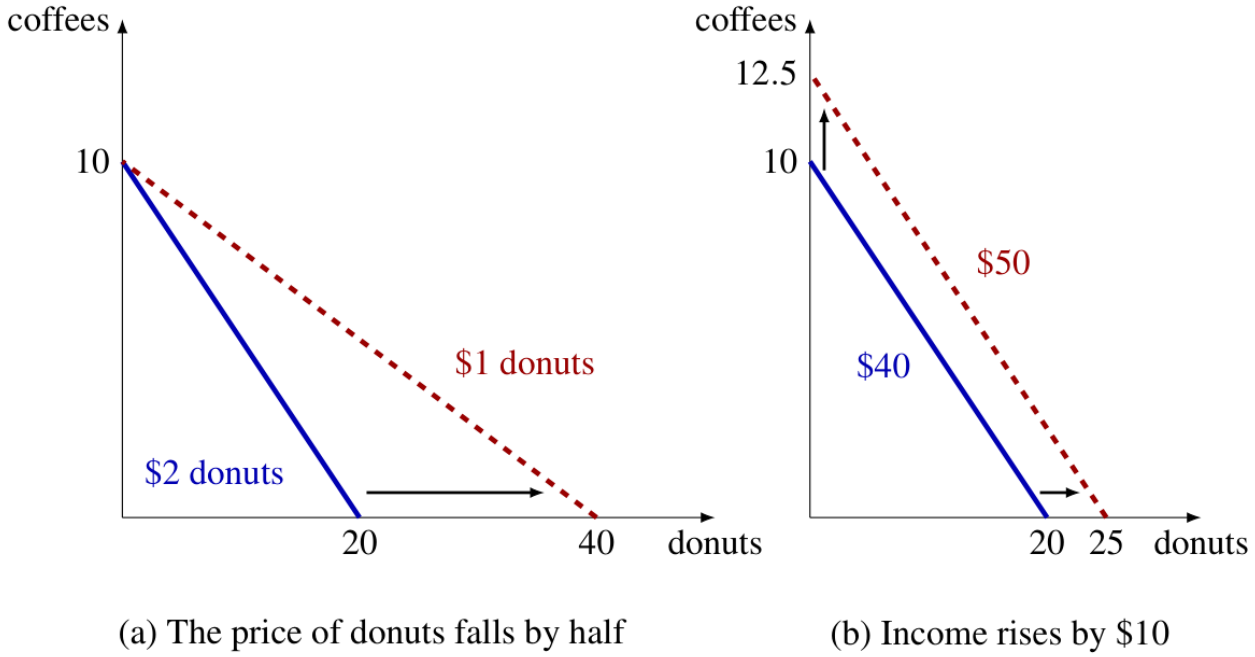


Figure 6.2 A fall in the price of donuts (left) and a rise in income (right).

- Left: the donut price falls from $2 to $1; the line pivots out to 40 donuts, and the coffee intercept stays at 10.
- Right: income rises from $40 to $50; the line shifts out in parallel, to 12.5 cups and 25 donuts.

6.3

The Best Affordable Bundle

- The highest reachable indifference curve
- MRS equals the price ratio
- The “bang for the buck” rule

Reaching the highest curve she can afford

- **The best affordable bundle (point E)**

  Combining the budget line with the preferences of Chapter 5, the consumer wants the highest indifference curve she can afford. That is bundle E, with 8 donuts and 6 cups of coffee, where her budget line just touches the highest indifference curve she can reach, IC1.

- Any bundle on a higher curve, such as IC2, is unaffordable.
- Any affordable bundle on a lower curve, such as IC0, leaves her less satisfied than she could be.

Where the budget line just touches the curve

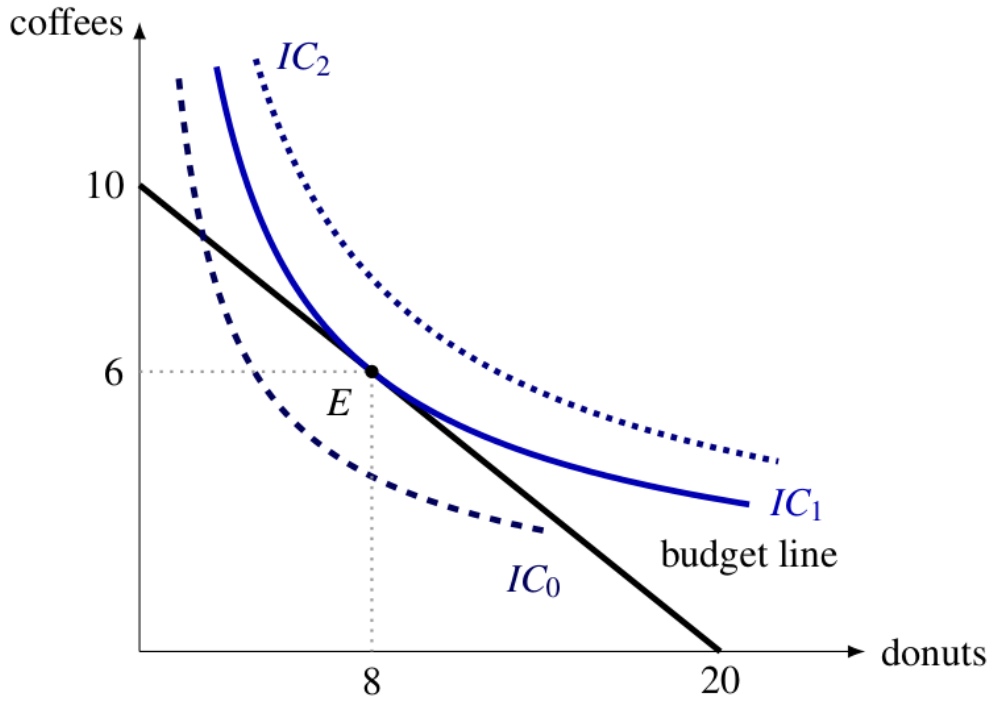


Figure 6.3 The best affordable bundle E.

- At E the budget line is tangent to the indifference curve.
- No affordable bundle she prefers exists; this is the optimum.

Explore Figure 6.3: intro-micro-companion-website.netlify.app/consumer-choice

At the optimum, MRS equals the price ratio

THE CONDITION

At the best bundle, the consumer's willingness to trade equals the market's: her marginal rate of substitution (the ratio of the two goods' marginal utilities) equals the price ratio: MRS = (price of donuts) ÷ (price of coffee).

INTUITIVELY

It is a simple arbitrage. If she would give up a whole cup of coffee for one more donut (her MRS), but the market only asks half a cup (the price ratio), donuts are a bargain and she buys more. She keeps adjusting until her personal valuation and the market price line up; only then is no improvement possible.

The “bang for the buck” rule

“Bang for the buck” rule

A rational consumer should get the same satisfaction from the last dollar she spends on each good. If a dollar on coffee brought more extra utility than a dollar on donuts, she could do better by moving a dollar to coffee; she shifts until marginal utility per dollar is equal across all goods.

SAME POINT, TWO WAYS

Equal marginal utility per dollar is just the “bang for the buck” version of “MRS equals the price ratio.” The two conditions are identical.

From the MRS to “bang for the buck”

- Since the MRS is the ratio of marginal utilities: (MU of donuts) ÷ (MU of coffee) = (price of donuts) ÷ (price of coffee).
- Multiply both sides by the MU of coffee, and divide both sides by the price of donuts:
- (MU of donuts) ÷ (price of donuts) = (MU of coffee) ÷ (price of coffee).

THE RULE

The marginal utility per dollar must be the same for both goods: the last dollar spent on each brings the same extra utility.

Bang for the buck guides spending

Coffee: cup	MU	MU per $	Donut	MU	MU per $
1	40	40/$4 = 10	1	20	20/$2 = 10
2	32	32/$4 = 8	2	16	16/$2 = 8
3	24	24/$4 = 6	3	12	12/$2 = 6

Table 6.1 Marginal utility per dollar guides the consumer's spending.

Bang for the buck guides spending

HOW SHE SPENDS

With $12 to spend, she buys in order of bang for the buck: the first cup and first donut each give 10 units per dollar, the second of each 8.

Two cups ($8) and two donuts ($4) use the $12; MU per dollar is 8 for both, so no reallocation helps: the condition of bundle E in Figure 6.3.

6.4

How the Best Bundle Responds to Income and Prices

- Income changes: normal vs. inferior
- A price fall: two effects
- Giffen goods previewed
- From choices to the demand curve

When income rises

Normal goods

With income \$40 → \$50, her best bundle moves from E (8 donuts, 6 coffees) to E′ (11 donuts, 7 coffees): she buys more of both.

Inferior goods

She buys less of them (instant noodles, bus rides) and switches to preferred alternatives.

When income rises

WHY THIS MATTERS

The real payoff of the model is that it predicts how purchases change when circumstances do: the comparative statics of Chapter 1.

A rise in income: both goods are normal

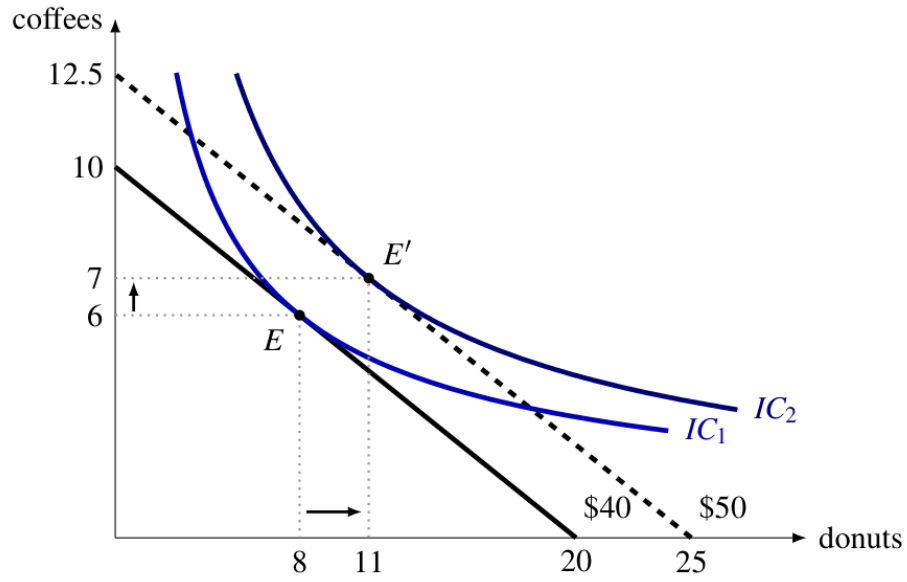


Figure 6.4 A rise in income from $40 to $50.

- Prices stay at $4 a cup and $2 a donut.
- E (8 donuts, 6 coffees) on IC1 moves to E′ (11 donuts, 7 coffees) on IC2.
- E′ costs $2 × 11 + $4 × 7 = $50. She buys more of both: both are normal goods.

A fall in the price of donuts

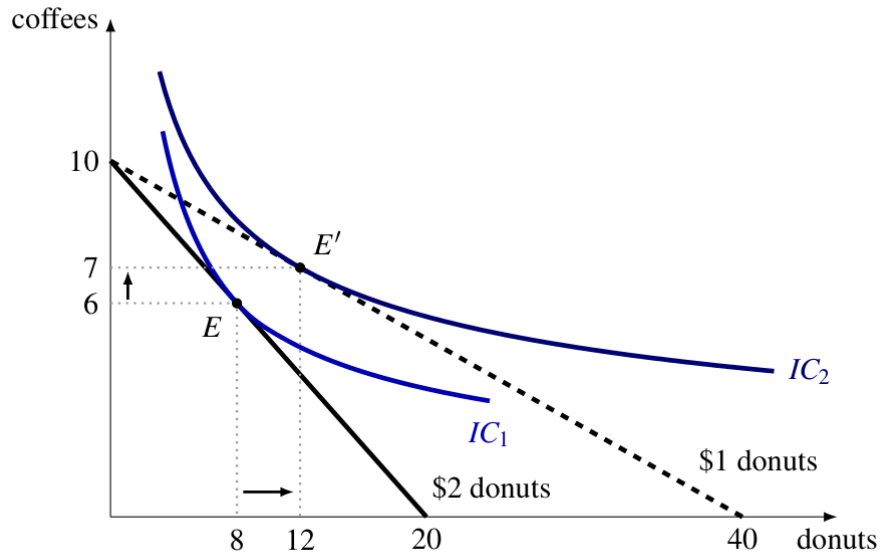


Figure 6.5 A fall in the price of donuts from $2 to $1.

- Donuts $2 → $1; income stays at $40 and coffee at $4.
- The best bundle moves from E (8 donuts, 6 coffees) to E′ (12 donuts, 7 coffees).
- E′ costs $1 × 12 + $4 × 7 = $40. She buys 4 more donuts.

A price fall splits into two effects

1. Substitution effect

The good is now less expensive relative to others, so the consumer substitutes toward it. This effect always raises the quantity bought of the good whose price fell.

A price fall splits into two effects

- **2. Income effect**

  A lower price stretches income: she can now afford more overall. For a normal good this reinforces the substitution effect; for an inferior good it works against it.

Explore Section 6.4.2: intro-micro-companion-website.netlify.app/income-substitution

Splitting the 4 extra donuts

- Substitution effect: 8 → 11 donuts. At the new prices, the least expensive way to stay on her original indifference curve is 11 donuts and 5 coffees, costing $1 × 11 + $4 × 5 = $31.
- Income effect: the lower price frees up $40 − $31 = $9, which adds 1 more donut, for 12 in total.

BOTH EFFECTS RAISE DONUTS

Donuts are a normal good for her, so the income effect reinforces the substitution effect: 3 + 1 = 4 more donuts.

Why demand slopes down, and the one exception

THE USUAL CASE

For most goods both effects push the same way, which is why demand curves slope downward.

THE GIFFEN EXCEPTION

Only for a strongly inferior good, where a large income effect overwhelms the substitution effect, could a lower price lead to less being bought (an upward-sloping demand curve); such a good is called a Giffen good. Whether such goods exist was open for over a century (Section 6.6).

Normal, inferior, and Giffen goods

Type of good	Substitution effect	Income effect	Total effect
Normal	buys more	buys more	buys more
Inferior	buys more	buys less, but by less than the substitution effect	buys more
Giffen	buys more	buys less, and by more than the substitution effect	buys less

Table 6.2 Substitution and income effects of a price fall.

Optimal choice builds the demand curve

- Lower the price of coffee step by step, holding income and the donut price fixed.
- At each price the consumer finds her best affordable bundle; as coffee becomes less expensive she chooses more of it.
- Plotting price against the quantity chosen builds up her individual demand curve (bottom panel of Figure 6.6), sloping downward, as the law of demand requires.

WHERE DEMAND COMES FROM

Adding up every consumer's demand gives the market demand curve of Chapter 2. Preferences, budgets, and optimal choice are the machinery behind that curve.

Building the demand curve for coffee

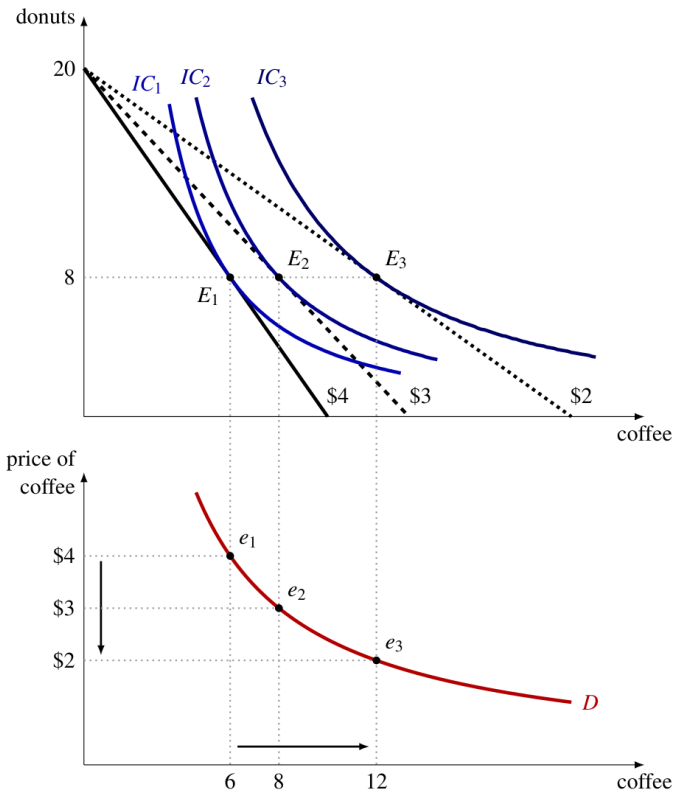


Figure 6.6 Building the demand curve for coffee: the best bundles (top) and the demand curve (bottom).

- Income $40, donuts $2; coffee falls from $4 to $3 to $2.
- The best bundle moves from E1 (6 cups, 8 donuts) to E2 (8 cups) to E3 (12 cups); each costs $40, as in $\$4 \times 6 + \$2 \times 8 = \$40$.

6.5

Consumer Surplus: Measuring the Gains from Buying

- What consumer surplus is
- A worked concert-ticket example
- The area below demand
- Why surplus matters

Measuring how much better off a buyer is

- **Consumer surplus**

  Buying is voluntary, so a buyer who makes a purchase must be at least as well off as before. Consumer surplus measures how much better off: the difference between the most a buyer would have been willing to pay and the price she actually pays.

It turns "the buyer gained" into a number we can add up and compare.

A worked example: five concert fans

THE SETUP

Five fans value a ticket at $50, $40, $30, $20, and $10. The price is $25.

The first three buy (each values it at $25 or more); the last two do not.

Fan value	Buys?	Surplus
$50	yes	**$50 − $25 = $25**
$40	yes	**$40 − $25 = $15**
$30	yes	**$30 − $25 = $5**
$20	no	—
$10	no	—

Total consumer surplus = $25 + $15 + $5 = $45.

Surplus is the area below demand, above price

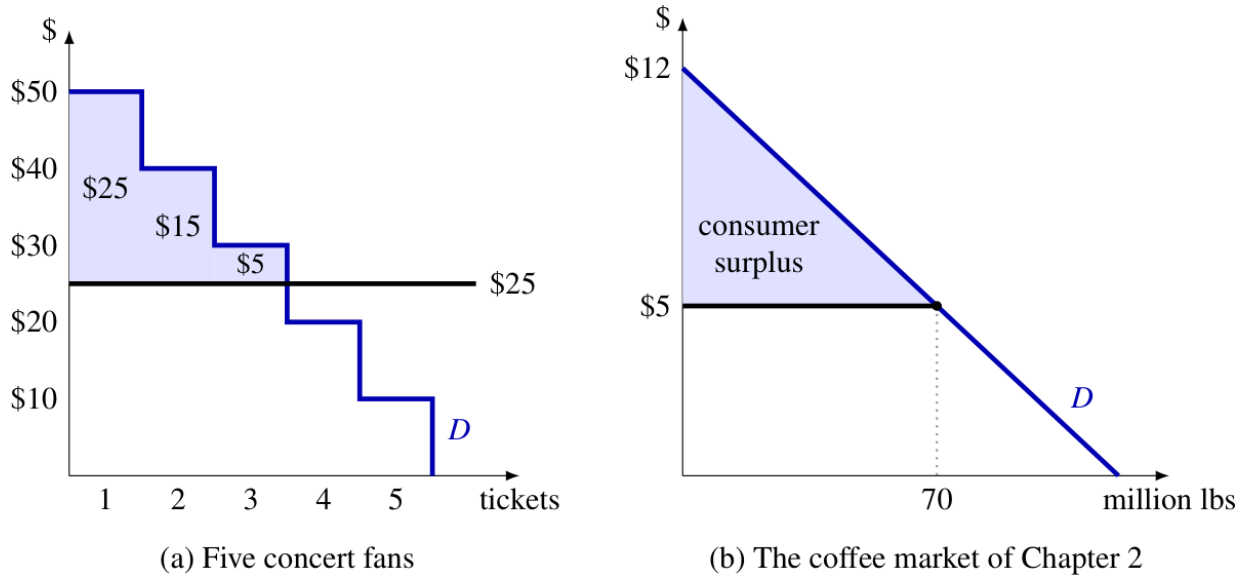


Figure 6.7 Consumer surplus for five concert fans (left) and for the coffee market (right).

Surplus is the area below demand, above price

- Left: the five fans' demand is a staircase; each shaded rectangle between a step and the $25 price is one fan's surplus: $25, $15, $5.
- Right: coffee market of Chapter 2 at $5: ½ × $7 × 70 million = $245 million.
- A price cut enlarges the area; a tax or price floor (Chapter 4) shrinks it, and the lost area is the harm to buyers.

A measure for judging policies

- When a tax raises a price, the loss of consumer surplus measures the harm to buyers.
- When competition or new technology lowers a price, the gain in surplus measures the benefit.
- Deadweight losses (Chapter 4) and the harm from monopoly (Chapter 11) are all measured as lost surplus.

6.5.1 WHY CONSUMER SURPLUS MATTERS

A measure for judging policies

TOTAL SURPLUS

Combined with producer surplus (Chapter 10), consumer surplus gives a single measure (total surplus) for judging whether a market makes the pie larger or smaller.

Self-check: a used textbook

SELF-CHECK

Four buyers value a used textbook at $80, $60, $40, and $20. The price is $45. Which buyers purchase, and what is the total consumer surplus?

The $80 and $60 buyers purchase (each value is at or above $45); the $40 and $20 buyers do not. Consumer surplus is ($80−$45) + ($60−$45) = $35 + $15 = $50.

6.6

Case Study: The Search for Giffen Goods

A century-old curiosity, finally found in real data.

6.6 CASE STUDY

A century-old puzzle

Giffen good

A good people buy more of when its price rises. For more than a century it was a theoretical curiosity no one had convincingly found in real data.

WHAT THE THEORY REQUIRED

The theory said it required a staple that takes up a large share of a poor household's budget and has few substitutes, so that the income effect is both large and negative.

Jensen and Miller (2008) finally found it

- A field experiment in two Chinese provinces randomly gave poor households vouchers subsidizing their main staple (rice in Hunan, wheat flour in Gansu).
- For poor households in Hunan (though not the very poorest, who had no room to substitute toward other foods), lowering the price of rice led them to consume less (equivalently, a higher price would have led them to consume more): clear Giffen behavior.

6.6 CASE STUDY

Jensen and Miller (2008) finally found it

WHY IT HAPPENED

The mechanism is the income effect of Section 6.4.2. Rice was such a large part of these families' budgets that the subsidy made them effectively richer; feeling richer, they spent some of the gain on meat and vegetables, reducing rice consumption even though rice had become less expensive.

6.6 CASE STUDY

Real, but only under precise conditions

THE VERDICT

The evidence was strong for rice in Hunan and weaker for wheat in Gansu, where wheat had closer substitutes in the diet and more households were so poor that they ate little besides staples. After a century of doubt, the Giffen good turned out to be real, but only under the precise conditions the theory specified.

6.7

Empirical Evidence: The Value of "Free" Digital Goods

Surplus that is enormous, and invisible in GDP.

How do you price something that's free?

THE PROBLEM

Consumer surplus is usually hidden, but it is especially large, and especially invisible, for digital goods priced at zero. How can we measure the benefit of a search engine or a social network that costs nothing?

- Brynjolfsson et al. (2019) ran large online experiments.
- They asked the smallest payment people would accept to give up a service for a period of time: a direct measure of the surplus it provides.

The numbers are large

$48
to give up Facebook for one month (median)

~$17,500/yr
needed to do without search engines

~$8,400/yr
needed to do without email

6.7 EMPIRICAL EVIDENCE

The numbers are large

WHY IT'S INVISIBLE

Because these services are free, none of this surplus shows up in the prices people pay, and so almost none appears in GDP. The chapter's two ideas at once: the gain from a good is its value minus its price, and when the price is zero that gain can be enormous.

6.8

A Look at the Experimental Evidence

Testing the theory of choice in the laboratory.

6.8 EXPERIMENTAL EVIDENCE

Does the theory of choice hold up?

THE ASSUMPTION

Consumer theory relies on a strong assumption: that people have stable, consistent preferences and choose the bundle they most prefer. Economists have tested it in the lab, with real money on the line.

THE VERDICT

The verdict is mixed in an interesting way: the theory describes most choices well, but it fails in systematic, repeatable ways that have reshaped the field.

Four experimental tests of standard choice

Study	What it tested	Method	Key result
Harbaugh et al. (2001)	Choices are consistent, even for children	Children aged 7 and 11, and college students, chose bundles of juice and chips from 11 menus	Inconsistencies averaged 4.3 at age 7 and about 2 at age 11, as for adults, far fewer than random choice (about 9)
Andreoni and Miller (2002)	Generous choices are also consistent	Subjects split tokens with another person under 8 budgets	Over 98 percent chose consistently; only a quarter were purely selfish
Kahneman et al. (1990; Section 5.8)	Value is independent of ownership	Half a group was given a mug, then a market was opened	Owners demanded about twice what buyers would pay (the endowment effect)
Iyengar and Lepper (2000; Section 5.7)	More options make choosers better off	A store displayed 6 or 24 jams to shoppers	The larger display drew more interest but one-tenth the purchase rate

Table 6.3 Four experimental tests of the standard theory of consumer choice.

Boundaries, not a refutation

THE MODEL STILL WORKS

The first two experiments support the theory: most choices, even by 11-year-olds and even when people are being generous, are consistent with stable preferences and the best-affordable-bundle rule of Section 6.3.

- The last two mark boundaries of the model: preferences can depend on what one already owns and on how many options are presented.
- Both are part of the behavioral economics that complements the standard model throughout this book.

CHAPTER 6

Key takeaways

1. **Budget constraint**

   Income sets the line's position; relative prices set its slope. Income shifts it; a price pivots it.

2. **Best bundle**

   Where the budget line touches the highest indifference curve; MRS equals the price ratio.

3. **“Bang for the buck” rule**

   Equalize marginal utility per dollar across goods; the same point, stated with marginal utility.

Key takeaways

4 **Income & substitution**

A price fall splits into both effects; together they make demand slope down (Giffen is the exception).

5 **From choice to demand**

Tracing the best bundle as price falls builds the demand curve of Chapter 2.

6 **Consumer surplus**

Value minus price (the area below demand and above price), a measure of welfare; Giffen, free goods, and lab tests.

7

CHAPTER 7

Choice under Risk and Uncertainty

Why the same person buys insurance and lottery tickets, and why losing $100 stings more than finding $100 feels good.

CHAPTER 7

What this chapter covers

7.2–7.3 Describing risk

- Expected value & spread
- Expected utility
- Attitudes toward risk

7.4–7.5 Measuring & managing

- Risk premium & CE
- Why insurance exists
- Diversification

7.6–7.8 Puzzles & evidence

- Why people gamble
- Paying to avoid small risks
- Prospect theory

CHAPTER 7

What this chapter covers

THE BIG IDEA

Once outcomes cannot be predicted in advance, choices gain a new dimension. This chapter builds the tools to understand choice when the future is risky.

7.1

Introduction

- A sure $50, or a coin flip for $100 or nothing?
- Why buy lottery tickets that pay back less than they cost?
- Why losing hurts more than winning pleases

Three puzzles about risky choice

The coin flip

Would you rather have $50 for certain, or flip a coin for $100 or nothing?

Insure and gamble

Many people buy lottery tickets that pay back far less than their price, and also buy insurance.

Losses sting more

Why does losing $100 hurt more than finding $100 feels good?

Three puzzles about risky choice

WHAT TIES THEM TOGETHER

Risk means outcomes a person cannot predict in advance. Allowing for risk gives choice a new dimension, and this chapter builds the tools to understand it.

7.2

Describing Risk: Expected Value and Spread

- A lottery: any situation with uncertain outcomes
- Expected value: the average payoff
- Spread: how risky the gamble is

Two numbers summarize a risky prospect

- **Lottery**

  Any situation with uncertain outcomes, each occurring with some probability: a coin flip, an investment, a crop that depends on the weather.

- **Expected value (EV)**

  The average payoff, weighting each outcome by its probability. A coin flip paying \$100 on heads and \$0 on tails has EV = ½(\$100) + ½(\$0) = \$50.

Expected value is not the whole story

SAME AVERAGE, DIFFERENT RISK

A sure $50 and the coin flip both have an expected value of $50, but the coin flip is risky and the sure thing is not. The EV tells us what a gamble pays on average, not how risky it is.

- For risk we look at the spread of outcomes.
- Two gambles with the same expected value can carry very different risk.
- A person's choice between them reveals how she feels about risk.

Measuring spread: variance and standard deviation

- **Variance:** the probability-weighted average of the squared distance between each outcome and the expected value.
- For the coin flip, each outcome lies $50 from the expected value:

$$\text{variance} = \tfrac{1}{2}(\$100 - \$50)^2 + \tfrac{1}{2}(\$0 - \$50)^2 = 2{,}500$$

- **Standard deviation:** its square root, in dollars: $\sqrt{2{,}500} = \$50$. A sure $50 has a variance and standard deviation of zero.

Two gambles, same average, different spread

Gamble A

Pays $45 or $55 on a coin flip. Outcomes cluster tightly around $50, so you end up close to the average regardless of the outcome.

Gamble B

Pays $0 or $100 on a coin flip. Outcomes spread far apart: you might walk away with nothing, or with twice the average.

Two gambles, same average, different spread

THE TAKEAWAY

Both have an expected value of $50. The wider the spread of outcomes around the expected value, the riskier the prospect. Most people, offered A or B for free, prefer the tighter gamble A: a first clue that they dislike risk.

The variance of each gamble

- **Same expected value:** ½($45) + ½($55) = $50 and ½($0) + ½($100) = $50.
- **Gamble A:** each outcome lies only $5 from the expected value.

$$\text{variance} = \tfrac{1}{2}(\$45 - \$50)^2 + \tfrac{1}{2}(\$55 - \$50)^2 = 25$$

- Gamble A: standard deviation $\sqrt{25} = \$5$. Gamble B: variance 2,500 and standard deviation $50, 10 times larger.

7.3

Expected Utility and Attitudes Toward Risk

- Weighing utility, not dollars
- Diminishing marginal utility of wealth
- Risk averse, neutral, and loving

People weigh utility, not dollars

Expected utility

To choose among risky prospects, a person weighs not the dollar payoffs but the utility those dollars bring, averaging utility across outcomes.

THE KEY FACT

For most people each additional dollar brings a little less extra satisfaction than the last: the diminishing marginal utility of Chapter 5, now applied to wealth. This single fact drives everything that follows.

A person with diminishing marginal utility

Point	Wealth	Utility	Marginal utility per dollar
A	$0	0	—
B	$25	5	(5 − 0)/($25 − $0) = 0.20
C	$50	7	(7 − 5)/($50 − $25) = 0.08
D	$100	10	(10 − 7)/($100 − $50) = 0.06

Table 7.1 A person with diminishing marginal utility of wealth.

- The rows are not equally spaced, so the last column divides each increase in utility by the increase in wealth.
- The extra utility per dollar falls as wealth rises.

The utility of wealth

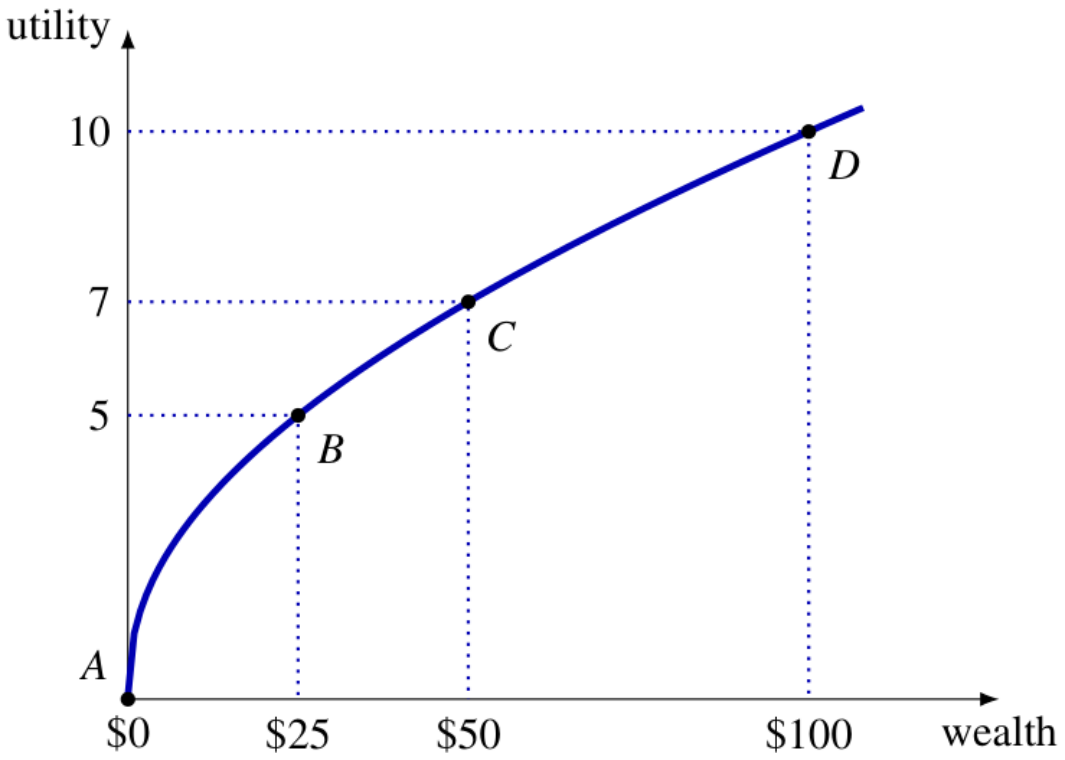


Figure 7.1 The utility of wealth in Table 7.1.

- Figure 7.1 plots the four points of Table 7.1.
- The curve is concave: it rises at a decreasing rate.
- Each extra dollar adds less utility than the one before.

Why this person prefers the sure thing

Option	Expected value	Expected utility
Sure $50	$50	7
Coin flip $0 / $100	$50	5

Sure $50 yields utility 7; the coin flip yields ½(0)+½(10)=5.

RISK AVERSION

Both have the same expected value of $50, but the sure thing gives more expected utility, so this person prefers it. The gain in utility from winning the extra $50 is smaller than the loss from ending up with nothing.

Risk aversion follows from diminishing marginal utility

utility of the sure \$50 = 7 > 5 = expected utility of the coin flip

- **Gain from winning:** from \$50 to \$100, utility rises by only $10 - 7 = 3$.
- **Loss from losing:** from \$50 to \$0, utility falls by $7 - 0 = 7$.
- In Table 7.1, each dollar between \$50 and \$100 adds only 0.06; each dollar below \$50 adds 0.08 or more.

Three attitudes toward risk

Risk averse

Prefers a sure amount to a risky prospect of equal expected value. Utility is concave (diminishing MU).

Risk neutral

Cares only about expected value; indifferent between the sure $50 and the coin flip. Utility is a straight line.

Risk loving

Actually prefers the gamble. Marginal utility of wealth is increasing; utility is convex.

Three attitudes toward risk

It Is All in the Curvature

The attitude shows up as the curvature of the utility-of-wealth curve: concave, straight, or convex.

Risk neutral and risk loving: Table 7.2

Risk neutral

Wealth	Utility	Marginal utility per dollar
$0	0	—
$25	2.5	(2.5 − 0)/($25 − $0) = 0.10
$50	5	(5 − 2.5)/($50 − $25) = 0.10
$100	10	(10 − 5)/($100 − $50) = 0.10

Risk loving

Wealth	Utility	Marginal utility per dollar
$0	0	—
$25	1	(1 − 0)/($25 − $0) = 0.04
$50	3	(3 − 1)/($50 − $25) = 0.08
$100	10	(10 − 3)/($100 − $50) = 0.14

Table 7.2 A risk-neutral person and a risk-loving person.

The sure $50 or the coin flip?

- **Risk neutral:** each dollar adds the same 0.10. The sure $50 gives utility 5, and the coin flip gives ½(0) + ½(10) = 5: she is indifferent.
- **Risk loving:** marginal utility rises from 0.04 to 0.08 to 0.14. The sure $50 gives utility 3, but the coin flip gives ½(0) + ½(10) = 5: she prefers the gamble.

Risk attitudes as the shape of the curve

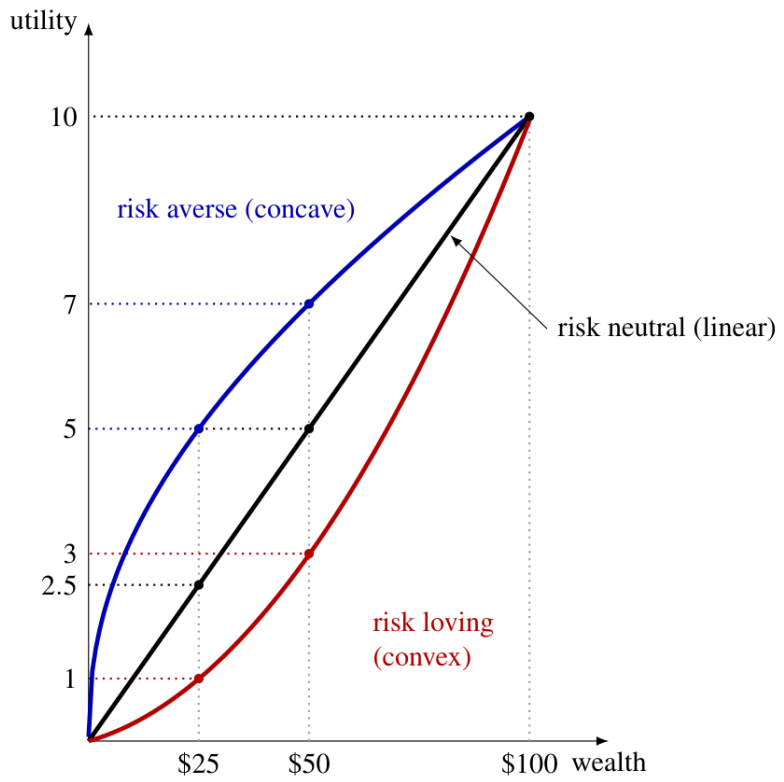


Figure 7.2 Risk attitudes and the shape of the utility-of-wealth curve.

Concave. Risk averse: each extra dollar adds less utility.

Linear. Risk neutral: utility rises in a straight line.

Convex. Risk loving: each extra dollar adds more utility.

7.4

Measuring Risk Aversion: Risk Premium and Certainty Equivalent

- Certainty equivalent: the sure amount that ties the gamble
- Risk premium: what she would give up to avoid risk

The certainty equivalent

Certainty equivalent

The sure amount of money that makes a person as well off as facing the gamble.

WORKED FROM TABLE 7.1

First, the coin flip yields expected utility ½(0) + ½(10) = 5, using points A and D of Table 7.1. Second, the sure wealth that also gives utility 5 is $25, at point B. So her certainty equivalent for the coin flip is $25: she would trade the gamble for a guaranteed $25.

The risk premium

Risk premium

The difference between the gamble's expected value and its certainty equivalent. Here $50 − $25 = $25.

