

PRINCIPLES OF MICROECONOMICS: SCARCITY AND SOCIAL PROVISIONING

PRINCIPLES OF MICROECONOMICS: SCARCITY AND SOCIAL PROVISIONING

Erik Dean, Justin Elardo, Mitch Green, Benjamin Wilson, and Sebastian
Berger



Principles of Microeconomics: Scarcity and Social Provisioning by Erik Dean, Justin Elardo, Mitch Green, Benjamin Wilson, Sebastian Berger is licensed under a Creative Commons Attribution 4.0 International License, except where otherwise noted.

This book is an adaptation of Principles of Economics by OpenStax, licensed under a Creative Commons Attribution 4.0 International License.

CONTENTS

<i>Preface</i>	xxi
<i>About OpenStax</i>	xxi
<i>About OpenStax's Resources</i>	xxi
<i>About Principles of Economics</i>	xxii
<i>About Our Team</i>	xxvii
<i>Ancillaries</i>	xxxix
<i>Copyright</i>	xxxix

CHAPTER 1. WELCOME TO ECONOMICS!

Introduction	2
1.1 What Is Economics, and Why Is It Important?	3
<i>The Problem of Scarcity</i>	3
<i>The Division of and Specialization of Labor</i>	5
<i>Why the Division of Labor Increases Production</i>	6
<i>Trade and Markets</i>	7
<i>Why Study Economics?</i>	7
<i>Key Concepts and Summary</i>	8
<i>References</i>	9
1.2 Microeconomics and Macroeconomics	10
<i>Microeconomics</i>	11
<i>Macroeconomics</i>	11
<i>Key Concepts and Summary</i>	11
1.3 How Economists Use Theories and Models to Understand Economic Issues	13
<i>Key Concepts and Summary</i>	15
1.4 How Economies Can Be Organized: An Overview of Economic Systems	18
<i>Regulations: The Rules of the Game</i>	21
<i>The Rise of Globalization</i>	21
<i>Key Concepts and Summary</i>	23
<i>References</i>	24

CHAPTER 2. CHOICE IN A WORLD OF SCARCITY

Introduction to Choice in a World of Scarcity	28
---	----

2.1 How Individuals Make Choices Based on Their Budget Constraint	30
<i>The Concept of Opportunity Cost</i>	31
<i>Identifying Opportunity Cost</i>	33
<i>Marginal Decision-Making and Diminishing Marginal Utility</i>	34
<i>Sunk Costs</i>	35
<i>From a Model with Two Goods to One of Many Goods</i>	35
<i>Key Concepts and Summary</i>	35
<i>References</i>	37
2.2 The Production Possibilities Frontier and Social Choices	39
<i>The Shape of the PPF and the Law of Diminishing Returns</i>	41
<i>Productive Efficiency and Allocative Efficiency</i>	42
<i>Why Society Must Choose</i>	43
<i>The PPF and Comparative Advantage</i>	43
<i>Key Concepts and Summary</i>	45
2.3 Confronting Objections to the Economic Approach	47
<i>First Objection: People, Firms, and Society Do Not Act Like This</i>	47
<i>Second Objection: People, Firms, and Society Should Not Act This Way</i>	48
<i>Key Concepts and Summary</i>	51
<i>References</i>	51
CHAPTER 3: DEFINING ECONOMICS: A PLURALISTIC APPROACH	
Defining Economics: A Pluralistic Approach	54
3.1 The Importance of Definitions	55
3.2 Multiple Perspectives Require Multiple Definitions	56
3.3 A Brief Synopsis of Different Economic Perspectives	57
3.4 Deconstructing the Orthodox Definition of Economics	60
3.5 A Critical Examination of the Orthodox Definition of Economics and its Resultant Impacts	62
<i>Glossary</i>	72
3.6 An Alternative Approach to Defining Economics	73
CHAPTER 4. DEMAND AND SUPPLY	
Introduction to Demand and Supply	76
4.1 Demand, Supply, and Equilibrium in Markets for Goods and Services	79
<i>Demand for Goods and Services</i>	79
<i>Supply of Goods and Services</i>	81
<i>Equilibrium—Where Demand and Supply Intersect</i>	82
<i>Key Concepts and Summary</i>	84
<i>References</i>	86

4.2 Shifts in Demand and Supply for Goods and Services	88
<i>What Factors Affect Demand?</i>	88
<i>The Ceteris Paribus Assumption</i>	89
<i>How Does Income Affect Demand?</i>	89
<i>Other Factors That Shift Demand Curves</i>	91
<i>Summing Up Factors That Change Demand</i>	93
<i>How Production Costs Affect Supply</i>	94
<i>Other Factors That Affect Supply</i>	96
<i>Summing Up Factors That Change Supply</i>	98
<i>Key Concepts and Summary</i>	99
<i>References</i>	101
4.3 Changes in Equilibrium Price and Quantity: The Four-Step Process	104
<i>Good Weather for Salmon Fishing</i>	105
<i>Newspapers and the Internet</i>	106
<i>The Interconnections and Speed of Adjustment in Real Markets</i>	107
<i>A Combined Example</i>	107
<i>Key Concepts and Summary</i>	111
<i>References</i>	113
4.4 Price Ceilings and Price Floors	114
<i>Price Ceilings</i>	114
<i>Price Floors</i>	116
<i>Key Concepts and Summary</i>	117
4.5 Demand, Supply, and Efficiency	120
<i>Consumer Surplus, Producer Surplus, Social Surplus</i>	120
<i>Inefficiency of Price Floors and Price Ceilings</i>	122
<i>Demand and Supply as a Social Adjustment Mechanism</i>	123
<i>Key Concepts and Summary</i>	124
 CHAPTER 5. LABOR AND FINANCIAL MARKETS	
Introduction to Labor and Financial Markets	128
5.1 Demand and Supply at Work in Labor Markets	130
<i>Equilibrium in the Labor Market</i>	130
<i>Shifts in Labor Demand</i>	132
<i>Shifts in Labor Supply</i>	133
<i>Technology and Wage Inequality: The Four-Step Process</i>	134
<i>Price Floors in the Labor Market: Living Wages and Minimum Wages</i>	136
<i>The Minimum Wage as an Example of a Price Floor</i>	137
<i>Key Concepts and Summary</i>	139
<i>References</i>	140

5.2 Demand and Supply in Financial Markets	142
<i>Who Demands and Who Supplies in Financial Markets?</i>	142
<i>Equilibrium in Financial Markets</i>	144
<i>Shifts in Demand and Supply in Financial Markets</i>	144
<i>The United States as a Global Borrower</i>	145
<i>Price Ceilings in Financial Markets: Usury Laws</i>	147
<i>Key Concepts and Summary</i>	148
<i>References</i>	150
5.3 The Market System as an Efficient Mechanism for Information	152
<i>Key Concepts and Summary</i>	154

CHAPTER 6. ELASTICITY

Introduction to Elasticity	162
6.1 Price Elasticity of Demand and Price Elasticity of Supply	165
<i>Calculating Price Elasticity of Demand</i>	166
<i>Calculating the Price Elasticity of Supply</i>	169
<i>Key Concepts and Summary</i>	171
6.2 Polar Cases of Elasticity and Constant Elasticity	178
<i>Key Concepts and Summary</i>	181
6.3 Elasticity and Pricing	183
<i>Does Raising Price Bring in More Revenue?</i>	185
<i>Can Costs Be Passed on to Consumers?</i>	186
<i>Elasticity and Tax Incidence</i>	188
<i>Long-Run vs. Short-Run Impact</i>	190
<i>Key Concepts and Summary</i>	191
6.4 Elasticity in Areas Other Than Price	194
<i>Income Elasticity of Demand</i>	194
<i>Cross-Price Elasticity of Demand</i>	195
<i>Elasticity in Labor and Financial Capital Markets</i>	195
<i>Expanding the Concept of Elasticity</i>	196
<i>Key Concepts and Summary</i>	198
<i>References</i>	199

CHAPTER 7. CONSUMER CHOICES

Introduction to Consumer Choices	204
7.1 Consumption Choices	206
<i>Total Utility and Diminishing Marginal Utility</i>	207
<i>Choosing with Marginal Utility</i>	209
<i>A Rule for Maximizing Utility</i>	211
<i>Measuring Utility with Numbers</i>	212
<i>Key Concepts and Summary</i>	213
<i>References</i>	215

7.2 How Changes in Income and Prices Affect Consumption Choices	218
<i>How Changes in Income Affect Consumer Choices</i>	218
<i>How Price Changes Affect Consumer Choices</i>	220
<i>The Foundations of Demand Curves</i>	221
<i>Applications in Government and Business</i>	222
<i>Key Concepts and Summary</i>	224
7.3 Labor-Leisure Choices	226
<i>The Labor-Leisure Budget Constraint</i>	227
<i>Applications of Utility Maximizing with the Labor-Leisure Budget Constraint</i>	229
<i>Key Concepts and Summary</i>	231
7.4 Intertemporal Choices in Financial Capital Markets	233
<i>Using Marginal Utility to Make Intertemporal Choices</i>	234
<i>Applications of the Model of Intertemporal Choice</i>	237
<i>The Unifying Power of the Utility-Maximizing Budget Set Framework</i>	238
<i>Behavioral Economics: An Alternative Viewpoint</i>	238
<i>Key Concepts and Summary</i>	241
<i>References</i>	241
 CHAPTER 8. CHALLENGING THE ROLE OF UTILITARIANISM	
The Role of Value(s) in the Economics Discipline	244
8.1 Economics and Value	245
8.2 Utilitarianism: The Philosophy Behind Orthodox Economics	247
8.3 Utility and Pareto Optimality: The Orthodox Economic View of Social Welfare	251
8.4 Abandoning the Normative Constraints of Utilitarianism	255
<i>Glossary</i>	260
8.5 Conclusion	261
 CHAPTER 9. AN INSTITUTIONAL ANALYSIS OF MODERN CONSUMPTION	
Introduction to An Institutional Analysis of Modern Consumption	264
9.1 Institutional Analysis	267
9.2 Conspicuous Consumption	269
9.3 The Complex World of Modern Consumption	272
9.4 Conclusion	275
 CHAPTER 10. COST AND INDUSTRY STRUCTURE	
Introduction to Cost and Industry Structure	278
10.1 Explicit and Implicit Costs, and Accounting and Economic Profit	281
<i>Key Concepts and Summary</i>	283
<i>References</i>	284

10.2 The Structure of Costs in the Short Run	285
<i>Fixed and Variable Costs</i>	285
<i>Average Total Cost, Average Variable Cost, Marginal Cost</i>	287
<i>Lessons from Alternative Measures of Costs</i>	289
<i>A Variety of Cost Patterns</i>	290
<i>Key Concepts and Summary</i>	291
10.3 The Structure of Costs in the Long Run	295
<i>Choice of Production Technology</i>	295
<i>Economies of Scale</i>	297
<i>Shapes of Long-Run Average Cost Curves</i>	298
<i>The Size and Number of Firms in an Industry</i>	300
<i>Shifting Patterns of Long-Run Average Cost</i>	303
<i>Key Concepts and Summary</i>	304
<i>Problems</i>	305
 CHAPTER 11. PERFECT COMPETITION	
Introduction to Perfect Competition	308
11.1 Perfect Competition and Why It Matters	310
<i>Key Concepts and Summary</i>	311
11.2 How Perfectly Competitive Firms Make Output Decisions	313
<i>Determining the Highest Profit by Comparing Total Revenue and Total Cost</i>	313
<i>Comparing Marginal Revenue and Marginal Costs</i>	315
<i>Profits and Losses with the Average Cost Curve</i>	318
<i>The Shutdown Point</i>	322
<i>Short-Run Outcomes for Perfectly Competitive Firms</i>	325
<i>Marginal Cost and the Firm's Supply Curve</i>	326
<i>Key Concepts and Summary</i>	329
11.3 Entry and Exit Decisions in the Long Run	333
<i>How Entry and Exit Lead to Zero Profits in the Long Run</i>	334
<i>The Long-Run Adjustment and Industry Types</i>	335
<i>Key Concepts and Summary</i>	335
11.4 Efficiency in Perfectly Competitive Markets	338
<i>Key Concepts and Summary</i>	340
<i>References</i>	340
 CHAPTER 12. MONOPOLY	
Introduction to a Monopoly	344

12.1 How Monopolies Form: Barriers to Entry	347
<i>Natural Monopoly</i>	347
<i>Control of a Physical Resource</i>	349
<i>Legal Monopoly</i>	349
<i>Promoting Innovation</i>	349
<i>Intimidating Potential Competition</i>	351
<i>Summing Up Barriers to Entry</i>	351
<i>Key Concepts and Summary</i>	352
12.2 How a Profit-Maximizing Monopoly Chooses Output and Price	355
<i>Demand Curves Perceived by a Perfectly Competitive Firm and by a Monopoly</i>	355
<i>Total Cost and Total Revenue for a Monopolist</i>	357
<i>Marginal Revenue and Marginal Cost for a Monopolist</i>	358
<i>Illustrating Monopoly Profits</i>	361
<i>The Inefficiency of Monopoly</i>	364
<i>Key Concepts and Summary</i>	365
<i>References</i>	367

CHAPTER 13. MONOPOLISTIC COMPETITION AND OLIGOPOLY

Introduction to Monopolistic Competition and Oligopoly	370
13.1 Monopolistic Competition	373
<i>Differentiated Products</i>	373
<i>Perceived Demand for a Monopolistic Competitor</i>	374
<i>How a Monopolistic Competitor Chooses Price and Quantity</i>	375
<i>Monopolistic Competitors and Entry</i>	377
<i>Monopolistic Competition and Efficiency</i>	379
<i>The Benefits of Variety and Product Differentiation</i>	380
<i>Key Concepts and Summary</i>	381
<i>References</i>	383
13.2 Oligopoly	385
<i>Why Do Oligopolies Exist?</i>	385
<i>Collusion or Competition?</i>	386
<i>The Prisoner's Dilemma</i>	387
<i>The Oligopoly Version of the Prisoner's Dilemma</i>	388
<i>How to Enforce Cooperation</i>	389
<i>Tradeoffs of Imperfect Competition</i>	390
<i>Key Concepts and Summary</i>	392
<i>References</i>	394

CHAPTER 14. THE RISE OF BIG BUSINESS

Introduction to the Rise of Big Business	398
<i>The Joint-Stock Corporation and Long Distance Trade</i>	399
<i>Mercantilism</i>	400

14.1 Big Business in American History	402
<i>Business Enterprise and Property Rights</i>	402
<i>Colonial legal institutions</i>	403
<i>The Supreme Court</i>	406
<i>Canals, Steamboats and Railroads</i>	407
<i>Glossary</i>	410
14.2 Industrialization and the Factory	412
<i>Large-scale technologies that make up the core of the economic system</i>	412
<i>Integrated chains of production that link markets and industries</i>	412
14.3 Big Business and Organized Labor	415
<i>The Knights of Labor</i>	415
<i>The Great Railway Strike of 1877</i>	416
14.4 Regulation of Big Business	418
<i>The Sherman Act</i>	419
14.5 Conclusion	420
 CHAPTER 15. COSTS AND PRICES: THE EVIDENCE	
Introduction to Costs and Prices	422
15.1 Testing the Neoclassical Theory of the Firm	423
<i>Testing the Theory</i>	424
<i>Designing the Average Total Cost Curve</i>	424
15.2 Costing and Pricing: A Heterodox Alternative	428
<i>Depreciation and the Going Concern</i>	428
<i>Prices from Pricing</i>	429
15.3 Comparing Neoclassical and Heterodox Theory	433
 CHAPTER 16. THE MEGACORP	
Introduction to the Megacorp	436
16.1 The Imperatives of Technology	439
16.2 Business Models, Plural: Aims and Methods of the Megacorp	443
16.3 Stabilizing Unstable Markets	448
16.4 The High Price of College Textbooks	450
 CHAPTER 17. MONOPOLY AND ANTITRUST POLICY	
Introduction to Monopoly and Antitrust Policy	454

17.1 Corporate Mergers	457
<i>Regulations for Approving Mergers</i>	457
<i>The Four-Firm Concentration Ratio</i>	459
<i>The Herfindahl-Hirshman Index</i>	460
<i>New Directions for Antitrust</i>	461
<i>Key Concepts and Summary</i>	463
17.2 Regulating Anticompetitive Behavior	466
<i>Restrictive Practices</i>	466
<i>Key Concepts and Summary</i>	468
17.3 Regulating Natural Monopolies	470
<i>The Choices in Regulating a Natural Monopoly</i>	470
<i>Cost-Plus versus Price Cap Regulation</i>	473
<i>Key Concepts and Summary</i>	473
17.4 The Great Deregulation Experiment	476
<i>Doubts about Regulation of Prices and Quantities</i>	476
<i>The Effects of Deregulation</i>	477
<i>Key Concepts and Summary</i>	478
<i>References</i>	479
 CHAPTER 18. ENVIRONMENTAL PROTECTION AND NEGATIVE EXTERNALITIES	
 Introduction to Environmental Protection and Negative Externalities	482
18.1 The Economics of Pollution	484
<i>Externalities</i>	485
<i>Pollution as a Negative Externality</i>	485
<i>Key Concepts and Summary</i>	488
<i>References</i>	490
18.2 Command-and-Control Regulation	492
<i>Key Concepts and Summary</i>	493
18.3 Market-Oriented Environmental Tools	495
<i>Pollution Charges</i>	495
<i>Marketable Permits</i>	497
<i>Better-Defined Property Rights</i>	498
<i>Applying Market-Oriented Environmental Tools</i>	499
<i>Key Concepts and Summary</i>	500
<i>Self-Check Questions</i>	500
<i>References</i>	501
18.4 The Benefits and Costs of U.S. Environmental Laws	503
<i>Benefits and Costs of Clean Air and Clean Water</i>	503
<i>Ecotourism: Making Environmentalism Pay</i>	504
<i>Marginal Benefits and Marginal Costs</i>	504
<i>Key Concepts and Summary</i>	506
<i>References</i>	508