WHAT IT MEANS

It is the amount she would willingly give up to avoid the risk: she values a coin flip worth $50 on average at only $25, and would pay as much as $25 to swap it for the sure thing. The more risk averse the person, the larger her risk premium.

Both measures on the concave curve

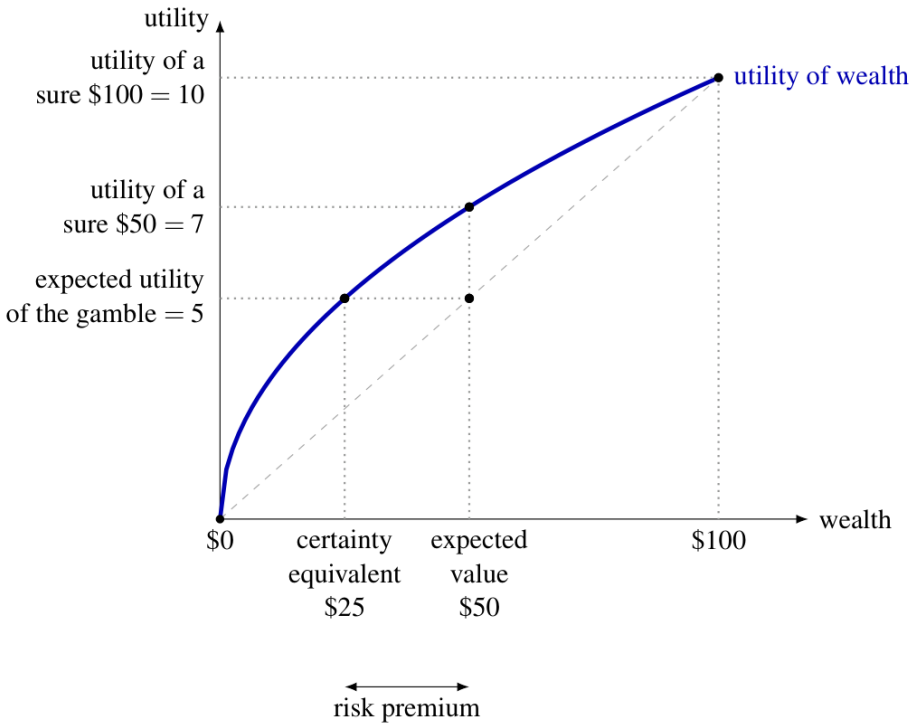


Figure 7.3 The certainty equivalent and the risk premium of a 50–50 gamble between $0 and $100.

- The certainty equivalent ($25) is the sure wealth giving the same utility as the gamble.
- The risk premium ($25) is the gap between expected value ($50) and the certainty equivalent.

Explore Figure 7.3: intro-micro-companion-website.netlify.app/risk-premium

Self-check: a quick calculation

SELF-CHECK

A coin flip pays $100 on heads and $0 on tails. What is its expected value? If a risk-averse person would trade it for a sure $30, what is her risk premium?

Expected value = ½($100) + ½($0) = $50. Her certainty equivalent is $30, so the risk premium is $50 − $30 = $20: what she would give up to avoid the risk.

7.5

Why Insurance Exists

- Risk aversion explains insurance
- A premium just above the expected loss
- Why both sides gain

Risk aversion explains insurance

- A risk-averse person facing a possible large loss (a house fire, a car accident, a costly illness) is willing to pay more than the expected loss to be rid of the risk: up to the expected loss plus her risk premium.
- An insurer that covers many independent risks faces little overall risk (Section 7.5.1 explains why).
- It can charge a premium slightly above the expected loss, cover its costs, and still leave the customer better off.

Risk aversion explains insurance

BOTH SIDES GAIN

Both sides gain. The customer converts an uncertain large loss into a small certain one. The same logic underlies health insurance, crop insurance, and appliance warranties.

Insurance with numbers: the uninsured position

- The person of Table 7.1 has wealth \$100, but a fire destroys all of it with probability ½.
- **Step 1.** Expected loss: ½(\$100) + ½(\$0) = \$50.
- **Step 2.** Expected utility without insurance: ½(0) + ½(10) = 5.
- **Step 3.** A sure \$25 also gives utility 5 (point B): being uninsured is as good as a sure \$25.

The most she would pay for insurance

- **Step 4.** A policy covering the whole loss leaves her a sure \$100 minus the premium. She buys it if this sure amount is at least her certainty equivalent of \$25: for any premium up to \$100 − \$25 = \$75.
  - A \$60 premium leaves a sure \$40, more than \$25: she buys.
  - An \$80 premium leaves only \$20: she does not.
- **Step 5.** The \$75 is the expected loss plus her risk premium: \$50 + \$25 = \$75.

7.5.1 DIVERSIFICATION & RISK POOLING

Don't put all your eggs in one basket

- **Diversification**

  Spreading savings across many companies so the unrelated ups and downs largely cancel out. The expected return is the same, but the spread (the risk) is far smaller.

- **Risk pooling**

  The same logic for insurers: in any year a predictable fraction of houses burn or cars crash, so the insurer's total payout is steady even though each customer faces a frightening individual risk.

Two faces of the same idea

THE UNIFYING PRINCIPLE

Combining many independent risks shrinks the risk of the whole. Risk pooling and diversification are two faces of one idea, and together they underpin insurance markets, mutual funds, and the advice every financial adviser gives.

7.6

Case Study: Why Do People Buy Lottery Tickets?

If people are risk averse, why buy pure risk?

7.6 CASE STUDY

Lotteries are, on average, terrible bets

~67¢

returned in prizes per $1 wagered in U.S. state lotteries

~$1

expected value of a $2 Powerball ticket before taxes (only ~32¢ from the smaller prizes)

THE PUZZLE

Powerball pays out about half of its sales as prizes, a far worse return than a casino. So why do so many people buy tickets that are pure risk?

7.6 CASE STUDY

Three forces reconcile this with theory

1

Tiny stakes

For the price of a coffee the loss is trivial, while the prize is life-changing; people may be locally risk loving over a small chance of a huge gain.

Three forces reconcile this with theory

2

Overweighting

Players overweight very small probabilities, treating one-in-292-million as larger than it is (Section 7.8).

3

Entertainment

The ticket buys the pleasure of anticipation, not just a financial bet.

7.6 CASE STUDY

Three forces reconcile this with theory

A NOTE ON WHO PLAYS

Lottery spending takes a larger share of income from lower-income households, which is why economists call lotteries a regressive way to raise revenue.

Explore the Data: look up the current Powerball jackpot and odds, and compute the expected value of a $2 ticket.

How can one person do both?

THE PUZZLE

The same person who pays for insurance (a sign of risk aversion) may also buy a lottery ticket (a sign of risk loving). How can one person be both?

Friedman and Savage (1948). A person's attitude toward risk may differ across ranges of wealth.

Over most wealth, risk averse. She willingly insures against losses that would set her back.

For a tiny sum, risk loving. She gambles a small sum, such as the $2 of a lottery ticket, on a remote chance of a life-changing jackpot.

Not contradictory, but complementary

THE RESOLUTION

She insures the wealth she has and gambles a trivial amount on the remote chance of wealth she could never otherwise reach.

7.7

Empirical Evidence: How Much Do People Pay to Avoid Small Risks?

Sydnor (2010) on homeowners' insurance deductibles.

A lower deductible is a small insurance policy

- If gambling shows people accepting risk, insurance choices show many of the same people paying a high price to avoid it, sometimes far too high a price.
- Sydnor (2010) examined the deductibles homeowners chose on their insurance policies.
- A lower deductible works as a small insurance policy: pay a higher premium now, and the insurer covers more of any future claim.

Paying five dollars to avoid one dollar of expected loss

83%

chose a deductible below the available maximum

≈5×

the extra premium for a lower deductible, relative to its expected value

Paying five dollars to avoid one dollar of expected loss

THE FINDING

The extra premium homeowners paid for the lower deductible was about five times its expected value (the extra claim payments they could expect to receive). Paying about five dollars to avoid one dollar of expected loss implies a degree of risk aversion over modest sums that standard expected-utility models cannot plausibly explain.

A numerical example: two deductibles

- A homeowner chooses between a $1,000 and a $500 deductible.
- In Sydnor's data, the lower deductible cost about $100 more per year.
- The chance of a claim in a year was below 5 percent: take 4 percent.
- **Step 1.** If a claim occurs, the lower deductible makes the insurer pay $1,000 − $500 = $500 more.

The deductible example, continued

- **Step 2.** A claim occurs with probability 0.04, so the lower deductible is worth, on average, $0.04 \times \$500 = \$20$ per year.
- **Step 3.** She pays \$100 for coverage with an expected value of \$20: \$100/\$20 = 5 times its expected value.
- Her risk premium for this small risk is at least $\$100 - \$20 = \$80$.

A broad pattern that points to behavior

THE BIGGER PICTURE

Faced with small but salient risks, people are willing to overpay dramatically for peace of mind, a pattern that points toward the behavioral ideas we turn to next.

7.8

A Look into Behavioral Economics: Prospect Theory and Loss Aversion

Kahneman and Tversky (1979): gains and losses from a reference point.

From final wealth to gains and losses

- **Prospect theory (Kahneman and Tversky, 1979)**

  The expected-utility model assumes people evaluate outcomes by their final level of wealth. Prospect theory instead has people evaluate gains and losses relative to a reference point, usually where they are now.

THREE DEPARTURES

Prospect theory departs from the model in three ways: a reference point; risk aversion over gains, risk seeking over losses, and loss aversion; and probability weighting.

The prospect-theory value function

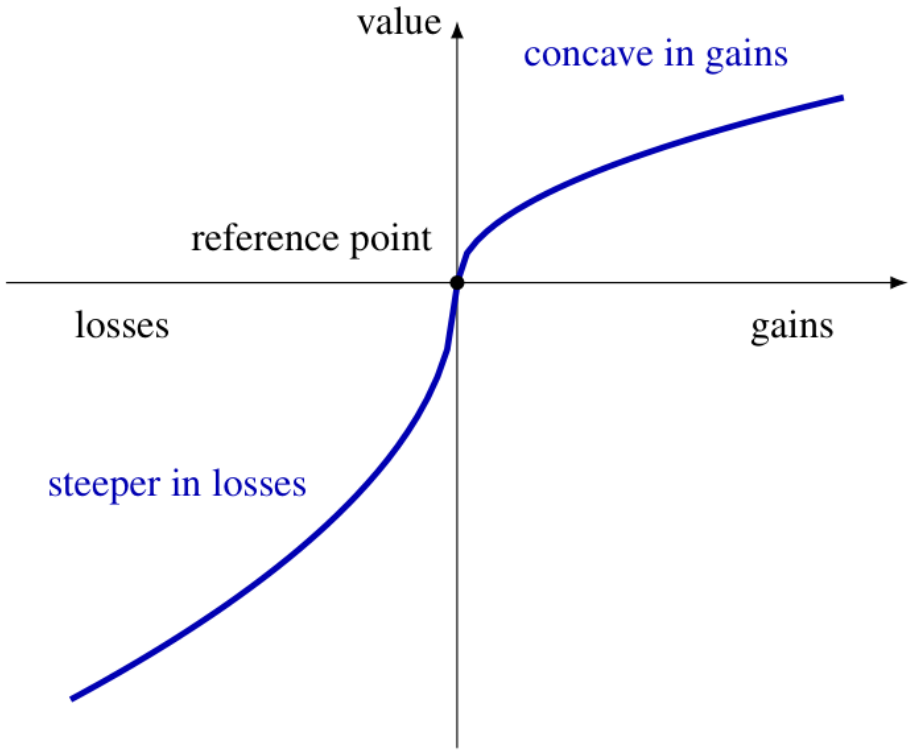


Figure 7.4 The prospect-theory value function.

Concave over gains. Diminishing sensitivity to bigger wins.

Convex over losses. Diminishing sensitivity to bigger losses.

Steeper on the loss side. The same dollar loss outweighs the gain.

Nobel 2002. Prospect theory earned Daniel Kahneman the prize.

The reference point and framing

- People evaluate gains and losses relative to a reference point, usually where they are now.
- In Figure 7.4, the reference point is the origin: gains lie to its right and losses to its left.
- **Framing:** because the reference point depends on how a choice is described, the same facts presented as a gain or as a loss can lead to different decisions.

Loss aversion and probability weighting

Loss aversion

Losses hurt more than equal gains please. A loss looms about 2.25× as large as an equivalent gain (Tversky and Kahneman, 1992).

Probability weighting

People overweight rare events and underweight near-certain ones.

Loss aversion and probability weighting

WHAT THEY EXPLAIN

Loss aversion explains why people refuse a 50–50 bet to win $110 or lose $100, and why 82% of the survey respondents in Chapter 2 judged the snow-shovel price increase unfair. Probability weighting explains both lottery jackpots and the overpurchase of insurance against unlikely losses (Section 7.7).

CHAPTER 7

Key takeaways

1. **Describing risk**
   Summarize a gamble by its expected value (average payoff) and the spread of its outcomes.

2. **Expected utility**
   People weigh utility, not dollars; diminishing marginal utility makes most people risk averse.

3. **Three attitudes**
   Risk averse (concave utility), risk neutral (linear), risk loving (convex).

CHAPTER 7

Key takeaways

4 **Measuring it**

Certainty equivalent and risk premium gauge how strongly someone dislikes risk.

5 **Managing it**

Insurance, diversification, and risk pooling all shrink risk by combining independent risks.

6 **Behavior**

Lotteries, over-insuring (Sydnor, 2010), and prospect theory's loss aversion and probability weighting.

8

CHAPTER 8

Production

How firms turn inputs into output, from a food truck to a tens-of-billions-of-dollars chip foundry.

CHAPTER 8

What this chapter covers

8.2–8.3 Short run

- The production function
- Total, marginal, average product
- Diminishing returns

8.4–8.5 Long run

- Isoquants & substitution
- Returns to scale
- Scale vs. diminishing product

8.6–8.8 Progress

- Kinds of technological progress
- Case: chip foundry (TSMC)
- Empirical: Moore's law

CHAPTER 8

What this chapter covers

THE BIG IDEA

Production is how firms turn inputs into output: the foundation for the analysis of costs in Chapter 9.

8.1 Introduction

- Why some goods need vast input combinations
- Why output eventually rises more slowly
- How chips became millions of times more powerful

Three questions about production

Scale

Why does a chip factory cost tens of billions while a food truck needs one cook and a griddle?

Crowding

When a restaurant crowds more cooks into one kitchen, why does output eventually rise more slowly?

Progress

How have computer chips become millions of times more powerful within a single lifetime?

Three questions about production

WHAT TIES THEM TOGETHER

These questions are all about production (how firms turn inputs into output), and they set up the analysis of costs in Chapter 9.

8.2

The Production Function and the Short Run

- Inputs: labor and capital
- The short run vs. the long run

Inputs in, output out

- **Production function**

  Describes the maximum output a firm can produce from each combination of inputs.

- **Two broad kinds of input**

  Labor: the workers. Capital: the machines, buildings, and equipment.

8.2 PRODUCTION FUNCTION

Two time horizons

Short run

At least one input is fixed, typically capital, since a firm cannot build a new factory overnight.

Long run

Every input can be varied: the firm can build plants, install machines, and hire freely.

THE KEY DISTINCTION

A restaurant can add cooks tonight but cannot enlarge its kitchen until next year.

Fixed and variable inputs

- **Fixed input:** one the firm cannot change in the short run, like the restaurant's kitchen.
- **Variable input:** one the firm can adjust right away, like the number of cooks.
- In the long run, every input can be varied.

8.3

Total, Marginal, and Average Product

- The production table
- Marginal & average product
- Diminishing marginal product
- The shape of the curves

Adding cooks to a fixed kitchen

Workers	Total product	Marginal product	Average product
0	0	—	—
1	10	10 − 0 = 10	10/1 = 10.0
2	24	24 − 10 = 14	24/2 = 12.0
3	39	39 − 24 = 15	39/3 = 13.0
4	50	50 − 39 = 11	50/4 = 12.5
5	58	58 − 50 = 8	58/5 = 11.6
6	63	63 − 58 = 5	63/6 = 10.5

Table 8.1 Total, marginal, and average product as a kitchen adds cooks.

- In the short run, with capital fixed, the firm raises output (the total product) by adding labor.

Two derived numbers do most of the work

- **Marginal product of labor**

  The extra output from one more worker: the third cook adds $39 - 24 = 15$ meals.

- **Average product**

  Total output divided by the number of workers: with three cooks producing 39 meals, the average is $39/3 = 13$ meals per cook.

The law of diminishing marginal product

THE LAW

As more of one input is added to fixed amounts of others, the extra output from each additional unit eventually shrinks.

- With the kitchen fixed, the fourth, fifth, and sixth cooks crowd the same stoves and counters.
- So each adds less output than the one before: marginal product falls from 15 to 11 to 8 to 5.

Total, marginal, and average product

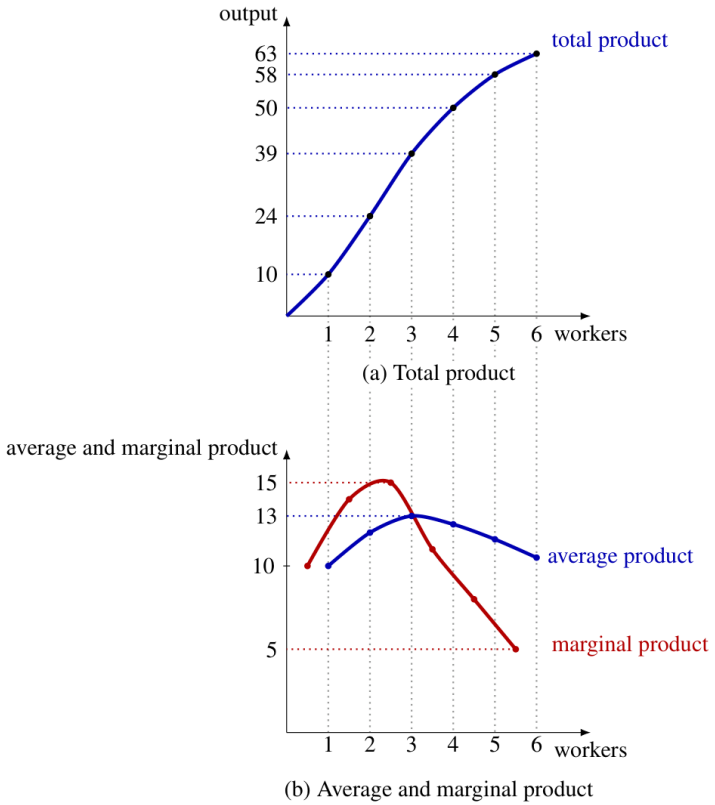


Figure 8.1 Total product (top) and average and marginal product (bottom), from Table 8.1.

- **Top:** total product steepens, then flattens.
- **Bottom:** marginal product rises, then falls.

Explore Figure 8.1: intro-micro-companion-website.netlify.app/production

Marginal product cuts average product at its peak

- In the bottom panel of Figure 8.1, the marginal product curve cuts the average product curve at the latter's peak.
- When the next worker is more productive than the current average, the average rises; when less, it falls.
- The dotted vertical lines link total product (top) with average product (bottom) at each number of workers; each marginal product point lies halfway between two numbers of workers, since it measures the change from one to the next.

Why the curve steepens, then flattens

Early: rising

In the top panel of Figure 8.1, specialization makes the marginal product of the first cooks rise, so the total product curve steepens.

Later: falling

Then crowding of the fixed inputs (too many cooks in the kitchen) makes marginal product fall, and the curve flattens: an inverted U shape for marginal and average product.

Why the curve steepens, then flattens

THE EXTREME CASE

A kitchen so crowded that workers trip over one another could even push marginal product to zero or below: an extra cook then lowers total output.

Self-check: reading Table 8.1

SELF-CHECK

The fifth cook raises output from 50 to 58 meals. What is the marginal product of the fifth cook, and the average product with five cooks?

Marginal product = 58 − 50 = 8 meals. Average product = 58 ÷ 5 = 11.6 meals per cook.

8.4

The Long Run: Substituting Between Inputs

- The same output from different input mixes
- Isoquants
- The marginal rate of technical substitution

Different mixes, same output

- **Isoquant (“equal quantity”)**
  Shows all the input combinations that produce a given level of output.

THE LONG-RUN CHOICE

In the long run a firm can vary both labor and capital, and the same output can usually be produced with different mixes: a car can be assembled by many workers with simple tools, or a few workers with many robots.

Reading an isoquant map

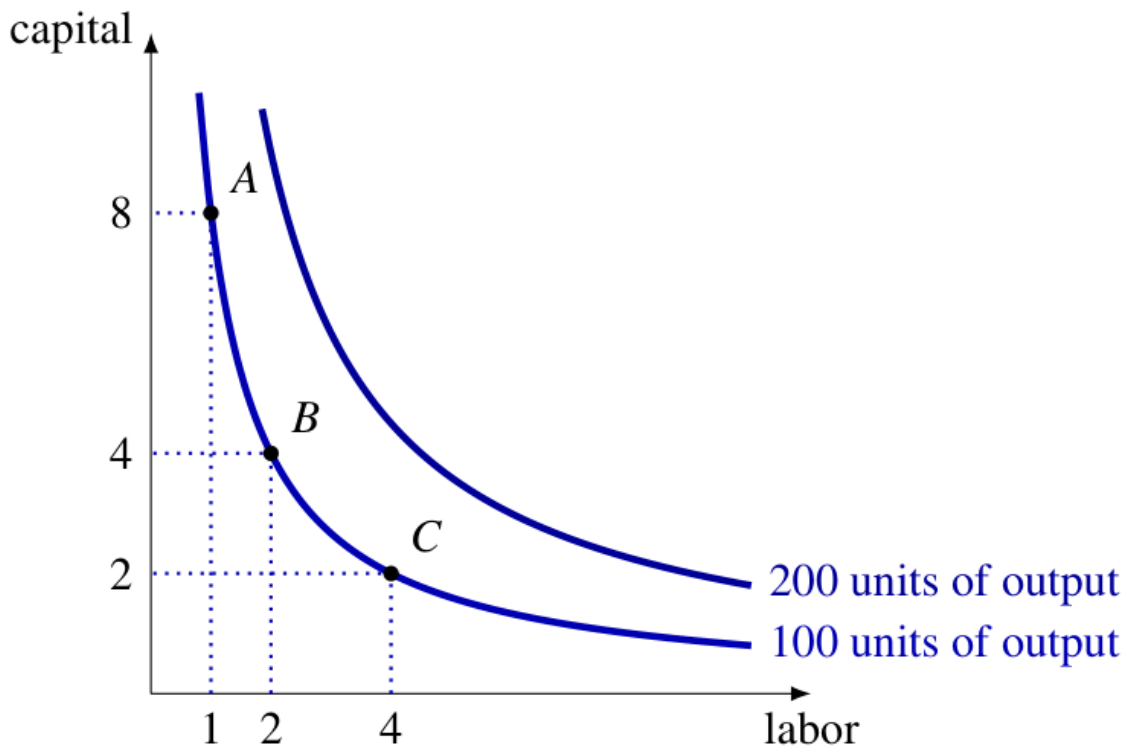


Figure 8.2 Isoquants for 100 and 200 units of output.

- Points A, B, and C produce 100 units: 1 worker and 8 machines, 2 and 4, or 4 and 2.
- Isoquants slope downward and bow toward the origin, like the indifference curves of Chapter 5.

The marginal rate of technical substitution

- **Marginal rate of technical substitution**

  The rate at which a firm can trade one input for another while holding output constant.

WHY IT DIMINISHES

It diminishes: a firm with many machines and few workers can give up several machines for one more worker, but as workers grow plentiful each substitutes for fewer and fewer machines.

The MRTS along the isoquant of Figure 8.2

- **From A to B:** one more worker replaces $8 - 4 = 4$ machines.
- **From B to C:** two more workers replace only $4 - 2 = 2$ machines, that is, $2/2 = 1$ machine per worker.
- Each additional worker substitutes for fewer and fewer machines: the MRTS diminishes.

8.5

Returns to Scale

- Scaling all inputs up together
- Increasing, constant, decreasing
- Scale vs. diminishing marginal product

What if a firm doubles every input?

↑

Increasing

Doubling all inputs more than doubles output. Common at low volumes: larger operations allow specialization and bigger, more efficient machines.

What if a firm doubles every input?

=

Constant

Doubling all inputs doubles output: for example, building a second identical factory beside the first.

What if a firm doubles every input?

↓

Decreasing

Doubling all inputs less than doubles output, often because very large organizations become harder to coordinate and manage.

Returns to scale ask what happens when all inputs scale up together, not just one.

8.5.1 TWO IDEAS, KEPT APART

Returns to scale vs. diminishing marginal product

	Diminishing marginal product	Returns to scale
Inputs changed	one input; others fixed	all inputs together
Example	more cooks, same kitchen	bigger kitchen and more cooks
When	always sets in eventually	may increase, be constant, or decrease
Horizon	the short run	the long run

The two are entirely compatible; do not confuse them.

The two are compatible

NO CONTRADICTION

A firm can have increasing returns to scale (doubling everything more than doubles output) and still face diminishing marginal product (adding cooks alone, kitchen fixed, runs into crowding).

- The first describes the long run, where the firm chooses the whole bundle of inputs; the second describes the short run, where one input is stuck.

8.6

Technological Progress

- Shifting the production function outward
- Three kinds of progress
- Progress and workers

Shifting the production function

Technological progress

So far, the same inputs always yielded the same output. Over time, progress shifts the production function: the firm produces more output from the same inputs.

WHY IT MATTERS

Technological progress is the main reason living standards rise over the long run, and in no industry has it been faster or more measurable than in semiconductors.

Three kinds of progress

Neutral

Raises output from labor and capital in equal proportion: a better recipe that lets the same cooks and ovens make more meals.

Labor-saving

Raises capital's productivity relative to labor, so firms substitute machines for workers (the industrial robots doing work once done by hand).

Three kinds of progress

Capital-saving

Does the reverse, making labor relatively more productive.

The differences matter for workers, and for the automation debate.

Progress and workers

THE TENSION

Most automation anxiety focuses on labor-saving progress, which can displace workers from particular jobs even as it makes the whole economy more productive.

- Economics offers no guarantee that everyone gains from technological change in the short run.
- What it shows: rising productivity expands the total that can be shared; how it is shared is a question for policy and politics.

8.7

Case Study: Inside a Chip Foundry

TSMC, the economics of production, made vivid.

What is a foundry?

- **Foundry**

  A firm that manufactures chips designed by other companies, such as Apple, Nvidia, and AMD, from smartphone processors to the graphics chips behind artificial intelligence.

~60%

of the global foundry market held by TSMC in 2023

Most

of the world's most advanced chips, produced by TSMC

Extremely capital intensive

- In 2024 alone, TSMC spent about $30 billion on new plants and equipment (TrendForce, 2025), more than three times the roughly $9 billion it paid its employees in wages and benefits (Taiwan Semiconductor Manufacturing Company, 2025).
- Production is overwhelmingly weighted toward capital rather than labor.
- In the language of Section 8.4, TSMC operates at the far northwest end of an isoquant like those in Figure 8.2.

8.8

Empirical Evidence: Moore's Law

The best-known quantitative record of technological progress.

Moore's law

- **Moore's law**

  In 1965 Gordon Moore noticed transistors per chip had been doubling about every year; in 1975 he revised the pace to a doubling roughly every two years.

~2 yrs

a doubling of transistors per chip

~40%

growth per year, exponential

Half a century of doubling

- The prediction held for about half a century, turning chips with thousands of transistors into chips with tens of billions.
- Each doubling was the production function of Section 8.6 shifting outward: the same silicon wafer, processed with better methods, yielded vastly more computing power.

A RECENT SLOWDOWN

The pace has recently slowed as transistors approach the size of atoms.

CHAPTER 8

Key takeaways

1. **Production function**
   Maximum output from each combination of inputs (labor and capital).

2. **Short vs. long run**
   Short run: at least one input fixed. Long run: every input varies.

3. **Marginal & average product**
   Diminishing marginal product always sets in as one input is added to fixed others.

Key takeaways

4 **Isoquants**

Show input mixes for a given output; the MRTS diminishes.

5 **Returns to scale**

Scaling all inputs may give increasing, constant, or decreasing returns (different from diminishing marginal product).

6 **Technological progress**

Shifts the function outward; Moore's law is its best-known record.

9

CHAPTER 9

The Costs of Production

How a firm's costs behave as it produces more, and why the answer shapes the size and number of firms in an industry.

CHAPTER 9

What this chapter covers

9.2–9.3 The measures

- Types of cost
- Accounting vs. economic profit
- Average & marginal cost

9.4–9.6 The shapes

- Short run vs. long run
- Economies of scale & scope
- The learning curve

9.7–9.8 Evidence

- Case: airlines & full planes
- Empirical: falling solar cost

CHAPTER 9

What this chapter covers

THE BIG IDEA

A flight to London costs nearly the same with 100 or 200 passengers; a solar panel that cost $100 per watt now costs under a dollar. Both are facts about cost, the subject of this chapter.

9.1 Introduction

- Three everyday facts about cost
- Building on production (Chapter 8)
- Why cost shapes industry structure

Three facts about cost

Airlines

A flight to London costs about the same with 100 or 200 passengers: fuel, crew, and aircraft are nearly fixed.

Solar

A panel that cost roughly $100 per watt in the mid-1970s costs well under a dollar today.

Opportunity

A student who turns down a $60,000 job to start a business gives up that salary, though no check changes hands.

Three facts about cost

THE QUESTION

Building on the production of Chapter 8, we ask how a firm's costs behave as it produces more.

9.2

Types of Cost

- Fixed, variable, and total cost
- Explicit vs. implicit costs
- Accounting vs. economic profit
- Sunk costs

Fixed, variable, and total cost

- **Fixed costs**

  Do not vary with output: the rent on a factory, the cost of an aircraft. They must be paid even if the firm produces nothing.

- **Variable costs**

  Rise with output: the materials, fuel, and labor used to make each unit.

Fixed, variable, and total cost

TOTAL COST

Total cost is simply the sum of fixed cost and variable cost.

Explicit vs. implicit costs

- **Explicit costs**

  Out-of-pocket payments: wages, rent, supplies. The costs an accountant records.

- **Implicit costs**

  The opportunity costs of resources the firm already owns: the salary the owner gives up by running the business, or the interest she forgoes by investing her own savings.

Accounting vs. economic profit

- **Economic cost = explicit + implicit**

  Adding implicit to explicit costs gives economic cost; revenue minus economic cost gives economic profit, a stricter, more meaningful measure than accounting profit, which ignores implicit costs.

THE KEY DISTINCTION

A business can show an accounting profit yet earn zero economic profit, meaning it merely covers what its resources could have earned elsewhere.

One cost to ignore: the sunk cost

- **Sunk cost**

  Already spent and impossible to recover. The money spent on a failed advertising campaign is gone regardless of what the firm does next.

THE DECISION RULE

A rational manager bases the next decision only on costs that still lie ahead. Letting a sunk cost sway the next decision, “throwing good money after bad,” is a common mistake: the sunk-cost fallacy.

Maria opens a café

Item	Amount
Revenue	$200,000
Explicit costs (rent, coffee, wages)	$130,000
Accounting profit	$200,000 − $130,000 = $70,000
Implicit: salary given up	$50,000
Implicit: interest forgone	$5,000
Economic profit	**$70,000 − $55,000 = $15,000**

Maria quits a $50,000 job and invests $100,000 of savings.

Maria opens a café

THE ARITHMETIC

Accounting profit = $200,000 − $130,000 = $70,000.

Economic profit subtracts $55,000 of opportunity cost: $70,000 − $55,000 = $15,000. Positive: the café is worth more to Maria than her next-best alternative.

Interpreting economic profit

- **Step 4.** Economic profit is positive, so the café is worth more to Maria than her next best alternative: she is ahead by $15,000 a year.
- Had the café earned only $50,000 in accounting profit, its economic profit would have been $50,000 − $55,000 = −$5,000.
- Economic profit, not accounting profit, should drive the decision.

Self-check

SELF-CHECK

An owner earns $90,000 in revenue, pays $60,000 in explicit costs, and gave up a $40,000 salary. Find accounting and economic profit.

Accounting profit = $90,000 − $60,000 = $30,000. Economic profit also subtracts the $40,000 implicit cost: $30,000 − $40,000 = −$10,000. The negative figure tells her she would have done better at her old job.

9.3

Average and Marginal Cost

- The full cost table
- The U-shaped average total cost
- Why marginal cost crosses average at its minimum

Two derived measures

- **Average total cost**

  Total cost divided by output: the cost per unit.

- **Marginal cost**

  The extra cost of producing one more unit.

Table 9.1 builds them from a firm with $100 of fixed cost, together with average variable cost (variable cost divided by output), lowest at $40 at outputs of 3 and 4.

Table 9.1: building the cost measures

Output	Fixed cost	Variable cost	Total cost	Average variable cost	Average total cost	Marginal cost
0	100	0	100 + 0 = 100	—	—	—
1	100	50	100 + 50 = 150	50/1 = 50	150/1 = 150	150 − 100 = 50
2	100	90	100 + 90 = 190	90/2 = 45	190/2 = 95	190 − 150 = 40
3	100	120	100 + 120 = 220	120/3 = 40	220/3 ≈ 73	220 − 190 = 30
4	100	160	100 + 160 = 260	160/4 = 40	260/4 = 65	260 − 220 = 40
5	100	220	100 + 220 = 320	220/5 = 44	320/5 = 64	320 − 260 = 60
6	100	300	100 + 300 = 400	300/6 = 50	400/6 ≈ 67	400 − 320 = 80

Table 9.1 Building average and marginal cost from fixed and variable cost.

Why average total cost is U-shaped

Falling at low output

The $100 fixed cost is spread over more units; the same force lets an airline lower its cost per passenger by filling seats.

Rising at high output

Diminishing marginal product (Chapter 8) makes each extra unit more expensive to produce.

Why average total cost is U-shaped

MARGINAL COST TOO

Marginal cost, after an initial fall, rises for the same reason: diminishing returns. In Table 9.1 it bottoms near an output of 3, then climbs.

Marginal cost cuts the average curves at their lowest points

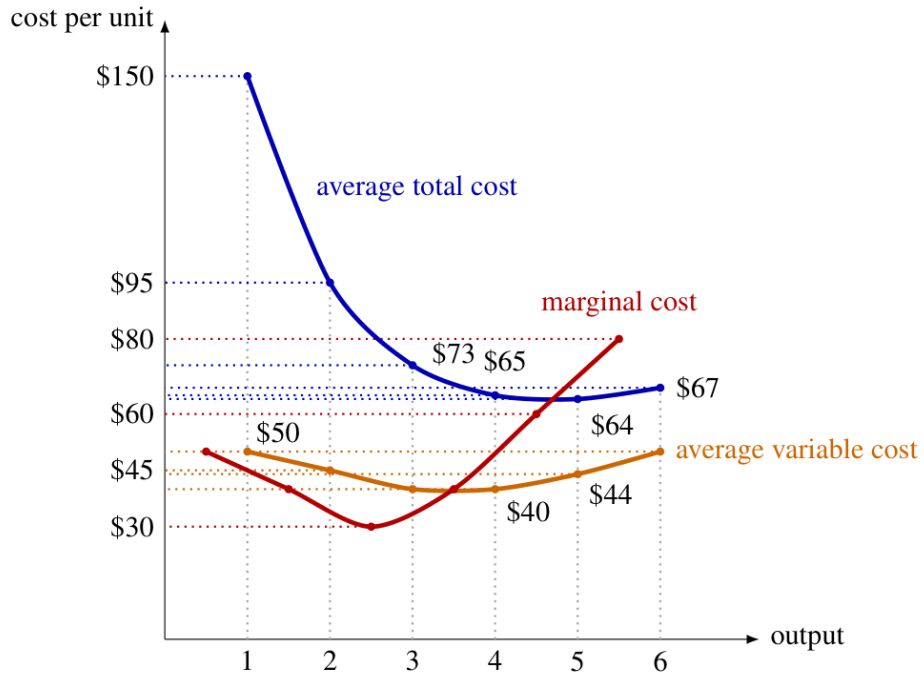


Figure 9.1 Marginal, average variable, and average total cost, from Table 9.1.

- Marginal cost crosses average total cost at its minimum, near output 5, where ATC bottoms at $64.
- When the next unit costs less than the average, it pulls the average down; when more, it pushes it up.

Marginal cost cuts the average curves at their lowest points

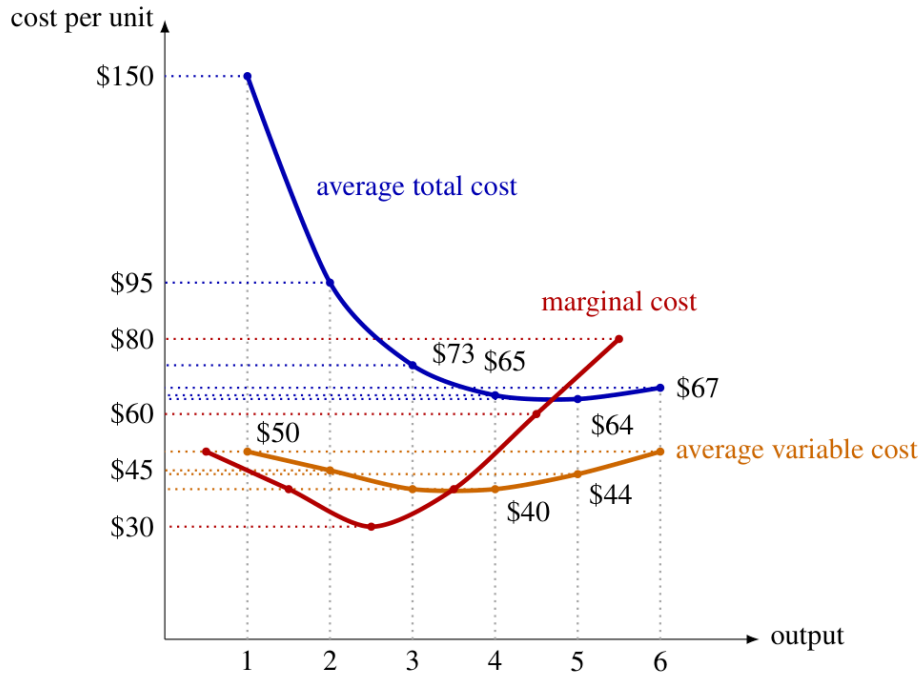


Figure 9.1 Marginal, average variable, and average total cost, from Table 9.1.

- So the curves can only cross where the average is flat, at its lowest point.
- Each marginal cost is plotted halfway between two levels of output, since it measures the change in total cost from one to the next.

Explore Figure 9.1: intro-micro-companion-website.netlify.app/cost-curves

9.3.1 HOW THE COST CURVES FIT TOGETHER

The curves are tied by simple arithmetic

Average fixed cost

Always falls as output rises: the same fixed cost spread over more units. This is “spreading the overhead.”

Average variable cost

U-shaped: falls at first, then rises as diminishing returns set in.

Average total cost

The sum of the two, also U-shaped. The gap above AVC narrows as AFC keeps falling.

The curves are tied by simple arithmetic

THE MARGINAL COST THREAD

Running through all of them, marginal cost cuts both average curves at their lowest points: it can only equal an average where that average is momentarily flat.