18.5 International Environmental Issues	510
<i>Key Concepts and Summary</i>	511
<i>References</i>	512
18.6 The Tradeoff between Economic Output and Environmental Protection	514
<i>Key Concepts and Summary</i>	515
CHAPTER 19. POSITIVE EXTERNALITIES AND PUBLIC GOODS	
Introduction to Positive Externalities and Public Goods	520
19.1 Why the Private Sector Under Invests in Innovation	523
<i>The Positive Externalities of New Technology</i>	524
<i>Why Invest in Human Capital?</i>	525
<i>Other Examples of Positive Externalities</i>	526
<i>Key Concepts and Summary</i>	527
<i>References</i>	530
19.2 How Governments Can Encourage Innovation	532
<i>Intellectual Property Rights</i>	532
<i>Policy #1: Government Spending on Research and Development</i>	534
<i>Policy #2: Tax Breaks for Research and Development</i>	535
<i>Policy #3 Cooperative Research</i>	535
<i>Key Concepts and Summary</i>	535
<i>References</i>	536
19.3 Public Goods	538
<i>The Definition of a Public Good</i>	538
<i>The Free Rider Problem of Public Goods</i>	539
<i>The Role of Government in Paying for Public Goods</i>	540
<i>Common Resources and the “Tragedy of the Commons”</i>	541
<i>Positive Externalities in Public Health Programs</i>	542
<i>Key Concepts and Summary</i>	543
<i>References</i>	544
CHAPTER 20. POVERTY AND ECONOMIC INEQUALITY	
Introduction to Poverty and Economic Inequality	548
20.1 Drawing the Poverty Line	550
<i>Key Concepts and Summary</i>	553
<i>References</i>	554
20.2 The Poverty Trap	556
<i>Key Concepts and Summary</i>	560

20.3 The Safety Net	564
<i>Temporary Assistance for Needy Families</i>	564
<i>The Earned Income Tax Credit (EITC)</i>	565
<i>Supplemental Nutrition Assistance Program (SNAP)</i>	566
<i>Medicaid</i>	567
<i>Key Concepts and Summary</i>	569
<i>References</i>	570
20.4 Income Inequality: Measurement and Causes	572
<i>Measuring Income Distribution by Quintiles</i>	573
<i>Lorenz Curve</i>	573
<i>Causes of Growing Inequality: The Changing Composition of American Households</i>	576
<i>Causes of Growing Inequality: A Shift in the Distribution of Wages</i>	576
<i>Key Concepts and Summary</i>	578
<i>References</i>	580
20.5 Government Policies to Reduce Income Inequality	584
<i>Redistribution</i>	585
<i>The Ladder of Opportunity</i>	585
<i>Inheritance Taxes</i>	586
<i>The Tradeoff between Incentives and Income Equality</i>	586
<i>Key Concepts and Summary</i>	588
<i>References</i>	589
 CHAPTER 21. ISSUES IN LABOR MARKETS: UNIONS, DISCRIMINATION, IMMIGRATION	
Introduction to Issues in Labor Markets: Unions, Discrimination, Immigration	592
21.1 Unions	595
<i>Facts about Union Membership and Pay</i>	595
<i>Higher Wages for Union Workers</i>	596
<i>The Decline in U.S. Union Membership</i>	599
<i>Key Concepts and Summary</i>	602
<i>References</i>	604
21.2 Employment Discrimination	606
<i>Earnings Gaps by Race and Gender</i>	606
<i>Investigating the Female/Male Earnings Gap</i>	607
<i>Investigating the Black/White Earnings Gap</i>	608
<i>Competitive Markets and Discrimination</i>	609
<i>Public Policies to Reduce Discrimination</i>	610
<i>An Increasingly Diverse Workforce</i>	611
<i>Key Concepts and Summary</i>	611
<i>References</i>	613

21.3 Immigration	616
<i>Historical Patterns of Immigration</i>	616
<i>Economic Effects of Immigration</i>	617
<i>Proposals for Immigration Reform</i>	618
<i>Key Concepts and Summary</i>	619
<i>References</i>	620

CHAPTER 22. INFORMATION, RISK, AND INSURANCE

Introduction to Information, Risk, and Insurance	622
22.1 The Problem of Imperfect Information and Asymmetric Information	624
<i>“Lemons” and Other Examples of Imperfect Information</i>	625
<i>How Imperfect Information Can Affect Equilibrium Price and Quantity</i>	626
<i>When Price Mixes with Imperfect Information about Quality</i>	626
<i>Mechanisms to Reduce the Risk of Imperfect Information</i>	627
<i>Key Concepts and Summary</i>	630
<i>References</i>	631
22.2 Insurance and Imperfect Information	633
<i>How Insurance Works</i>	634
<i>Government and Social Insurance</i>	635
<i>Risk Groups and Actuarial Fairness</i>	635
<i>The Moral Hazard Problem</i>	636
<i>The Adverse Selection Problem</i>	637
<i>U.S. Health Care in an International Context</i>	638
<i>Government Regulation of Insurance</i>	639
<i>The Patient Protection and Affordable Care Act</i>	641
<i>Key Concepts and Summary</i>	642
<i>References</i>	644

CHAPTER 23. FINANCIAL MARKETS

Introduction to Financial Markets	648
23.1 How Businesses Raise Financial Capital	651
<i>Early Stage Financial Capital</i>	651
<i>Profits as a Source of Financial Capital</i>	652
<i>Borrowing: Banks and Bonds</i>	652
<i>Corporate Stock and Public Firms</i>	653
<i>How Firms Choose between Sources of Financial Capital</i>	654
<i>Key Concepts and Summary</i>	655
<i>References</i>	657

23.2 How Households Supply Financial Capital	659
<i>Expected Rate of Return, Risk, and Actual Rate of Return</i>	659
<i>Bank Accounts</i>	660
<i>Bonds</i>	662
<i>Stocks</i>	664
<i>Mutual Funds</i>	667
<i>Housing and Other Tangible Assets</i>	668
<i>The Tradeoffs between Return and Risk</i>	671
<i>Key Concepts and Summary</i>	672
<i>References</i>	674
23.3 How to Accumulate Personal Wealth	676
<i>Why It Is Hard to Get Rich Quick: The Random Walk Theory</i>	676
<i>Getting Rich the Slow, Boring Way</i>	677
<i>How Capital Markets Transform Financial Flows</i>	679
<i>Key Concepts and Summary</i>	680
<i>References</i>	681

CHAPTER 24. PUBLIC ECONOMY

Introduction to Public Economy	684
24.1 Voter Participation and Costs of Elections	686
<i>Key Concepts and Summary</i>	688
<i>References</i>	689
24.2 Special Interest Politics	691
<i>Identifiable Winners, Anonymous Losers</i>	693
<i>Pork Barrels and Logrolling</i>	693
<i>Key Concepts and Summary</i>	694
<i>References</i>	696
24.3 Flaws in the Democratic System of Government	697
<i>Where Is Government's Self-Correcting Mechanism?</i>	699
<i>A Balanced View of Markets and Government</i>	700
<i>Key Concepts and Summary</i>	701
<i>References</i>	703

CHAPTER 25. MONEY AND THE THEORY OF THE FIRM

Introduction to Money and the Theory of the Firm	706
25.1 The Metallist and the Barter Myth	709
<i>Endnotes</i>	711
<i>Glossary</i>	712
25.2 Smith, Marx, Keynes, Chartalism and Modern Money Theory	713
<i>Endnotes</i>	717
<i>Glossary</i>	717

25.3 The Money Hierarchy and the False Duality of the State and Market	718
<i>Endnotes</i>	722
25.4 Local Currency Systems: Social Money and Community Currencies	723
<i>Endnotes</i>	724
CHAPTER 26. INTERNATIONAL TRADE	
Introduction to International Trade	728
26.1 Absolute and Comparative Advantage	730
<i>A Numerical Example of Absolute and Comparative Advantage</i>	731
<i>Gains from Trade</i>	733
<i>Key Concepts and Summary</i>	736
<i>References</i>	737
26.2 What Happens When a Country Has an Absolute Advantage in All Goods	739
<i>Production Possibilities and Comparative Advantage</i>	739
<i>Mutually Beneficial Trade with Comparative Advantage</i>	740
<i>How Opportunity Cost Sets the Boundaries of Trade</i>	742
<i>Comparative Advantage Goes Camping</i>	743
<i>Key Concepts and Summary</i>	744
<i>References</i>	746
26.3 Intra-industry Trade between Similar Economies	747
<i>The Prevalence of Intra-industry Trade between Similar Economies</i>	747
<i>Gains from Specialization and Learning</i>	748
<i>Economies of Scale, Competition, Variety</i>	749
<i>Dynamic Comparative Advantage</i>	750
<i>Key Concepts and Summary</i>	750
<i>References</i>	752
26.4 The Benefits of Reducing Barriers to International Trade	754
<i>From Interpersonal to International Trade</i>	755
<i>Key Concepts and Summary</i>	756
<i>References</i>	757
CHAPTER 27. GLOBALIZATION AND PROTECTIONISM	
Introduction to Globalization and Protectionism	760
27.1 Protectionism: An Indirect Subsidy from Consumers to Producers	762
<i>Demand and Supply Analysis of Protectionism</i>	763
<i>Who Benefits and Who Pays?</i>	768
<i>Key Concepts and Summary</i>	769
<i>References</i>	770

27.2 International Trade and Its Effects on Jobs, Wages, and Working Conditions	772
<i>Fewer Jobs?</i>	772
<i>Trade and Wages</i>	774
<i>Labor Standards and Working Conditions</i>	775
<i>Key Concepts and Summary</i>	776
<i>References</i>	777
27.3 Arguments in Support of Restricting Imports	779
<i>The Infant Industry Argument</i>	779
<i>The Anti-Dumping Argument</i>	781
<i>The Environmental Protection Argument</i>	782
<i>The Unsafe Consumer Products Argument</i>	784
<i>The National Interest Argument</i>	784
<i>Key Concepts and Summary</i>	786
<i>References</i>	788
27.4 How Trade Policy Is Enacted: Globally, Regionally, and Nationally	790
<i>The World Trade Organization</i>	790
<i>Regional Trading Agreements</i>	791
<i>Trade Policy at the National Level</i>	792
<i>Long-Term Trends in Barriers to Trade</i>	793
<i>Key Concepts and Summary</i>	793
<i>References</i>	794
27.5 The Tradeoffs of Trade Policy	796
<i>Key Concepts and Summary</i>	797
<i>References</i>	799
 CHAPTER 28. THE ECONOMICS OF GLOBALIZATION AND TRADE: A PLURALISTIC APPROACH	
Introduction to Globalization and Trade from a Pluralistic Perspective	802
28.1 The Orthodox Story of Trade: A Synopsis	804
28.2 A Critical Examination of the Orthodox Depiction of Free Trade	807
28.3 Challenging Functionality: A More Penetrating Critique	813
<i>Glossary</i>	818
28.4 An Alternative Presentation of International Trade: Path Dependency	819

PREFACE

Principles of Economics is designed for a two-semester principles of economics sequence. It is traditional in coverage, including introductory economics content, microeconomics, macroeconomics and international economics. At the same time, the book includes a number of innovative and interactive features designed to enhance student learning. Instructors can also customize the book, adapting it to the approach that works best in their classroom.

Welcome to *Principles of Economics*, an OpenStax resource. This textbook has been created with several goals in mind: accessibility, customization, and student engagement—all while encouraging students toward high levels of academic scholarship. Instructors and students alike will find that this textbook offers a strong foundation in economics in an accessible format.

ABOUT OPENSTAX

OpenStax is a non-profit organization committed to improving student access to quality learning materials. Our free textbooks go through a rigorous editorial publishing process. Our texts are developed and peer-reviewed by educators to ensure they are readable, accurate, and meet the scope and sequence requirements of today's college courses. Unlike traditional textbooks, OpenStax resources live online and are owned by the community of educators using them. Through our partnerships with companies and foundations committed to reducing costs for students, OpenStax is working to improve access to higher education for all. OpenStax is an initiative of Rice University and is made possible through the generous support of several philanthropic foundations.

ABOUT OPENSTAX'S RESOURCES

OpenStax resources provide quality academic instruction. Three key features set our materials apart from others: they can be customized by instructors for each class, they are a “living” resource that grows online through contributions from science educators, and they are available free or for minimal cost.

CUSTOMIZATION

OpenStax learning resources are designed to be customized for each course. Our textbooks provide a solid foundation on which instructors can build, and our resources are conceived and written with flexibility in mind. Instructors can select the sections most relevant to their curricula and create a textbook that speaks directly to the needs of their classes and student body. Teachers are encouraged to expand on existing examples by adding unique context via geographically localized applications and topical connections.

Principles of Economics can be easily customized using our online platform (<http://cnx.org/content/col11613/>). Simply select the content most relevant to your current semester and create a textbook

that speaks directly to the needs of your class. *Principles of Economics* is organized as a collection of sections that can be rearranged, modified, and enhanced through localized examples or to incorporate a specific theme of your course. This customization feature will ensure that your textbook truly reflects the goals of your course. *Principles of Economics* is also available in two volumes, one covering microeconomics and one covering macroeconomics principles.

CURATION

To broaden access and encourage community curation, *Principles of Economics* is “open source” licensed under a Creative Commons Attribution (CC-BY) license. The economics community is invited to submit examples, emerging research, and other feedback to enhance and strengthen the material and keep it current and relevant for today’s students. You can submit your suggestions to info@openstaxcollege.org.

COST

Our textbooks are available for free online, and in low-cost print and e-book editions.

ABOUT PRINCIPLES OF ECONOMICS

Principles of Economics is designed for a two-semester principles of economics sequence. The text has been developed to meet the scope and sequence of most introductory courses. At the same time, the book includes a number of innovative features designed to enhance student learning. Instructors can also customize the book, adapting it to the approach that works best in their classroom.

COVERAGE AND SCOPE

To develop *Principles of Economics*, we acquired the rights to Timothy Taylor’s second edition of *Principles of Economics* and solicited ideas from economics instructors at all levels of higher education, from community colleges to Ph.D.-granting universities. They told us about their courses, students, challenges, resources, and how a textbook can best meet the needs of both instructors and students.

The result is a book that covers the breadth of economics topics and also provides the necessary depth to ensure the course is manageable for instructors and students alike. And to make it more applied, we have incorporated many current topics. We hope students will be interested to know just how far-reaching the recent recession was (and still is), for example, and why there is so much controversy even among economists over the Affordable Care Act (Obamacare). The Keystone Pipeline, Occupy Wall Street, minimum wage debates, and the appointment of the United States’ first female Federal Reserve chair, Janet Yellen, are just a few of the other important topics covered.

The pedagogical choices, chapter arrangements, and learning objective fulfillment were developed and vetted with feedback from educators dedicated to the project. They thoroughly read the material and offered critical and detailed commentary. The outcome is a balanced approach to micro and macro economics, to both Keynesian and classical views, and to the theory and application of economics concepts. New 2015 data are incorporated for topics that range from average U.S. household consumption in Chapter 2 to the total value of all home equity in Chapter 17. Current events are treated in a politically-balanced way as well.

The book is organized into eight main parts:

- **What is Economics?** The first two chapters introduce students to the study of economics with a focus on making choices in a world of scarce resources.
- **Supply and Demand**, Chapters 3 and 4, introduces and explains the first analytical model in economics: supply, demand, and equilibrium, before showing applications in the markets for labor and finance.
- **The Fundamentals of Microeconomic Theory**, Chapters 5 through 10, begins the microeconomics portion of the text, presenting the theories of consumer behavior, production and costs, and the different models of market structure, including some simple game theory.
- **Microeconomic Policy Issues**, Chapters 11 through 18, cover the range of topics in applied micro, framed around the concepts of public goods and positive and negative externalities. Students explore competition and antitrust policies, environmental problems, poverty, income inequality, and other labor market issues. The text also covers information, risk and financial markets, as well as public economy.
- **The Macroeconomic Perspective and Goals**, Chapters 19 through 23, introduces a number of key concepts in macro: economic growth, unemployment and inflation, and international trade and capital flows.
- **A Framework for Macroeconomic Analysis**, Chapters 24 through 26, introduces the principal analytic model in macro, namely the Aggregate Demand/Aggregate Supply Model. The model is then applied to the Keynesian and Neoclassical perspectives. The Expenditure-Output model is fully explained in a stand-alone appendix.
- **Monetary and Fiscal Policy**, Chapters 27 through 31, explains the role of money and the banking system, as well as monetary policy and financial regulation. Then the discussion switches to government deficits and fiscal policy.
- **International Economics**, Chapters 32 through 34, the final part of the text, introduces the international dimensions of economics, including international trade and protectionism.