9.4

The Short Run and the Long Run

- Short run: fixed costs are fixed
- Long run: all costs variable
- The long-run average cost curve as a lower boundary

Choosing the size of the plant

Short run

Fixed costs are truly fixed: the firm cannot change the size of its factory.

Long run

The firm can choose the size of its factory, so all costs become variable.

Choosing the size of the plant

- **Long-run average cost**

  Traces the lowest cost per unit achievable at each output when the firm is free to build the best-sized plant. It forms a lower boundary beneath the short-run curves.

Three plant sizes for a bakery

Loaves per day	Small plant	Medium plant	Large plant	Long-run average cost
100	240/100 = 2.40	350/100 = 3.50	650/100 = 6.50	2.40 (small)
200	440/200 = 2.20	460/200 = 2.30	720/200 = 3.60	2.20 (small)
300	750/300 = 2.50	600/300 = 2.00	810/300 = 2.70	2.00 (medium)
400	1,200/400 = 3.00	820/400 = 2.05	920/400 = 2.30	2.05 (medium)
500	1,750/500 = 3.50	1,100/500 = 2.20	1,050/500 = 2.10	2.10 (large)

Table 9.2 Short-run average total cost of three plant sizes, in dollars per loaf.

Three plant sizes for a bakery

- Each cell divides the plant's total cost per day by the loaves it bakes. A bigger plant has a higher fixed cost: expensive at low output, less expensive at high output.

Reading Table 9.2: short run and long run

- **Step 1. The short run:** once a plant is built, each column is its short-run average total cost. For the small plant: $2.40, $2.20, then up to $3.50, the U shape of Table 9.1.
- **Step 2. The long run:** the bakery picks the plant with the lowest average cost: small for 100 or 200 loaves, medium for 300 or 400, large for 500.

Reading Table 9.2: the long-run curve

- **Step 3. The lower boundary:** long-run average cost is never above any entry in its row. With the small plant, 400 loaves cost $3.00 each; with the medium plant, $2.05.
- **Step 4. The shape:** long-run average cost falls from $2.40 to $2.00 as output rises to 300 loaves, then rises to $2.10 at 500 loaves (Section 9.5).

9.5

Economies and Diseconomies of Scale

- The shape of long-run average cost
- Minimum efficient scale
- Where the forces come from

What the shape of long-run cost tells us

Economies of scale

Long-run average cost falls as output rises: bigger is less costly per unit (specialization, large machines, spread fixed costs).

Constant returns to scale

Long-run average cost is flat: size does not change cost per unit.

What the shape of long-run cost tells us

Diseconomies of scale

Long-run average cost rises.

Figure 9.2 shows a typical long-run average cost curve.

The long-run average cost curve

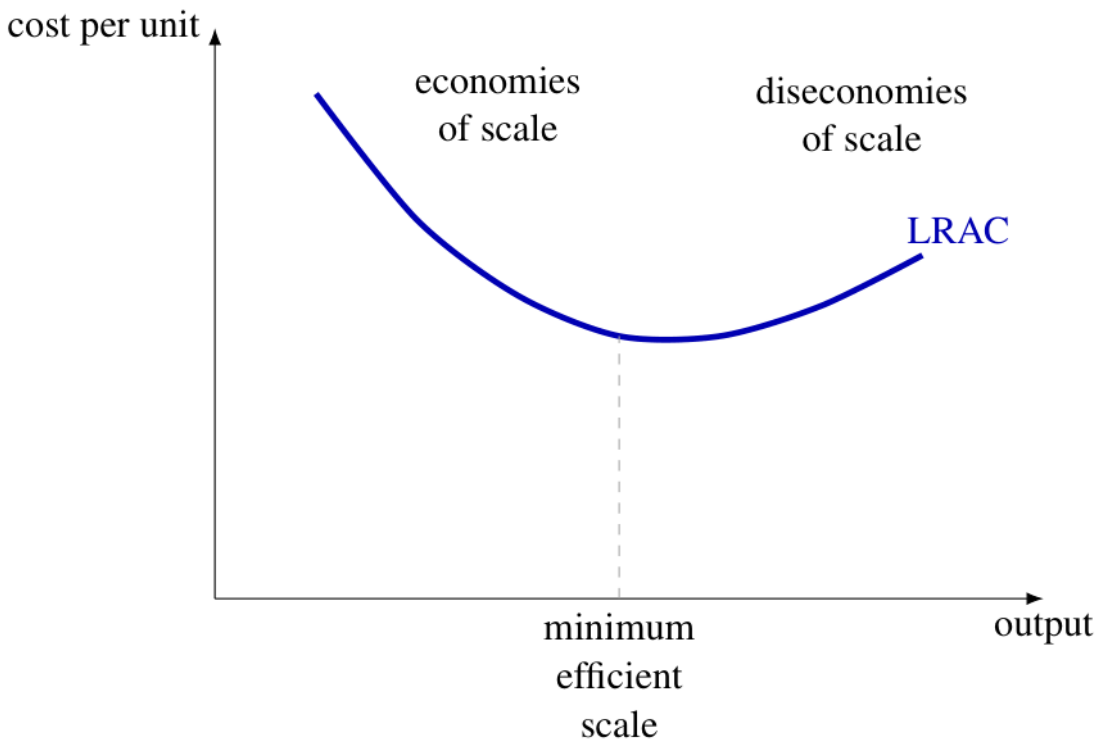


Figure 9.2 The long-run average cost curve.

- **Minimum efficient scale**

  The output at which average cost first reaches its minimum. Its size relative to the whole market helps determine how many firms an industry can support.

Minimum efficient scale shapes the industry

Small MES → many firms

Where minimum efficient scale is small (restaurants, hair salons), many firms coexist.

Large MES → few firms

Where it is enormous relative to demand (aircraft, semiconductors), only a few firms can operate efficiently.

A point we develop in the market-structure chapters ahead.

The forces behind the U-shape

- **Economies of scale (downward slope)**

  Specialization (each worker and machine focuses on one task); large, indivisible equipment (a blast furnace or assembly line); and spreading fixed costs like research, design, and advertising over many units (Section 9.5.2).

- **Diseconomies of scale (upward slope)**

  The growing difficulty of managing a very large organization: communication slows, coordination falters, and layers of management multiply.

Economies of scale in chip foundries

- The chip foundries of Section 8.7: the up-front cost of designing and building a fab is the same whether it produces a few chips or millions.
- So larger volumes sharply lower the cost per chip.
- This is why a handful of giant firms manufacture the world's advanced semiconductors, and why TSMC's scale is so difficult for rivals to match.

AI data centers: the numbers

- Training a foundation model (the large statistical model behind a chatbot like ChatGPT) requires racks of specialized chips running in parallel for weeks or months.
- An Nvidia H100 sold in 2024 for roughly $25,000 to $40,000; a training cluster can contain tens of thousands of them.
- A cluster costs from several hundred million to several billion dollars, before a single query is served.

The economics of the chip fab, amplified

- The up-front cost is enormous, but the marginal cost of one more training experiment is small.
- A large cluster spreads the up-front cost across many models and users; a smaller entrant recovers the same fixed cost over a smaller base.
- So the frontier of AI research has settled into a few well-funded laboratories, heavy on capital and very light on labor.

9.6

Economies of Scope and the Learning Curve

- Producing several products together
- The learning curve
- Scale vs. cumulative experience

Two related ideas

Economies of scope

It costs less to produce several products together than separately: a dairy that makes both milk and cheese, or a company that offers both search and email from the same data centers.

The learning curve

Cost per unit falls as a firm accumulates total production experience over time, learning to produce more efficiently.

Scale vs. the learning curve

Economies of scale

Depend on the rate of output: how much the firm produces per period.

The learning curve

Depends on cumulative output to date: total experience accumulated over time.

WHERE IT MATTERS MOST

The learning curve is most dramatic in high-technology manufacturing, as Section 9.8 shows for solar power.

9.7

Case Study: Why Airlines Chase Full Planes

The spreading of fixed cost of Section 9.3, operating flight by flight.

9.7 CASE STUDY

Air travel: fixed costs at work

- Once a flight is scheduled, most of its cost (the aircraft, the fuel, the pilots and cabin crew, the landing fees) is fixed regardless of how many board.
- The marginal cost of one more passenger on a flight departing anyway is very small: a bit of fuel for the added weight, a meal, and the cost of processing the ticket.

THE LOAD FACTOR

Because cost per passenger is total flight cost divided by passengers, filling more seats spreads the large fixed cost over more travelers: the spreading of fixed cost of Section 9.3, flight by flight.

9.7 CASE STUDY

The obsession with the load factor

~83%

of seats filled on U.S. carriers' domestic flights in 2023 (U.S. Bureau of Transportation Statistics, 2026)

≈ 0

extra cost of a seat flown full vs. empty

9.7 CASE STUDY

The obsession with the load factor

WHY IT MATTERS

Airlines use elaborate pricing, frequent-flyer programs, and last-minute discounts to fill seats that would otherwise fly empty. The strategy has pushed load factors steadily upward: about 71% in 2000, 77% in 2005, and 83% in 2023.

9.8

Empirical Evidence: The Falling Cost of Solar Power

The learning curve of Section 9.6 at its clearest.

A learning curve in action

- The price of a solar photovoltaic module (its price per watt of generating capacity) has fallen sharply over half a century.
- From roughly $100 per watt in the mid-1970s to well under a dollar per watt today.

~$100
per watt in the mid-1970s

< $1
per watt today

A learning curve in action

THE MECHANISM

Cost falling with accumulated experience, not with the current rate of output. A standard learning curve.

Swanson's law

SWANSON'S LAW

Each time the cumulative quantity of solar modules ever produced doubles, their cost per watt falls by about 20 percent.

- It has held for decades as manufacturers refined their materials, processes, and scale.
- The same fixed percentage decline per doubling of cumulative output (Wright's law) was first documented in aircraft manufacturing in the 1930s.

CHAPTER 9

Key takeaways

1. **Types of cost**
   Fixed + variable = total; explicit + implicit = economic cost.

2. **Economic profit**
   Subtracts implicit (opportunity) costs, a stricter measure than accounting profit.

3. **Cost curves**
   Average total cost is U-shaped; marginal cost crosses it at its minimum.

Key takeaways

4 **Short vs. long run**

In the long run all costs are variable; long-run average cost is a lower boundary.

5 **Scale, scope, learning**

Economies of scale set the minimum efficient scale; scope and the learning curve cut cost further.

6 **Evidence**

Airlines chase full planes; solar follows Swanson's law, falling 20% per doubling.

10

CHAPTER 10

Perfect Competition

The benchmark market structure: many tiny price takers, free entry, and a price set by the cost of production.

CHAPTER 10

What this chapter covers

10.2–10.4 The firm

- What perfect competition is
- The output rule: P = MC
- Profit, loss, and shutting down

10.5–10.6 Long run & efficiency

- Free entry & exit
- Long-run supply
- Surplus & efficiency
- Price supports in farming

10.7–10.8 Evidence

- Case: wheat & the world price
- The vanishing American farm

CHAPTER 10

What this chapter covers

THE BIG IDEA

Perfect competition is the first extreme: sellers entirely at the mercy of the market price. It is the benchmark against which every other market structure in the book is measured.

10.1 Introduction

- A Kansas wheat farmer who cannot budge the price
- Price takers vs. price makers
- Perfect competition as the benchmark

The farmer who cannot set the price

- A Kansas wheat farmer can plant more acres or fewer, but she cannot budge the price of wheat; it is set on global exchanges by millions of buyers and sellers.
- Charge a penny above the going price and she sells nothing: identical wheat is available everywhere else.
- Why are some sellers at the mercy of the market price, while a drug company holding a patent can set its own? This chapter studies the first extreme.

10.2

What Is Perfect Competition?

- Four conditions
- Price taking
- P = MR = firm's demand

Four conditions define a perfectly competitive market

1

1. Many buyers & sellers

Each one too small to affect the market price.

2

2. Identical product

Buyers have no reason to prefer one seller's output to another's.

Four conditions define a perfectly competitive market

3

3. Free entry & exit

Firms enter when there is money to be made and leave when there is not.

4

4. Good information

Buyers know the prices on offer.

The main consequence: each firm is a price taker

Price taker

A firm that can sell as much as it wants at the going market price, but nothing at all above it. The market's supply and demand set the price; the individual firm simply accepts it.

THE KEY PICTURE

The market demand curve slopes down, but the demand curve facing a single firm is a flat line at the market price.

Market demand slopes down; the firm's is flat

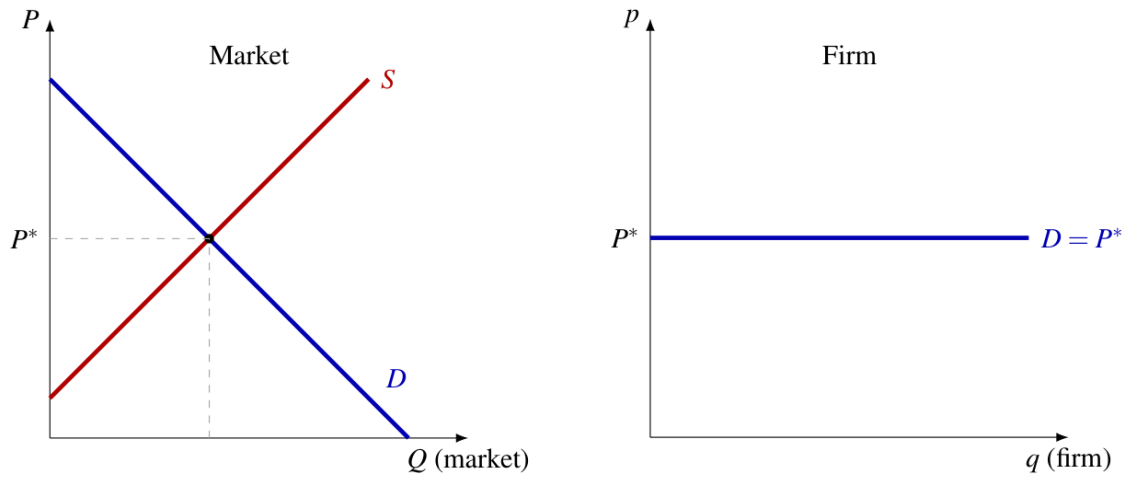


Figure 10.1 The market (left) and an individual firm (right).

- The market sets the price through total supply and demand.
- At that price the firm can sell any quantity: its demand curve is horizontal.

10.3

The Firm's Output Decision

- Marginal revenue = price
- Produce where P = MC
- Profit as a rectangle

Marginal revenue equals the market price

- **Marginal revenue (for a price taker)**

  The revenue from selling one more unit is simply the market price, since the firm can sell each unit at that price.

FROM CHAPTER 1

Expand output as long as the extra revenue exceeds the extra cost, and stop where they are equal. For a price taker this gives a simple rule.

The output rule

THE COMPETITIVE OUTPUT RULE

Produce the quantity at which price equals marginal cost.

- Producing less would mean passing up units whose price exceeds their marginal cost.
- Producing more would mean making units that cost more than the price to produce.

Table 10.1: profit at a price of $80

Output	Total cost	Average variable cost	Average total cost	Marginal cost	Profit at a price of $80
0	$100 + 0 = 100$	—	—	—	$0 \times 80 - 100 = -100$
1	$100 + 50 = 150$	$50/1 = 50$	$150/1 = 150$	$150 - 100 = 50$	$1 \times 80 - 150 = -70$
2	$100 + 90 = 190$	$90/2 = 45$	$190/2 = 95$	$190 - 150 = 40$	$2 \times 80 - 190 = -30$
3	$100 + 120 = 220$	$120/3 = 40$	$220/3 \approx 73$	$220 - 190 = 30$	$3 \times 80 - 220 = 20$
4	$100 + 160 = 260$	$160/4 = 40$	$260/4 = 65$	$260 - 220 = 40$	$4 \times 80 - 260 = 60$
5	$100 + 220 = 320$	$220/5 = 44$	$320/5 = 64$	$320 - 260 = 60$	$5 \times 80 - 320 = 80$
6	$100 + 300 = 400$	$300/6 = 50$	$400/6 \approx 67$	$400 - 320 = 80$	$6 \times 80 - 400 = 80$

Table 10.1 The costs of Table 9.1 and the firm's profit at a price of $80.

Reading Table 10.1

- The first five units cost $50, $40, $30, $40, and $60, all less than the $80 each brings in: profit rises from −$100 to $80 at 5 units.
- The sixth unit costs $80, as much as it brings in: profit stays at $80. The firm expands until marginal cost reaches $80, at 6 units.

Profit at 6 units

- **Revenue:** 6 × $80 = $480. Total cost: $400.
- **Profit:** $480 − $400 = $80, about $13 on each unit (average total cost is about $67).
- Producing more would mean making units that cost more than $80 to produce.

Profit per unit is price minus average total cost

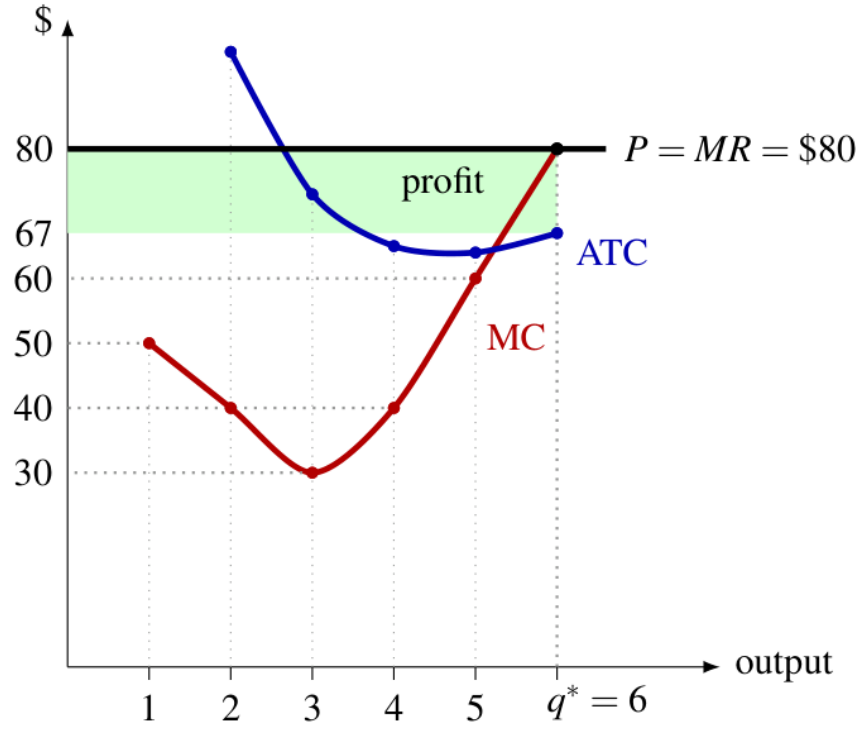


Figure 10.2 The competitive firm's output and profit at a price of $80.

- Output is set where the flat price line crosses the MC curve.
- The profit rectangle = (P − ATC) × quantity.

Explore Figure 10.2: intro-micro-companion-website.netlify.app/competitive-firm

Same firm, three outcomes, set only by the price

Price	Output (P = MC)	Average total cost there	Outcome
$80	6 units	about $67	profit: 6 × $80 − $400 = $80
$64	5 units	$64	revenue 5 × $64 = $320 = total cost: breaks even
$50	4 units	$65	loss: $260 − 4 × $50 = $60

- From Table 10.1: minimum average total cost is $64, at 5 units; minimum average variable cost is $40, at 3 and 4 units.
- In each case the firm produces where price equals marginal cost.

Why a firm at $50 keeps operating

- At $50 the price is below average total cost ($65): the firm loses $260 − $200 = $60.
- Its revenue, $200, covers the variable cost of 4 units, $260 − $100 = $160, and leaves $200 − $160 = $40 toward the $100 fixed cost.
- Shutting down would lose the whole $100 fixed cost; operating loses only $60.

10.4

When Does a Firm Shut Down?

- Short run: shut down below minimum AVC
- Long run: leave below minimum ATC
- The short-run supply curve

Can the firm avoid its fixed cost?

- If the price is low enough, the firm may do better to produce nothing.
- The answer depends on whether the firm can avoid its fixed cost, which is what separates the short run (Section 10.4.1) from the long run (Section 10.4.2).

Operate while the price covers average variable cost

- **Shutdown rule**

  In the short run, fixed costs must be paid no matter what. The firm keeps operating as long as the price covers its average variable cost: at least it earns something toward the fixed cost it owes anyway. If price falls below minimum AVC ($40 in Table 10.1), it loses less by producing zero.

Operate while the price covers average variable cost

WHY

A ski resort closes in summer and a food stand shuts on a slow day, not because they are bankrupt, but because the price fails to cover even the variable cost of operating.

The firm's short-run supply curve

SHORT-RUN SUPPLY

The firm's short-run supply curve is its marginal-cost curve, above the minimum of average variable cost.

The firm's short-run supply curve

SELF-CHECK

A firm has minimum AVC of $40 and minimum ATC of $64. Does it operate or shut down if the price is (a) $50, (b) $30?

(a) At $50, above the $40 minimum AVC, it operates at a loss, covering variable cost and part of fixed cost. (b) At $30, below $40, it shuts down.

Shutting down in the long run

- In the long run the firm can avoid its fixed cost too: it can sell its equipment, let its lease run out, and leave the industry.
- **Long-run rule:** leave the industry if the price is below the minimum average total cost ($64 in Table 10.1).

Back to the price of $50

- In the short run, the firm keeps producing 4 units and loses $60, less than the $100 fixed cost.
- If the price is expected to stay at $50, the firm leaves the industry once its fixed commitments end, and its loss falls to zero.

10.5

Profit, Loss, and the Long Run

- Free entry erodes profit
- Zero economic profit = normal return
- The long-run supply curve

Free entry erodes the profit away

Profits attract entry

New firms enter; their output shifts market supply right, pushing the price down until price reaches minimum ATC and profit is zero.

Losses cause exit

Some firms exit; supply falls and the price rises until the survivors break even.

Free entry erodes the profit away

WHERE IT STOPS

Entry continues until the profit that attracted it is gone: that is, until price has fallen to the minimum of average total cost ($64 in Table 10.1).

Free entry: the market and a typical firm

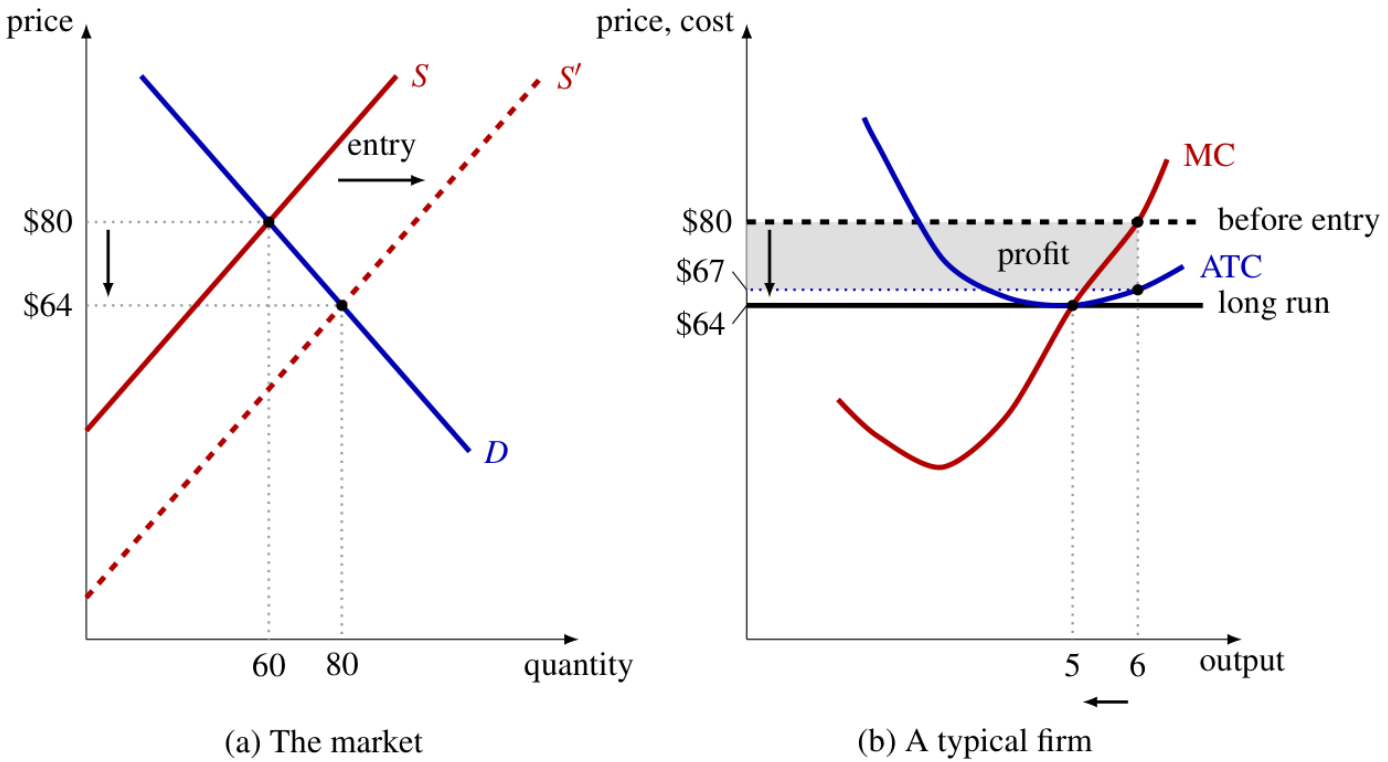


Figure 10.3 Free entry: the market (left) and a typical firm (right).

Explore Figure 10.3: intro-micro-companion-website.netlify.app/long-run-entry

Reading Figure 10.3

- **Left:** entry shifts market supply from S to S′, and the price falls from $80 to $64.
- **Right, at $80:** the firm produces 6 units and earns about $13 a unit above its average total cost of about $67 (the shaded area).
- **Right, at $64:** the minimum of average total cost; the firm produces 5 units and its economic profit is zero.

Table 10.2: market demand behind Figure 10.3

Price	Quantity demanded
$88	50
$80	60
$72	70
$64	80
$56	90

Table 10.2 Market demand in Figure 10.3.

- **At $80:** buyers want 60 units, which 10 firms supply by producing 6 units each.
- **At $64:** buyers want 80 units and each firm produces 5, so entry stops at 80 ÷ 5 = 16 firms, 6 more than before.

Zero economic profit is not zero earnings

- **What zero economic profit means**

  Economic cost includes the owner's opportunity costs (Chapter 9). Zero economic profit means the firm earns what its owners could have earned elsewhere (in another industry or business): a normal return, no more and no less.

This is the long-run resting point of a competitive industry.

The long-run supply curve is flat at minimum ATC

- In the short run, with a fixed number of firms, market supply slopes upward, as in Chapter 2.
- In the long run, a price above minimum ATC draws entry until output pushes it back down; a price below it triggers exit until the price recovers.
- With identical firms, the long-run supply curve is flat at the minimum average total cost.

The long-run supply curve is flat at minimum ATC

THE LESSON

Demand then determines the quantity but not the price, which is set by the firms' lowest achievable cost. This is why a competitively produced good's price tracks its cost so closely, and why permanent windfalls are rare.

10.6

Producer Surplus and Efficiency

- Producer surplus
- Why P = MC maximizes total surplus
- Regulating a competitive market

Producer surplus, and total surplus

- **Producer surplus**

  The difference between the price a seller receives and the lowest price at which it would have been willing to sell, given by the marginal-cost curve.

- **Total surplus**

  Consumer surplus (Chapter 6) plus producer surplus: the total value the market creates.

Producer surplus: five ticket sellers

- Five people would each sell a concert ticket for no less than $10, $20, $30, $40, and $50; the market price is $35.
- The first three sell; the last two do not. Their producer surplus:

$$(\$35 - \$10) + (\$35 - \$20) + (\$35 - \$30) = \$25 + \$15 + \$5 = \$45$$

Producer surplus in two markets

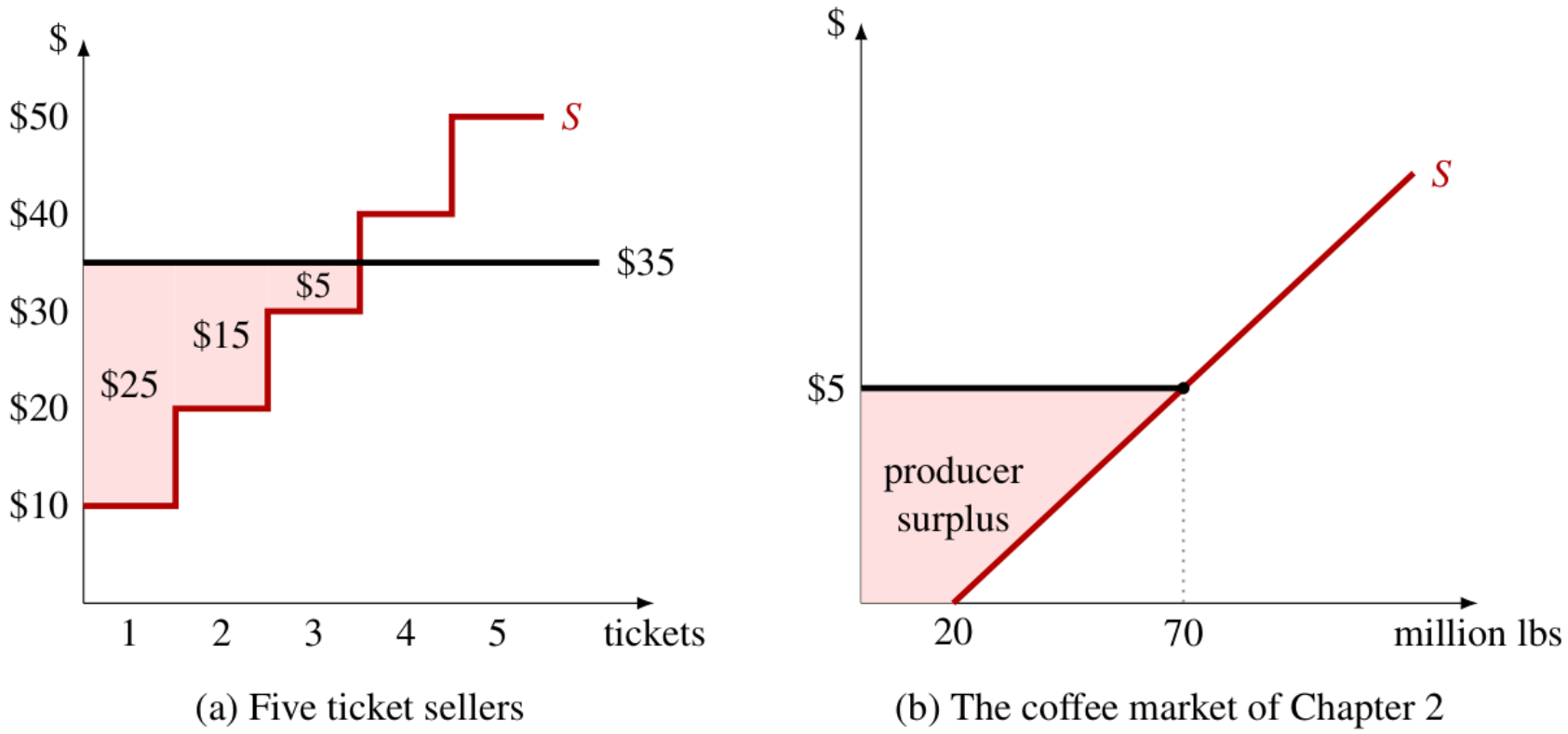


Figure 10.4 Producer surplus for five ticket sellers (left) and for the coffee market (right).

Reading Figure 10.4: the ticket sellers

- Producer surplus is the area above the supply curve and below the price.
- **Left panel:** the supply curve is a staircase, one step per seller; each shaded rectangle between the $35 price and a step is one seller's surplus: $25, $15, and $5.

Reading Figure 10.4: the coffee market

- **Right panel:** the coffee market of Chapter 2. At the $5 price, sellers bring 70 million pounds; supply starts at a price of zero with 20 million pounds.
- The shaded area is a rectangle plus a triangle:

$$\$5 \times 20 \text{ million} + \tfrac{1}{2} \times \$5 \times 50 \text{ million} = \$100 \text{ million} + \$125 \text{ million} = \$225 \text{ million}$$

Competition maximizes total surplus

THE EFFICIENCY RESULT

The competitive equilibrium has an important property: it maximizes total surplus.

- Output expands to where price (what the marginal unit is worth to buyers) equals marginal cost (what it costs to make).
- Every unit worth more than it costs gets produced; no unit costing more than it is worth does.
- No reallocation could make society better off. Economists call this the “invisible hand,” borrowing Adam Smith’s phrase.

The efficiency of competition, in surplus terms

- At equilibrium P = MC, so the value of the last unit to buyers (willingness to pay, off the demand curve) equals its cost to producers (off the supply curve).
- Every unit worth more than it costs gets produced; none costing more than it is worth does, so total surplus is as large as it can possibly be.

Why Competition Is the Benchmark

Any force that pushes price away from marginal cost (a tax, Chapter 4; a price control; or the monopoly of Chapter 11) shrinks total surplus, leaving a deadweight loss. This is the precise sense of Adam Smith's invisible hand.

Total surplus in the coffee market

- At the competitive equilibrium ($5, 70 million pounds): consumer surplus $245 million (Figure 6.7), producer surplus $225 million (Figure 10.4).

$$\text{total surplus} = \$245 + \$225 = \$470 \text{ million}$$

- The $2 tax of Chapter 4 cut the quantity from 70 to 60 million pounds and destroyed $10 million of total surplus.

Agricultural price supports: a regulated benchmark market

- **Price support**

  A price floor set above the competitive equilibrium (Chapter 4). At the support price, the quantity farmers wish to sell exceeds the quantity buyers will purchase, leaving a surplus that the government buys up and stores to keep the price from falling back.

THE PARADOX

Agriculture is the standard example of perfect competition, yet it is among the most heavily regulated sectors in the world: a paradox worth examining.

A price support holds the price above equilibrium

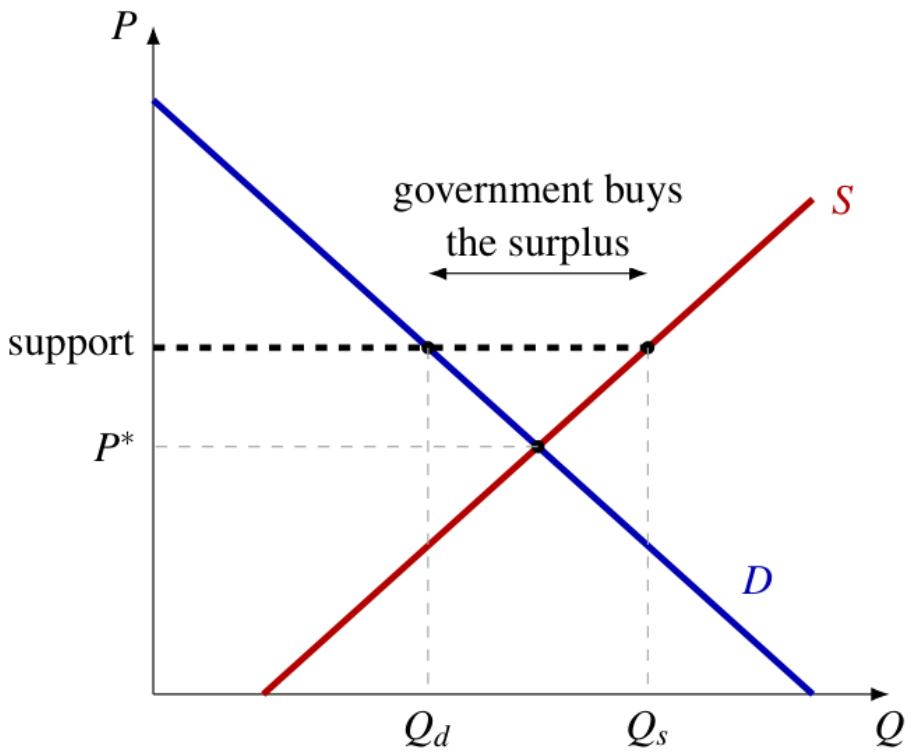


Figure 10.5 An agricultural price support.

- At the support price, quantity supplied exceeds quantity demanded.
- The gap is a surplus the government must buy and store.

Price supports in practice

European Community (1980s)

The Common Agricultural Policy guaranteed “intervention prices”; surpluses piled into the “butter mountains” and “wine lakes” of the 1980s.

United States

A web of price supports, subsidized crop insurance, and direct payments; direct government payments alone came to about $28 billion in 2025.

Price supports in practice

THE TRADE-OFF

These programs raise farm incomes, but, as Figure 10.5 predicts, by holding prices above the market-clearing level and generating surpluses the government must buy, store, or dispose of, at a cost borne by taxpayers and consumers.

The same policy in the developing world: India's MSP

23 crops

covered by India's minimum support price (MSP)

Far above

the buffer India needs: public stocks of wheat and rice

The same policy in the developing world: India's MSP

THE APPEAL AND THE TRAP

The MSP lifted and steadied millions of farmers' incomes, but produced the surplus that the model predicts: much of it stored at great expense, some left to rot. Three 2020 laws opening crop trading to private buyers outside the regulated markets met a year of protests by millions of farmers, afraid of losing the price floor that protected them, that forced the laws' repeal in 2021.

10.7

Case Study: Wheat Farming and the World Price

Few markets come as close to perfect competition as a basic crop.

10.7 CASE STUDY

Wheat meets the conditions of perfect competition

Identical product

Wheat of a given grade is the same from Kansas, Ukraine, or Australia; buyers have no reason to favor one farm's output.

Numerous, tiny sellers

Even the U.S., the world's 4th-largest producer, accounts for ~a tenth of global trade, grown by tens of thousands of farms.

10.7 CASE STUDY

Wheat meets the conditions of perfect competition

THE RESULT

No single farmer produces enough to move the world price: the price-taking assumption of Section 10.2.

10.7 CASE STUDY

Price takers compete only on cost

- The price of wheat is set on global commodity exchanges by total world supply and demand; each farmer treats it as given, deciding only how much to plant and harvest.
- A drought in one major region cuts the global harvest, so the world price rises for everyone: the supply shift of Chapter 2.
- Farmers compete not by setting prices, which they cannot do, but by lowering their costs: the only margin a price taker controls.

10.8

Empirical Evidence: The Vanishing American Farm

The long-run force of entry and exit, over a century.

10.8 EMPIRICAL EVIDENCE

Entry, exit, and the vanishing American farm

6.81 M
U.S. farms at the 1935 peak

1.88 M
U.S. farms in 2024, lowest in 100+ years

- Land farmed fell far less than the number of farms (to 876 million acres in 2024), so average farm size grew.
- This is the long-run force of Section 10.5: entry when returns are high, exit when they are low.

10.8 EMPIRICAL EVIDENCE

The competitive long run, at work over decades

- Rapid productivity growth (better seeds, machines, agricultural science, and farming methods) repeatedly lowered costs and pushed crop prices down.
- With thin margins near minimum average cost, farmers earning less than they could elsewhere left for better-paying nonfarm jobs, the exit the model predicts.
- The survivors bought or rented the land they left and grew larger.