Chapter 1 Welcome to Economics!

Chapter 2 Choice in a World of Scarcity

Chapter 3 Demand and Supply

Chapter 4 Labor and Financial Markets

Chapter 5 Elasticity

Chapter 6 Consumer Choices

Chapter 7 Cost and Industry Structure

Chapter 8 Perfect Competition

Chapter 9 Monopoly

Chapter 10 Monopolistic Competition and Oligopoly

Chapter 11 Monopoly and Antitrust Policy

Chapter 12 Environmental Protection and Negative Externalities

Chapter 13 Positive Externalities and Public Goods

Chapter 14 Poverty and Economic Inequality

Chapter 15 Issues in Labor Markets: Unions, Discrimination, Immigration

Chapter 16 Information, Risk, and Insurance

Chapter 17 Financial Markets

Chapter 18 Public Economy

Chapter 19 The Macroeconomic Perspective

Chapter 20 Economic Growth

Chapter 21 Unemployment

Chapter 22 Inflation

Chapter 23 The International Trade and Capital Flows

Chapter 24 The Aggregate Demand/Aggregate Supply Model

Chapter 25 The Keynesian Perspective

Chapter 26 The Neoclassical Perspective

Chapter 27 Money and Banking

Chapter 28 Monetary Policy and Bank Regulation

Chapter 29 Exchange Rates and International Capital Flows

Chapter 30 Government Budgets and Fiscal Policy

Chapter 31 The Impacts of Government Borrowing

Chapter 32 Macroeconomic Policy Around the World

Chapter 33 International Trade

Chapter 34 Globalization and Protectionism

Appendix A The Use of Mathematics in Principles of Economics

Appendix B Indifference Curves

Appendix C Present Discounted Value

Appendix D The Expenditure-Output Model

Alternate Sequencing

Principles of Economics was conceived and written to fit a particular topical sequence, but it can be used flexibly to accommodate other course structures. One such potential structure, which will fit reasonably well with the textbook content, is provided. Please consider, however, that the chapters were not written to be completely independent, and that the proposed alternate sequence should be carefully considered for student preparation and textual consistency.

Chapter 1 Welcome to Economics!

Chapter 2 Choice in a World of Scarcity

Chapter 3 Demand and Supply

Chapter 4 Labor and Financial Markets

Chapter 5 Elasticity

Chapter 6 Consumer Choices

Chapter 33 International Trade

Chapter 7 Cost and Industry Structure

Chapter 12 Environmental Protection and Negative Externalities

Chapter 13 Positive Externalities and Public Goods

Chapter 8 Perfect Competition

Chapter 9 Monopoly

Chapter 10 Monopolistic Competition and Oligopoly

Chapter 11 Monopoly and Antitrust Policy

Chapter 14 Poverty and Economic Inequality

Chapter 15 Issues in Labor Markets: Unions, Discrimination, Immigration

Chapter 16 Information, Risk, and Insurance

Chapter 17 Financial Markets

Chapter 18 Public Economy

Chapter 19 The Macroeconomic Perspective

Chapter 20 Economic Growth

Chapter 21 Unemployment

Chapter 22 Inflation

Chapter 23 The International Trade and Capital Flows

Chapter 24 The Aggregate Demand/Aggregate Supply Model

Chapter 25 The Keynesian Perspective

Chapter 26 The Neoclassical Perspective

Chapter 27 Money and Banking

Chapter 28 Monetary Policy and Bank Regulation

Chapter 29 Exchange Rates and International Capital Flows

Chapter 30 Government Budgets and Fiscal Policy

Chapter 31 The Impacts of Government Borrowing

Chapter 32 Macroeconomic Policy Around the World

Chapter 34 Globalization and Protectionism

Appendix A The Use of Mathematics in Principles of Economics

Appendix B Indifference Curves

Appendix C Present Discounted Value

Appendix D The Expenditure-Output Model

PEDAGOGICAL FOUNDATION

Throughout the OpenStax version of *Principles of Economics*, you will find new features that engage the students in economic inquiry by taking selected topics a step further. Our features include:

- **Bring It Home:** This added feature is a brief case study, specific to each chapter, which connects the chapter's main topic to the real world. It is broken up into two parts: the first at the beginning of the chapter (in the Intro module) and the second at chapter's end, when students have learned what's necessary to understand the case and "bring home" the chapter's core concepts.
- **Work It Out:** This added feature asks students to work through a generally analytical or computational problem, and guides them step-by-step to find out how its solution is derived.

- **Clear It Up:** This boxed feature, which includes pre-existing features from Taylor’s text, addresses common student misconceptions about the content. Clear It Ups are usually deeper explanations of something in the main body of the text. Each CIU starts with a question. The rest of the feature explains the answer.
- **Link It Up:** This added feature is a very brief introduction to a website that is pertinent to students’ understanding and enjoyment of the topic at hand.

QUESTIONS FOR EACH LEVEL OF LEARNING

The OpenStax version of *Principles of Economics* further expands on Taylor’s original end of chapter materials by offering four types of end-of-module questions for students.

- **Self-Checks:** Are analytical self-assessment questions that appear at the end of each module. They “click-to-reveal” an answer in the web view so students can check their understanding before moving on to the next module. Self-Check questions are not simple look-up questions. They push the student to think a bit beyond what is said in the text. Self-Check questions are designed for formative (rather than summative) assessment. The questions and answers are explained so that students feel like they are being walked through the problem.
- **Review Questions:** Have been retained from Taylor’s version, and are simple recall questions from the chapter and are in open-response format (not multiple choice or true/false). The answers can be looked up in the text.
- **Critical Thinking Questions:** Are new higher-level, conceptual questions that ask students to *demonstrate their understanding by applying* what they have learned in different contexts. They ask for outside-the-box thinking, for *reasoning* about the concepts. They push the student to places they wouldn’t have thought of going themselves.
- **Problems:** Are exercises that give students additional practice working with the analytic and computational concepts in the module.

UPDATED ART

Principles of Economics includes an updated art program to better inform today’s student, providing the latest data on covered topics.

ABOUT OUR TEAM

SENIOR CONTRIBUTING AUTHOR

Timothy Taylor, Macalester College

Timothy Taylor has been writing and teaching about economics for 30 years, and is the Managing Editor of the *Journal of Economic Perspectives*, a post he’s held since 1986. He has been a lecturer for The Teaching Company, the University of Minnesota, and the Hubert H. Humphrey Institute of Public Affairs, where students voted him Teacher of the Year in 1997. His writings include numerous pieces for journals such as the *Milken Institute Review* and *The Public Interest*, and he has been an editor on many projects, most notably for the Brookings Institution and the World Bank, where he was Chief Outside Editor for the *World Development Report 1999/2000, Entering the 21st Century: The Changing Development Landscape*. He also blogs four to five times per week at <http://conversableeconomics.com>.

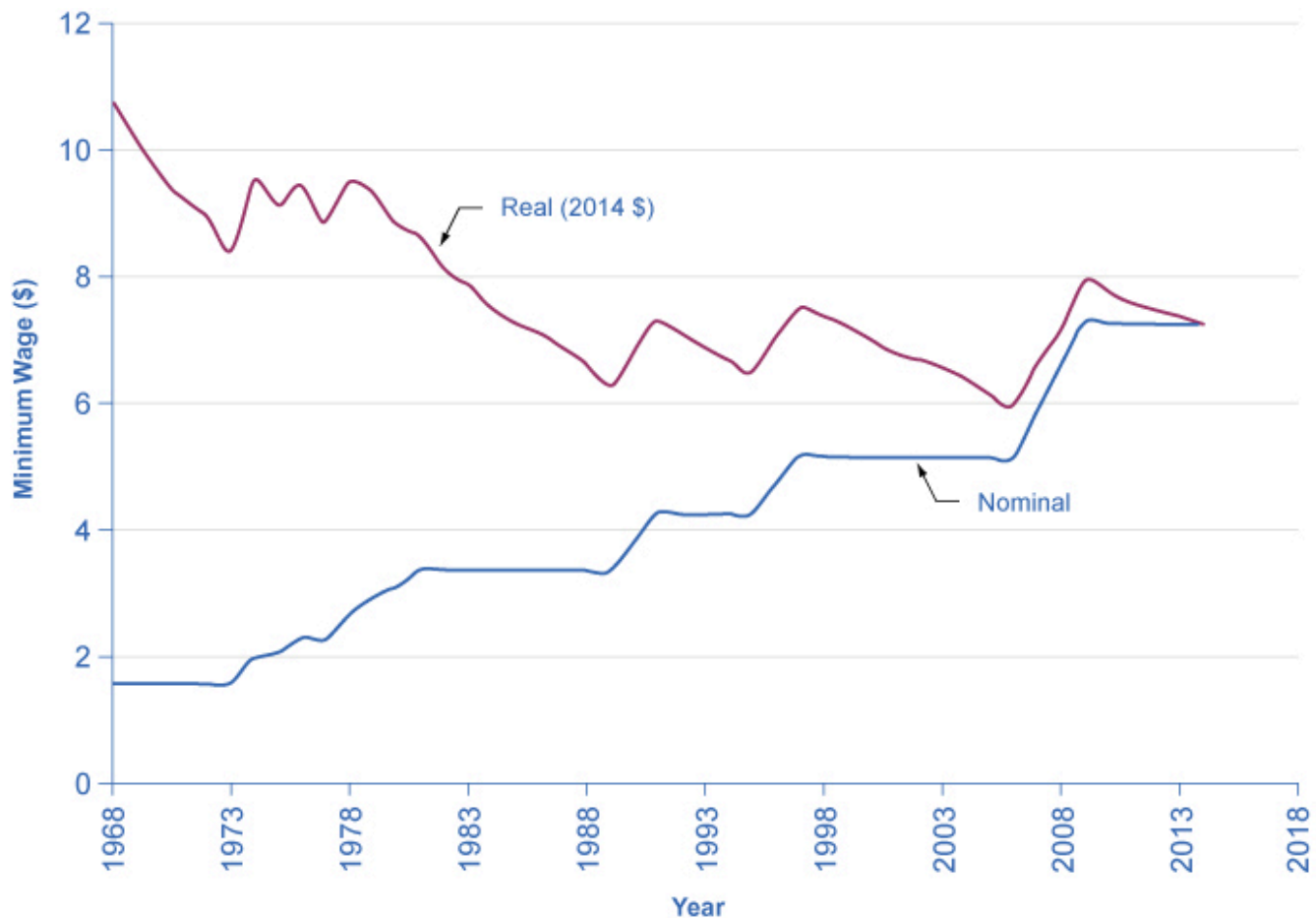


Figure 1. U.S. Minimum Wage and Inflation. After adjusting for inflation, the federal minimum wage dropped more than 30 percent from 1967 to 2010, even though the nominal figure climbed from \$1.40 to \$7.25 per hour. Increases in the minimum wage in 2007, 2008, and 2009 kept the decline from being worse—as it would have been if the wage had remained the same as it did from 1997 through 2006. (Sources: <http://www.dol.gov/whd/minwage/chart.htm>; <http://data.bls.gov/cgi-bin/surveymost?cu>)

omist.blogspot.com. Timothy Taylor lives near Minneapolis with his wife Kimberley and their three children.

Steven A. Greenlaw, University of Mary Washington

Steven Greenlaw has been teaching principles of economics for more than 30 years. In 1999, he received the Grellet C. Simpson Award for Excellence in Undergraduate Teaching at the University of Mary Washington. He is the author of *Doing Economics: A Guide to Doing and Understanding Economic Research*, as well as a variety of articles on economics pedagogy and instructional technology, published in the *Journal of Economic Education*, the *International Review of Economic Education*, and other outlets. He wrote the module on Quantitative Writing for *Starting Point: Teaching and Learning Economics*, the web portal on best practices in teaching economics. Steven Greenlaw lives in Alexandria, Virginia with his wife Kathy and their three children.

CONTRIBUTING AUTHORS

Eric Dodge	Hanover College
Cynthia Gamez	University of Texas at El Paso
Andres Jauregui	Columbus State University
Diane Keenan	Cerritos College
Dan MacDonald	California State University San Bernardino
Amyaz Moledina	The College of Wooster
Craig Richardson	Winston-Salem State University
David Shapiro	Pennsylvania State University
Ralph Sonenshine	American University

EXPERT REVIEWERS

Bryan Aguiar	Northwest Arkansas Community College
Basil Al Hashimi	Mesa Community College
Emil Berendt	Mount St. Mary's University
Zena Buser	Adams State University
Douglas Campbell	The University of Memphis
Sanjukta Chaudhuri	University of Wisconsin – Eau Claire
Xueyu Cheng	Alabama State University
Robert Cunningham	Alma College
Rosa Lea Danielson	College of DuPage
Steven Deloach	Elon University
Debbie Evercloud	University of Colorado Denver
Sal Figueras	Hudson County Community College
Reza Ghorashi	Richard Stockton College of New Jersey
Robert Gillette	University of Kentucky
Shaomin Huang	Lewis-Clark State College
George Jones	University of Wisconsin-Rock County
Charles Kroncke	College of Mount St. Joseph
Teresa Laughlin	Palomar Community College
Carlos Liard-Muriente	Central Connecticut State University
Heather Luea	Kansas State University
Steven Lugauer	University of Notre Dame
William Mosher	Nashua Community College
Michael Netta	Hudson County Community College
Nick Noble	Miami University
Joe Nowakowski	Muskingum University
Shawn Osell	University of Wisconsin, Superior
Mark Owens	Middle Tennessee State University
Sonia Pereira	Barnard College
Brian Peterson	Central College
Jennifer Platania	Elon University
Robert Rycroft	University of Mary Washington
Adrienne Sachse	Florida State College at Jacksonville
Hans Schumann	Texas AM
Gina Shamshak	Goucher College
Chris Warburton	John Jay College of Criminal Justice, CUNY
Mark Witte	Northwestern
Chiou-nan Yeh	Alabama State University

ANCILLARIES

OpenStax projects offer an array of ancillaries for students and instructors. Please visit <http://openstaxcollege.org> and view the learning resources for this title.

COPYRIGHT

© May 18, 2016 OpenStax Economics. Textbook content produced by OpenStax Economics is licensed under a Creative Commons Attribution License 4.0 license. Under this license, any user of this textbook or the textbook contents herein must provide proper attribution as follows: **The OpenStax College name, OpenStax College logo, OpenStax College book covers, OpenStax CNX name, and OpenStax CNX logo are not subject to the creative commons license and may not be reproduced without the prior and express written consent of Rice University.** For questions regarding this license, please contact partners@openstaxcollege.org. If you use this textbook as a bibliographic reference, then you should cite it as follows: OpenStax Economics, Principles of Economics. OpenStax CNX. May 18, 2016 <http://cnx.org/contents/69619d2b-68f0-44b0-b074-a9b2bf90b2c6@11.330>. If you redistribute this textbook in a print format, then you must include on every physical page the following attribution:

“Download for free at <http://cnx.org/contents/69619d2b-68f0-44b0-b074-a9b2bf90b2c6@11.330>.” If you redistribute part of this textbook, then you must retain in every digital format page view (including but not limited to EPUB, PDF, and HTML) and on every physical printed page the following attribution:

“Download for free at <http://cnx.org/contents/69619d2b-68f0-44b0-b074-a9b2bf90b2c6@11.330>.”

Book Name:
Principles of Economics

Author:
OpenStax

Keywords/Tags:
Economics, Microeconomics, Macroeconomics, Introductory

Copyright: 2016 by Rice University.



Principles of Economics by Rice University is licensed under a Creative Commons Attribution 4.0 International License, except where otherwise noted.

This textbook is available for free at open.bccampus.ca

CHAPTER 1. WELCOME TO ECONOMICS!

INTRODUCTION



Figure 1. Do You Use Facebook? Economics is greatly impacted by how well information travels through society. Today, social media giants Twitter, Facebook, and Instagram are major forces on the information super highway. (Credit: Johan Larsson/Flickr)

DECISIONS ... DECISIONS IN THE SOCIAL MEDIA AGE

To post or not to post? Every day we are faced with a myriad of decisions, from what to have for breakfast, to which route to take to class, to the more complex—"Should I double major and add possibly another semester of study to my education?" Our response to these choices depends on the information we have available at any given moment; information economists call "imperfect" because we rarely have all the data we need to make perfect decisions. Despite the lack of perfect information, we still make hundreds of decisions a day.

And now, we have another avenue in which to gather information—social media. Outlets like Facebook and Twitter are altering the process by which we make choices, how we spend our time, which movies we see, which products we buy, and more. How many of you chose a university without checking out its Facebook page or Twitter stream first for information and feedback?