THE TAKEAWAY

The disappearance of millions of small farms is the competitive long run in action: free exit driving economic profit toward zero and reallocating labor to where it is more valuable.

Why farms grew larger

- Kislev and Peterson (1982) traced the growth in farm size to the attraction of better-paying nonfarm work.
- Rising nonfarm wages, relative to the price of farm machinery, explain virtually all of the growth in U.S. farm size from 1930 to 1970.

Key takeaways

1. **Price taking**
   Many tiny firms and an identical product make each firm a price taker; its demand curve is flat at the market price.

2. **P = MC**
   The competitive firm produces where price equals marginal cost: the single output rule.

3. **Shutdown rule**
   Operate while price covers average variable cost; below minimum AVC, shut down. Supply = MC above minimum AVC.

CHAPTER 10

Key takeaways

4 **The long run**

Free entry and exit drive price to minimum ATC, where economic profit is zero: a normal return.

5 **Efficiency**

P = MC maximizes total surplus; taxes, price controls, and monopoly create deadweight loss (the invisible hand).

6 **Evidence**

Price supports and surpluses, wheat at the world price, and the century-long decline in U.S. farms.

11

CHAPTER 11

Monopoly and Market Power

Why one seller can raise its price more than sixfold when the wheat farmer cannot raise hers by a penny.

CHAPTER 11

What this chapter covers

11.2–11.3 The seller

- Why monopolies arise
- The output decision (MR equals MC)
- Marginal revenue worked out

11.4–11.6 Welfare & power

- Welfare cost vs. competition
- Markup, Lerner index, patents
- Monopsony; natural monopoly

11.7–11.8 Evidence

- Case: EpiPen & De Beers
- Empirical: patents & generics
- Regulation in practice

CHAPTER 11

What this chapter covers

THE BIG IDEA

Market power is the ability to set price above marginal cost. We study its purest form: monopoly, a market served by a single seller.

11.1 Introduction

- EpiPen: $94 to over $600
- De Beers and the price of diamonds
- What sets the monopolist apart

When markets fail

- Chapter 10: competitive markets produce every unit buyers value above its cost, so total surplus is as large as it can be.
- **Market failure:** the market outcome no longer maximizes total surplus.
- Sources studied ahead: a single firm (Chapters 11 and 12), a few firms (Chapter 15), asymmetric information (Chapter 16), and effects on others (Chapter 18). Chapter 17 studies a related problem: how a seller who does not know buyers' values can sell through an auction.

Two sellers who could set their own price

EpiPen, 2007–2016

A two-pack that stops a life-threatening allergic reaction rose from about $94 to more than $600, though the device barely changed.

De Beers

For most of the twentieth century one company controlled the world's diamond supply and shaped the price of every engagement ring.

Two sellers who could set their own price

THE PUZZLE

How can one seller raise its price more than sixfold, or hold prices high for decades, when a wheat farmer cannot? The answer is market power.

11.2

Why Monopolies Arise

- Control of a key resource
- Economies of scale
- Patents & franchises
- Network effects

A monopoly survives only behind a barrier

Barriers to entry

A monopoly can persist only if something keeps rivals out. These barriers take several forms: control of a resource, economies of scale, legal protection, or network effects.

THE COMMON THREAD

Without a barrier, the profits a monopolist earns would attract entrants, and the price would be competed back down toward marginal cost.

Five barriers to entry (1–3)

1

1. Key resource

A firm owning the essential input can exclude others, as De Beers did with the world's diamond mines.

2

2. Economies of scale

When average cost falls over the whole market, one firm serves it at lower cost: a natural monopoly (water, electricity).

Five barriers to entry (1–3)

3

3. Patents & copyrights

A temporary legal monopoly to reward innovation: a patent lasts 20 years from filing (U.S. and Japan); a copyright, the author's life plus 70 years.

Five barriers to entry (4–5)

4

4. Government franchise

The state may grant one firm the exclusive right to operate, or keep it for itself: postal services, national defense, U.S. state lotteries, Sweden's Systembolaget stores.

5

5. Network effects

A product can grow more valuable as more people use it, making it hard for newcomers to break in, as with some digital platforms.

Five barriers to entry (4–5)

IN COMMON

Each barrier blocks entry in a different way, but all do the same job: they keep the single seller alone in the market.

11.3

The Monopolist's Output Decision

- Facing the whole demand curve
- Marginal revenue below price
- MR equals MC, then price above MC
- A numeric example

The one decisive difference

- **The monopolist faces the entire market demand curve**

  To sell one more unit it must lower its price, not just on that unit, but on every unit it sells.

MARGINAL REVENUE BELOW PRICE

So the extra revenue from one more sale (the marginal revenue) is less than the price: the gain from the extra unit is partly offset by the lower price now received on all the others. The MR curve lies below demand.

Marginal revenue below price, in numbers

Quantity	Price	Total revenue (price times quantity)	Marginal revenue
1	$11	$11 × 1 = $11	$11 − $0 = $11
2	$10	$10 × 2 = $20	$20 − $11 = $9
3	$9	$9 × 3 = $27	$27 − $20 = $7
4	$8	$8 × 4 = $32	$32 − $27 = $5
5	$7	$7 × 5 = $35	$35 − $32 = $3
6	$6	$6 × 6 = $36	$36 − $35 = $1

Table 11.1 Price, total revenue, and marginal revenue along a straight-line demand curve.

Reading Table 11.1

- Selling 2 units requires cutting the price to $10 on both: total revenue $20. The second unit adds its own $10 but loses the $1 cut on the first: $10 − $1 = $9.
- A third unit: $9 on all three, total revenue $27, marginal revenue $9 − 2 × $1 = $7.
- Each new unit adds its own price but subtracts the price cut on every earlier unit: marginal revenue is below the price, and the gap widens.

Marginal revenue lies below the demand curve

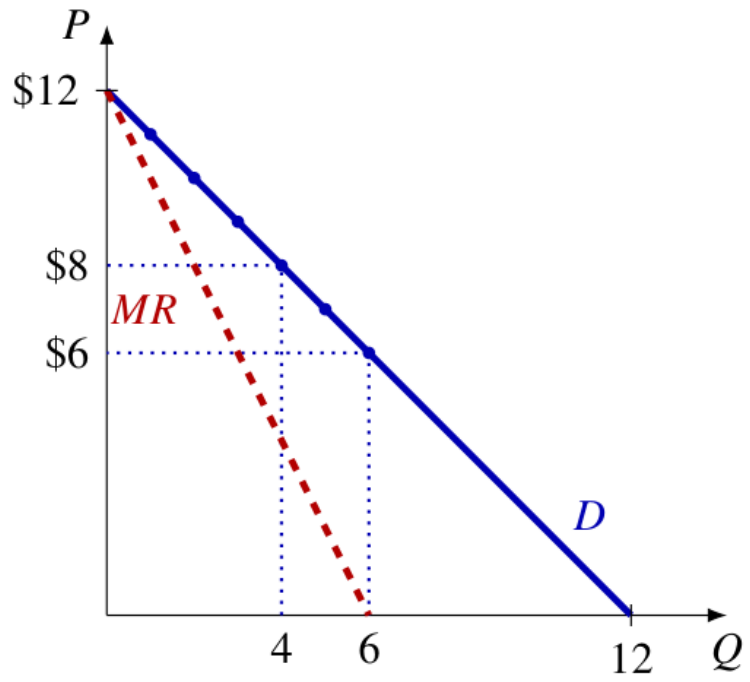


Figure 11.1 Demand and marginal revenue for a monopolist.

- Each point on demand shows the price buyers will pay for that quantity.
- Straight-line demand: MR starts at the same price ($12) but is twice as steep, reaching zero at 6 units instead of 12.

The rule: produce where MR equals MC

- **Profit-maximizing rule (unchanged)**

  Produce where marginal revenue equals marginal cost. But then charge the highest price the demand curve will bear for that quantity: a price above marginal cost.

The rule: produce where MR equals MC

A WORKED CHOICE

With the demand of Table 11.1 and a marginal cost of $4, MR falls twice as fast as the price: it starts at $12 and falls by $2 for each extra unit, so it reaches the $4 marginal cost at 4 units. Demand shows buyers pay $8, double marginal cost.

MR equals MC at 4 units, price $8

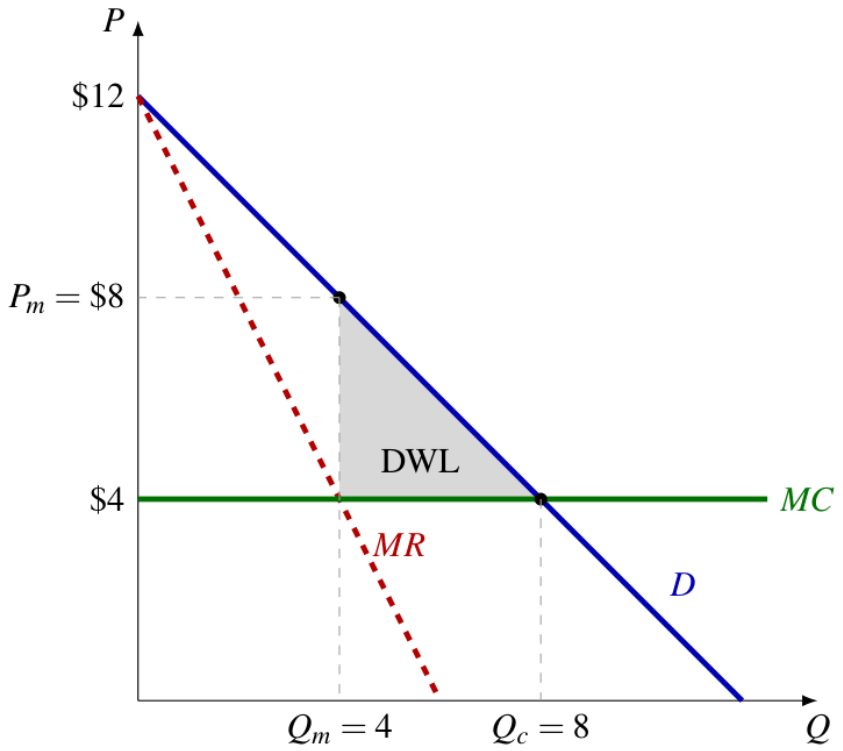


Figure 11.2 The monopolist's choice of output and price.

Quantity. 4 units, where MR meets the $4 marginal cost.

Price. $8, given by the demand curve.

Loss. A deadweight-loss triangle versus the competitive output.

Explore Figure 11.2: intro-micro-companion-website.netlify.app/monopoly

Choosing output with a marginal cost of $4

Quantity	Total revenue	Marginal revenue	Marginal cost	Profit (total revenue minus total cost)
1	$11	$11	$4	$11 − $4 × 1 = $7
2	$20	$9	$4	$20 − $4 × 2 = $12
3	$27	$7	$4	$27 − $4 × 3 = $15
4	$32	$5	$4	$32 − $4 × 4 = $16
5	$35	$3	$4	$35 − $4 × 5 = $15
6	$36	$1	$4	$36 − $4 × 6 = $12

Table 11.2 Marginal revenue, marginal cost, and profit, with a marginal cost of $4.

Reading Table 11.2

- The fourth unit adds $5, more than its $4 marginal cost: profit rises from $15 to $16.
- The fifth adds only $3, less than $4: profit falls from $16 to $15. The monopolist stops at 4 units, as in Figure 11.2.

11.4

Monopoly versus Competition: The Welfare Cost

- Less output, higher price
- Transfer to the monopolist
- The deadweight loss
- Markup, Lerner index, patents

Less output, higher price, and a deadweight loss

	Monopoly	Perfect competition	Difference
Quantity	4	8	4 − 8 = −4
Price	$8	$4	$8 − $4 = +$4
Consumer surplus	½ × 4 × ($12 − $8) = $8	½ × 8 × ($12 − $4) = $32	$8 − $32 = −$24
Producer surplus	4 × ($8 − $4) = $16	8 × ($4 − $4) = $0	$16 − $0 = +$16
Total surplus	$8 + $16 = $24	$32 + $0 = $32	$24 − $32 = −$8
Deadweight loss	½ × (8 − 4) × ($8 − $4) = $8	$0	$8 − $0 = +$8

Table 11.3 Monopoly versus perfect competition.

Reading Table 11.3

- A competitive industry with the same costs would sell 8 units at $4; the monopolist sells only 4 units at $8.
- Consumer surplus is the triangle above the price and below demand; producer surplus is the rectangle above marginal cost and below the price.
- Consumer surplus falls by $24: $16 is a transfer to the monopolist, and the last $8 is the deadweight loss.

The harm is two-part

A transfer

Consumers pay more and consume less, so some consumer surplus becomes the monopolist's profit.

A deadweight loss

The units between 4 and 8, which buyers value above the $4 marginal cost, are never produced. Those mutually beneficial trades simply vanish.

The harm is two-part

WHY IT MATTERS

The deadweight loss is the same kind of loss a tax created in Chapter 4.

Two effects, two different concerns

- **Transfer**

  By charging above marginal cost, the monopolist takes for itself surplus that under competition went to consumers ($16 in Table 11.3). Society's total surplus is redistributed, not destroyed: a question of fairness and distribution.

Two effects, two different concerns

- **Deadweight loss**

  The surplus from the units never produced because output is restricted is lost ($8 in Table 11.3). These are trades buyers valued above their cost; the surplus they would have created vanishes for everyone. This is what economists mean by inefficient.

Self-check: the deadweight-loss triangle

SELF-CHECK

With the demand of Table 11.1 and a marginal cost of $4, a monopolist sells 4 units at $8. What would a competitive industry sell, and what is the deadweight loss?

Competition sets price equal to marginal cost: the price falls to $4 at 8 units. The lost output is the 4 units between 4 and 8, so the deadweight loss is the triangle ½ × 4 × $4 = $8.

Elasticity determines the size of the markup

- **Demand not very elastic:** buyers have few alternatives, so the firm can raise the price far above marginal cost; the markup is large.
- **Very elastic demand:** the markup must be small.
- Perfect competition (zero markup) and a monopolist facing demand that is not very elastic (the EpiPen) are the two ends of a spectrum; most real firms lie in between.

Measuring market power: the Lerner index

- **Lerner index (Lerner, 1934)**

  The markup as a share of the price: the gap between price and MC, divided by the price. In our example, the $4 gap divided by the $8 price, or 0.5: half of the price is markup. Competition gives 0; values near 1 mean strong market power.

Measuring market power: the Lerner index

THE LINK TO ELASTICITY

A profit-maximizing firm sets the Lerner index equal to 1 divided by the elasticity at its price. At $8, a 1% price rise cuts quantity by 2%: elasticity 2, so the index is 1 divided by 2, or 0.5. A monopolist always prices where demand is elastic (elasticity above 1).

The markup and the elasticity of demand

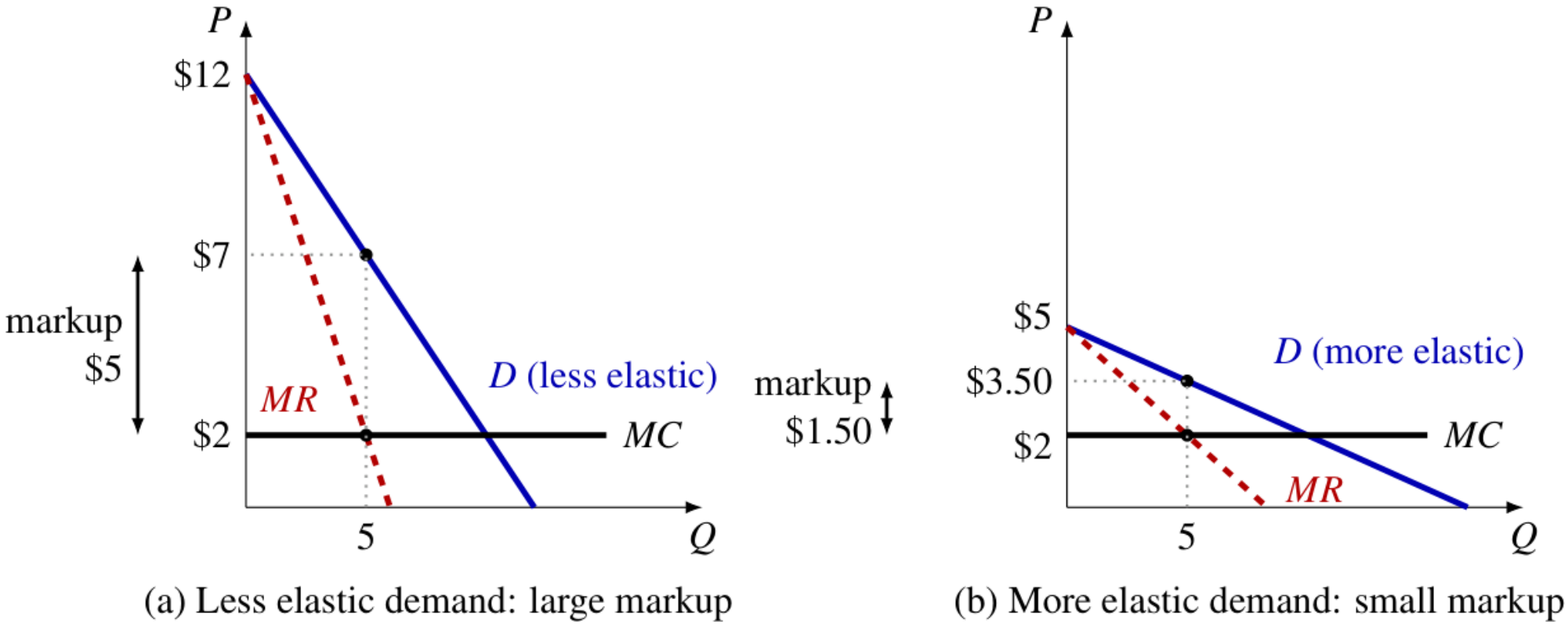


Figure 11.3 The markup and the elasticity of demand, with a marginal cost of $2.

Reading Figure 11.3

- **(a) Less elastic:** the price, $7, lies far above the $2 marginal cost, a $5 markup.
- **(b) More elastic:** the price, $3.50, stays closer to marginal cost, a $1.50 markup.
- With an elasticity of 4 instead of 2, the Lerner index would fall to 0.25: only a quarter of the price would be markup.

Why the monopolist prices where demand is elastic

Quantity	Price	Total revenue	Marginal revenue	Profit (marginal cost of $4)	Demand
4	$8	$8 × 4 = $32	$32 − $27 = $5	$32 − $4 × 4 = $16	elastic
5	$7	$7 × 5 = $35	$35 − $32 = $3	$35 − $4 × 5 = $15	elastic
6	$6	$6 × 6 = $36	$36 − $35 = $1	$36 − $4 × 6 = $12	elastic
7	$5	$5 × 7 = $35	$35 − $36 = −$1	$35 − $4 × 7 = $7	inelastic
8	$4	$4 × 8 = $32	$32 − $35 = −$3	$32 − $4 × 8 = $0	inelastic
9	$3	$3 × 9 = $27	$27 − $32 = −$5	$27 − $4 × 9 = −$9	inelastic
10	$2	$2 × 10 = $20	$20 − $27 = −$7	$20 − $4 × 10 = −$20	inelastic

Table 11.4 Total revenue, marginal revenue, and profit along the demand of Table 11.1.

Reading Table 11.4: elastic and inelastic

- Total revenue rule (Section 3.3): a price cut that raises total revenue means demand is elastic; one that lowers it means demand is inelastic.
- Up to 6 units, each price cut raises total revenue: demand is elastic.
- From 6 to 7 units, total revenue falls from $36 to $35: demand becomes inelastic at the midpoint, a price of $6.

A monopolist never produces where demand is inelastic

- At 8 units and $4, cutting output to 7 and raising the price to $5 raises total revenue by $35 − $32 = $3 and lowers total cost by $4: profit rises from $0 to $7.
- Because every unit costs something, the monopolist stops while marginal revenue is still positive, where it equals marginal cost.
- The monopoly price of $8 lies above the midpoint, $6, in the elastic portion.

A temporary monopoly as a reward for innovation

- Many products exist only because their inventors expected to earn a monopoly profit for a while.
- A new medicine: a median cost of $985 million per new drug, counting the drugs that fail (Wouters et al., 2020).
- Without a patent, a rival could copy the molecule and price near marginal cost; a patent gives 20 years of exclusive sales.

The deal has two sides

- **The cost:** while the patent lasts, the inventor prices above marginal cost, and society bears a deadweight loss like the one in Table 11.3.
- **The benefit:** the prospect of profit pays for research that would not otherwise take place; once the patent expires, rivals copy the product and its price falls toward marginal cost (Section 11.8).

GLP-1 drugs show both sides

- Semaglutide (Ozempic, Wegovy) and tirzepatide (Mounjaro, Zepbound) are patented; U.S. list prices have run to about $1,000 or more a month.
- At approval in November 2023, Zepbound listed at about $1,060 a month, against $1,349 for Wegovy (KFF Health News, 2023).
- Semaglutide's patent expired in India and Brazil in March 2026; generics entered India within days at a fraction of the price (CNBC, 2026).

Patents and other barriers

- The EpiPen (Section 11.7) shows the cost side alone: its price rose more than sixfold for a device that had barely changed.
- Nvidia's patented chip designs and its CUDA software (Section 11.7.1) helped its revenue reach $130.5 billion in fiscal 2025, up 114 percent (Nvidia Corporation, 2025).
- A patent does not make monopoly harmless: governments accept the deadweight loss for a limited time in exchange for innovation.

11.5

Monopsony: A Single Buyer

- Mirror image of monopoly
- MCL above the wage
- Hires where MCL equals MRP; wage given by supply
- Minimum wage can raise employment

Market power can rest with the buyer

Monopsony

A market with a single buyer: the only large employer in a small town, or a dominant supermarket chain buying from local farmers.

THE MIRROR IMAGE

The logic mirrors monopoly. Just as a monopolist restricts how much it sells to raise the price it receives, a monopsonist restricts how much it buys to lower the price it pays.

The mechanism in a labor market

- **Marginal cost of labor (MCL) lies above the wage**

  A monopsonist faces the entire upward-sloping labor supply curve of the town. To attract one more worker it must post a higher wage (each additional worker must be drawn away from a better alternative), and pay that higher wage to every worker already on payroll. The extra cost of one more worker is therefore above the wage.

The mechanism in a labor market

THE HIRING RULE

The firm hires up to where MCL equals MRP (the value of an extra worker's output) and pays the wage that the labor supply curve gives at that quantity. The result: fewer workers hired and a lower wage than a competitive labor market would yield.

A worked example: one employer, six workers

Workers	Wage needed to attract them	Total labor cost	Marginal cost of labor	Value of an extra worker's output
1	$12	1 × $12 = $12	$12 − $0 = $12	$24
2	$14	2 × $14 = $28	$28 − $12 = $16	$22
3	$16	3 × $16 = $48	$48 − $28 = $20	$20
4	$18	4 × $18 = $72	$72 − $48 = $24	$18
5	$20	5 × $20 = $100	$100 − $72 = $28	$16
6	$22	6 × $22 = $132	$132 − $100 = $32	$14

Table 11.5 Hiring decisions of a monopsonist.

Monopsony versus competition

- The second worker costs $16, not $14: the $14 wage plus the $2 raise paid to the first worker.
- **Monopsony:** the third worker adds $20 and costs $20; the fourth would add $18 but cost $24. It hires 3 workers at $16.
- **Competition:** the wage equals the value of an extra worker's output: 4 workers at $18 (a fifth would need $20 but add only $16).

The labor market of Table 11.5

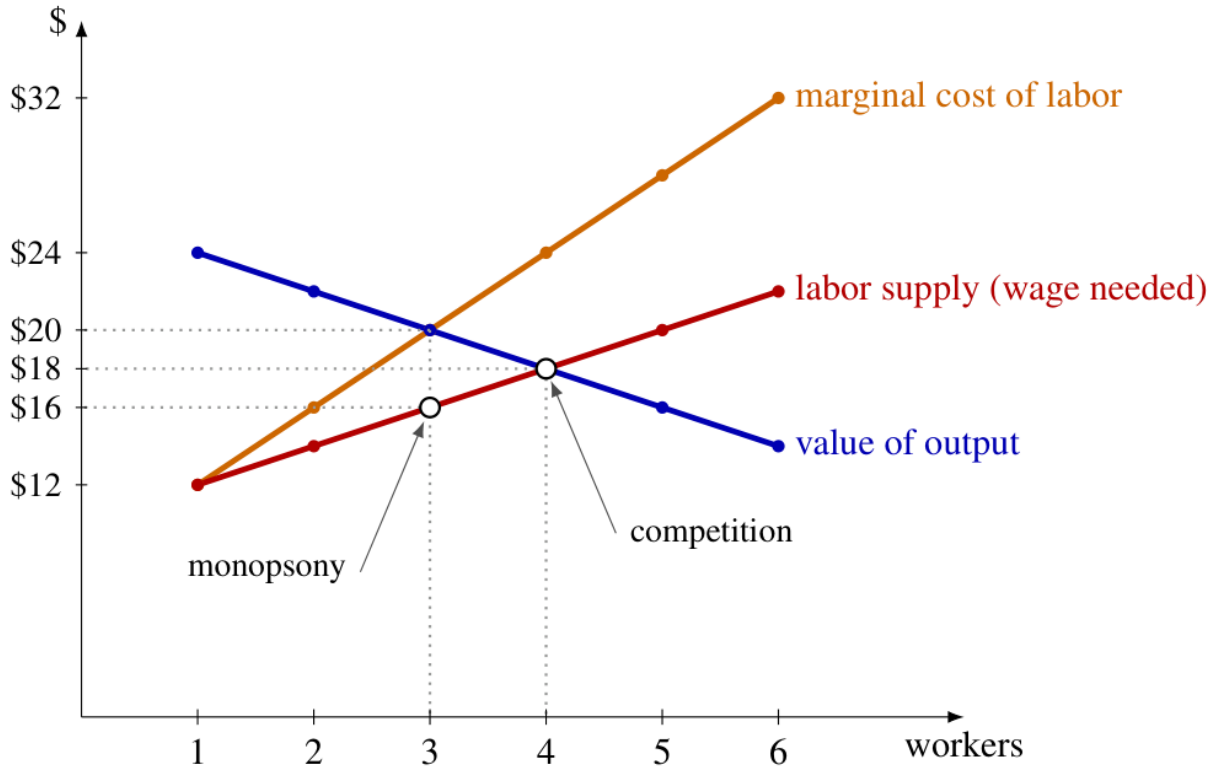


Figure 11.4 The labor market of Table 11.5.

- **Monopsony:** 3 workers at a wage of $16.
- **Competition:** 4 workers at a wage of $18.

A minimum wage of $18

- Up to 4 workers are willing to work at $18, so each of the first 4 costs the firm $18.
- The firm hires the fourth worker, who adds $18 of output at a cost of $18.
- Employment rises from 3 to 4 and the wage from $16 to $18: the competitive outcome.

Why a minimum wage can sometimes raise employment

- Monopsony is one reason economists watch employer concentration in local labor markets: a handful of dominant employers can push wages below the competitive level even without collusion.
- In a competitive labor market (Chapter 4) a binding minimum wage reduces employment.
- In a monopsonized labor market, a minimum wage set just above the monopsony wage can raise employment as well as wages: the floor forces the employer to stop suppressing the wage.

Why a minimum wage can sometimes raise employment

THE EVIDENCE

Monopsony is the second explanation, mentioned in Section 4.7, for why the studies of that section, starting with Card and Krueger (1994), found small job losses, or none at all, from moderate minimum-wage increases.

Monopsony in practice

- Warehouse workers where an Amazon fulfillment center is the largest hirer, nurses in a city served by a single hospital system, and meatpacking workers in a small plant town.
- Azar et al. (2020), using online job postings across US commuting zones: about 60 percent of local labor markets are highly concentrated.
- The fewer the large employers, the closer the wage lies to the monopsony benchmark.

11.6

Natural Monopoly and the Case for Regulation

- Falling average cost over the whole market
- Regulator's dilemma: MC vs. AC pricing
- Bell System and América Móvil

When competition cannot fix the problem

- **Natural monopoly**

  Arises when a single firm can serve the whole market at lower cost than two or more firms could, because average cost keeps falling over the entire range of demand (Chapter 9).

When competition cannot fix the problem

WHY ONE FIRM

Water pipes, electricity grids, and gas networks: a large fixed cost and a low marginal cost. Two networks would each pay the whole fixed cost while serving only part of the market. The very economies of scale that make one firm efficient also leave it with monopoly power.

A natural monopoly with numbers

- The demand of Table 11.1, a constant marginal cost of $2, and a fixed cost of $16.
- Average cost is $2 plus $16 divided by the quantity, and it falls as the quantity grows:

$$\$2 + \$16/5 = \$5.20 \text{ at 5 units;} \quad \$2 + \$16/10 = \$3.60 \text{ at 10 units}$$

The regulator's dilemma: which price?

Marginal-cost pricing

A price equal to MC reaches the efficient competitive output. But because MC lies below average cost throughout, the firm cannot cover its average cost and loses money; it needs a subsidy to survive.

Average-cost pricing (the usual compromise)

Price equals AC, so the firm just breaks even (zero economic profit), while output is far closer to the efficient level. This is the principle behind utility rate-setting in most countries.

Regulating a natural monopoly

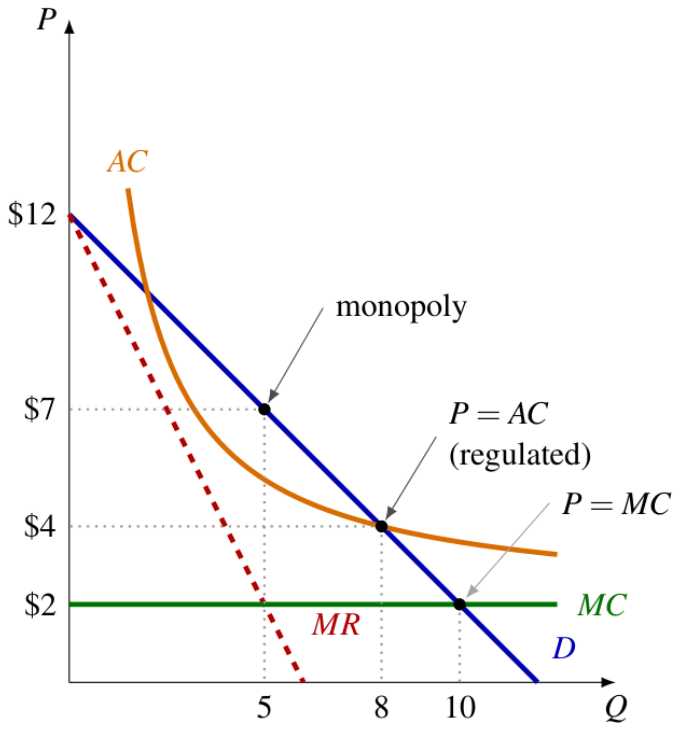


Figure 11.5 Regulating a natural monopoly.

Monopoly. 5 units at $7 (MR equals MC, unregulated).

AC pricing. 8 units at $4 (break even).

MC pricing. 10 units at $2 (efficient, needs subsidy).

Explore Figure 11.5: intro-micro-companion-website.netlify.app/natural-monopoly

Table 11.6: three prices for a natural monopoly

	Unregulated monopoly	Marginal cost pricing	Average cost pricing
Quantity	5	10	8
Price	$12 − $5 = $7	$12 − $10 = $2	$12 − $8 = $4
Average cost	$2 + $16/5 = $5.20	$2 + $16/10 = $3.60	$2 + $16/8 = $4
Profit	($7 − $5.20) × 5 = $9	($2 − $3.60) × 10 = −$16	($4 − $4) × 8 = $0
Deadweight loss	½ × 5 × $5 = $12.50	$0	½ × 2 × $2 = $2

Table 11.6 A natural monopoly under three prices.

Reading Table 11.6

- The deadweight loss is a triangle: base = the output gap from the efficient 10 units; height = the price minus the $2 marginal cost.
- The unregulated monopoly leaves a deadweight loss of $12.50; marginal cost pricing removes all of it, but only with a $16 subsidy.
- Average cost pricing removes all but $2 of it and costs taxpayers nothing.

Society's two usual responses

Regulate

Cap the price near average cost, so the firm covers its costs without exploiting customers: the model for public utilities in most countries.

Own and operate

The government may run the network, replacing the private monopolist entirely.

Society's two usual responses

THE POLICY QUESTION

Natural monopoly is a case where the competitive benchmark of Chapter 10 is simply unavailable. The question becomes how best to restrain a monopoly that cannot be competed away.

A rich-country case: the Bell System

- For most of the twentieth century, U.S. telephone service was a single regulated monopoly: AT&T, the Bell System.
- In 1974 the Department of Justice sued under the antitrust laws; in 1982 AT&T agreed to break itself up.
- The 1984 settlement split it into a long-distance company (keeping the AT&T name) and seven regional "Baby Bells."

A rich-country case: the Bell System

THE RESULT

The breakup opened long-distance service to competition and was followed by a steep fall in long-distance prices: the standard case of dismantling a monopoly.

A developing-country case: América Móvil

- After Mexico privatized Telmex in 1990, it and its mobile arm América Móvil (Carlos Slim) came to dominate the market.
- A 2013 constitutional reform created an independent regulator, the IFT, with strong new powers. (In 2025 Mexico dissolved the IFT and moved its powers to new government agencies.)
- It named América Móvil a "preponderant economic agent" and required it to share its network and offer free interconnection.

A developing-country case: América Móvil

−27%

communications prices in Mexico, June 2013 to April 2020

73 → 50%

América Móvil's telecom market share, 2013–2020

THE LESSON

A breakup in a rich country, an independent regulator forcing access in a developing one: the tools differ, the goal is the same.

11.7

Case Study: The EpiPen and De Beers

Two monopolies, built on different barriers.

11.7 CASE STUDY

EpiPen: patents and regulatory barriers

- After acquiring the product in 2007, Mylan raised the list price of a two-pack from about $94 to more than $600 by 2016.
- Mylan held over 80%, at times near 100%, of the U.S. market for epinephrine auto-injectors.
- Patents and the difficulty of winning regulatory approval kept competitors out.

~550%
price increase, 2007–2016

80%+
U.S. market share held by Mylan

11.7 CASE STUDY

EpiPen: patents and regulatory barriers

WHY IT STUCK

With no close substitute available to buyers, demand was far less elastic than for most products, and the markup was enormous: as the logic of Section 11.4.2 predicts.

De Beers: control of a key resource

- Founded in 1888, De Beers controlled an estimated 90% of the world's production of rough diamonds by 1900 (Gemological Institute of America, n.d.); for most of the twentieth century it sold the bulk of the world's rough diamonds through its own sales channel.
- It restricted supply, held diamonds off the market in downturns, and its "A Diamond Is Forever" campaign discouraged resale, keeping a second-hand supply from undercutting it.
- Its control loosened once other producers began to sell their diamonds on their own.

11.7 CASE STUDY

De Beers: control of a key resource

THE TAKEAWAY

De Beers maintained high prices the way the model predicts: by restricting supply.

As other producers began to sell on their own, its grip on the price weakened with its share.

How De Beers's control loosened

- Australia: the Argyle mine left the De Beers sales channel in 1996.
- Canada: the country's first diamond mine, opened in 1998, marketed most of its output independently.
- Russia: Alrosa, required by the European Commission to scale back its sales to De Beers, built its own list of buyers.
- De Beers's share of the world's rough diamonds, by value, fell to about 64% by 2000 and roughly 30% by 2019.

Nvidia and the AI chip stack

- By 2024 Nvidia accounted for about 80 percent of AI-training accelerator sales worldwide; in 2025 it became the first company valued at $4 trillion and then $5 trillion.
- **Barrier 1, the chip:** a leading-edge accelerator takes years and billions of dollars to design.
- **Barrier 2, CUDA:** two decades of libraries, tools, and trained engineers are built on this software layer.

The effect the model predicts

- With demand far less elastic than for most products, Nvidia sold its H100 chips (Section 9.5.2) well above estimated manufacturing cost.
- Its 2020 attempt to buy Arm for about $40 billion was abandoned in 2022 after U.S., U.K., and EU authorities objected.
- By 2024 several authorities were investigating whether Nvidia tied scarce chips to its other products.

11.8

Empirical Evidence: Patent Expirations and Generic Entry

A natural experiment every time a drug loses its patent.

Patent expiry: monopoly becomes competition

- While the patent lasts, the maker holds a legal monopoly and prices far above marginal cost.
- When the patent expires, generic manufacturers, selling a chemically identical product, are free to enter.
- The market then moves toward the competitive benchmark of Chapter 10.

Patent expiry: monopoly becomes competition

Drugs About to Lose Their Patent

Upcoming U.S. patent expirations: Merck's Keytruda (2028), Novo Nordisk's Ozempic (2031), Bristol Myers Squibb / Pfizer's Eliquis (2028), and Regeneron / Bayer's Eylea (composition patent expired in 2023, biosimilars launching in 2024). List prices during exclusivity ran to thousands, and for some of these drugs tens of thousands, of dollars per patient per year.

Prices collapse as generics multiply

Generic competitors	Price vs. original brand
1 (first generic)	about 39% lower
2	about 54% lower
4	about 79% lower
6 or more	more than 95% lower

FDA data: the more generics enter, the closer price falls toward marginal cost.

Prices collapse as generics multiply

THE PROGRESSION

With a single rival, competition is still limited: the generic price is about 39% below the brand's.

As more generics enter, the price falls much further. Reiffen and Ward (2005): prices stay above long-run marginal cost until there are eight or more producers.

CHAPTER 11

Key takeaways

1. **Why monopolies arise**

   A barrier to entry keeps rivals out: a key resource, scale, patents, franchise, or network effects.

2. **MR below price**

   Selling more means cutting the price on all units, so MR is below the price; the firm produces where MR equals MC.

3. **The welfare cost**

   Less output, higher price: a transfer to the monopolist plus a deadweight loss from trades that never happen.

Key takeaways

4 **The markup**

How far price exceeds MC depends on elasticity; less elastic demand sustains larger markups; Lerner index equals 1 divided by the elasticity (the EpiPen).

5 **Monopsony**

A single buyer hires fewer workers at a lower wage; a minimum wage just above the monopsony wage can raise both.

6 **Natural monopoly**

Falling AC over the whole market; regulators use average-cost pricing to let the firm break even.

CHAPTER 11

Key takeaways

7 **Evidence**

EpiPen and De Beers; Bell System and América Móvil; drug prices fall 95%+ as generics enter.

12

CHAPTER 12

Pricing Strategies: Price Discrimination and Bundling

How firms with market power charge different buyers different prices for essentially the same thing, and how they get away with it.

CHAPTER 12

What this chapter covers

12.2–12.3 Foundations

- When discrimination is possible
- First-degree (perfect)
- Is it good or bad?