As you will see in this course, what happens in economics is affected by how well and how fast information is disseminated through a society, such as how quickly information travels through Facebook. "Economists love nothing better than when deep and liquid markets operate under conditions of perfect information," says Jessica Irvine, National Economics Editor for News Corp Australia.

This leads us to the topic of this chapter, an introduction to the world of making decisions, processing information, and

understanding behavior in markets —the world of economics. Each chapter in this book will start with a discussion about current (or sometimes past) events and revisit it at chapter’s end—to “bring home” the concepts in play.

CHAPTER OBJECTIVES

Introduction

In this chapter, you will learn about:

- What Is Economics, and Why Is It Important?
- Microeconomics and Macroeconomics
- How Economists Use Theories and Models to Understand Economic Issues
- How Economies Can Be Organized: An Overview of Economic Systems

What is economics and why should you spend your time learning it? After all, there are other disciplines you could be studying, and other ways you could be spending your time. As the Bring it Home feature just mentioned, making choices is at the heart of what economists study, and your decision to take this course is as much an economic decision as anything else.

Economics is probably not what you think. It is not primarily about money or finance. It is not primarily about business. It is not mathematics. What is it then? It is both a subject area and a way of viewing the world.

1.1 WHAT IS ECONOMICS, AND WHY IS IT IMPORTANT?

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Discuss the importance of studying economics
- Explain the relationship between production and division of labor
- Evaluate the significance of scarcity

Economics is the study of how humans make decisions in the face of scarcity. These can be individual decisions, family decisions, business decisions or societal decisions. If you look around carefully, you will see that scarcity is a fact of life. **Scarcity** means that human wants for goods, services and resources exceed what is available. Resources, such as labor, tools, land, and raw materials are necessary to produce the goods and services we want but they exist in limited supply. Of course, the ultimate scarce resource is time- everyone, rich or poor, has just 24 hours in the day to try to acquire the goods they want. At any point in time, there is only a finite amount of resources available.

Think about it this way: In 2015 the labor force in the United States contained over 158.6 million workers, according to the U.S. Bureau of Labor Statistics. Similarly, the total area of the United States is 3,794,101 square miles. These are large numbers for such crucial resources, however, they are limited. Because these resources are limited, so are the numbers of goods and services we produce with them. Combine this with the fact that human wants seem to be virtually infinite, and you can see why scarcity is a problem.

If you still do not believe that scarcity is a problem, consider the following: Does everyone need food to eat? Does everyone need a decent place to live? Does everyone have access to healthcare? In every country in the world, there are people who are hungry, homeless (for example, those who call park benches their beds, as shown in Figure 1), and in need of healthcare, just to focus on a few critical goods and services. Why is this the case? It is because of scarcity. Let's delve into the concept of scarcity a little deeper, because it is crucial to understanding economics.

THE PROBLEM OF SCARCITY

Think about all the things you consume: food, shelter, clothing, transportation, healthcare, and entertainment. How do you acquire those items? You do not produce them yourself. You buy them. How do you afford the things you buy? You work for pay. Or if you do not, someone else does on your behalf.



Figure 1. Scarcity of Resources. Homeless people are a stark reminder that scarcity of resources is real. (Credit: “daveynin”/Flickr Creative Commons)

Yet most of us never have enough to buy all the things we want. This is because of scarcity. So how do we solve it?

Visit this website to read about how the United States is dealing with scarcity in resources.



Every society, at every level, must make choices about how to use its resources. Families must decide whether to spend their money on a new car or a fancy vacation. Towns must choose whether to put more of the budget into police and fire protection or into the school system. Nations must decide whether to devote more funds to national defense or to protecting the environment. In most cases, there just isn't enough money in the budget to do everything. So why do we not each just produce all of the things we consume? The simple answer is most of us do not know how, but that is not the main reason. (When you study economics, you will discover that the obvious choice is not always the right answer—or at least the complete answer. Studying economics teaches you to think in a different of way.) Think back to pioneer days, when individuals knew how to do so much more than we do today, from building their homes, to growing their crops, to hunting for food, to repairing their equipment. Most of us do not know how to do all—or any—of those things. It is not because we could not learn.

Rather, we do not have to. The reason why is something called *the division and specialization of labor*, a production innovation first put forth by **Adam Smith**, Figure 2, in his book, *The Wealth of Nations*.

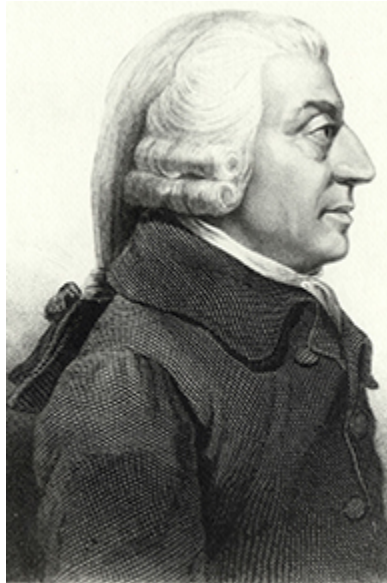


Figure 2. Adam Smith. Adam Smith introduced the idea of dividing labor into discrete tasks. (Credit: Wikimedia Commons)

THE DIVISION OF AND SPECIALIZATION OF LABOR

The formal study of economics began when Adam Smith (1723–1790) published his famous book *The Wealth of Nations* in 1776. Many authors had written on economics in the centuries before Smith, but he was the first to address the subject in a comprehensive way. In the first chapter, Smith introduces the **division of labor**, which means that the way a good or service is produced is divided into a number of tasks that are performed by different workers, instead of all the tasks being done by the same person.

To illustrate the division of labor, Smith counted how many tasks went into making a pin: drawing out a piece of wire, cutting it to the right length, straightening it, putting a head on one end and a point on the other, and packaging pins for sale, to name just a few. Smith counted 18 distinct tasks that were often done by different people—all for a pin, believe it or not!

Modern businesses divide tasks as well. Even a relatively simple business like a restaurant divides up the task of serving meals into a range of jobs like top chef, sous chefs, less-skilled kitchen help, servers to wait on the tables, a greeter at the door, janitors to clean up, and a business manager to handle paychecks and bills—not to mention the economic connections a restaurant has with suppliers of food, furniture, kitchen equipment, and the building where it is located. A complex business like a large manufacturing factory, such as the shoe factory shown in Figure 3, or a hospital can have hundreds of job classifications.



Figure 3. Division of Labor. Workers on an assembly line are an example of the divisions of labor. (Credit: Nina Hale/Flickr Creative Commons)

WHY THE DIVISION OF LABOR INCREASES PRODUCTION

When the tasks involved with producing a good or service are divided and subdivided, workers and businesses can produce a greater quantity of output. In his observations of pin factories, Smith observed that one worker alone might make 20 pins in a day, but that a small business of 10 workers (some of whom would need to do two or three of the 18 tasks involved with pin-making), could make 48,000 pins in a day. How can a group of workers, each specializing in certain tasks, produce so much more than the same number of workers who try to produce the entire good or service by themselves? Smith offered three reasons.

First, **specialization** in a particular small job allows workers to focus on the parts of the production process where they have an advantage. (In later chapters, we will develop this idea by discussing **comparative advantage**.) People have different skills, talents, and interests, so they will be better at some jobs than at others. The particular advantages may be based on educational choices, which are in turn shaped by interests and talents. Only those with medical degrees qualify to become doctors, for instance. For some goods, specialization will be affected by geography—it is easier to be a wheat farmer in North Dakota than in Florida, but easier to run a tourist hotel in Florida than in North Dakota. If you live in or near a big city, it is easier to attract enough customers to operate a successful dry cleaning business or movie theater than if you live in a sparsely populated rural area. Whatever the reason, if people specialize in the production of what they do best, they will be more productive than if they produce a combination of things, some of which they are good at and some of which they are not.

Second, workers who specialize in certain tasks often learn to produce more quickly and with higher quality. This pattern holds true for many workers, including assembly line laborers who build cars, stylists who cut hair, and doctors who perform heart surgery. In fact, specialized workers often know their jobs well enough to suggest innovative ways to do their work faster and better.

A similar pattern often operates within businesses. In many cases, a business that focuses on one or a

few products (sometimes called its “**core competency**”) is more successful than firms that try to make a wide range of products.

Third, specialization allows businesses to take advantage of **economies of scale**, which means that for many goods, as the level of production increases, the average cost of producing each individual unit declines. For example, if a factory produces only 100 cars per year, each car will be quite expensive to make on average. However, if a factory produces 50,000 cars each year, then it can set up an assembly line with huge machines and workers performing specialized tasks, and the average cost of production per car will be lower. The ultimate result of workers who can focus on their preferences and talents, learn to do their specialized jobs better, and work in larger organizations is that society as a whole can produce and consume far more than if each person tried to produce all of their own goods and services. The division and specialization of labor has been a force against the problem of scarcity.