12.4–12.6 The toolkit

- Third-degree (group) pricing
- Second-degree (menu) pricing
- Bundling & two-part tariffs

12.7–12.8 Evidence

- Case: airline yield management
- Empirical: price dispersion
- Online & personalized pricing

CHAPTER 12

What this chapter covers

THE BIG IDEA

A monopolist charging a single price forgoes profit. Price discrimination is the attempt to recover the surplus left by some buyers and to reach others who were priced out.

12.1 Introduction

- Different buyers, different prices
- Capturing more of the surplus
- A road map of the strategies

The same good, many prices

Flights

One passenger pays $200 for a seat; the traveler beside her pays $650.

Cinema

A student and a retiree are charged less than the adult next to them.

Software

A suite that would cost hundreds bought piece by piece sells for far less as one bundle.

The same good, many prices

THE QUESTIONS

Why do sellers charge different buyers different prices for essentially the same thing? This chapter studies price discrimination and the related practice of bundling: strategies to capture more of the surplus from Chapter 6.

12.2

When Is Price Discrimination Possible?

- Market power
- The ability to separate buyers
- No resale (no arbitrage)

What price discrimination means

Price discrimination

Selling the same good to different buyers at different prices, for reasons unrelated to cost.

WHY IT MATTERS

When the three conditions on the next slide hold, a seller can practice one of three degrees of price discrimination: perfect, group, or menu pricing.

Three conditions must hold

1

1. Market power

The seller must be a price setter, not a price taker; a perfectly competitive firm has no room to discriminate.

2

2. Separate buyers

The seller can tell buyers apart by willingness to pay or elasticity: by age, location, timing, or the choices they make.

3

3. No resale

Low-price buyers cannot resell to high-price buyers (no arbitrage).

Three conditions must hold

NO-RESALE, IN PRACTICE

This is why discrimination is common for services that cannot be resold (a haircut, a flight) and rare for easily traded goods.

12.3

First-Degree (Perfect) Price Discrimination

- Each buyer pays her maximum
- No consumer surplus remains
- Output reaches the competitive level

Charging each buyer her maximum

- **First-degree (perfect) price discrimination**

  The seller charges each buyer her maximum willingness to pay. Every buyer pays the most she would, so no consumer surplus remains; the seller captures all of it as profit.

Why Output Reaches the Efficient Level

Output expands all the way to the competitive level, where price equals marginal cost. No longer forced to drop the price on earlier units to sell more, the seller serves every buyer whose value exceeds marginal cost.

The seller captures the entire surplus

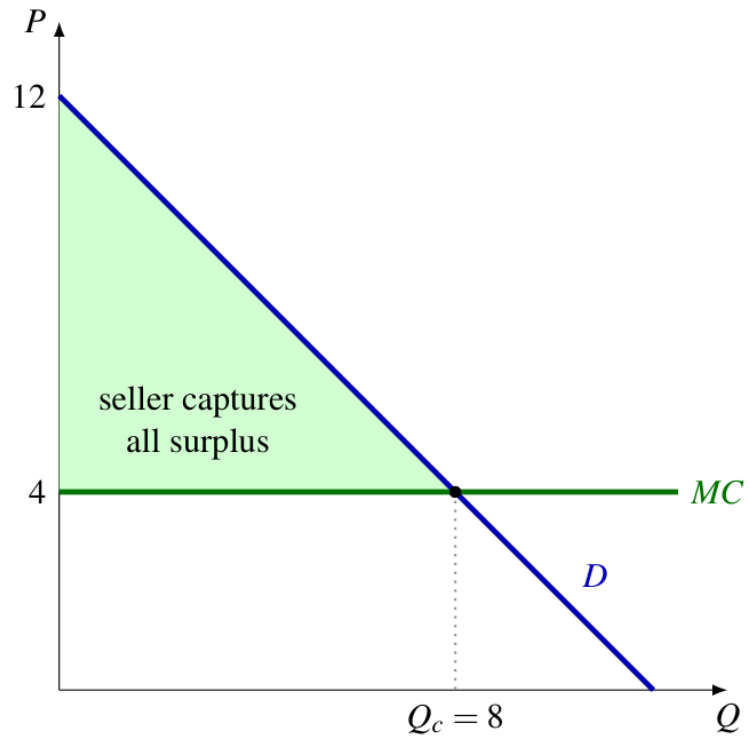


Figure 12.1 Perfect price discrimination.

- The seller takes the entire area between the demand curve and marginal cost.
- Output reaches the competitive level: demand runs from $12 down to 12 units and MC is $4, so the seller sells 8 units.
- Rare in practice: sellers seldom know each buyer's valuation.

Where we see it approximated

Car dealers

Haggling each customer down to her personal limit.

Financial aid

Colleges charge each family what it can pay.

Online sellers

Increasingly use personal data to tailor prices to each buyer.

Each comes close to charging every buyer her own price.

12.3.1

Is Price Discrimination Good or Bad?

A mixed verdict the economics explains.

12.3.1 GOOD OR BAD?

A genuine trade-off

The cost: surplus shifts

It transfers surplus from buyers to the seller. Under perfect discrimination consumers keep none of the gains from trade, which strikes many as unfair.

The gain: more access

The seller serves any buyer who pays above marginal cost, reaching low-value customers a single-price monopolist would shut out.

12.3.1 GOOD OR BAD?

A genuine trade-off

THE VERDICT

It often raises the quantity sold: the student with the discounted ticket, the patient who gets the lower-priced drug, the early-booking traveler. Whether any case is desirable is a normative question on which reasonable people differ.

12.4 Third-Degree Price Discrimination

- Different prices to identifiable groups
- The elasticity rule
- Two markets, two prices
- A worked two-market example

Charging groups different prices

Third-degree price discrimination

The most common form: charging different prices to identifiable groups of buyers.

THE ELASTICITY RULE

The rule follows from elasticity: charge the higher price to the group with the less elastic demand, and the lower price to the more elastic group.

Group pricing all around us

- Student, senior, and child discounts: these groups have lower or more price-sensitive demand.
- Matinee versus evening movie prices; peak versus off-peak fares.
- Lower drug prices in poorer countries.

COUPONS

Coupons are a variant of this rule: by clipping one, a shopper signals she is price sensitive enough to spend the effort and gets a lower price; the less price-sensitive shopper, who does not bother, pays full price.

Two markets, two prices

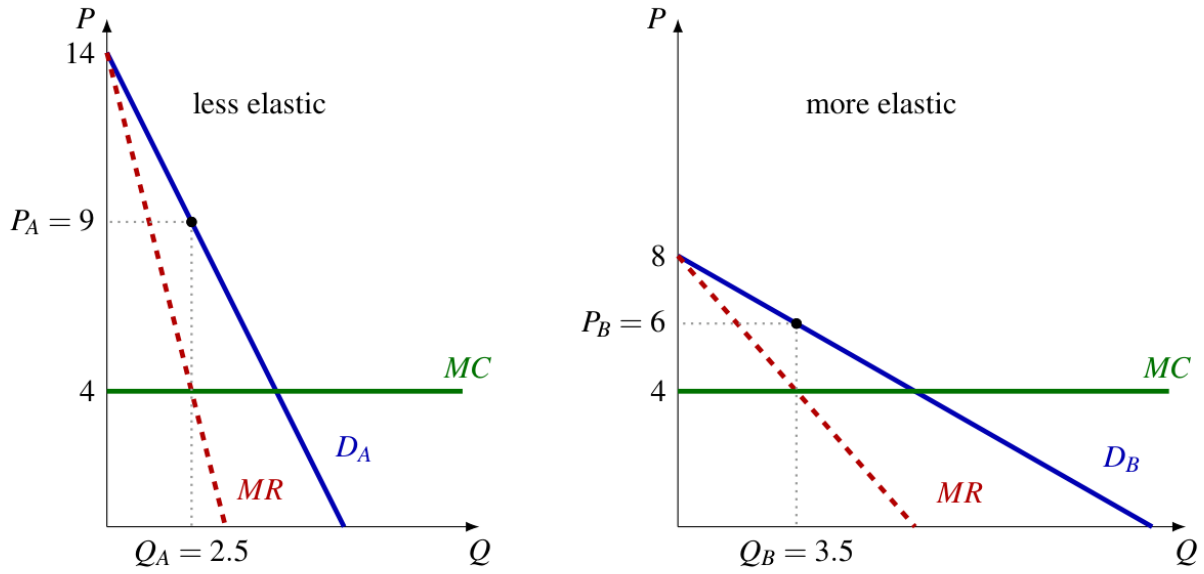


Figure 12.2 Third-degree price discrimination in a less elastic and a more elastic market.

Explore Figure 12.2: intro-micro-companion-website.netlify.app/price-discrimination

- The firm splits buyers into two markets with different elasticities.
- Higher price where demand is less elastic; lower price where it is more elastic.
- With MC = $4 in both: 2.5 units at $9 in the less elastic market, 3.5 units at $6 in the more elastic one.

Software: businesses vs. students

Group	Demand	Will pay	Price set
Businesses (100)	inelastic	up to $200	$200
Students (300)	elastic	at most $30	$30

A software firm tells the two groups apart (student ID) and resale is hard.

Software: businesses vs. students

WHY IT EXCEEDS ONE PRICE

A single price forces a painful choice: set it high and lose all the students, or low and forgo the revenue businesses would pay.

Charging $200 to businesses and $30 to students collects the most each will bear: more than any single price could earn. The higher price goes to the less elastic group.

One price or two? The numbers

- The customers: 100 businesses will pay up to $200 each and 300 students at most $30 each; an extra copy costs the firm nothing.
- A single price of $200: only the businesses buy, and revenue is 100 × $200 = $20,000.
- A single price of $30: all 400 customers buy, and revenue is only 400 × $30 = $12,000.
- Two prices: $200 to businesses and $30 to students earn $20,000 + 300 × $30 = $29,000, more than either single price.

Everyone served, the firm earns more

THE OUTCOME

Everyone is served, the firm earns more, and the students get software they could not have afforded at the business price: the same mixed verdict as in Section 12.3.1.

Everyone served, the firm earns more

SELF-CHECK

A cinema sells to students (elastic demand) and adults (inelastic demand). Which group should pay the higher price, and why?

Adults. With less elastic demand they cut back little when the price rises, so the profit-maximizing markup is larger for them; price-sensitive students get the discount. The firm charges the lower price in the more elastic market.

12.5 Second-Degree Price Discrimination

- A menu, not a label
- Quantity discounts
- Versioning into tiers

Let buyers reveal themselves

- **Second-degree price discrimination**

  When the seller cannot tell groups apart, it offers a menu and lets buyers reveal themselves through their choices.

Let buyers reveal themselves

Quantity discounts

A lower price per unit in bulk.

Versioning

A basic, standard, and premium tier: economy / premium economy / business class; free / pro / enterprise editions.

Versioning with numbers

- A software firm sells a basic and a premium version to 50 power users and 50 casual users. It cannot tell them apart; each version costs nothing to supply.

	Number of buyers	Basic version	Premium version
Power users	50	$40	$100
Casual users	50	$30	$40

Table 12.1 Buyers' valuations of two versions of a program.

Selling only the premium version

- At $100, only the power users buy: revenue is 50 × $100 = $5,000.
- At $40, everyone buys: revenue is 100 × $40 = $4,000.
- With a single version, the best the firm can do is $5,000.

Selling both versions

- The basic version sells at $30, the most casual users will pay for it.
- A power user who bought the basic version would keep $40 − $30 = $10, so the premium price P must leave her at least as much: $100 − P ≥ $10, that is, P ≤ $90.
- At $90 and $30, revenue is 50 × $90 + 50 × $30 = $4,500 + $1,500 = $6,000.

What the menu achieves

- The menu raises revenue from $5,000 to $6,000.
- The basic version brings in the casual users, but power users get a $10 discount on the premium version so that they do not switch.
- With many types of buyers and many versions, finding the best menu requires contract theory, beyond this book.

12.6

Bundling and Two-Part Tariffs

- Selling several goods together
- The two-part tariff
- Why bundling raises profit

Two related strategies

Bundling

Selling several goods together for a single price: a cable package, an office software suite, a season ticket. It raises profit when buyers' valuations differ but partly offset one another.

Two-part tariff

A fixed fee plus a per-unit price: a warehouse club's membership plus groceries, a theme park's entry plus in-park spending, a printer plus its ink. The fixed fee captures surplus; the per-unit price keeps the buyer buying.

Step 1: selling the channels separately

Customer	Values sports	Values movies
Sports fan	$15	$3
Film fan	$2	$14

A cable company sells a sports channel and a movie channel to two customers.

Step 1: selling the channels separately

BEST SINGLE PRICES

Sports channel: charge $15, sell one (to the sports fan) for $15; charge $2, sell two for 2 × $2 = $4. So $15 is best.

Movie channel: earns most at $14, again selling one (charging $3 sells two for 2 × $3 = $6).

Separate selling: $15 + $14 = $29.

Step 2: bundle the two channels

Customer	Values the bundle
Sports fan	$15 + $3 = $18
Film fan	$2 + $14 = $16

Price the bundle at $16 and both customers buy it.

Step 2: bundle the two channels

BUNDLING WINS

Pricing the bundle at $16 sells it to both: revenue = 2 × $16 = $32.

That exceeds the $29 from separate selling.

It works because valuations are negatively correlated: each loves one channel and shrugs at the other, so the bundle smooths out their differences.

The lesson: negatively correlated values

THE GENERAL LESSON

When buyers' valuations are negatively correlated, the bundle lets the firm capture more of the total they are collectively willing to pay than selling each item alone.

Two-part tariffs

- A fixed fee plus a per-unit price: a warehouse club's membership fee plus groceries, a theme park's entry ticket plus in-park spending, a printer plus its ink cartridges (Chapter 5).
- The fixed fee captures the buyer's surplus while the per-unit price keeps her buying.

A tennis club: one price per hour

- Each member's monthly demand for court time is $P = 10 - Q$ (Q hours, P the price per hour); each hour costs the club $2.
- Like the monopolist of Chapter 11, the club sets $MR = 10 - 2Q$ equal to marginal cost: $10 - 2Q = 2$ gives $Q = 4$ hours at $\$10 - \$4 = \$6$.
- **Profit:** $(\$6 - \$2) \times 4 = \$16$ per member.

A tennis club: a two-part tariff

- The club charges \$2 per hour, its marginal cost: each member plays $Q = 10 - 2 = 8$ hours.
- A member's consumer surplus is the triangle

$$\tfrac{1}{2} \times 8 \times (\$10 - \$2) = \$32$$

- A monthly membership fee of \$32 collects all of it; since each hour sells at cost, profit is the \$32 fee.

What the two-part tariff achieves

- Profit doubles, from $16 to $32 per member.
- Members play the efficient 8 hours, as under first-degree price discrimination (Section 12.3).
- When members' demands differ, a single fee cannot capture every member's full surplus, but the club still gains from a low per-hour price plus a fee.

12.7

Case Study: How Airlines Charge Many Prices for One Cabin

Yield management: price discrimination at its most refined.

12.7 CASE STUDY

Two groups, very different demand

Business travelers

Trips are urgent, expensable, and price-insensitive: inelastic demand.

Leisure travelers

They plan ahead and shop hard on price: elastic demand.

THE GOAL

Passengers in a single economy cabin have often paid very different fares, even though their seats are nearly identical. The airline's goal: separate these two groups (a practice they call yield management).

12.7 CASE STUDY

Rules that make travelers sort themselves

- The airline cannot ask which type you are, so it designs rules: a blend of the group and menu pricing of Sections 12.4 and 12.5.
- Low fares require booking weeks ahead and, classically, a Saturday-night stay: easy for a leisure traveler, impossible for one flying out Monday, back Thursday.
- Refundable, last-minute, and business-class fares cost far more and are bought mostly by those with inelastic demand.

The same seat, sold under different restrictions, fetches the high price from those who will pay it and the low price from those who would otherwise not fly.

12.8

Empirical Evidence: Price Dispersion and Online Pricing

How much do prices for the same good actually vary?

Fare dispersion points to discrimination

- Borenstein and Rose (1994) studied the airline industry.
- They asked how much fares differ between two passengers on the same route.
- Dispersion was greater on more competitive routes.

36%

expected fare gap between two passengers, as a share of the average ticket price

More

dispersion on more competitive routes

Fare dispersion points to discrimination

WHY IT'S DISCRIMINATION

Cost differences cannot explain greater dispersion on more competitive routes; the authors attribute the pattern to price discrimination.

The internet refines the sorting

THE ORBITZ EPISODE

In 2012 the Wall Street Journal reported that the travel site Orbitz had found that customers browsing on Apple computers spent as much as 30% more per night on hotels than Windows users, and had begun steering Mac users toward pricier options (Mattioli, 2012).

- Orbitz stressed it showed different choices, not different prices for the identical room.
- Still, the data buyers leave behind (device, location, history) let firms approximate the personalized pricing of Section 12.3.

The internet refines the sorting

- Online price discrimination stays limited by competition and by consumer backlash when it seems unfair: the fairness concern from Chapter 2.

Ride-hailing and the destination

- Chang et al. (2022) collected Uber fares for rides from the Los Angeles, New York, and San Francisco airports to hotels in those cities.
- For rides with the same distance, duration, and time of travel, fares were higher for travelers heading to more expensive hotels.
- The room rate signals willingness to pay: third-degree price discrimination based on the route. The authors estimate it raises travelers' welfare on average.

CHAPTER 12

Key takeaways

1. **When it's possible**

   Market power, the ability to separate buyers, and no resale.

2. **First degree**

   Each buyer pays her maximum; no consumer surplus, but output reaches the efficient level.

3. **Third degree**

   Group pricing by elasticity: the higher price goes to the less elastic group.

Key takeaways

4 **Second degree**

A menu (quantity discounts and versioning) lets buyers sort themselves.

5 **Bundling & tariffs**

Bundling pays when valuations are negatively correlated; two-part tariffs split a fixed fee from a per-unit price.

6 **Evidence**

Airline yield management, 36% fare dispersion (Borenstein–Rose), and online personalized pricing.

13

CHAPTER 13

Game Theory I: Simultaneous-Move Games

When the best choice for one player depends on what the others do: the language of games, and the tool behind most of the rest of this book.

CHAPTER 13

What this chapter covers

13.2–13.4 Foundations

- What a game is
- Dominant strategies & the prisoner's dilemma
- Nash equilibrium

13.5–13.6 More games

- Games with multiple equilibria
- Coordination & anticoordination
- Mixed strategies

13.7–13.9 Evidence

- Case: Blu-ray vs. HD DVD
- Empirical: penalty kicks
- Experimental tests

What this chapter covers (continued)

THE BIG IDEA

A strategic situation is one in which your best move depends on what others do. Game theory is the tool for studying them, and it underlies most of the remaining chapters of this book.

13.1 Introduction

- Two gas stations matching cuts
- Two disc formats, one survivor
- Penalty kicks that match a prediction

Three puzzles, one common thread

Gas stations

Two across the street match every price cut: drivers gain, the stations lose.

Format war

Studios split between two disc formats for almost two years; one vanished weeks after a studio switched.

Penalty kicks

Pros aim left or right in proportions that closely match a math prediction.

Three puzzles, one common thread (continued)

WHAT THEY SHARE

Each is a strategic situation: the best choice for one player depends on what the others do. Most interesting markets lie between competition and monopoly: a few firms whose fortunes depend on one another.

13.2

What Is a Game?

- Players, strategies, payoffs
- The payoff matrix
- Reading a cell

Every game has three ingredients

1

Players

The decision makers: the people or firms who choose.

2

Strategies

The actions each player can take.

3

Payoffs

What each player earns from every combination of actions.

Every game has three ingredients (continued)

THE PAYOFF MATRIX

When players choose simultaneously, without seeing the others' choices, we summarize the game in a payoff matrix: rows are one player's strategies, columns the other's, cells the payoffs to each.

Reading a cell

Convention

In each cell, the first number is the row player's payoff and the second is the column player's. In the matrices, the row player's strategies and payoffs are blue and the column player's are red.

THE BEHAVIORAL ASSUMPTION

We assume each player chooses the strategy that maximizes her own payoff, given what she expects the other to do.

13.3

Dominant Strategies and the Prisoner's Dilemma

- A best move no matter what
- The prisoner's dilemma
- Individually rational, collectively worse

A dominant strategy

Dominant strategy

A strategy that is best no matter what the other player does. When a player has one, we can predict she will use it, regardless of what her opponent does.

THE CLASSIC CASE

The best-known example is the prisoner's dilemma. Two suspects are questioned separately; each may stay silent or confess. Higher payoffs mean shorter sentences.

The prisoner's dilemma

		Suspect 2	
	Suspect 1	Silent	Confess
Suspect 1	Silent	3, 3	0, 5
	Confess	5, 0	1, 1

Matrix 13.1 The prisoner's dilemma.

- **Both silent:** 3 each.
- **Both confess:** 1 each.
- **Only one confesses:** she earns 5; the silent one earns 0.
- Higher payoffs mean shorter sentences.

Confessing is a dominant strategy

- **If Suspect 2 stays silent:** Suspect 1 earns 5 by confessing and 3 by staying silent.
- **If Suspect 2 confesses:** Suspect 1 earns 1 by confessing and 0 by staying silent.
- Either way confessing is better: it is a dominant strategy for Suspect 1.
- Going over the rows, the same holds for Suspect 2 (5 rather than 3, and 1 rather than 0).

Individually rational, collectively worse

THE TRAGEDY

Both confess and each earns 1, even though both staying silent would have given each 3. Individually rational choices lead to a collectively worse outcome.

- This single structure explains a wide range of behavior.
- Arms races, overfishing, and cartels, whose members are each tempted to produce more (Chapter 15).

Individually rational, collectively worse (continued)

SELF-CHECK

Each player can Confess or Stay silent, and Confess pays more no matter what. The dominant strategy is Confess for both, so both confess, even though both staying silent would leave each better off. The pursuit of individual advantage produces a jointly worse outcome.

13.4 Nash Equilibrium

- Most games lack a dominant strategy
- Mutual best responses
- Finding it: the best-response method

A game with no dominant strategy

Firm 1 \ Firm 2	Standard A	Standard B
Standard A	3, 3	0, 0
Standard B	0, 0	2, 2

Matrix 13.2 A coordination game.

- Two firms adopt Standard A or B; both gain only by matching.
- **Firm 2 picks A:** Firm 1 prefers A (3 rather than 0).
- **Firm 2 picks B:** Firm 1 prefers B (2 rather than 0).
- So neither firm has a dominant strategy.

A more general solution concept

Nash equilibrium

A combination of strategies in which each player is doing the best she can, given what the other is doing, so no player can gain by changing her strategy alone. A point of mutual best responses.

WHY IT MATTERS

It is a natural resting point: only there does no one regret her choice. Most games have no dominant strategy, so we need this more general idea.

Nash equilibrium in the prisoner's dilemma

(Confess, Confess) is Nash

Given that the other confesses, each does best to confess too. No one can gain by deviating alone.

(Silent, Silent) is not

It would make both better off, but each is tempted to deviate and confess, so it is not an equilibrium.

Nash equilibrium in the prisoner's dilemma (continued)

A KEY WARNING

A Nash equilibrium need not be the best outcome for the players. This gap between what is individually stable and what is collectively best recurs throughout economics.

Finding a Nash equilibrium in three steps

- **Step 1.** For each column, underline the row player's highest payoff.
- **Step 2.** For each row, underline the column player's highest payoff.
- **Step 3.** A cell with both payoffs underlined is a Nash equilibrium; we shade it.
- The method works even when no dominant strategy exists.

The method in the prisoner's dilemma

		Suspect 2	
	Suspect 1	Silent	Confess
Suspect 1	Silent	3, 3	0, <u>5</u>
	Confess	<u>5</u>, 0	<u>1</u>, <u>1</u>

Matrix 13.3 The prisoner's dilemma: best responses.

- **Step 1.** Suspect 1's best response is to confess in both columns.
- **Step 2.** Suspect 2's best response is to confess in both rows.
- **Step 3.** Only (Confess, Confess) has both payoffs underlined.

Explore Matrix 13.3: intro-micro-companion-website.netlify.app/game-solver

What the method reveals

- Some games have **more than one** equilibrium, such as the coordination game of Section 13.5.
- Others have **none** in pure strategies, such as the penalty kick of Section 13.6.
- Then no cell has both payoffs underlined, and we turn to mixed strategies.

13.5 Games with Multiple Equilibria

- Coordination: choices should align
- Anticoordination: choices should differ
- Two pure-strategy Nash equilibria in each

Coordination games: two equilibria

		Firm 2
Firm 1	Standard A	Standard B
Standard A	3, 3	0, 0
Standard B	0, 0	2, 2

Matrix 13.4 A coordination game: best responses.

- Players do better when their choices align.
- Each firm's best response is to match the other's standard.
- Two shaded cells: (A, A) and (B, B) are both Nash equilibria.

Coordination in real life

- **Driving:** everyone drives on the same side, the left (UK, Japan) or the right (United States).
- **Standards:** firms gain from a common plug shape or file format.
- **Meeting up:** two friends gain only if they go to the same place.
- Which equilibrium is reached may depend on history, expectations, or a focal point.

Two variants with their own names

- **Battle of the sexes:** the players prefer different equilibria (a concert or a game), yet must coordinate; Section 13.7 studies an example.
- **Stag hunt:** one equilibrium is safe but low; the other pays both more but requires each to trust the other.

Anticoordination: the game of chicken

- Players do better when their choices **differ**.
- Two drivers speed toward each other; each can swerve or go straight.
- **Both swerve:** 3 each. **Both go straight:** they crash, 0 each.
- **One goes straight:** she earns 4; the one who swerves (the “chicken”) earns 2.

Chicken has two equilibria

Driver 1 \ Driver 2	Swerve	Straight
Swerve	3, 3	2, 4
Straight	4, 2	0, 0

(a) The game

Driver 1 \ Driver 2	Swerve	Straight
Swerve	3, 3	2, 4
Straight	4, 2	0, 0

(b) Best responses

Matrix 13.5 The game of chicken.

- In each equilibrium, one driver swerves and the other goes straight.

Best responses in chicken

- **If Driver 2 swerves,** Driver 1 goes straight (4 rather than 3).
- **If Driver 2 goes straight,** Driver 1 swerves (2 rather than 0).
- Equilibria: (Swerve, Straight) and (Straight, Swerve).
- Each prefers the one in which she goes straight, so committing not to back down is valuable.

Chicken in real life

- **Labor strikes:** union and firm each prefer that the other gives in first; a long strike hurts both.
- **International crises:** each country prefers that the other backs down; the worst outcome is a war neither wants.
- **Market entry:** room for one firm; each prefers that the rival stays out, and both lose if both enter.

13.6 Mixed Strategies

- Games with no pure-strategy equilibrium
- Randomizing to avoid being outguessed
- Where mixed strategies appear

When there is no pure-strategy equilibrium

THE PENALTY KICK

Consider a penalty kick: the kicker aims left or right; the keeper dives left or right. The kicker wants to go where the keeper does not; the keeper wants to go where the kicker does.

- Regardless of what the players are doing, one of them always wants to switch.
- So there is no resting point in pure strategies: no cell where both play a best response.

A penalty kick

Kicker	Goalkeeper: Dive left	Goalkeeper: Dive right
Aim left	0, 1	1, 0
Aim right	1, 0	0, 1

(a) The game

Kicker	Goalkeeper: Dive left	Goalkeeper: Dive right
Aim left	0, <u>1</u>	<u>1</u>, 0
Aim right	<u>1</u>, 0	0, <u>1</u>

(b) Best responses

Matrix 13.6 A penalty kick.

- No cell has both payoffs underlined: there is no pure-strategy equilibrium.

No player wants to be predictable

- If the kicker always aims right, the goalkeeper dives right and stops the shot.
- If the goalkeeper always dives left, the kicker aims right and scores.
- So each player randomizes: a **mixed strategy**.

The kicker must be indifferent

- Let q be the probability that the goalkeeper dives left.
- Aim left: $EU_K(\text{Left}) = q \times 0 + (1 - q) \times 1 = 1 - q$
- Aim right: $EU_K(\text{Right}) = q \times 1 + (1 - q) \times 0 = q$
- Indifference: $1 - q = q$, so $q = ½$.

The goalkeeper must be indifferent

- Let p be the probability that the kicker aims left.
- Dive left: $EU_G(\text{Left}) = p \times 1 + (1 - p) \times 0 = p$
- Dive right: $EU_G(\text{Right}) = p \times 0 + (1 - p) \times 1 = 1 - p$
- Indifference: $p = 1 - p$, so $p = ½$.

The solution: randomize

Mixed strategy

Each player randomizes, choosing left or right with carefully chosen probabilities. In this symmetric example, each goes left half the time and right half the time, so neither can be outguessed.

Explore Matrix 13.6: intro-micro-companion-website.netlify.app/mixed-strategies

They describe real behavior

Tennis serve

A server who always aimed at the same corner would be returned every time, so good servers randomize.

Tax audits

An authority that audited on a fixed schedule would be gamed, so it audits at random.

They describe real behavior (continued)

Military timing

A force that always attacked at dawn would be ambushed, so it varies its timing.

Firm promotions

Unpredictable sales keep rivals and bargain-hunters guessing.

The common thread

THE PATTERN

Each is a situation where one side wants to be where the other is not (server vs. returner, auditor vs. cheater, attacker vs. defender), so any predictable pattern can be exploited.

Game theory predicts not a single action but a probability distribution over actions. As Section 13.8 shows with penalty kicks, skilled players come close to the predicted mix.

13.7

Case Study: The Format War Between Blu-ray and HD DVD

A coordination game, settled by one large player.

Two incompatible formats

- **2006:** Blu-ray (backed by Sony) and HD DVD (backed by Toshiba) reach consumers.
- A disc in one format cannot play on the other's player, so everyone gains from a single format.
- **End of 2007:** Sony, Disney, and Fox release only on Blu-ray; Universal and Paramount only on HD DVD; Warner Bros. on both.

The format war as a game

- **Both choose Blu-ray:** Studio 1 earns 3 and Studio 2 earns 2.
- **Both choose HD DVD:** the payoffs are reversed.
- **Different formats:** buyers hold off, and both earn 0.
- A coordination game of the battle-of-the-sexes type (Section 13.5.1).

Two equilibria, and each studio prefers a different one

Studio 1 \ Studio 2	Blu-ray	HD DVD
Blu-ray	3, 2	0, 0
HD DVD	0, 0	2, 3

(a) The game

Studio 1 \ Studio 2	Blu-ray	HD DVD
Blu-ray	**3**, **2**	0, 0
HD DVD	0, 0	**2**, **3**

(b) Best responses

Matrix 13.7 A format war.

- Equilibria: both on Blu-ray, or both on HD DVD; the game does not say which one is reached.

How the format war ended

- For almost two years, the industry reached neither equilibrium.
- **January 4, 2008:** Warner Bros. announces Blu-ray only (Mitchell, 2008).
- Blu-ray becomes the best response for most studios and retailers; on **February 19, 2008,** Toshiba drops HD DVD (Toshiba Corporation, 2008).
- One move by a large player took the industry to one of the two equilibria.

13.8

Empirical Evidence: Penalty Kicks in Soccer

A sharp, testable prediction, tested in high-stakes sport.

Do professionals really randomize?

THE PREDICTION

Mixed strategies make a sharp prediction: skilled players should randomize in specific proportions, and an outside observer should not be able to predict their next move.

- Palacios-Huerta (2003) analyzed 1,417 penalty kicks from professional soccer matches.
- Theory predicts how often kickers and goalkeepers choose each side, given that players differ in their stronger foot.

Do professionals really randomize? (continued)

1,417
penalty kicks analyzed

Close
match between the predicted and the observed left/right choices

The data match the theory

Right proportions

Professional kickers and keepers chose left and right in very nearly the predicted proportions.

Serially unpredictable

A player was no more likely to go left after going right than after going left, just as a true mixed strategy requires.

The data match the theory (continued)

THE TAKEAWAY

Players who spend their careers in these duels, with millions at stake, behave very much as game theory says they should, even if they have never heard of Nash.

Mixed strategies in other sports

- **Soccer:** the mixed-strategy equilibrium predicts kick and dive directions better than other benchmarks (Azar and Bar-Eli, 2011).
- **American football:** it explains the choice between running and passing (McGarrity and Linnen, 2010).
- **Tennis:** in nearly half a million serves, winning a point depends very little on the serve's direction (Gauriot et al., 2023).
- **A review:** Bar-Eli et al. (2020).

13.9

A Look at the Experimental Evidence

Do people play as the theory predicts?

Lab tests confirm the theory, with limits

- **Ultimatum bargaining:** theory predicts the smallest offer, accepted; yet offers near half are common, and low offers are often rejected (Güth and Kocher, 2014).
- **One-shot prisoner's dilemma:** theory predicts that both defect; yet a substantial minority cooperate (Mengel, 2018: 96 studies, more than 3,500 participants).
- Camerer (2003) surveys the evidence for many other games.

Split or Steal: cooperation at high stakes

- Golden Balls: in the final round of this British TV show (2007 to 2009), two contestants each secretly choose to split or steal a jackpot.
- The rules: if both split, they share it; if one steals, the stealer takes it all; if both steal, both leave with nothing. Stealing is a weakly dominant strategy.
- The evidence: studying 574 contestants in 287 episodes, van den Assem et al. (2012) found that players split 53% of the time, with jackpots averaging over $20,000; cooperation declined as stakes rose but leveled off at around 45% for medium and large jackpots.

A consistent picture, with limits

Theory predicts well

People recognize dominant strategies and settle on Nash equilibria in simple games.

But fairness intrudes

They lean toward fairness and cooperation more than pure self-interest implies: rejecting unfair offers, cooperating in dilemmas.

A consistent picture, with limits (continued)

SOCIAL PREFERENCES

These are the social preferences we first met in Chapter 5. Run it yourself: play the prisoner's dilemma live in Veconlab's "Matrix Game" (veconlab.com) and compare how often students cooperate with the prediction.

Key takeaways

- **Games:** players, strategies, and payoffs, summarized in a payoff matrix (row player's payoff first).
- **Dominant strategy:** best no matter what; in the prisoner's dilemma both confess, though both would gain by staying silent.
- **Nash equilibrium:** no player gains by deviating alone; find it by underlining best responses.
- **Coordination and anticoordination:** choices should align or differ; each game has two pure-strategy equilibria.

Key takeaways (continued)

- **Mixed strategies:** each player's mix makes the other indifferent; in the penalty kick, each goes left with probability ½.
- **Penalty kicks:** professionals choose sides in very nearly the predicted proportions (Palacios-Huerta, 2003).
- **Format war:** studios stayed split until Warner Bros. chose Blu-ray in January 2008; Toshiba then dropped HD DVD.
- **Experiments:** people find dominant strategies and Nash equilibria, but lean toward fairness and cooperation.

14

CHAPTER 14

Game Theory II: Sequential and Repeated Games

Timing changes everything: when players move in sequence, or meet one another again and again.

CHAPTER 14

What this chapter covers

14.2–14.4 Sequence

- Game trees
- Backward induction
- Credible threats & commitment

14.5 Repetition

- When cooperation survives
- Grim trigger & tit-for-tat
- Tacit collusion

14.6–14.7 Evidence

- Airline entry battles
- OPEC as a repeated game
- Repeated games in the lab

What this chapter covers (continued)

THE BIG IDEA

These questions all turn on timing: players move in sequence, or meet one another repeatedly. This chapter extends the game theory of Chapter 13 to both settings.

14.1 Introduction

- An airline slashes fares, then raises them again
- Rivals who meet week after week
- Cartel members who meet again and again

Three puzzles about timing

Empty threats?

An incumbent slashes fares when a rival enters, then raises them once it retreats. Is the threat always believable?

Why repetition helps

Two rival supermarkets facing each other week after week sustain high prices that two one-time sellers could not.

Cartels that meet again

OPEC members hold output below what each would like for long periods, then fall into price wars.

Three puzzles about timing (continued)

WHAT THEY SHARE

Players move in sequence, or meet one another repeatedly, and that is enough to change what is possible.

14.2

Sequential Games and Game Trees

- When a payoff matrix is not enough
- Reading a game tree
- The entry game

When players move in sequence, we use a tree

Game tree

When the later mover sees what the earlier one did, a payoff matrix no longer captures the situation. The tree branches at each decision point, showing who moves, what they can choose, and the payoffs at the end of each path.

The entry game. A potential entrant first decides whether to enter the market.

Stay out. The incumbent keeps the market to itself, earning $10 million.

Enter. The incumbent must accommodate (each firm earns $5 million) or fight a price war (each firm loses $2 million).

The entry game as a game tree

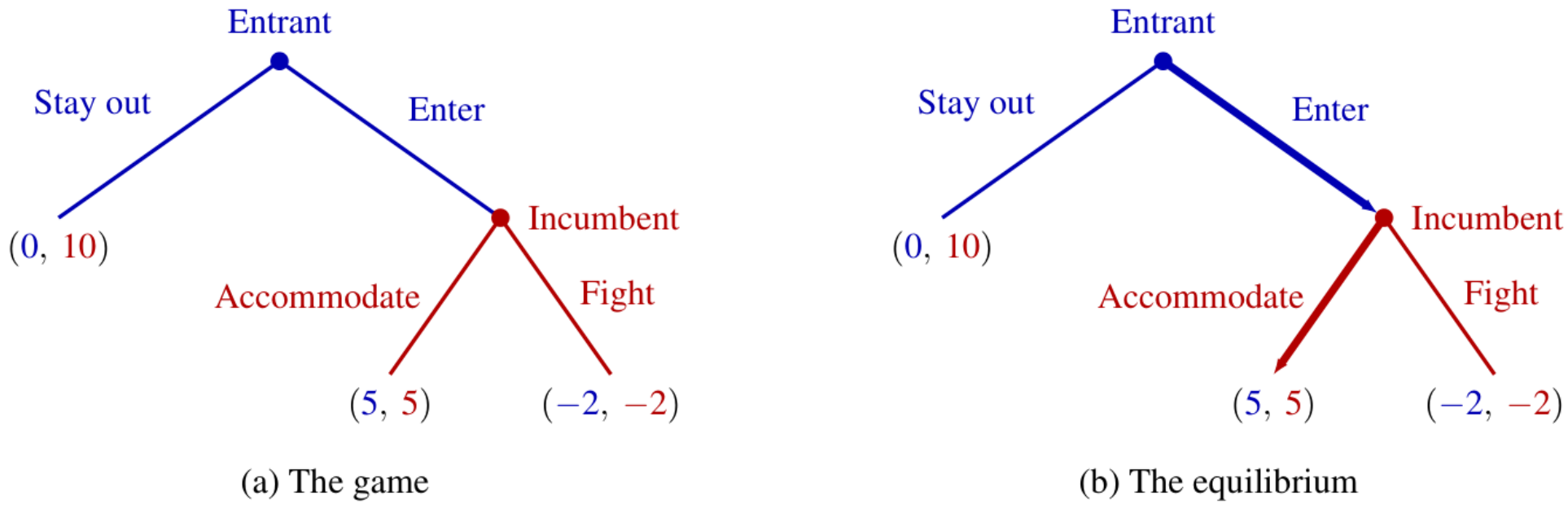


Figure 14.1 The entry game, with payoffs (entrant, incumbent).

Explore Figure 14.1: intro-micro-companion-website.netlify.app/backward-induction

The entry game as a game tree (continued)

- The entrant moves first; if it enters, the incumbent accommodates or fights.
- As in the payoff matrices of Chapter 13, the first mover's choices and payoffs are blue and the second mover's are red.

14.3

Backward Induction and Subgame-Perfect Equilibrium

- Reasoning from the last move backward
- Why sequence changes everything

Solve from the end of the tree back to the start

- Look first at the incumbent's choice, which comes second.
- If entry has occurred, it compares accommodating (5) with fighting (−2). It accommodates (the thick arrow in the right panel of Figure 14.1).
- Knowing this, the entrant looks ahead: entering yields 5, which exceeds the 0 from staying out. So it enters.

Solve from the end of the tree back to the start (continued)

- **Subgame-perfect equilibrium**

  The strategy combination produced by backward induction, in which each player acts optimally at every decision point (every node of the tree). Also called rollback.

The entry game as a simultaneous game

- Now both firms choose at the same moment: the incumbent decides without knowing whether the entrant entered.
- **Entrant stays out:** (0, 10), whatever the incumbent chose.
- **Entrant enters:** (5, 5) if the incumbent accommodates, (−2, −2) if it fights.

Two Nash equilibria with simultaneous moves

Entrant	Incumbent: Accommodate	Incumbent: Fight
Stay out	0, 10	0, 10
Enter	5, 5	−2, −2

(a) The game

Entrant	Incumbent: Accommodate	Incumbent: Fight
Stay out	0, <u>10</u>	<u>0</u>, <u>10</u>
Enter	<u>5</u>, <u>5</u>	−2, −2

(b) Best responses

Matrix 14.1 The entry game as a simultaneous game.

- Both payoffs are underlined in (Enter, Accommodate) and in (Stay out, Fight).

Best responses in the simultaneous game

- **Incumbent accommodates:** the entrant enters (5 rather than 0).
- **Incumbent fights:** the entrant stays out (0 rather than −2).
- **Entrant stays out:** the incumbent earns 10 either way, so both strategies are best responses.
- **Entrant enters:** the incumbent accommodates (5 rather than −2).

Sequence removes the empty threat

- In (Stay out, Fight), the threat to fight keeps the entrant out, and it costs the incumbent nothing.
- With sequential moves, only (Enter, Accommodate) survives: once entry occurs, fighting only hurts the incumbent.
- Backward induction asks what a player will actually do when her turn arrives, not what she threatens beforehand.

Self-check: is the threat credible?

SELF-CHECK

In a sequential entry game, the incumbent threatens a price war if a rival enters, but fighting would lose it money. By backward induction, is the threat credible, and does the rival enter?

ANSWER

The threat is not credible: once entry has occurred, fighting only hurts the incumbent, so it accommodates instead. Anticipating this, the rival enters. Only a binding commitment could make the threat believable.

14.4

Credible Threats and Commitment

- Why an empty bluff fails
- Commitment that ties your own hands
- Commitment devices in the real world
- Regulatory hold-up

A threat backward induction exposes

Noncredible threat

As Section 14.3.1 showed, the incumbent's threat to fight is not credible: once entry has occurred, it prefers to accommodate.

THE REMEDY: COMMITMENT

What could make the threat credible? A prior action that changes the incumbent's own later incentives, so that fighting actually becomes its preferred response to entry.

Commitment changes the incumbent's incentives

- Before any entry, the incumbent builds extra capacity, signs contracts, or invests in a reputation for toughness.
- In Figure 14.2, this commitment costs the incumbent 2 and raises its payoff from fighting by 8.
- Fighting then becomes its preferred response to entry.

The entry game with commitment

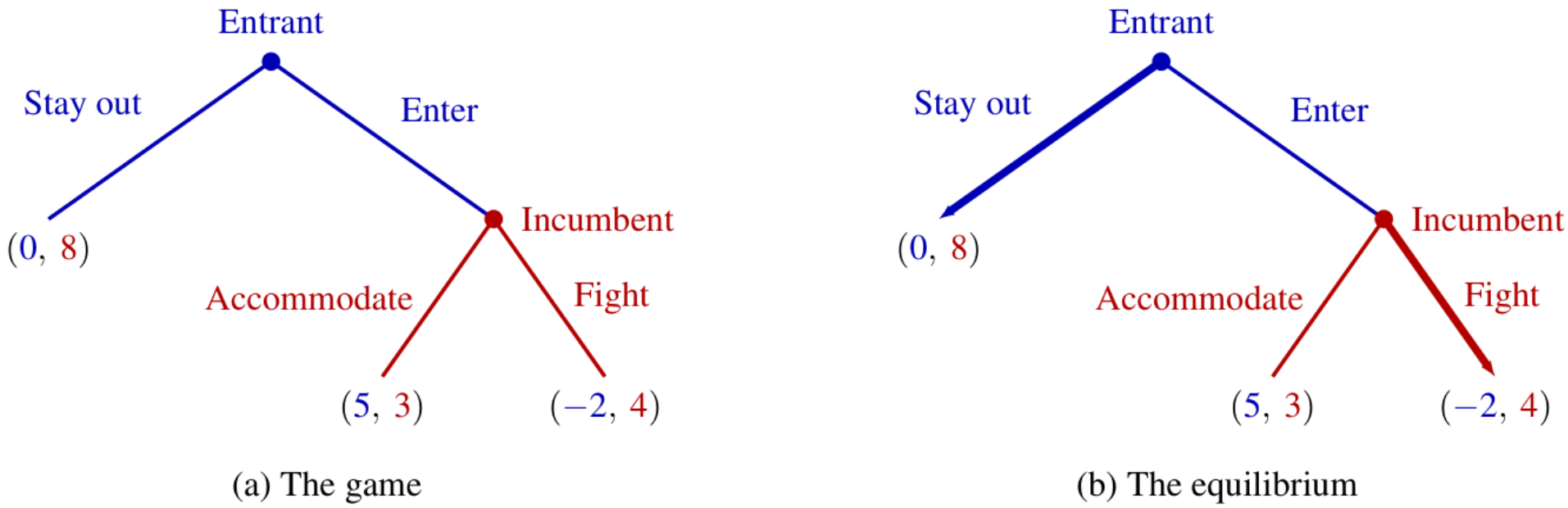


Figure 14.2 The entry game with a commitment, with payoffs (entrant, incumbent).

- The threat to fight is now credible, and the entrant stays out.

The numbers behind Figure 14.2

- The commitment: costs the incumbent 2 and raises its payoff from fighting by 8.
- After entry: fighting now earns 4 (−2 plus 8, minus the cost of 2) and accommodating only 3 (5 minus 2), so the threat is credible.
- The entrant: foresees a fight that would give it −2, so it stays out.
- The incumbent: earns 8 (10 minus the cost of 2), more than the 5 it earned without the commitment.

How people and firms tie their own hands

Commitment device

Any prior, hard-to-reverse action that changes one's own future incentives, so that a threat or promise becomes believable.

How people and firms tie their own hands (continued)

Firms

Build a large factory or sign long-term supply contracts, committing to high volume and warning entrants that they will fight.

Savers & quitters

Lock money in a plan you cannot touch; bet a friend $1,000 you will quit smoking, raising the cost of backsliding.

Announcing a goal

Even a public announcement is a mild device: breaking it now carries the extra cost of embarrassment.

Throwing away an option

- A general who burns the bridge behind his army cannot retreat, which makes his vow to fight credible and may deter the attack.
- By visibly throwing away an option, you change what others (or your later self) expect you to do, and that works in your favor.

Sunk investment invites hold-up

- Many valuable projects (a power plant, a water system, a fiber-optic network) need a huge, irreversible investment first.
- Once the money is spent, the government is tempted to cap the firm's prices low.
- The firm chooses whether to invest; if it does, the government honors the deal or reneges.

The hold-up problem, as a sequential game

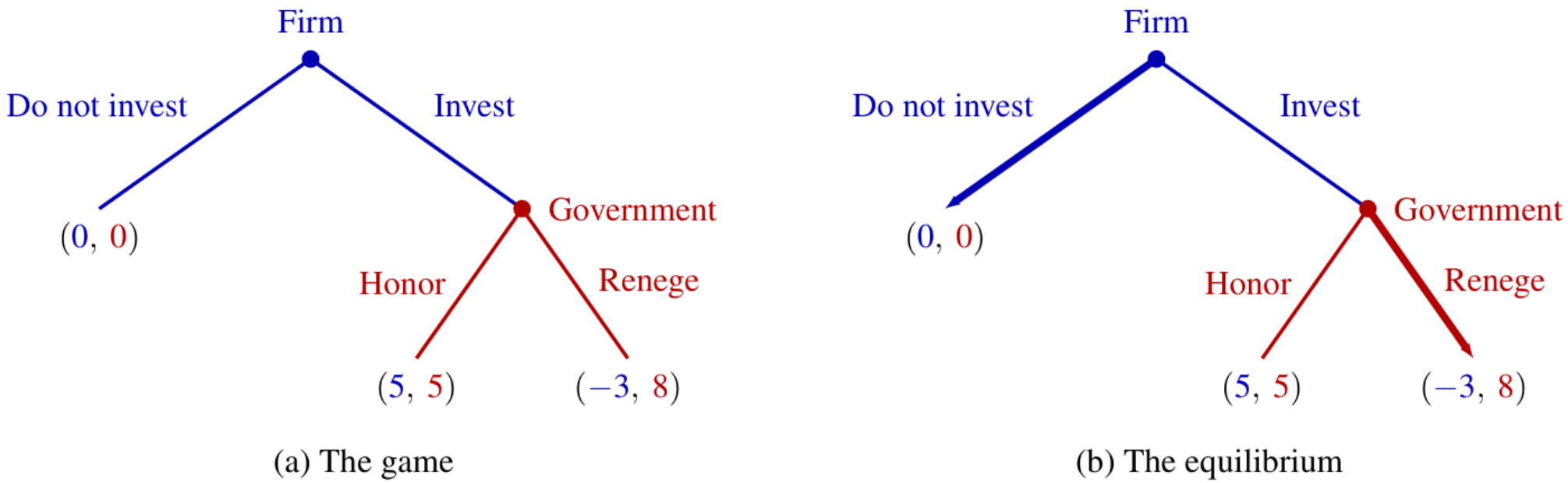


Figure 14.3 The regulatory hold-up game, with payoffs (firm, government).

- The government would renege (8 against 5), so the firm does not invest.

The hold-up problem, and its cure

Hold-up problem

Anticipating that the government will renege after the investment is sunk, the firm refuses to invest, and the project (worth 5 to each side) never happens.

INDEPENDENT REGULATORS

The remedy is the commitment of Section 14.4: the government must credibly tie its own hands so that reneging becomes costlier, for instance through independent regulators. Britain created Ofgas (gas, 1986, now part of Ofgem) and Ofwat (water, 1989), setting prices by formulas fixed years in advance.

Where commitment is weak, hold-up is sharper

- Guasch (2004) studied more than 1,000 Latin American infrastructure concessions, mid-1980s to 2000.
- About 30% were renegotiated after the fact, far more in some sectors.
- Sometimes governments squeezed firms once capital was sunk; sometimes firms bid low, then demanded better terms.

55%
of transport concessions renegotiated

74%
of water concessions renegotiated

Where commitment is weak, hold-up is sharper (continued)

THE STAKES

The average concession was renegotiated just 2.2 years after signing. Weak commitment raised the cost of infrastructure and deterred the very investment these countries most needed.

14.5 Repeated Games and Cooperation

- Why finite repetition unravels
- Indefinite horizons sustain cooperation
- Grim trigger and tit-for-tat
- Tacit collusion and price wars

The horizon matters

Repeated games

Many situations recur: the same firms compete month after month, the same neighbors interact for years. Players can reward cooperation and punish defection over time.

The horizon matters (continued)

The Folly of Finite Repetition

If a prisoner's dilemma is repeated a known, finite number of times, cooperation unravels. In the last round there is no future to protect, so both defect; knowing that, they defect in the second-to-last round too, and so on back to the first.

An indefinite horizon: an example

- Players never know for sure when the game will end.
- Say that they meet again next round with probability 0.7.
- **Two more rounds:** $0.7 \times 0.7 = 0.49$.
- **30 more rounds:** 0.7^{30}, about 23 in a million.

Indefinite repetition can sustain cooperation

INDEFINITE HORIZON

If the game is repeated indefinitely (players never know for sure when it will end), cooperation can be sustained. The key is a strategy that makes cheating unprofitable by threatening to withdraw future cooperation.

THE CONDITION

If players value the future enough, the loss of future cooperation outweighs the one-time gain from cheating, and cooperation holds.

Two punishment strategies

Grim trigger

Cooperate until the other defects even once, then defect forever after. A harsh, unforgiving punishment.

Tit-for-tat

Cooperate first, then simply copy the other player's previous move. Milder: it punishes, but forgives as soon as the other cooperates again.

Two punishment strategies (continued)

WHY IT MATTERS

This is why ongoing business rivals can sustain high prices (tacit collusion), and why price-matching guarantees (a public promise to match any rival's lower price) can keep prices high without any explicit agreement.

Keeping prices high without an agreement

Tacit collusion

Rival firms keep prices high without ever signing an illegal agreement. Two supermarkets facing each other week after week are playing a repeated pricing game: each could cut its price today, but the rival would match tomorrow, touching off a price war that leaves both worse off.

Keeping prices high without an agreement (continued)

The Shadow of the Future

So each refrains, held high not by a contract but by the shadow of future punishment: the grim-trigger and tit-for-tat logic. A few firms can behave almost like a monopoly even while competing fiercely on the surface.

Price-matching guarantees and price wars

Price-matching, paradox

A public promise to match any rival's lower price can keep prices up: it predicts that any cut will be instantly undone, removing the reward for cutting.

When it breaks down

A new entrant ignores the unwritten rules, or a firm in trouble breaks ranks, and the result is the price war the firms had been avoiding.

Price-matching guarantees and price wars (continued)

WHY REGULATORS CARE

Competition authorities watch for tacit collusion precisely because its effects can resemble those of an illegal cartel, even when no law has been broken.

14.6

Case Study: Entry Battles in the Airline Industry

The entry game of Figure 14.1, played for real.

14.6 CASE STUDY

American Airlines and the Dallas/Fort Worth hub

- In the 1990s, small low-cost carriers began serving routes out of American's Dallas/Fort Worth hub.
- American sharply cut fares and added flights on those routes: Dallas to Kansas City, Wichita, Colorado Springs, and Long Beach.
- When the low-cost carriers withdrew, American raised fares again.

14.6 CASE STUDY

US v. AMR: predation or competition?

THE ALLEGATION

In 1999 the Department of Justice sued, alleging American had priced below cost to drive out entrants and "fight" in the sense of our game tree, planning to recoup losses later through higher fares once the threat of entry was gone.

- To be illegal predation, pricing had to be below American's own cost, with a realistic chance of recouping the losses.
- The courts ruled for American, finding the evidence of below-cost pricing insufficient.

14.7

Empirical Evidence: Cooperation in Repeated Games

OPEC, and repeated games in the laboratory.

OPEC plays a repeated game

- OPEC produces close to a third of the world's crude oil (31% in 2025, not counting the UAE, which left the group in May 2026).
- Each period is a prisoner's dilemma: if all limit output, the price stays high, but each earns more by producing above its quota.
- Members meet month after month, so cooperation can survive only if cheating is punished.

The early 1980s: Saudi Arabia absorbs the cuts

- After the price increases of the 1970s, demand for oil fell.
- Many members kept producing close to their quotas or above them; Saudi Arabia, the largest producer, absorbed most of the cuts.
- **Saudi output:** about 9.8 million barrels a day in early 1980, about 2.5 million in the third quarter of 1985.

Late 1985: the swing producer gives up

- Saudi Arabia stopped playing the role of “swing producer” and began selling all the oil it could.
- From August 1985 to mid-1986, OPEC output rose by about 4 million barrels a day, more than half of it from Saudi Arabia.
- Crude oil fell to about $12 a barrel in the first half of 1986 (Gately, 1986).

Flooding the market as punishment

- **Griffin and Neilson (1994):** Saudi Arabia gave up its role once its profits fell below what it would earn if each member chose its own output.
- It then appears to have followed tit-for-tat, punishing excessive cheating by the other members.
- Later history fits the same logic: members exceed their quotas, and quota disputes have led members to leave.

The shadow of the future in the lab

- **Dal Bó (2005):** participants played the prisoner's dilemma for money, in matches that ended at random.
- A higher chance of another round significantly reduced opportunistic play, as Section 14.5 predicts.
- **Dal Bó and Fréchette (2011):** with experience, cooperation can take hold, but only when the gain from cooperating is large relative to the risk of being exploited.

When the last round is known

- **Embrey et al. (2018):** participants cooperate until a certain round and defect afterward.
- With experience, the first defection moves earlier, as backward induction predicts, although slowly.
- **Dal Bó and Fréchette (2018):** cooperation is more likely when it can be supported in equilibrium, that is, when players value future rounds enough.

Run it yourself

- Run Veconlab's repeated prisoner's-dilemma session (veconlab.com).
- See whether your group sustains cooperation with strategies such as tit-for-tat.

Key takeaways

- **Sequential games:** a game tree, solved by backward induction; the result is a subgame-perfect equilibrium.
- **Entry game:** the threat to fight is not credible, so the entrant enters and the incumbent accommodates.
- **Commitment:** a prior, hard-to-reverse action can make a threat believable; limiting your options can help you.
- **Hold-up:** a government free to renege leads firms not to invest; independent regulators and long-term contracts help.

Key takeaways (continued)

- **Repeated games:** finite repetition unravels; indefinitely, grim trigger or tit-for-tat can sustain cooperation.
- **Airlines:** American cut fares when low-cost carriers entered and raised them after they left; the courts ruled for American.
- **OPEC:** when Saudi Arabia stopped absorbing cheating in late 1985, oil fell to about $12 a barrel in 1986.

15

CHAPTER 15

Oligopoly and Imperfect Competition

A few large firms, each watching the others, where the tools of market structure meet game theory.

CHAPTER 15

What this chapter covers

15.2–15.5 The models

- Measuring concentration (HHI)
- Cournot: quantities
- Bertrand: prices
- Stackelberg: moving first

15.6–15.8 Behavior

- Collusion & why it fails
- Algorithmic pricing
- Monopolistic competition
- Merger control

15.9–15.11 Evidence

- Case: global beer
- Empirical: lysine cartel
- Experiment: oligopoly in the lab

CHAPTER 15

What this chapter covers (continued)

THE BIG IDEA

An oligopoly is dominated by a few large firms whose every decision depends on what their rivals will do: the natural home of game theory.

15.1 Introduction

- Markets ruled by a few giants
- Neither monopoly nor perfect competition
- Market structure meets game theory

A few firms dominate, and each watches the others

Mobile service

Verizon, AT&T, and T-Mobile sell almost all U.S. mobile-phone service.

Beer

AB InBev brews about one of every four beers drunk worldwide.

Passenger jets

Airbus and Boeing delivered about 90% of the world's new commercial jets in 2024.

15.1 INTRODUCTION

A few firms dominate, and each watches the others

WHAT THEY SHARE

These markets are oligopolies: dominated by a few large firms whose every decision depends on what rivals will do. Because each firm's best move depends on its rivals', oligopoly is the natural home of game theory.

15.2

Measuring Market Power: Concentration

- The concentration ratio
- The Herfindahl–Hirschman Index (HHI)
- A worked calculation

Two ways to gauge concentration

- **Concentration ratio**

  Adds up the market shares of the largest firms; the four-firm ratio (the combined share of the top four) is the most common.

- **Herfindahl–Hirschman Index (HHI)**

  Sums the squares of all firms' market shares (in percentage points). Squaring gives extra weight to large firms, so the HHI rises as a market becomes more lopsided.

What the HHI looks like at the extremes

1,000

10 equal firms

10×10^2

5,000

two equal firms

2×50^2

10,000

pure monopoly

100^2

THE MERGER THRESHOLD

Competition authorities use the HHI to judge mergers. Under the 2023 U.S. Merger Guidelines, a market above 1,800 is highly concentrated, and a merger that raises its HHI by more than 100 points is presumed to harm competition.

15.2.1 A WORKED HHI CALCULATION

Same firms, different shares

Market	Shares (%)	HHI	Calculation
A (even)	25, 25, 25, 25	2,500	4 × 625
B (lopsided)	70, 10, 10, 10	5,200	**4,900 + 300**

Both markets have four firms, yet the HHI judges B far more concentrated.

Same firms, different shares (continued)

WHY SQUARING MATTERS

Both markets have the same four-firm concentration ratio of 100%. But squaring the shares gives heavy weight to B's dominant firm; that is the HHI's advantage over a simple concentration ratio.

Self-check: compute the HHI

SELF-CHECK

A market has four firms with shares 40, 30, 20, and 10 percent. What is the HHI, and would U.S. regulators call this market concentrated?

The HHI is $40^2 + 30^2 + 20^2 + 10^2 = 1{,}600 + 900 + 400 + 100 = 3{,}000$. Because it exceeds 1,800, the U.S. merger guidelines classify the market as highly concentrated.

15.3

Cournot Competition: Choosing Quantities

- Firms choose how much to produce
- Best response functions and Nash equilibrium
- More firms, lower price

Competing by choosing quantities

Cournot model

A few firms making an identical product each choose how much to produce, simultaneously, and the combined output determines the market price.

Best response function

Each firm picks its best output given what it expects its rivals to produce. The list of these best outputs, one for each output the rival might choose, is the firm's best response function.

A numerical example: the monopoly benchmark

- Price equals 12 minus the total units sold: $P = 12 - Q$ (so 8 units sell for \$4).
- Production costs are negligible (zero, for simplicity).
- A monopolist sets $MR = 12 - 2Q$ equal to its zero marginal cost: $12 - 2Q = 0$ gives $Q = 6$.
- **Price:** \$12 − \$6 = \$6. **Profit:** 6 × \$6 = \$36.

Firm 1's best response

- Two firms produce q_1 and q_2 units, so $Q = q_1 + q_2$.
- Given q_2, firm 1 faces the leftover demand $P = (12 - q_2) - q_1$.
- As a monopolist on it, firm 1 sets $(12 - q_2) - 2q_1 = 0$.
- **Best response function:** $q_1 = 6 - \frac{q_2}{2}$.

Firm 1's best response, output by output

Firm 2's output, q_2	Firm 1's leftover demand	Firm 1's best response, $q_1=6-\frac{q_2}{2}$
0	$P=(12-0)-q_1=12-q_1$	$6-\frac{0}{2}=6$
2	$P=(12-2)-q_1=10-q_1$	$6-\frac{2}{2}=5$
4	$P=(12-4)-q_1=8-q_1$	$6-\frac{4}{2}=4$
6	$P=(12-6)-q_1=6-q_1$	$6-\frac{6}{2}=3$
8	$P=(12-8)-q_1=4-q_1$	$6-\frac{8}{2}=2$
10	$P=(12-10)-q_1=2-q_1$	$6-\frac{10}{2}=1$

Table 15.1 Firm 1's best response to firm 2's output.

Best response functions cross at the Nash equilibrium

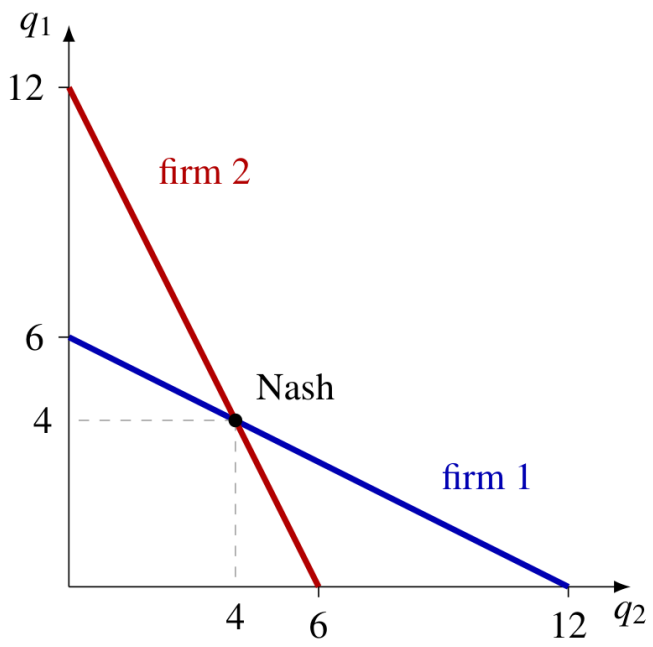


Figure 15.1 Cournot best response functions and the Nash equilibrium.

- Firm 1's vertical intercept, 6, is its monopoly output ($q_2 = 0$).
- Each extra unit of q_2 cuts q_1 by half a unit: **strategic substitutes**.
- Firm 2: $q_2 = 6 - \frac{q_1}{2}$, starting at 6 on the horizontal axis.

Explore Figure 15.1: intro-micro-companion-website.netlify.app/cournot

The Cournot equilibrium

- Identical firms produce the same output: $q = 6 - \frac{q}{2}$, so $q = 4$ (as Table 15.1 confirms).
- Total output $2 \times 4 = 8$; price $\$12 - \$8 = \$4$; each firm earns $4 \times \$4 = \$16$.
- Versus monopoly: more output (8 rather than 6 units) and a lower price ($4 rather than $6).
- Combined profit, $2 \times \$16 = \$32$, falls short of the monopoly profit of $36.

More firms

- With three firms, each faces the demand left by its two rivals, who produce $2q$.
- Best response: $q = 6 - \frac{2q}{2} = 6 - q$, so $q = 3$.
- Total output $3 \times 3 = 9$; price $\$12 - \$9 = \$3$.
- With n firms, each produces $\frac{12}{n+1}$ units.

Cournot competition with more firms

Number of firms	Output per firm	Total output	Price	Profit per firm
1	$\frac{12}{2}=6$	$1\times6=6$	$\$12-\$6=\$6$	$6\times\$6=\$36$
2	$\frac{12}{3}=4$	$2\times4=8$	$\$12-\$8=\$4$	$4\times\$4=\$16$
3	$\frac{12}{4}=3$	$3\times3=9$	$\$12-\$9=\$3$	$3\times\$3=\$9$
5	$\frac{12}{6}=2$	$5\times2=10$	$\$12-\$10=\$2$	$2\times\$2=\$4$
11	$\frac{12}{12}=1$	$11\times1=11$	$\$12-\$11=\$1$	$1\times\$1=\$1$

Table 15.2 Cournot competition with more firms.

Toward perfect competition

- As the number of firms grows, total output approaches 12 units.
- The price approaches $0, the marginal cost: the outcome approaches perfect competition.

15.4

Bertrand Competition: Choosing Prices

- Firms set prices, not quantities
- The Bertrand paradox
- Why two firms can be enough

Competing in prices: a sharp result

- Firms set prices at the same time; with identical products, consumers buy from the firm charging the lowest price (equal prices split the market).
- Each firm has an incentive to undercut the other by a penny to capture the whole market.
- Undercutting continues until the price is driven all the way down to marginal cost (the competitive outcome) with only two firms.

THE BERTRAND PARADOX

The Bertrand paradox: two price-setting firms can be enough to eliminate market power entirely.

Why two firms rarely compete the price to cost

Differentiation

Products differ (a Coke is not a Pepsi), so the least expensive does not win every buyer.

Capacity limits

The model lets the lowest-price firm serve all demand: plausible for streaming, hard for cars, appliances, or furniture.

Repetition

Firms meet repeatedly (Chapter 14), giving them ways to avoid ruinous price wars.

Why two firms rarely compete the price to cost (continued)

COURNOT OR BERTRAND?

Whether oligopoly looks more like Cournot (prices above marginal cost) or Bertrand (prices driven down to marginal cost) depends on these features.

15.5

Stackelberg: Moving First

- A leader chooses quantity first
- The follower responds
- The first-mover advantage

Moving first and committing to a large output

Stackelberg model

One firm, the leader, chooses its quantity first; the follower observes it and plays its best response, the function of Table 15.1. By backward induction, the leader anticipates this response.

6
leader's output

3
follower's output

Explore Section 15.5: intro-micro-companion-website.netlify.app/stackelberg

The leader's choice

Leader's output, q_L	Follower's output, $6-\frac{q_L}{2}$	Total output	Price	Leader's profit	Follower's profit
0	$6-\frac{0}{2}=6$	0+6=6	\$12−\$6=\$6	0×\$6=\$0	6×\$6=\$36
2	$6-\frac{2}{2}=5$	2+5=7	\$12−\$7=\$5	2×\$5=\$10	5×\$5=\$25
4	$6-\frac{4}{2}=4$	4+4=8	\$12−\$8=\$4	4×\$4=\$16	4×\$4=\$16
6	$6-\frac{6}{2}=3$	6+3=9	\$12−\$9=\$3	6×\$3=\$18	3×\$3=\$9
8	$6-\frac{8}{2}=2$	8+2=10	\$12−\$10=\$2	8×\$2=\$16	2×\$2=\$4
10	$6-\frac{10}{2}=1$	10+1=11	\$12−\$11=\$1	10×\$1=\$10	1×\$1=\$1

Table 15.3 The Stackelberg leader's choice.

The leader's best choice

- The leader's profit is highest, $18, when it produces 6 units.
- The follower responds with 3 units; total output is 9 and the price is $3.
- **Versus Cournot** (the row with $q_L = 4$): the leader produces more (6 rather than 4 units) and earns more ($18 rather than $16).
- The follower produces less (3 rather than 4 units) and earns less ($9 rather than $16).

The first-mover advantage

FIRST-MOVER ADVANTAGE

By committing to a large output, the leader earns more than it would in simultaneous Cournot competition.

- This is a direct application of commitment: by moving first and visibly, the leader forces the follower to accommodate.
- It works much like the entry-deterring incumbent of Chapter 14, who committed early to shape a rival's response.

15.6

Collusion: Acting Like a Monopoly

- Agreeing to restrict output
- A prisoner's dilemma
- Why collusion is hard to sustain

15.6 COLLUSION

The temptation to act like a monopoly

THE TEMPTATION

Oligopolists earn less than a monopolist because they compete. So they are tempted to collude: agree to restrict output and split the monopoly profit.

THE NUMBERS

If the two Cournot firms jointly produced the monopoly output of 6 (three each) instead of 8, the price would rise from \$4 to \$6, and their combined profit from $2 \times \$16 = \$32$ to $6 \times \$6 = \$36$, or \$18 each.

But collusion is a prisoner's dilemma

- At the collusive output, each firm is tempted to cheat by producing extra at the high price.
- If both cheat, the agreement collapses back to competition.
- Repeated interaction can sustain collusion: if one firm produced more, the others would expand their output in later periods, pushing the price down for all.

WHY IT IS POLICED

This is why competition authorities watch oligopolies so closely, and why explicit price-fixing agreements are illegal.

Price fixing reaches the grocery aisle

- Canned tuna: StarKist was sentenced in 2019 to pay a $100 million criminal fine, the maximum allowed by law, for conspiring with its rivals to fix tuna prices.
- Prison: the former chief executive of rival Bumble Bee was sentenced in 2020 to 40 months in prison.
- Chicken: in 2021, Pilgrim's Pride was fined $107.9 million for fixing the prices of broiler chickens (U.S. Department of Justice).

Beyond the United States

- **Chile (2014):** the three largest chicken producers (Agrosuper, Ariztía, Don Pollo) limited production and assigned quotas; heavy fines, trade association dissolved, upheld in 2015.
- **Germany (2014):** the three largest sugar producers were fined about 280 million euros for agreeing on territories, quotas, and prices.
- **Canada (2023):** Canada Bread was fined 50 million Canadian dollars for fixing wholesale bread prices with Weston Foods.

Three forces that break a cartel

1

More firms

The more firms there are, the harder it is to detect and punish cheating, and the larger each firm's temptation to defect.

2

Demand shocks

They make it hard to tell a cheater from an unlucky rival whose sales simply fell.

3

No legal enforcement

Explicit price-fixing is illegal, so members cannot enforce the cartel in court.

Three forces that break a cartel (continued)

IN PRACTICE

These pressures explain why cartels like OPEC constantly struggle with quota-busting, and why the lysine conspiracy eventually came apart.

Pricing algorithms can learn to collude

- Many oligopolists (online retailers, hotels, gasoline stations) let algorithms read rival prices and set their own.
- **Calvano et al. (2020):** learning algorithms, with no instruction to collude and no way to communicate, learned to price above the competitive benchmark.
- They punished deviations with temporary price wars, the pattern of Section 15.6.1.

Algorithmic pricing in real markets

- **Assad et al. (2024):** in Germany's retail gasoline market, adopting the software raised a station's margin only when its neighbor also adopted it.
- For competition authorities: there may be no explicit agreement to prosecute, yet the market outcome looks like a cartel's.

15.7

Monopolistic Competition: Many Firms, Differentiated Products

- Many firms, differentiated products
- Some market power each
- Welfare: variety vs. efficiency

Many firms selling differentiated products

- **Monopolistic competition**

  Many firms sell differentiated products: restaurants, coffee shops, hair salons, clothing brands. Each has some market power because its product is different, so it faces a downward-sloping demand curve and prices above marginal cost, like a small monopolist.

Many firms selling differentiated products (continued)

Free Entry Drives Profit to Zero

But entry is free: when firms earn profits, new ones enter with their own variety until economic profit is competed away to zero, as under perfect competition.

A trade-off: variety vs. efficiency

The inefficiency

Price stays above marginal cost, and each firm produces below its lowest-cost scale (excess capacity), like a rarely-full restaurant or a salon with idle chairs.

The benefit

Variety. Consumers choose among dozens of restaurants, brands, and coffee shops, each slightly different, and that diversity is valuable.

A trade-off: variety vs. efficiency (continued)

THE VERDICT

A modest inefficiency in exchange for abundant choice: worthwhile if consumers value variety enough.

15.8

Regulating Oligopoly: Merger Control

How governments review mergers, and two real cases.

Governments regulate by policing mergers

- Fewer firms tend to charge higher prices, so governments block or condition mergers that would make a market too concentrated.
- The main screen uses two numbers: the HHI after the merger, and the increase in the HHI that the merger causes.

Merger screens

HHI after the merger	Increase in the HHI: 100 points or less	Increase in the HHI: more than 100 points
Highly concentrated (above 1,800)	Rarely raises concern	Presumed to harm competition
Moderately concentrated (1,000 to 1,800)	Rarely raises concern	May be examined; no presumption
Unconcentrated (below 1,000)	Rarely raises concern	Rarely raises concern

Figure 15.2 Merger screens under the U.S. Merger Guidelines.

- Red cell: highly concentrated, and an increase of more than 100 points.

Reading the merger screens

- **Rows:** HHI after the merger: below 1,000, 1,000 to 1,800, or above 1,800.
- **Columns:** increase in the HHI: 100 points or less, or more than 100 points.
- **Red cell:** presumed to harm competition; the merging firms must show otherwise.
- **Yellow cell:** may be examined, but no presumption; **green cells:** rarely raise concern.

Two details of the guidelines

- The 2023 U.S. Merger Guidelines define only the 1,800 threshold; the 1,000 cutoff follows earlier U.S. guidelines.
- A second trigger: a merger creating a firm with more than 30% of the market, and raising the HHI by more than 100 points, is also presumed to harm competition.

Two real merger reviews

- **Siemens–Alstom (EU, 2019)**

  In 2017 Siemens and Alstom agreed to merge their rail businesses, backed by Germany and France as a “European champion” to rival China's state-owned train maker. The European Commission blocked it, ruling it would harm competition in signaling and high-speed trains and raise prices: competition policy over industrial-policy ambitions.

Two real merger reviews (continued)

- **Walmart–Massmart (South Africa, 2010–11)**

  Little threat to competition, but South African law weighs public-interest factors. The Tribunal approved on conditions: no layoffs for two years, rehire 503 workers, honor labor agreements, and a 100-million-rand supplier fund.

15.9

Case Study: The Global Beer Oligopoly

Oligopoly forming through merger, then competing on brands.

15.9 CASE STUDY

Oligopoly built through mega-merger

THE DEAL

In 2016, Anheuser-Busch InBev, the product of earlier mega-mergers, acquired its largest rival, SABMiller, for about $107 billion.

~26%

of all beer on earth brewed by AB InBev (2024)

~13%

next-largest brewer, Heineken

15.9 CASE STUDY

Oligopoly built through mega-merger (continued)

THE GIANTS

Four giants (AB InBev, Heineken, China Resources Snow, and Carlsberg) together brew about half of the world's beer.

15.9 CASE STUDY

Beer illustrates several chapter themes at once

Rivals' reactions

Concentrated enough that each giant must weigh rivals' reactions, as in Cournot or Stackelberg competition.

Differentiation

They own hundreds of brands (Budweiser, Stella Artois, Corona), competing on variety, softening Bertrand price pressure.

Merger review

The wave of mergers that built these giants is what concentration measures like the HHI are designed to single out for antitrust review.

15.9 CASE STUDY

Measuring a merger's effect: MillerCoors

- **2008:** Miller and Coors, the second- and third-largest U.S. brewers, combine their U.S. operations in MillerCoors.
- **Miller and Weinberg (2017):** retail beer prices rose sharply after the deal, for MillerCoors and for its main rival, Anheuser-Busch.
- Prices ended up 6 to 8% higher than with continued competition, linked to easier price coordination, the tacit collusion of Section 14.5.1.

15.10 Empirical Evidence: The Lysine Cartel

Collusion is not just theory: it happens, it is illegal, and it gets caught.

A classic cartel acting as a single monopolist

- Lysine is an animal-feed additive, made by only a handful of firms worldwide.
- 1992–1995: Archer Daniels Midland and four rivals secretly agreed to fix the price and divide the world market.
- An informant (an ADM executive) cooperated with the FBI, which secretly recorded the cartel's meetings.

+70%
price rise within the first nine months

$100M
ADM fine (lysine and citric acid), then the largest U.S. criminal antitrust fine

A classic cartel acting as a single monopolist (continued)

THE COLLAPSE

The cartel collapsed in 1995 and several executives went to prison. The case shows collusion can raise prices substantially, and why the law works so hard to detect and deter it.

Why cartels come apart

- **The informant:** an ADM executive cooperated with the FBI, which recorded executives treating competitors as friends and customers as enemies (Connor, 2001).
- **Levenstein and Suslow (2006):** the biggest threats to cartels are entry and the need to adjust the agreement to changing conditions, more than cheating.

15.11

A Look at the Experimental Evidence: Oligopoly in the Laboratory

Reproducing oligopoly with real payoffs to test the models directly.

Testing the models in the lab

Study	Setting	Theory's prediction	Experimental result
Huck et al. (2004)	Cournot, with 2 to 5 firms	More firms push output toward the competitive level	Two firms partly collude; three behave near Cournot-Nash; four or five are at or above Cournot
Dufwenberg and Gneezy (2000)	Bertrand (price) competition	Two firms suffice to reach marginal cost pricing	With two firms prices stay above cost; with three or four they fall close to it
Fonseca and Normann (2012)	Repeated interaction with communication	—	Allowing rival firms to communicate sharply increases collusion

Table 15.4 Experimental tests of oligopoly models.

What the experiments teach

- Fewer firms yield higher prices; more firms push markets toward the competitive outcome.
- **Suetens and Potters (2007):** tacit collusion is more common when firms set prices than quantities, so the Bertrand outcome is harder to reach in the lab.
- **The warning:** one of the strongest predictors of collusion is whether firms can talk to one another; that is why communication among rivals about prices is illegal.
- **A survey:** Potters and Suetens (2013).
- **Run it yourself:** Veconlab's "Cournot" market (veconlab.com): does output settle near Cournot, or drift toward collusion or competition?

CHAPTER 15

Key takeaways

- **HHI:** sums squared market shares; above 1,800 a U.S. market is highly concentrated, and a rise of more than 100 points is presumed harmful.
- **Cournot:** output lies between the monopoly and competitive levels; the price approaches marginal cost as firms multiply.
- **Bertrand:** undercutting drives the price to marginal cost with two firms, unless products differ, capacity is limited, or firms meet repeatedly.

Key takeaways (continued)

- **Stackelberg:** the leader commits to a large output and earns more than under Cournot ($18 from 6 units rather than $16 from 4).
- **Collusion:** a prisoner's dilemma, sustained by the threat that rivals expand output later; algorithms can learn it too (Calvano et al., 2020).

Key takeaways (continued)

- **Monopolistic competition:** prices above marginal cost and zero long-run profit; excess capacity, but more variety.
- **Evidence:** the lysine cartel raised the price about 70% in nine months; in the lab, letting rivals communicate sharply increases collusion.

16

CHAPTER 16

Asymmetric Information

When one party to a deal knows something the other does not, and what markets do about it.

CHAPTER 16

What this chapter covers

16.2–16.4 Hidden type

- The market for lemons
- Adverse selection in insurance
- Signaling & screening

16.5–16.6 Hidden action

- Moral hazard
- The principal-agent problem
- Incentive contracts & regulation

16.7–16.9 Evidence

- Case: online platforms
- Empirical: RAND experiment
- Lab: lemons in the laboratory

CHAPTER 16

What this chapter covers (continued)

THE BIG IDEA

Asymmetric information is so important that the three economists who first analyzed it (Akerlof, Spence, and Stiglitz) shared the 2001 Nobel Prize.

16.1 Introduction

- Two forms: hidden type vs. hidden action
- Adverse selection and its cures
- Moral hazard and its cures

Three everyday puzzles

Used cars

Why pay full price when the seller is likelier unloading a dud than a gem?

Insurance

Why does a healthy young person skip coverage, leaving a sicker, costlier pool?

Online sellers

Why does a seller with thousands of good reviews fetch a higher price?

Three everyday puzzles (continued)

THE NEW ASSUMPTION

In every market so far, buyers and sellers knew what they were trading. This chapter drops that assumption: one party knows something the other does not.

Two forms of asymmetric information

- **Adverse selection: a hidden type**

  Something known before the deal: the quality of a car, the health of an applicant. The cure is to reveal it through signaling and screening.

- **Moral hazard: a hidden action**

  Something done after the deal: how carefully the insured behaves, how hard an employee works. The cure is to design contracts that restore incentives.

16.2

The Market for Lemons

Akerlof (1970): how hidden quality unravels a market.

Akerlof's used-car example

THE SETUP

Half of all used cars are good, half are lemons. Only the seller knows which kind a given car is; the buyer cannot tell them apart.

Car	Share of cars	Value to its owner	Value to a buyer
Good car	½	$10,000	$12,000
Lemon	½	$4,000	$5,000

Table 16.1. The used-car market: every car is worth more to a buyer than to its owner.

A "lemon" is the slang term for a bad car.

The unraveling, step by step

- **Full information:** every car trades, a good car at $10,000 to $12,000 and a lemon at $4,000 to $5,000.
- **Step 1, the price buyers offer:** the average value to a buyer, ½($12,000) + ½($5,000) = $6,000 + $2,500 = $8,500.
- **Step 2, who sells:** lemon owners ($4,000 < $8,500) sell; owners of good cars ($10,000 > $8,500) refuse.

The unraveling, step by step (continued)

- **Step 3, buyers revise their offer:** any car offered at $8,500 must be a lemon, so buyers pay only $5,000; lemon owners still sell ($4,000 < $5,000).
- **Step 4, the gains that are lost:** each good car is worth $12,000 − $10,000 = $2,000 more to a buyer, yet none trades; only lemon trades survive, each creating $5,000 − $4,000 = $1,000.

This is adverse selection

Adverse selection

Because the hidden quality is correlated with who chooses to sell, the bad drives out the good, and mutually beneficial trades in good cars never happen.

WHY IT MATTERS

The good cars are worth more to buyers than to owners, yet they never trade. Hidden information destroys gains both sides would have shared.

The market unravels

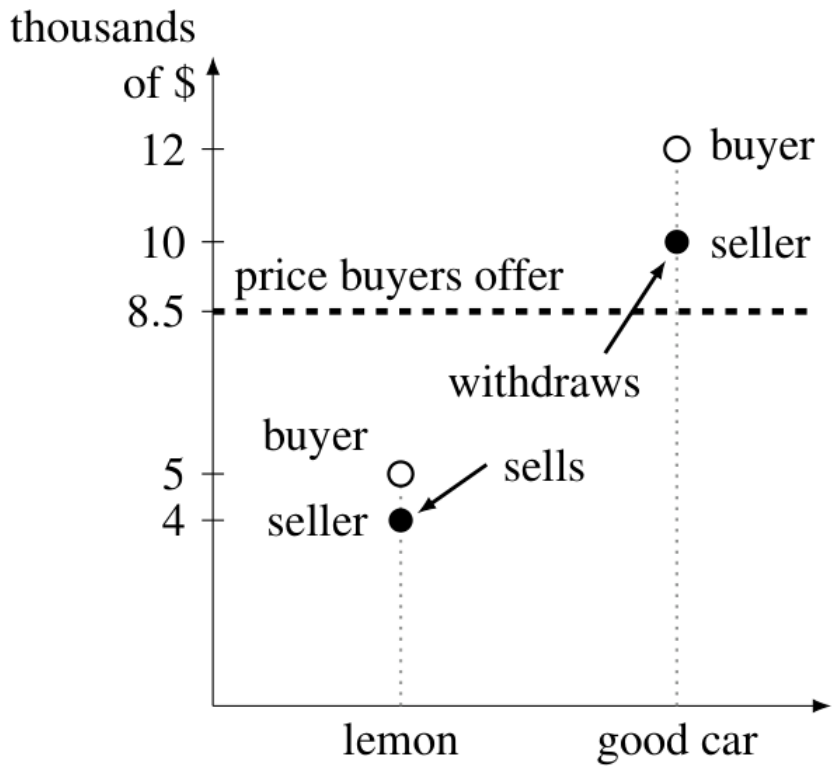


Figure 16.1 Adverse selection, in thousands of dollars.

- At the average price ($8,500), good-car owners refuse to sell.
- Only lemons remain, so buyers will pay only $5,000.
- The market unravels to lemons only.

Explore Figure 16.1: intro-micro-companion-website.netlify.app/adverse-selection

When does the market unravel?

- **Share of good cars:** let q be the share of good cars, so 1 − q is the share of lemons.
- **Price buyers offer:** the average value, q($12,000) + (1 − q)($5,000) = $5,000 + q($7,000).
- **Good cars stay only if** $5,000 + q($7,000) ≥ $10,000, that is, q ≥ 5/7 ≈ 0.71: at least 5 of every 7 cars must be good.

Two examples, and Table 16.1

- **80 percent good:** buyers pay $5,000 + 0.8($7,000) = $10,600, owners of good cars stay, and every car trades.
- **60 percent good:** buyers pay $5,000 + 0.6($7,000) = $9,200, owners of good cars withdraw, and the market unravels.
- **Table 16.1:** the share of good cars is ½, below 5/7, so the good cars disappear: lemons, only half of the cars, destroy the market for the other half.

Check your understanding

SELF-CHECK

Buyers value good used cars at $8,000 and lemons at $2,000, and owners of good cars value them at $7,000. Buyers cannot tell the two apart and believe half are each, so they offer $5,000. What happens to the good cars, and what is this called?

Owners of good cars will not sell for $5,000, so they withdraw; only lemons remain and the price falls further. Hidden quality drives the good cars out of the market: adverse selection, the lemons problem.

16.3

Adverse Selection in Insurance

The same logic, with the roles reversed.

Now the buyer holds the hidden information

- People who know they are high-risk (in poor health, or reckless drivers) are the most eager to buy insurance.
- The healthy and careful are tempted to go without.
- An insurer setting premiums for the average customer attracts a worse-than-average pool.
- Premiums rise, low-risk and then even average-risk customers drop out, and the pool worsens again.

Why insurers do what they do

Screen applicants

Medical questionnaires and driving records reveal hidden risk.

Tie to employment

Group plans capture healthy and sick workers together.

Require everyone in

Some countries require everyone to be covered, so the healthy cannot opt out.

Why insurers do what they do (continued)

THE COMMON THREAD

Each device fights the same problem: stop the healthy from opting out and leaving only the costly behind.

A death spiral in numbers

- **The pool:** an insurer sells one-year health policies to 300 people, 100 in each of three groups.
- **Risk averse:** each person would pay more than her expected cost for full coverage (the risk premium of Chapter 7).
- **Hidden type:** only each person knows her group, so the insurer charges everyone one premium, set to cover the average expected cost of those who buy.

16.3.1 A DEATH SPIRAL IN NUMBERS

Three groups of customers

Group	People	Expected cost	Most she would pay	Gain from coverage
Low risk	100	$1,000	$1,500	$1,500 − $1,000 = $500
Medium risk	100	$3,000	$4,500	$4,500 − $3,000 = $1,500
High risk	100	$8,000	$9,500	$9,500 − $8,000 = $1,500

Table 16.2. Three groups of customers. Every person values coverage above its expected cost.

The spiral, round by round

- **Step 1:** all 300 buy, so the premium is ($1,000 + $3,000 + $8,000)/3 = $12,000/3 = $4,000; low-risk customers (at most $1,500) drop out.
- **Step 2:** the premium rises to ($3,000 + $8,000)/2 = $11,000/2 = $5,500; medium-risk customers (at most $4,500) drop out.
- **Step 3:** the premium rises to $8,000; high-risk customers (up to $9,500) stay, and the spiral stops.

The gains that are lost

- **Step 4:** the 100 low-risk and 100 medium-risk customers end up uninsured, although each values coverage above its cost.
- **Lost gains:** 100($500) + 100($1,500) = $50,000 + $150,000 = $200,000.
- **The cause:** not recklessness or a low value of insurance, but asymmetric information and the ability to opt out.

The loop behind the spiral

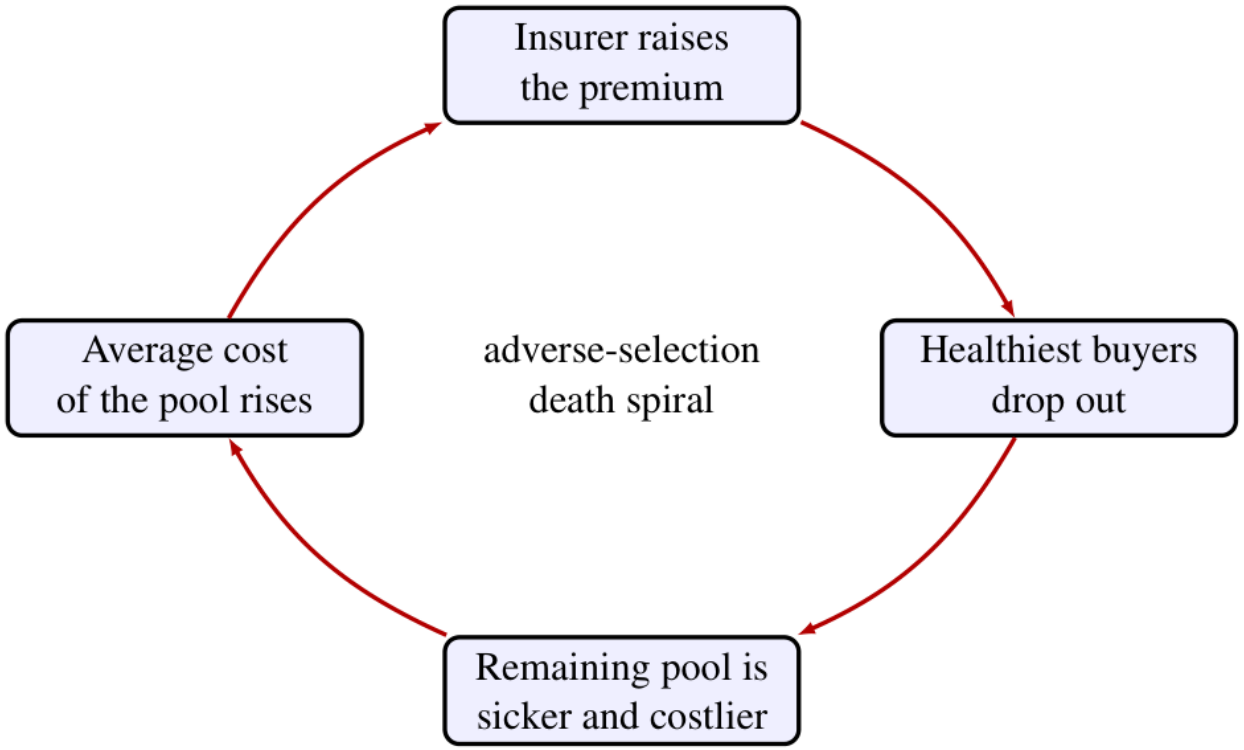


Figure 16.2 The adverse-selection death spiral.

- A higher premium pushes out the healthiest customers.
- The pool that remains is costlier.
- Its higher average cost calls for a higher premium.

Explore Figure 16.2: intro-micro-companion-website.netlify.app/death-spiral

What happens under a mandate?

- **All 300 must buy:** the premium stays at $4,000, the average expected cost of the whole pool, and the insurer covers its costs.
- **Gains:** 100($500) + 100($1,500) + 100($1,500) = $350,000, against 100($1,500) = $150,000 when only the high-risk group is insured.
- **The cost:** each low-risk member pays $4,000 for coverage she values at $1,500, so part of the others' gain comes from her pocket.

A real death spiral: Harvard's health plans

- **The plans:** in the early 1990s, about 10,000 employees chose between a generous PPO (Blue Cross/Blue Shield of Massachusetts) and less expensive HMOs.
- **The switch:** in 1995 (1996 for the rest of the staff), Harvard paid the same amount toward every plan, so PPO members paid its full extra cost.
- **The evidence:** Cutler and Zeckhauser (1998), in 1996 dollars for a family policy.

Three years to collapse

- **1995:** the PPO charge rose from about $500 to roughly $1,000; enrollment fell from 20 to 15 percent; leavers averaged age 46, stayers 51.
- **1996:** the PPO lost money and raised its premium; the charge topped $2,000, enrollment fell to about 9 percent, and leavers were 20 percent healthier than average.
- **1997:** only another steep increase could break even, and the PPO was discontinued.

What the spiral cost

- **Cutler and Reber (1998):** adverse selection cost 2 to 4 percent of Harvard's baseline health spending.
- **A partial offset:** stronger competition among plans also cut Harvard's premiums by 5 to 8 percent.
- **The lesson:** within three years, the most generous plan had disappeared, as in the rounds of Section 16.3.1.

The same mechanism everywhere

Credit markets

Borrowers willing to accept very high rates are often the riskiest, so banks ration credit and demand collateral.

Labor markets

At a fixed wage, those most likely to accept may have the fewest options, so firms screen with interviews and probation.

Company shares

Sellers may know more about a firm's prospects, so buyers discount what they will pay.

The same mechanism everywhere (continued)

AKERLOF'S LOGIC

In every case the mechanism is Akerlof's: hidden quality, correlated with who participates, drives out the good.

16.4 Signaling and Screening

- Signaling: the informed party reveals its type
- Screening: the uninformed party draws it out
- What makes a signal credible
- Education as a signal

If hidden information is the problem, revealing it is the cure

- **Signaling: the informed party reveals its type**

  A worker earns a degree as a signal of ability, credible because it is easier for able workers to obtain (Spence, 1973). A seller offering a long warranty signals that the product is reliable.

- **Screening: the uninformed party draws it out**

  An insurer offers a menu of policies, knowing high-risk customers choose fuller coverage; a lender requires collateral that only serious borrowers will post.

Screening with a menu of policies

- **Rothschild and Stiglitz (1976):** a menu of insurance policies sorts customers.
- **Low-risk customers** accept a high deductible in exchange for a low premium.
- **High-risk customers,** who expect to pay the deductible more often, turn that trade down.

16.4.1 WHAT MAKES A SIGNAL CREDIBLE

A signal must be one the bad type cannot imitate

THE KEY PRINCIPLE

The signal must be costlier for the low type than for the high type. Otherwise everyone would send it and it would reveal nothing.

- A college degree signals ability because it is genuinely harder, and costlier, for a less able student to earn.
- A long warranty signals reliability because it would bankrupt a shoddy maker but costs a quality maker little.
- Heavy spending on a headquarters or advertising signals a firm expects to be around to recoup it.

Spence's principle

SPENCE'S PRINCIPLE

Credible signals are costly in a way that separates the types; signals that anyone could mimic convey nothing.

Education as a signal: the setup

- **Two types, equal numbers:** over a career, a high-ability worker produces output worth $1,200,000 for her employer; a low-ability worker, $800,000.
- **Hidden ability, visible degree:** employers cannot observe ability, but they can see a degree.
- **Cost of the degree:** $250,000 for a high-ability worker and $500,000 for a low-ability worker, for whom the coursework is harder.

A degree that only sorts

- **To isolate the signal,** the degree teaches nothing that raises productivity in this example.
- **This does not mean** that degrees do not raise productivity: even if they did not, a degree could reveal a worker's type.
- **The pay rule:** employers pay $1,200,000 to university graduates and $800,000 to everyone else.

The numbers, in thousands of dollars

Worker	Value to an employer	Cost of the degree	Earnings with the degree, minus its cost	Earnings without the degree
High ability	1,200	250	1,200 − 250 = 950	800
Low ability	800	500	1,200 − 500 = 700	800

Table 16.3. Education as a signal, in thousands of dollars.

Workers' choices confirm the pay rule

- **Step 1, high ability:** $1,200,000 − $250,000 = $950,000 exceeds $800,000, so she earns the degree.
- **Step 2, low ability:** $1,200,000 − $500,000 = $700,000 is less than $800,000, so she does not.
- **The signal works:** only high-ability workers hold degrees, so a firm that sees a degree infers high ability.

The signal's cost for society

- **Without degrees:** every worker is paid the average product, ½($1,200,000) + ½($800,000) = $1,000,000.
- **With the signal:** a high-ability worker keeps $950,000, less than $1,000,000; the $250,000 she spends only sorts workers.
- **No one stops alone:** without a degree she would be paid $800,000. A society can spend more on credentials than the information they reveal is worth.

Self-check: how costly must the degree be?

- **Self-check.** Keep the example of Table 16.3, but let the degree cost a low-ability worker \$250,000 + k, where $k \geq 0$.
- **The pay rule:** employers pay \$1,200,000 to graduates and \$800,000 to everyone else.
- **The question:** what is the smallest value of k that still allows the high-ability worker to signal her type to the firm?

Self-check: the answer is k = $150,000

- **High ability:** $1,200,000 − $250,000 = $950,000 exceeds $800,000, so she earns the degree for any k.
- **Low ability does not imitate if** $1,200,000 − ($250,000 + k) ≤ $800,000, that is, $950,000 − k ≤ $800,000, or k ≥ $150,000.
- **At k = $150,000,** her cost, $400,000, equals the $400,000 wage gain, so she is indifferent and does not earn it (we assume that an indifferent worker does not earn the degree); for any smaller k, she would imitate.

16.5

Moral Hazard

Hidden actions taken after a deal is struck.

Hidden actions, not hidden types

Moral hazard

The second problem concerns hidden actions taken after a contract is signed. Insurance is again the leading case.

- A person with full fire insurance has less incentive to install smoke detectors.
- A driver with full collision coverage may park less carefully.
- A patient whose insurer pays every bill has little reason to weigh whether a test is worth its cost.

Timing tells the two problems apart

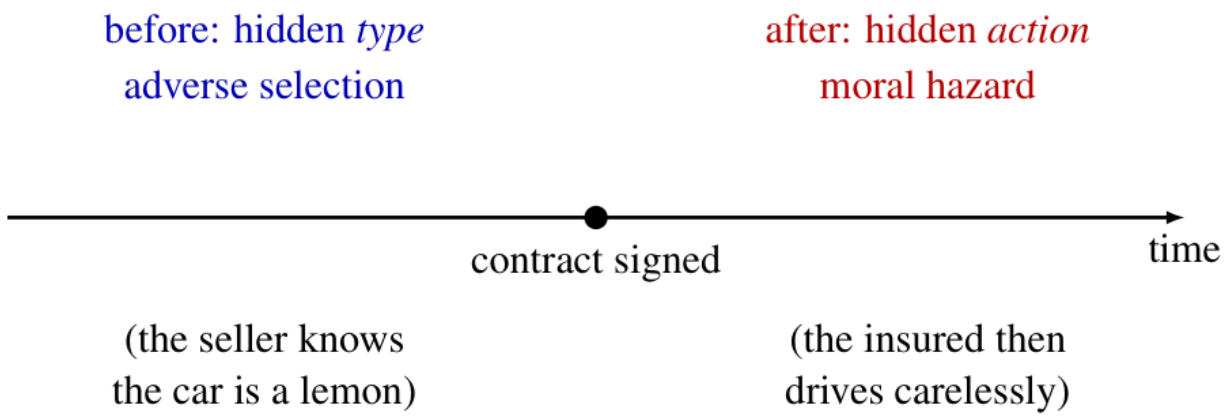


Figure 16.3 The timing of adverse selection and moral hazard.

Adverse selection A hidden type, known before the contract.

Moral hazard A hidden action, taken after the contract.

Both problems can occur together

- **Adverse selection:** a high-risk driver buys a policy designed for the average driver (with a low deductible).
- **Moral hazard:** after signing, the driver becomes even more reckless, expecting the insurance company to cover most damages.

16.5 MORAL HAZARD

Coverage dulls the incentive to take care

THE MECHANISM

The insured does not become reckless on purpose, but coverage dulls the incentive to take costly care, so insured losses run higher than they otherwise would.

Because the insurer cannot observe how careful each customer is, it cannot forbid carelessness; it must restore incentive through the contract.

Insurance and care: the setup

- **A $1,000 bicycle:** stolen during the year with probability 20 percent if she is careless (left outside unlocked), 5 percent if she is careful (heavy lock, indoors at night).
- **Care costs her** the equivalent of $30 a year in time and hassle.
- **The insurer** cannot see her care, and it sets the premium equal to its expected payout, given the care it expects her to take.

No insurance, then full coverage

- **Step 1, no insurance:** being careful costs \$30 + 0.05(\$1,000) = \$30 + \$50 = \$80; being careless costs 0.20(\$1,000) = \$200. She is careful.
- **Step 2, full coverage:** care costs her \$30, carelessness \$0, so she is careless; the premium is 0.20(\$1,000) = \$200, four times the 0.05(\$1,000) = \$50 for a careful customer.
- **Moral hazard:** coverage removes her incentive to take care and raises the expected loss.

A deductible restores care

- **Step 3, a deductible of D dollars:** being careful costs $30 + 0.05D and being careless 0.20D, so she is careful if $30 + 0.05D ≤ 0.20D, that is, D ≥ $30/(0.20 − 0.05) = $30/0.15 = $200.
- **Step 4, the $200 deductible:** she is careful; the insurer pays $1,000 − $200 = $800 with probability 5 percent and charges 0.05($800) = $40.
- **Her expected yearly cost:** $40 + $30 + 0.05($200) = $80.

Comparing the three policies

Policy	Is she careful?	Premium	Expected yearly cost
No insurance	yes: \$80 < \$200	\$0	\$30 + 0.05(\$1,000) = \$80
Full coverage	no: \$30 > \$0	0.20(\$1,000) = \$200	\$200
\$200 deductible	yes: \$40 = \$40	0.05(\$800) = \$40	\$40 + \$30 + 0.05(\$200) = \$80

Table 16.4. Insurance and care for a \$1,000 bicycle.

Why the $200 deductible is attractive

- **Same expected cost as no insurance, $80,** but a theft costs her at most $200 instead of $1,000, which a risk-averse student values (Chapter 7).
- **$120 a year less than full coverage,** whose $200 premium pays for the extra thefts that carelessness causes.

Step 3, drawn

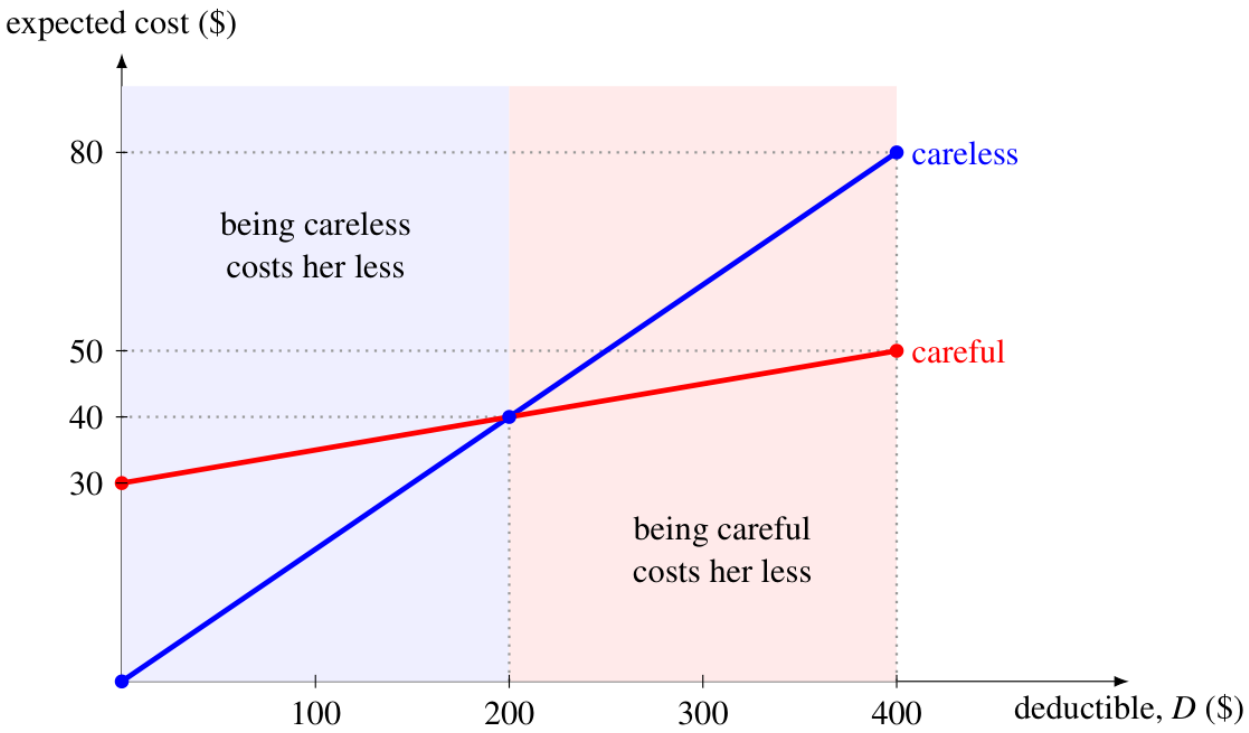


Figure 16.4 Expected out-of-pocket cost for each deductible D.

- Careless: the expected cost rises steeply with D.
- Careful: it rises slowly.
- They meet at D = $200; from there on, she is careful.

16.6 The Principal-Agent Problem

- When a principal cannot observe an agent's action
- Skin in the game through the contract
- The trade-off in incentive contracts

Moral hazard is one instance of a broader problem

The principal-agent problem

A principal (insurer, employer, shareholder) cannot fully observe the action of an agent (the insured, an employee, a manager) whose interests differ from its own.

THE SOLUTION

The solution: write a contract that ties the agent's reward to an observable outcome, giving the agent “skin in the game.”

Two ways to give agents skin in the game

Deductibles & co-payments

The customer bears part of any loss and keeps an incentive for care, as the $200 deductible of Section 16.5.1 does.

Performance pay

Commissions, bonuses, and stock mean workers gain when the firm does.

Two ways to give agents skin in the game (continued)

THE CATCH

Every such contract involves a trade-off between two goals: stronger incentives push the agent to act well but also load more risk onto her.

16.6.1 THE TRADE-OFF IN INCENTIVE CONTRACTS

Every contract walks a tightrope

Stronger incentive

Tie pay tightly to results (pure commission, mostly stock) and the agent works hard, keeping the fruits of her effort.

More risk

Results depend partly on luck: a recession, a bad year. A risk-averse agent (Ch. 7) must be paid extra to bear it.

Every contract walks a tightrope (continued)

THE BALANCE

The best contract balances the two: enough pay-for-performance to motivate, but not so much that the agent bears more risk than she can manage.

Why real contracts look the way they do

- Most workers are paid mostly a fixed salary with some bonus, rather than pure piece rates.
- Insurance comes with deductibles, rather than either full coverage or none.
- Executive pay mixes salary, bonus, and stock.

THE BOTTOM LINE

The principal cannot observe effort directly, so it designs a contract that makes the agent want to act well, and pays (in extra compensation for risk) for the privilege.

Performance pay in practice: Safelite Glass

- **The firm:** the largest installer of automobile glass in the United States.
- **The switch:** until 1994, installers were paid by the hour; then a piece rate per windshield installed, with a guaranteed minimum close to the old hourly wage.
- **The result:** output per worker rose 44 percent, and the firm's profits rose as well (Lazear, 2000).

Where the 44 percent came from

- **Incentives:** about half came from the same workers installing more windshields once their pay depended on it (Section 16.6.1).
- **Sorting:** the rest came from piece rates attracting and keeping faster installers, a form of screening (Section 16.4).
- **Risk:** the guaranteed minimum limited the risk installers bore in a slow month, the other side of the trade-off.

Regulation exists to fix these problems

- **The death spiral**

  In developed economies, a main concern is the insurance death spiral of Section 16.3.1 (Figure 16.2): higher premiums drive out the healthiest customers, until the market threatens to collapse.

THE CURE

The standard regulatory cure is to mandate participation, keeping the healthy in the pool.

Mandates and information rules

- **Individual mandate:** from 2014, the U.S. Affordable Care Act required most people to buy health insurance or pay a penalty (reduced to zero from 2019).
- **Universal coverage:** many countries achieve the same end through compulsory coverage.
- **Information rules:** lemon laws (car makers must replace or buy back defective new cars), used-car disclosure rules, food and drug labeling, and securities disclosure attack the lemons problem directly.

A developing-country example: group lending

- Lending to the poor is crippled by asymmetric information: hidden risk, unwatchable loan use, no collateral.
- The Grameen Bank (Yunus, Bangladesh: a project in 1976, a bank in 1983) introduced group lending.
- Members pick trustworthy partners (peer selection) and monitor one another to repay.
- Joint liability makes each member responsible for the others' loans.

A developing-country example: group lending (continued)

THE INSIGHT

Members choose trustworthy partners and pressure one another to repay, to keep a good standing in the community: private institutions at work where formal regulation is thin.

A developing-country example: group lending (continued)

~97%

repayment rate achieved by Grameen and similar lenders

2006

Yunus and Grameen shared the Nobel Peace Prize

What randomized evaluations find

- **Banerjee et al. (2015):** six evaluations of microcredit in six countries find modestly positive, but not large, effects on borrowers.
- **Giné and Karlan (2014):** in the Philippines, removing group liability, or lending to individuals rather than groups, did not increase default.

16.7

Case Study: How Online Platforms Solve the Lemons Problem

Reputation systems turn hidden quality into something observable.

16.7 CASE STUDY

The internet looks like an ideal setting for lemons

- Buyers purchase from sellers they have never met, sight unseen: the conditions of Section 16.2.
- That eBay, Airbnb, and Uber function at all is largely due to reputation systems.
- After each transaction, buyers and sellers rate one another, and ratings are public.
- This turns hidden quality into something observable: a modern form of signaling and screening.

16.7 CASE STUDY

The effect is measurable

~8%

price premium an eBay seller with a long positive record earns over a newcomer (Resnick et al., 2006)

Two-sided

Airbnb and ride-hailing apps: both sides rate each other

The effect is measurable (continued)

WHY IT WORKS

These systems do not eliminate asymmetric information, but by making reputations visible and valuable they keep the good from being driven out by the bad, allowing markets to exist where Akerlof's analysis predicts they might collapse.

16.7 CASE STUDY

Keeping ratings honest

- **The temptation:** the premium that a good record earns creates an incentive to fake one.
- **The FTC rule (August 2024, in force since October 2024):** bans fake reviews and testimonials (including AI-generated ones), paying for reviews, and suppressing negative ones, with civil penalties.
- **Like the disclosure rules of Section 16.6.3,** it protects the information on which buyers rely.

16.8

Empirical Evidence: The RAND Health Insurance Experiment

How large is moral hazard in practice?

One of the largest social experiments ever run

- Between 1974 and 1981, RAND randomly assigned more than 5,800 people to health plans.
- Plans ranged from completely free care to plans requiring people to pay most of their own bills up to a cap.
- Because assignment was random, any difference in care used could be attributed to the insurance plan.

5,800+
people randomly assigned to plans

1974–81
years the experiment ran

One of the largest social experiments ever run (continued)

WHY IT IS CLEAN

Random assignment gives a clean measure of moral hazard: the effect of the insurance, not of who chose it.

16.8 EMPIRICAL EVIDENCE

The results were decisive

−30%

spent on care by those who paid most of their own bills vs. those given free care

~0

average health improvement from the extra care

- The extra spending under free care came almost entirely from starting more episodes of care, not from more costly care once treatment had begun: the dulled incentive of Section 16.5.
- For the average participant the extra care produced little or no improvement in health.

An important exception

THE QUALIFICATION

The poorest and sickest patients did benefit from free care, for instance through better control of high blood pressure.

- Cost sharing curbs moral hazard, but it can also deter care that is genuinely worthwhile.
- The RAND findings still shape the deductibles and copayments in health plans today.

16.9

A Look at the Experimental Evidence: Lemons in the Laboratory

Does the lemons problem appear when real people trade?

Laboratory markets for lemons

- **Lynch et al. (1986),** in a study for the U.S. Federal Trade Commission.
- **The design:** sellers produce a “regular” unit or a costlier “super” unit; buyers value supers more but learn the grade only after buying.
- **The rules changed** from one market to the next, and the researchers measured how much of the largest possible gains from trade each market achieved.

Without reliable information, lemons

- **No identification, no enforcement:** 96 percent of the units sold (384 of 399) were regulars, and the markets achieved only about 72 percent of the possible gains.
- **Unenforced claims are cheap talk:** 61 of the 105 regular units sold, 58 percent, were advertised as supers.

Disclosure and warranties restore quality

- **Enforced disclosure:** 89 percent of the units sold (275 of 308) were supers, and the markets achieved on average more than 95 percent of the possible gains.
- **Optional enforced warranty:** 64 of the 65 supers sold carried one, a signal that a seller of regular units could not afford (Section 16.4.1).

More laboratory evidence

- **Miller and Plott (1985):** signaling appeared in most of their 11 markets, and how well it separated quality depended on its cost.
- **Dulleck et al. (2011):** with 936 participants, liability makes credence-good markets (such as car repairs) work well; reputation has little effect.
- **The lesson:** hidden quality drives out good quality, and credible information brings it back.

CHAPTER 16

Key takeaways (1 of 3)

1. **Two forms:** adverse selection comes from a hidden type known before the deal; moral hazard from a hidden action taken after it.

2. **Lemons:** buyers pay at most the average value, so good cars withdraw; in Table 16.1, the market for good cars survives only if at least 5/7 of the cars, about 71 percent, are good.

3. **Death spiral:** in Table 16.2, the premium rises from $4,000 to $5,500 to $8,000 until only the high-risk group is insured; Harvard's PPO disappeared within three years; a mandate keeps the healthy in.

Key takeaways (2 of 3)

4. **Signaling and screening:** a signal separates the types only if the gain it brings lies between its cost for the high type and for the low type (Table 16.3); screening offers a menu of choices.

5. **Moral hazard:** a deductible of at least $200 restores care (Table 16.4); contracts balance incentives against risk; Safelite's piece rates raised output per worker 44 percent.

6. **Private institutions:** group lending achieves repayment rates around 97 percent; on eBay, a long positive record earns a price premium of about 8 percent.

Key takeaways (3 of 3)

7. **RAND experiment:** people who paid most of their own bills spent about 30 percent less on care than those given free care; the extra care brought little or no improvement in health for the average participant.

8. **Laboratory:** hidden quality led to lemons (96 percent of units sold were of low quality); enforced truthful disclosure raised the share of high-quality units to 89 percent.

17

CHAPTER 17

Auctions and Bidding Strategies

A mechanism for selling things whose value is uncertain, and the art of bidding well in one.

CHAPTER 17

What this chapter covers

17.2–17.3 Formats & bidding

- Four standard auctions
- Strategic equivalences
- Bidding with private values

17.4–17.5 Revenue & curse

- Revenue equivalence
- Common values
- The winner's curse

17.6–17.7 Evidence

- Case: spectrum & search ads
- Empirical: oil-lease bidding

CHAPTER 17

What this chapter covers (continued)

THE BIG IDEA

When a seller does not know what buyers will pay, an auction lets competition among bidders reveal it. This chapter applies the game theory of Part V to auction design and bidding.

17.1 Introduction

- Selling the radio airwaves
- Choosing which ad to show
- Why winners regret oil leases

Three puzzles, one answer

Spectrum

How does a government sell radio waves worth tens of billions when it has no idea what they are worth to each company?

Three puzzles, one answer (continued)

Search ads

How does a search engine like Google decide, after a search but before the page loads, which advertiser's link to show on top and what to charge?

Oil leases

Why do oil companies that win the rights to a drilling tract often regret it?

Three puzzles, one answer (continued)

THE COMMON THREAD

The answer to all three is an auction: a mechanism for selling things whose value is uncertain or known only to the bidders.

Why auctions?

- **An auction**

  A mechanism for selling things whose value is uncertain or known only to the bidders. It solves a problem posted prices cannot: when a seller does not know what buyers are willing to pay (the related problem that Chapter 11 mentions after its list of market failures), competition among bidders reveals it.

Why auctions? (continued)

THE PLAN

We begin with the four standard formats, then study how to bid with private values, the surprising result that all four raise the same revenue, and the trap of the winner's curse when value is uncertain and shared.

17.2

Four Standard Auctions

- English (ascending)
- Dutch (descending)
- First-price sealed-bid
- Second-price (Vickrey)

The two open formats

English (ascending) auction

The price rises until only one bidder remains. The familiar auction used by auction houses such as Sotheby's and by eBay.

Dutch (descending) auction

The price starts high and falls until a bidder accepts, as in the Dutch flower markets that give it its name.

The two sealed formats

- **First-price sealed-bid auction**

  Each bidder submits one secret bid. The highest bidder wins and pays her own bid.

- **Second-price sealed-bid (Vickrey) auction**

  The highest bidder wins but pays only the second-highest bid.

The four formats side by side

Format	Bids	How the price is set	Winner pays
English (ascending)	Open	Rises until one bidder remains	≈ second-highest value
Dutch (descending)	Open	Starts high, falls until accepted	the price she accepts
First-price sealed-bid	Sealed	Highest secret bid wins	her own bid
Second-price (Vickrey)	Sealed	Highest secret bid wins	the second-highest bid

The four standard auction formats, by whether bids are public or sealed.

Public or sealed, a 2×2 map

Open (bids public)	Sealed (bids secret)
English (ascending) price rises until one bidder remains	*Second-price (Vickrey)* highest bid wins, pays the 2nd-highest bid
Dutch (descending) price falls until a bidder accepts	*First-price* highest secret bid wins, pays its own bid

Figure 17.1 The four standard auction formats, by whether bids are public or sealed.

Public or sealed, a 2×2 map (continued)

- Open (English, Dutch): bidders see the price move and react.
- Sealed (first- and second-price): one private bid, no chance to react.

Two pairs behave the same way

- **English ≈ second-price sealed-bid**

  Under independent private values (each bidder knows her own value and it does not depend on what others learn): you stay in until the price reaches your value, then drop out. The auction ends when the second-to-last bidder quits, so the winner pays roughly the second-highest value: the second-price rule.

Two pairs behave the same way (continued)

- **Dutch ≈ first-price sealed-bid**

  You must decide in advance the price at which you will jump in, and you pay it, just as in a first-price auction you submit one bid and pay it, with no chance to react to others.

Why this ties the chapter together

ONE SIMPLIFICATION

These equivalences let us reason about all four formats by analyzing just two bidding problems: the second-price logic, where truthful bidding is a dominant strategy, and the first-price logic, where you must shade your bid.

17.3

Bidding with Private Values

- Each bidder knows her own value
- Second-price: bid your true value
- First-price: shade your bid

What is a private value?

- **Private values**

  Each bidder knows her own value for the object (what it is worth to her), and these values differ from person to person. How should she bid?

The answer depends on the format. Two cases, very different logic.

Second-price: bid your true value

THE CLEAN ANSWER

Bidding your value is a dominant strategy: best no matter what others do.

- You pay the second-highest bid, not your own, so your bid sets whether you win, not how much you pay.
- Raising your bid above your value risks winning at a price above what the object is worth.
- Lowering it below your value only risks losing an auction you would have been glad to win.

No other bid does better

A NOBEL RESULT

The result is due to Vickrey (1961), who shared the 1996 Nobel Prize. It makes second-price auctions prized for their simplicity and honesty.

First-price: shade your bid below your value

- **Bid shading**

  Because you pay your own bid, bidding your full value would guarantee zero profit even if you win. So you shade your bid below your value, sacrificing some chance of winning in exchange for a profit when you do.

HOW MUCH TO SHADE

It depends on how many rivals you face and how high they bid: the more rivals, the less you shade (with common values, more rivals call for more shading: Section 17.5).

A worked example: the object is worth $100 to you

The Claim, Checked with Numbers

You value the object at $100 in a second-price auction. Start with a bid equal to your true value, $100. Does any other bid do better? Check the cases.

Bid $100, rival bids $80

You win and pay $80, for a $20 gain.

Bid $100, rival bids $120

You lose, which is fine, since winning would mean paying $120 for something worth $100.

Neither deviation can help

Bid more, say $130

Only changes the outcome when the top rival bid is between $100 and $130: you now win but pay more than the object is worth. A loss.

Bid less, say $70

Only changes the outcome when the top rival bid is between $70 and $100: you now lose an auction you would have won at a price below $100.

Neither deviation can help (continued)

WHY BOTH FAIL

Your own bid sets whether you win, but not how much you pay. So bidding your true value is a dominant strategy: Vickrey's clean result.

Bidding more or less than your value

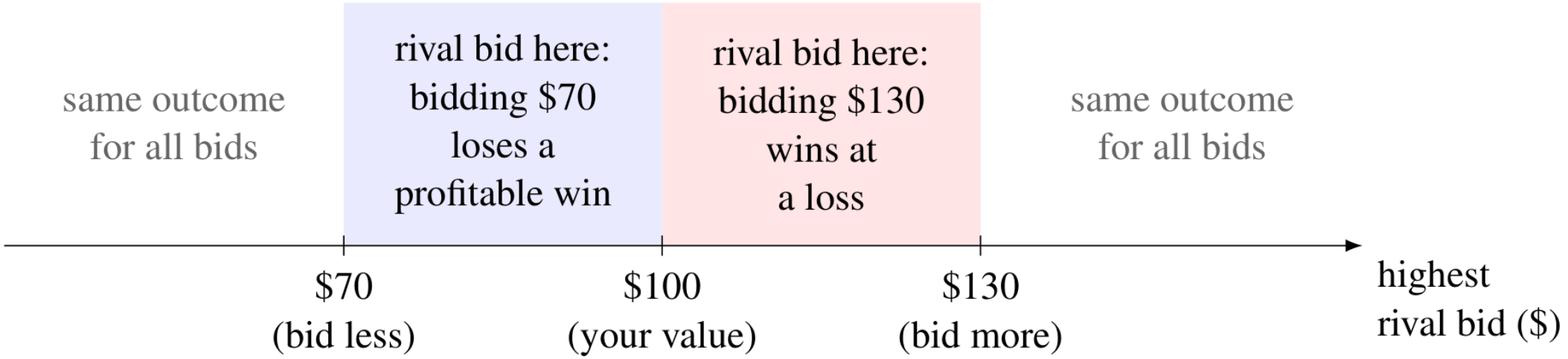


Figure 17.2 Bidding more or less than your $100 value in a second-price auction.

- Either deviation can only hurt you and never help: your bid sets whether you win, not how much you pay.

Self-check

SELF-CHECK

In a second-price (Vickrey) auction you value an item at $100. Why is bidding $100 your best strategy, no matter what others bid?

The winner pays the second-highest bid, not her own, so your bid only decides whether you win, never how much you pay. Bidding your true $100 wins whenever the highest rival bid is below your value and never overpays. Truthful bidding is a dominant strategy.

17.4

Revenue Equivalence

Which format makes the most money for the seller?

All four raise the same expected revenue

- **Revenue equivalence:** under ideal conditions, all four formats raise the same expected revenue.
- **The ideal conditions:** bidders who are alike, whose private values are drawn independently in the same way, and who care only about their expected profit.

Two bidders: what to expect

- **The setting:** two bidders whose values are equally likely to lie anywhere between $0 and $100.
- **Expected revenue:** the seller does not know the values, so we compare the formats by their average revenue.
- **Three equal pieces:** on average, two such values split $0 to $100 into thirds, so the lower value averages $100/3 ≈ $33 and the higher value 2 × $100/3 ≈ $67.

Both formats raise about $33

- **Second-price:** each bids her value, and the winner pays the lower value, about $33 on average.
- **First-price:** each bids half of her value (the equilibrium bid, derived with more advanced tools); the winner has the higher value, about $67, and pays half of it, ½ × (2 × $100/3) = $100/3 ≈ $33.
- **Why they match:** the first-price winner shades her bid, but she pays her own bid rather than the second-highest; on average the two effects cancel.

Explore Section 17.4: intro-micro-companion-website.netlify.app/auctions

A benchmark

- **When its assumptions fail** (bidders are risk averse or values are correlated), the formats raise different revenues.
- **Correlated values:** positively, if one bidder's high value makes high rival values more likely (a painting's resale price, an oil tract's geological report); negatively, if it makes low values more likely.
- **Auction design:** choosing the auction that maximizes the seller's expected revenue becomes the main problem.

17.5

Common Values and the Winner's Curse

- When the prize's value is shared but unknown
- Why winning is bad news
- How to avoid the curse

From private to common values

Common value

The object is worth the same to everyone, but its value is unknown at the time of bidding. An offshore oil tract holds a certain amount of oil; a company up for sale has a definite worth. Each bidder forms an estimate, but no one knows the true figure.

Estimates scatter around the true value, some too high, some too low.

Why winning is bad news

The winner's curse

If bids increase with estimates, the winner is, by definition, the bidder with the highest estimate, very likely an overestimate. Winning means you probably received the most optimistic estimate and have likely paid more than the object is worth.

THE CURE, IN ONE LINE

The cure: bid as though you had already won, recognize that winning signals you overestimated, and shade your bid downward accordingly.

Estimates scatter; the winner overestimates

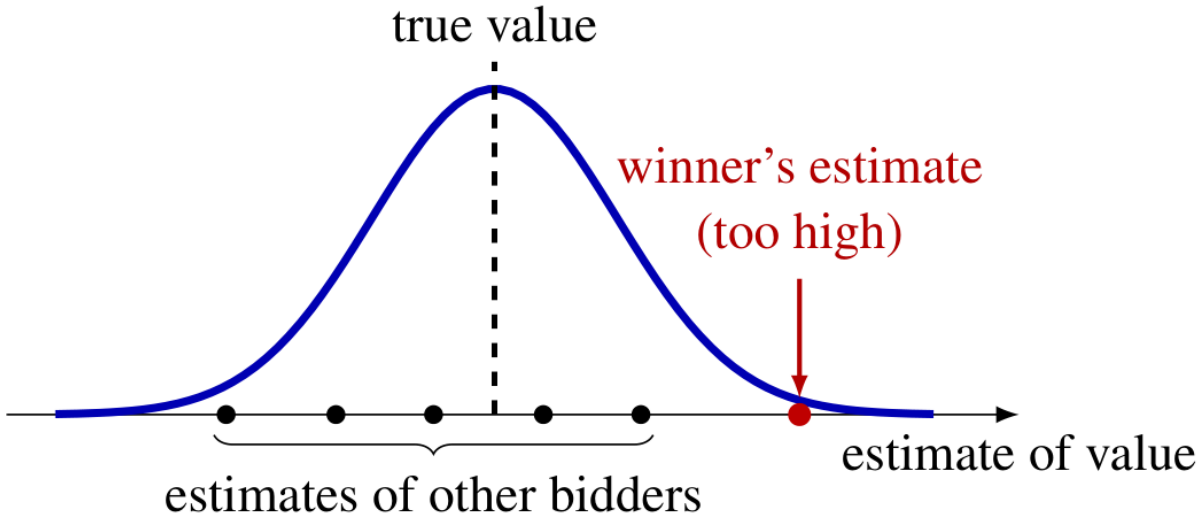


Figure 17.3 The winner's curse.

- Bidders' estimates cluster around the true value, but spread out on both sides.
- The winner draws the highest estimate, most likely above the truth.
- Win, and you have probably overpaid.

17.5.1 HOW TO AVOID THE WINNER'S CURSE

Bid as though you have already won

Where the Trap Comes From

The trap comes from a failure of imagination: a bidder forms her estimate, bids close to it, and forgets that winning carries information.

- Ask: “if my bid turns out to be the highest, what does that tell me about my estimate?” It was probably too optimistic.
- So shade your bid downward, below your raw estimate, by an amount that grows with the number of rivals.
- The more bidders there are, the further the highest estimate tends to lie above the true value, so the shading should be larger.

A reason to bid thoughtfully, not to stay away

THE TAKEAWAY

Experienced bidders in common-value auctions (oil companies, corporate acquirers, players in online ad markets) learn to do this. It is the lesson the oil industry drew from the study of Section 17.7.

17.6

Case Study: Selling the Airwaves and the Search Page

Two of the largest auction markets, built directly on this theory.

17.6 CASE STUDY

Selling the airwaves: spectrum auctions

- Governments auction the rights to use frequency bands for mobile phones and the internet.
- The formats are designed by economists to handle bidders who want particular combinations of licenses.
- Milgrom and Wilson won the 2020 Nobel Prize for improving auction theory and inventing new formats for complex goods like spectrum.

WORLDWIDE

These designs have raised hundreds of billions of dollars worldwide.

17.6 CASE STUDY

Selling the airwaves: spectrum auctions (continued)

~$81 bn

raised by the FCC's 2021 “C-band” auction, among the highest-grossing spectrum auctions ever

How the search-ad auction works

- **An auction in every search:** it runs in milliseconds to decide which ads appear and what each advertiser pays: the generalized second-price auction.
- **Ranking:** the highest bidder gets the top position, the second-highest bidder the second position, and so on.
- **Price per click:** not its own bid but the bid ranked just below it: the top position pays the second-highest bid, the second position the third-highest, and so on.

17.6 CASE STUDY

Second-price logic, with a twist

- **Several positions for sale:** bidding your true value is no longer always best; an advertiser can gain by bidding less and taking a lower, less expensive position (Edelman et al., 2007).
- **Scale:** run billions of times a day, these auctions generate more than half of the revenue of Alphabet, Google's parent (about $198 billion of its $350 billion in 2024).

17.7

Empirical Evidence: The Winner's Curse in Oil-Lease Bidding

First identified not in a laboratory but in the oil fields.

Capen et al. (1971)

- Three petroleum engineers noticed a puzzle in offshore lease auctions.
- Oil companies bidding for offshore leases in the Gulf of Mexico earned disappointingly low returns “year after year.”
- And yet the auctions were fiercely competitive.

THE DIAGNOSIS

Their explanation was the winner's curse. The oil's value was roughly the same for every company but unknown; each firm bid on its own geologists' estimate. The winner was typically the firm whose geologists were most optimistic, so the winning bid routinely exceeded the oil's true worth.

What the industry learned

- Companies that bid their raw estimate rather than shading it down systematically overpaid.
- The paper taught the industry to bid more cautiously.
- It gave economists their first real-world documentation of a phenomenon later confirmed in dozens of lab experiments (people bidding for a jar of coins reliably overpay).

Later evidence

- **Hendricks et al. (2003):** in federal offshore oil and gas lease auctions held between 1954 and 1970, bidders were aware of the curse and bid largely as the theory predicts.
- **Massey and Thaler (2013):** in the National Football League's draft, teams pay too much for the top picks, which on average yield less value per dollar of salary than later first-round picks.
- **Reviews:** Thaler (1988) and Kagel and Levin (2002).

CHAPTER 17

Key takeaways (1 of 3)

1. **Four formats:** English (ascending), Dutch (descending), first-price sealed-bid, and second-price (Vickrey). The English auction works like the second-price auction, and the Dutch like the first-price.

2. **Second-price:** bidding your true value is a dominant strategy: your bid decides only whether you win, and the price is the second-highest bid.

3. **First-price:** you pay your own bid, so you shade below your value; with private values, the more rivals you face, the less you shade.

Key takeaways (2 of 3)

4. **Revenue equivalence:** under ideal conditions all four formats raise the same expected revenue; with two bidders and values between $0 and $100, both the first- and second-price auctions raise about $33.

5. **Winner's curse:** with a common value, the winner has probably overestimated it; bid as though you have already won and shade down, the more so the more rivals you face.

CHAPTER 17

Key takeaways (3 of 3)

6 **Spectrum and search ads:** the FCC's 2021 C-band auction raised about $81 billion; search ads use a generalized second-price auction, where bidding your true value is not always best.

7 **Oil leases:** Capen et al. (1971) found low returns: the winner was typically the firm whose geologists had been most optimistic.

18

CHAPTER 18

Externalities, Public Goods, and Common Resources

When market incentives get the answer wrong (too much pollution, too little flood protection, too little restraint) and the policies to address them.

CHAPTER 18

What this chapter covers

18.2–18.4 Externalities

- Private vs. social cost
- Taxes, subsidies, permits
- The Coase theorem

18.5–18.6 The commons

- Public goods & free riding
- Common resources
- Tragedy of the commons

18.7–18.9 Evidence

- Case: congestion pricing
- Empirical: acid rain
- Behavioral: public goods

CHAPTER 18

What this chapter covers (continued)

THE BIG IDEA

Every failure here shares one root: a gap between private and social costs or benefits, so self-interested choices no longer serve the common good.

18.1 Introduction

- Smoke, flood barriers, and fisheries
- Where competitive markets fail
- A gap between private and social

18.1 INTRODUCTION

Three cases where the market gets it wrong

Too much pollution

A factory's smoke imposes costs on everyone downwind, costs it does not necessarily pay.

Too little flood protection

A flood barrier protects every home behind it, yet no homeowner wants to pay for one the neighbors can use for free.

Three cases where the market gets it wrong (continued)

Too little restraint

A fishery open to all is fished to collapse, though every fisher would gain by holding back.

WHY IT MATTERS

Chapter 10's competitive markets were efficient because each party bore the full costs and reaped the full benefits of their choices. These failures break that link.

18.2

Externalities: When Private and Social Costs Diverge

- What an externality is
- Negative externalities & overproduction
- Positive externalities & underproduction

A cost or benefit that falls on a bystander

- **Externality**

  A cost or benefit that an activity imposes on someone who is not part of it, so whoever decides on the activity does not fully enjoy its benefits or suffer its costs.

- **Negative externality**

  Like pollution: social cost exceeds the private cost the producer pays.

A polluting factory: the market overproduces

- The firm weighs only its private cost, so it produces where demand meets private marginal cost: the market quantity, $Q^* = 6$ units.
- But each unit imposes $4 of harm downstream, so the true social marginal cost lies above the private one.
- The socially optimal quantity, where demand meets social marginal cost, is lower, $Q^{SO} = 4$ units.

THE RESULT

Units 4 to 6, whose social cost exceeds their value, generate a deadweight loss. Polluting costs the firm too little, so we get too much of it.

A negative externality, drawn

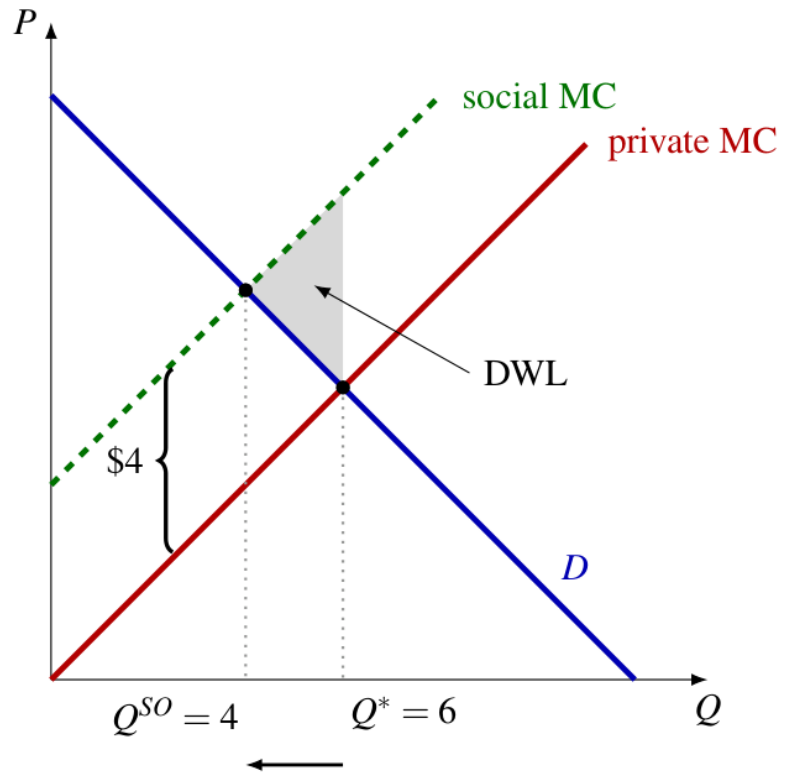


Figure 18.1 A negative externality.

- Social marginal cost lies above private marginal cost by the external harm.
- The market quantity, $Q^* = 6$, exceeds the socially optimal quantity, $Q^{SO} = 4$.
- The shaded triangle is the deadweight loss.

Explore Figure 18.1: intro-micro-companion-website.netlify.app/externality

Pollution damages by industry

- Muller et al. (2011): compared the damages from the air pollution that each U.S. industry emits with the industry's value added (the value of its output minus the cost of the materials and services it buys).
- Damages above value added: for several industries, including solid waste combustion, sewage treatment, stone quarrying, and oil- and coal-fired power plants.
- Coal-fired electricity: caused the largest damages of any industry, between 0.8 and 5.6 times its value added.

Beneficial spillovers: the market underproduces

- A person who gets vaccinated protects everyone she might have infected; an educated person becomes a better-informed voter and colleague; a beekeeper's bees pollinate the neighbor's orchard.
- Social benefit exceeds the private benefit buyers capture, so demand (private benefit) lies below the social-benefit curve.
- Guided by private benefit alone, the market produces too little: too few vaccinations, too little education, too few beehives.

18.2.1 POSITIVE EXTERNALITIES

Beneficial spillovers: the market underproduces

THE MIRROR IMAGE

Where a negative externality calls for a tax to discourage the activity, a positive externality calls for a subsidy to encourage it.

A positive externality, drawn

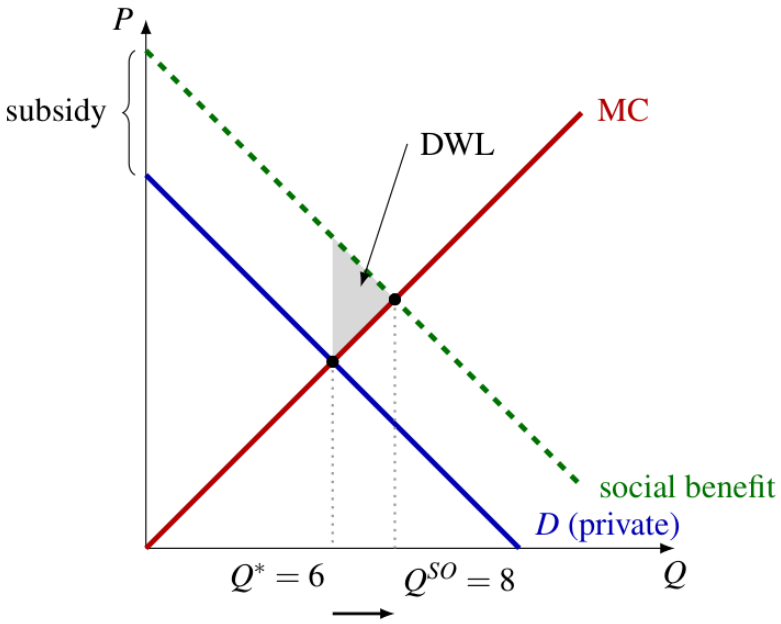


Figure 18.2 A positive externality.

- The social-benefit curve lies above private demand.
- The market stops at Q* = 6, short of the socially optimal Q^{SO} = 8.
- A subsidy shifts private demand up until it coincides with the social benefit.

18.3

Correcting Externalities: Taxes, Subsidies, and Permits

- Internalize the externality
- Pigouvian taxes & subsidies
- Tradable permits (cap and trade)

Three tools to internalize the externality

Corrective (Pigouvian) tax

Set equal to the external harm ($4 per unit), it raises private cost to social cost, cutting output to the socially optimal 4 units. (Cigarettes, carbon.)

Three tools to internalize the externality (continued)

Subsidy

The reverse for positive externalities: it pays for the extra social benefit, as with vaccines or education (Chapter 4).

Tradable permits

Cap and trade: set a legal limit on total emissions and let firms buy and sell the right to pollute.

Three tools to internalize the externality (continued)

THE UNIFYING IDEA

The policy is to make decision makers face the full social cost or benefit: to internalize the externality. (The tax is named for the economist Arthur Pigou, 1920.)

Cap-and-trade, step by step

- **The cap:** a legal limit on total emissions, divided among firms as permits (for instance, permits to emit 10 tons of carbon dioxide per firm per year).
- **The trade:** firms that can cut emissions cheaply do so and sell their permits; firms for which cutting is technologically costly buy them.
- **The price:** when the cap is set at the socially optimal level (4 units in Figure 18.1, if each unit of output requires one permit), the price of a permit settles at the $4 external harm.

Corrective taxes in practice

- **Cigarettes:** the U.S. federal tax is $1.01 per pack of 20, and states add their own, the highest being New York's $5.35 per pack.
- **Carbon in Switzerland:** since 2022, 120 Swiss francs per ton of carbon dioxide from heating oil and natural gas, about 30 centimes per liter of heating oil.
- **Carbon in Sweden:** among the highest carbon taxes in the world (Section 18.3.2).

Subsidies in practice

- **Pell Grants:** up to $7,395 per student per year for college education in the United States (2024–25).
- **Heat pumps:** the United Kingdom pays £7,500 toward each heat pump that replaces a fossil-fuel boiler (since October 2023).
- **Electric cars:** Germany paid up to 4,500 euros per new electric car (until December 2023); the United States offered a tax credit of up to $7,500 per new electric vehicle (until September 2025).

Subsidizing clean alternatives

- More than doubled: federal support for renewable energy in the United States rose from $7.4 billion in fiscal year 2016 to $15.6 billion in fiscal year 2022.
- A large share: 46% of all federal energy subsidies over these years went to renewables (U.S. Energy Information Administration, 2023).
- Indirect versus direct: such a subsidy lowers the private cost of a clean substitute, an indirect way to reduce pollution, whereas a corrective tax raises the polluter's cost directly.

Two ways to put a price on pollution

Tax: fixes the price

The government sets the charge per ton and firms decide how much to emit; the regulator knows the price but not the quantity.

Cap & trade: fixes the quantity

The government sets total allowed emissions and issues that many permits; it knows the outcome but not the price firms pay.

Two ways to put a price on pollution (continued)

THEIR SHARED VIRTUE

Both outperform rigid command-and-control: because firms choose how to respond, reductions happen wherever they are least costly, so society hits its goal at the lowest total cost. A tax gives price certainty; a cap gives quantity certainty.

18.3.2

Carbon and Pollution Pricing Around the World

No longer theoretical: governments across the income spectrum now price pollution.

18.3.2 PRICING AROUND THE WORLD

Four schemes, four approaches

Country / region	Scheme	Detail
Sweden	Carbon tax (1991)	Began at 250 kronor (~$25) per ton; 1,510 kronor (~$150) per ton in 2025, among the highest anywhere
European Union	Emissions Trading System (2005)	Cap-and-trade covering power plants and heavy industry across the continent
China	National ETS (2021)	Began with ~2,200 power plants (~40% of emissions); steel, cement, and aluminum added in 2025 (~60%); world's largest carbon market by volume
India (Gujarat/Surat)	Particulate cap-and-trade (2019)	World's first market for particulate (soot) air pollution

Selected carbon and pollution-pricing schemes worldwide.

India's Surat market: evidence from an RCT

99%

factories in the market met pollution limits (vs. 66% under conventional regulation)

20–30%

cut in particulate emissions, while lowering firms' compliance costs

India's Surat market: evidence from an RCT (continued)

THE FINDING

Greenstone et al. (2025) ran a randomized controlled trial (the gold standard of evidence). A striking demonstration that market-based instruments introduced in wealthy countries can work, and work at low cost, against the pollution that burdens the developing world.

Self-check: the right corrective tax

QUESTION

A factory's production imposes a $3-per-unit pollution cost on neighbors that it ignores. What is the efficient corrective (Pigouvian) tax, and what does it do to output?

SELF-CHECK

A tax of $3 per unit, equal to the external cost, makes the firm face the full social cost of what it produces. The firm then cuts output to the efficient level, where marginal social benefit equals marginal social cost.

18.4

The Coase Theorem: Bargaining to the Solution

- Sometimes no government action is needed
- Clear rights, equal information, low-cost bargaining
- Powerful, but limited

Why externalities arise: nobody owns the resource

Missing ownership

No one owns the river, so the factory can use it as a free dump, and the cost falls on people downstream who have no way to charge for it.

THE LINK TO COASE

If the people downstream owned the river, the factory would have to pay them for the right to pollute it, and the harm would become part of the factory’s costs. This observation is the starting point of the work of Ronald Coase.

Private bargaining can reach efficiency

THE COASE THEOREM

If property rights are clearly defined, the parties are equally informed about the costs and benefits at stake, and transaction costs (the costs of reaching and enforcing an agreement) are low, the parties to an externality negotiate their way to the efficient outcome on their own, and they reach the same efficient outcome regardless of who holds the right.

Private bargaining can reach efficiency (continued)

- Ronald Coase, “The Problem of Social Cost” (1960), which helped earn him the 1991 Nobel Prize in Economics.
- The assignment of rights determines who pays whom, but not what actually gets produced.
- So private negotiation can substitute for government action when the setting allows it.

A bar, a family, and late-night music

- **The harm:** the bar's live music after 11 p.m. costs the family next door $300 per month (the value they would place on quiet nights).
- **The benefit:** playing music after 11 p.m. is worth $150 per month to the bar.
- **Efficient:** stop the music at 11 p.m.; the family gains $300 and the bar loses $150, for a net gain of $300 − $150 = $150.

Either assignment of the right stops the music

- **The family has the right to quiet:** the bar would pay at most $150, and the family refuses any offer below $300; no deal is struck, and the music stops at 11 p.m.
- **The bar has the right to play:** the family pays the bar between $150 and $300 to stop; a deal is struck, and the music stops at 11 p.m.
- **Coase's point:** the assignment of rights determines who pays whom, but not what actually gets produced.

Three conditions, often violated

- **1. Clearly defined property rights:** someone must hold the right (the bar to play music late, or the family to have quiet), and the courts must enforce it; if it is unclear who holds it, bargaining cannot start.
- **2. Symmetric information:** if only the bar knows what the late music is worth to it, it can exaggerate that value, and the bargain may break down (Chapter 16).
- **3. Low transaction costs:** this fails with costly legal disputes, holdouts, or strategic bargaining, or when the affected parties are too many to sit at a table (the millions affected by carbon dioxide cannot negotiate with the millions of emitters).

When bargaining fails

- **Examples:** climate change, urban air pollution, fisheries with hundreds of boats.
- **Then** private bargaining alone will not reach the efficient outcome, and many economists turn to the tools of Section 18.3: corrective taxes, subsidies, and tradable permits.
- **Coase's critique of Pigou (1960):** taxes and regulations are also costly, so each arrangement, including leaving the market alone, should be judged by its total effect, case by case.

18.5

Public Goods and Free Riding

- Rival and excludable
- The four types of goods
- Public goods & the free-rider problem

18.5 PUBLIC GOODS

Two properties that decide who can supply a good

Rival

One person's use leaves less for others: a sandwich.

Excludable

Non-payers can be kept from using it: a ticketed concert.

WHY IT MATTERS

Considering these two properties gives four types of goods. Whether a good is rival and whether it is excludable determines whether markets can supply it.

The four types of goods (Figure 18.3)

	Excludable	Nonexcludable
Rival	*Private goods* food, clothing, a sandwich	*Common resources* fisheries, congested roads, grazing land
Nonrival	*Club goods* cable TV, toll roads	*Public goods* national defense, flood protection

Figure 18.3 Four types of goods, by whether they are rival and whether they are excludable.

Public goods and the free-rider problem

- **Public good**

  Both nonrival and nonexcludable: national defense, flood protection, clean air, basic research.

THE FREE-RIDER PROBLEM

Markets undersupply public goods because of the free-rider problem: since no one can be excluded, each person hopes to enjoy the good without paying, so too little is provided. This is why public goods are typically financed by government through taxes, which compel everyone to contribute.

Excludability depends on institutions

- **The lighthouse,** a classic example of a public good, was often financed privately in England and Wales.
- **Light dues:** owners charged ships dues at the ports, which made lighthouse services excludable.
- **Government nearby:** the Crown granted the rights and Parliament regulated the dues (Coase, 1974; Van Zandt, 1993; Bertrand, 2006).

Club goods: nonrival but excludable

- **Examples:** cable television, a streaming service, a toll road without traffic jams.
- **Private supply:** one more subscriber does not reduce what others enjoy, but nonpayers can be kept out, so firms can sell club goods.
- **The inefficiency:** one more user costs almost nothing, so excluding people who would pay something, but less than the price, wastes benefits (Chapter 11).

Common resources: rival but nonexcludable

- **Examples:** ocean fisheries, public grazing land, a congested road, the atmosphere's capacity to absorb carbon.
- **No one can be kept out,** but what one person uses is no longer available to others.
- **The result:** overuse, as Section 18.6 shows.

18.6

Common Resources and the Tragedy of the Commons

- Rival but nonexcludable
- Why they are overused
- Property, regulation, community

Rival but nonexcludable, so overused

Common resource

Rival but nonexcludable: ocean fisheries, public grazing land, a congested road, the atmosphere's capacity to absorb carbon.

The Tragedy of the Commons

Each fisher who takes one more fish gains the whole fish but bears only a small part of the cost of a depleted stock, which falls on all fishers, so all fish too much and the fishery collapses: the tragedy of the commons.

Three broad policies to address the commons

1

Property rights

Divide the resource into tradable catch quotas so each owner has reason to conserve the stock she now captures: New Zealand's Quota Management System (1986), Iceland's cod ITQs (1990), Chile's territorial user-rights fisheries (TURFs).

Three broad policies to address the commons (continued)

2

Government regulation

Catch limits, seasons, gear restrictions, or congestion charges: Pacific halibut season limits, U.S. Endangered Species Act, EU Common Fisheries Policy, Singapore's Vehicle Quota System.

Three broad policies to address the commons (continued)

3

Community self-governance

Users craft and enforce their own rules: the Swiss Törbel pastures (charter of 1483), Valencian huertas water tribunals, Maine lobster gangs, Japanese iriaichi common woodlands.

Three broad policies to address the commons (continued)

THE MAIN IDEA

The common thread: someone, somehow, must make each user bear the cost her use imposes on everyone else.

Explore Section 18.6: intro-micro-companion-website.netlify.app/commons

Elinor Ostrom: communities can solve it themselves

A THIRD SOLUTION

Elinor Ostrom (1990), who won the 2009 Nobel Prize, documented many cases (mountain pastures, irrigation systems, coastal fisheries) in which the users themselves crafted and enforced rules to manage a shared resource sustainably, without either privatizing it or handing it to the state.

- The tragedy of the commons, while real, is not inevitable.
- Communities with the right institutions, monitoring, and trust can solve it on their own.

18.7

Case Study: Pricing Congestion and Pollution

Two of the largest applications: congestion pricing and carbon markets.

18.7 CASE STUDY

A crowded road is a common resource

- Each driver who enters slows everyone else but pays nothing for the delay she causes, so the road is overused.
- The remedy is to charge for it.

~20%
drop in traffic entering London's zone after the 2003 congestion charge

~30%
fall in peak-period congestion delays

18.7 CASE STUDY

A crowded road is a common resource (continued)

AND ELSEWHERE

Revenue funded public transport. Stockholm adopted a similar scheme after a 2006 trial; Singapore has priced road use since the 1970s.

18.7 CASE STUDY

Congestion pricing elsewhere

- **Singapore:** electronic road pricing since 1998, with charges that vary by location and time of day.
- **Santiago, Chile:** higher tolls on its urban highways at peak and congested times.
- **City-center fees:** Milan (since 2012) and Gothenburg (since 2013); New York City began charging drivers entering Manhattan below 60th Street in January 2025.

18.7 CASE STUDY

Carbon markets: cap and trade for greenhouse gases

THE EU ETS

The European Union's Emissions Trading System, launched in 2005, was the world's first large carbon market, capping the emissions of thousands of power plants and factories and letting them trade allowances so that cuts occur where they are least costly.

Both policies put a price on something once free to dump or to crowd: the core prescription of Section 18.3.

18.7 CASE STUDY

Other cap-and-trade systems for greenhouse gases

- **New Zealand** (since 2008).
- **Regional Greenhouse Gas Initiative** among power plants in northeastern U.S. states (since 2009).
- **California** (since 2013, linked with Québec's since 2014); South Korea (since 2015); China’s national system (since 2021).

18.8

Empirical Evidence: The U.S. Acid Rain Program

Strong evidence that market-based environmental policy works.

The world's first large-scale cap-and-trade

- 1990 amendments to the Clean Air Act targeted the sulfur dioxide from power plants that was acidifying lakes and forests.
- The government capped total emissions and issued tradable allowances.
- Utilities that could cut at low cost sold allowances to those that could not.

The results vindicated the theory

>40%

fall in sulfur-dioxide emissions from regulated sources by 2007 (from the 1990 level), below the long-term cap three years ahead of schedule

low cost

a small fraction of what command-and-control would have required

The results vindicated the theory (continued)

THE LESSON

Because the market directed reductions to wherever they were least costly, costs were far below what regulation would have required. The acid-rain program became the model for later carbon markets: putting a price on pollution yields large environmental gains at low cost.

18.9

A Look into Behavioral Economics: Public-Good Experiments

The free-rider problem, tested in the laboratory many times.

The public-goods game

THE SETUP

Participants each receive a sum of money and may contribute any part of it to a common pot that is multiplied and shared equally. Pure self-interest predicts everyone contributes nothing and free-rides on others.

- Real people are more generous than that, but only up to a point.
- As Fehr and Gächter (2000) summarize, in earlier experiments participants initially contributed 40–60% of their money (more than the selfish prediction).
- Cooperation steadily eroded with repetition, until in the final round most contributed nothing, as free riding predicts (see also Ledyard, 1995).

The twist: let players punish free riders

A PARTIAL HUMAN SOLUTION

When players were allowed to punish free riders at a cost to themselves, cooperation no longer collapsed but held at high levels, with most participants eventually contributing their entire endowment.

People will sacrifice to punish those who do not pull their weight, and the threat of that punishment sustains cooperation.

Why cooperation erodes

- **Conditional cooperators:** Fischbacher et al. (2001) find that about half of the participants contribute more when they expect others to contribute more; about a third are free riders.
- **The decline:** as conditional cooperators see free riders contribute little, they cut back too.

Punishment does not always help

- **Herrmann et al. (2008):** the same experiment in 16 cities around the world.
- **Antisocial punishment:** in some cities, participants punished high contributors as well as free riders, more often in societies with weak norms of civic cooperation and weak rule of law.

What the experiments consistently find

- **Isaac and Walker (1988):** cooperation depends mainly on the return to contributing, and less on the size of the group.
- **Chaudhuri (2011):** beyond punishment, expressing disapproval of free riders, advice to new participants, and grouping cooperators together sustain contributions.

CHAPTER 18

Key takeaways (1 of 3)

1. **Externalities:** a negative externality makes social cost exceed private cost, so the market overproduces; a positive one makes social benefit exceed private benefit, so it produces too little; both create a deadweight loss.

2. **Correcting them:** a Pigouvian tax equal to the external harm, or a subsidy for positive externalities. A tax fixes the price of pollution, cap-and-trade the quantity; both cut emissions where it is least costly.

3. **Coase theorem:** with clear property rights, equally informed parties, and low transaction costs, bargaining reaches the same efficient outcome whoever holds the right; with unclear property rights, private information, or high transaction costs (for instance, too many parties), it does not.

Key takeaways (2 of 3)

4. **Public goods:** nonrival and nonexcludable, so markets undersupply them (the free-rider problem); the lighthouses of England and Wales show that excludability depends on institutions.

5. **Common resources:** rival but nonexcludable, so overused; property rights, direct regulation, and community self-governance (Ostrom, 1990) make users bear the cost of their use.

CHAPTER 18

Key takeaways (3 of 3)

6. **Acid rain:** the 1990 cap-and-trade program cut sulfur dioxide emissions from regulated sources by more than 40 percent by 2007, at far lower cost than command-and-control.

7. **Experiments:** contributions start at 40 to 60 percent of the money and erode to nothing; costly punishment of free riders keeps cooperation high (Fehr and Gächter, 2000).