TRADE AND MARKETS

Specialization only makes sense, though, if workers can use the pay they receive for doing their jobs to purchase the other goods and services that they need. In short, specialization requires trade.

You do not have to know anything about electronics or sound systems to play music—you just buy an iPod or MP3 player, download the music and listen. You do not have to know anything about artificial fibers or the construction of sewing machines if you need a jacket—you just buy the jacket and wear it. You do not need to know anything about internal combustion engines to operate a car—you just get in and drive. Instead of trying to acquire all the knowledge and skills involved in producing all of the goods and services that you wish to consume, the market allows you to learn a specialized set of skills and then use the pay you receive to buy the goods and services you need or want. This is how our modern society has evolved into a strong economy.

WHY STUDY ECONOMICS?

Now that we have gotten an overview on what economics studies, let’s quickly discuss why you are right to study it. Economics is not primarily a collection of facts to be memorized, though there are plenty of important concepts to be learned. Instead, economics is better thought of as a collection of questions to be answered or puzzles to be worked out. Most important, economics provides the tools to work out those puzzles. If you have yet to be bitten by the economics “bug,” there are other reasons why you should study economics.

- Virtually every major problem facing the world today, from global warming, to world poverty, to the conflicts in Syria, Afghanistan, and Somalia, has an economic dimension. If you are going to be part of solving those problems, you need to be able to understand them. Economics is crucial.
- It is hard to overstate the importance of economics to good citizenship. You need to be able to vote intelligently on budgets, regulations, and laws in general. When the U.S. government came close to a standstill at the end of 2012 due to the “fiscal cliff,” what were the issues involved? Did you know?
- A basic understanding of economics makes you a well-rounded thinker. When you read articles about economic issues, you will understand and be able to evaluate the writer’s argument. When you hear classmates, co-workers, or political candidates talking about

economics, you will be able to distinguish between common sense and nonsense. You will find new ways of thinking about current events and about personal and business decisions, as well as current events and politics.

The study of economics does not dictate the answers, but it can illuminate the different choices.

KEY CONCEPTS AND SUMMARY

Economics seeks to solve the problem of scarcity, which is when human wants for goods and services exceed the available supply. A modern economy displays a division of labor, in which people earn income by specializing in what they produce and then use that income to purchase the products they need or want. The division of labor allows individuals and firms to specialize and to produce more for several reasons: a) It allows the agents to focus on areas of advantage due to natural factors and skill levels; b) It encourages the agents to learn and invent; c) It allows agents to take advantage of economies of scale. Division and specialization of labor only work when individuals can purchase what they do not produce in markets. Learning about economics helps you understand the major problems facing the world today, prepares you to be a good citizen, and helps you become a well-rounded thinker.

SELF-CHECK QUESTIONS

1. What is scarcity? Can you think of two causes of scarcity?
2. Residents of the town of Smithfield like to consume hams, but each ham requires 10 people to produce it and takes a month. If the town has a total of 100 people, what is the maximum amount of ham the residents can consume in a month?
3. A consultant works for \$200 per hour. She likes to eat vegetables, but is not very good at growing them. Why does it make more economic sense for her to spend her time at the consulting job and shop for her vegetables?
4. A computer systems engineer could paint his house, but it makes more sense for him to hire a painter to do it. Explain why.

REVIEW QUESTIONS

1. Give the three reasons that explain why the division of labor increases an economy's level of production.
2. What are three reasons to study economics?

CRITICAL THINKING QUESTIONS

1. Suppose you have a team of two workers: one is a baker and one is a chef. Explain why the kitchen can produce more meals in a given period of time if each worker specializes in what they do best than if each worker tries to do everything from appetizer to dessert.

2. Why would division of labor without trade not work?
3. Can you think of any examples of *free* goods, that is, goods or services that are not scarce?

REFERENCES

Bureau of Labor Statistics, U.S. Department of Labor. 2015. "The Employment Situation—February 2015." Accessed March 27, 2015. <http://www.bls.gov/news.release/pdf/empisit.pdf>.

Williamson, Lisa. "US Labor Market in 2012." *Bureau of Labor Statistics*. Accessed December 1, 2013. <http://www.bls.gov/opub/mlr/2013/03/art1full.pdf>.

GLOSSARY

division of labor the way in which the work required to produce a good or service is divided into tasks performed by different workers

economics the study of how humans make choices under conditions of scarcity

economies of scale when the average cost of producing each individual unit declines as total output increases

scarcity when human wants for goods and services exceed the available supply

specialization when workers or firms focus on particular tasks for which they are well-suited within the overall production process

SOLUTIONS

Answers for Self-Check Questions

1. Scarcity means human wants for goods and services exceed the available supply. Supply is limited because resources are limited. Demand, however, is virtually unlimited. Whatever the supply, it seems human nature to want more.
2. $100 \text{ people} / 10 \text{ people per ham} = \text{a maximum of 10 hams per month}$ if all residents produce ham. Since consumption is limited by production, the maximum number of hams residents could consume per month is 10.
3. She is very productive at her consulting job, but not very productive growing vegetables. Time spent consulting would produce far more income than it what she could save growing her vegetables using the same amount of time. So on purely economic grounds, it makes more sense for her to maximize her income by applying her labor to what she does best (i.e. specialization of labor).
4. The engineer is better at computer science than at painting. Thus, his time is better spent working for pay at his job and paying a painter to paint his house. Of course, this assumes he does not paint his house for fun!

1.2 MICROECONOMICS AND MACROECONOMICS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe microeconomics
- Describe macroeconomics
- Contrast monetary policy and fiscal policy

Economics is concerned with the well-being of *all* people, including those with jobs and those without jobs, as well as those with high incomes and those with low incomes. Economics acknowledges that production of useful goods and services can create problems of environmental pollution. It explores the question of how investing in education helps to develop workers' skills. It probes questions like how to tell when big businesses or big labor unions are operating in a way that benefits society as a whole and when they are operating in a way that benefits their owners or members at the expense of others. It looks at how government spending, taxes, and regulations affect decisions about production and consumption.

It should be clear by now that economics covers a lot of ground. That ground can be divided into two parts: **Microeconomics** focuses on the actions of individual agents within the economy, like households, workers, and businesses; **Macroeconomics** looks at the economy as a whole. It focuses on broad issues such as growth of production, the number of unemployed people, the inflationary increase in prices, government deficits, and levels of exports and imports. Microeconomics and macroeconomics are not separate subjects, but rather complementary perspectives on the overall subject of the economy.

To understand why both microeconomic and macroeconomic perspectives are useful, consider the problem of studying a biological ecosystem like a lake. One person who sets out to study the lake might focus on specific topics: certain kinds of algae or plant life; the characteristics of particular fish or snails; or the trees surrounding the lake. Another person might take an overall view and instead consider the entire ecosystem of the lake from top to bottom; what eats what, how the system stays in a rough balance, and what environmental stresses affect this balance. Both approaches are useful, and both examine the same lake, but the viewpoints are different. In a similar way, both microeconomics and macroeconomics study the same economy, but each has a different viewpoint.

Whether you are looking at lakes or economics, the micro and the macro insights should blend with each other. In studying a lake, the micro insights about particular plants and animals help to under-

stand the overall food chain, while the macro insights about the overall food chain help to explain the environment in which individual plants and animals live.

In economics, the micro decisions of individual businesses are influenced by whether the macroeconomy is healthy; for example, firms will be more likely to hire workers if the overall economy is growing. In turn, the performance of the macroeconomy ultimately depends on the microeconomic decisions made by individual households and businesses.

MICROECONOMICS

What determines how households and individuals spend their budgets? What combination of goods and services will best fit their needs and wants, given the budget they have to spend? How do people decide whether to work, and if so, whether to work full time or part time? How do people decide how much to save for the future, or whether they should borrow to spend beyond their current means?

What determines the products, and how many of each, a firm will produce and sell? What determines what prices a firm will charge? What determines how a firm will produce its products? What determines how many workers it will hire? How will a firm finance its business? When will a firm decide to expand, downsize, or even close? In the microeconomic part of this book, we will learn about the theory of consumer behavior and the theory of the firm.

MACROECONOMICS

What determines the level of economic activity in a society? In other words, what determines how many goods and services a nation actually produces? What determines how many jobs are available in an economy? What determines a nation's standard of living? What causes the economy to speed up or slow down? What causes firms to hire more workers or to lay workers off? Finally, what causes the economy to grow over the long term?

An economy's macroeconomic health can be defined by a number of goals: growth in the standard of living, low unemployment, and low inflation, to name the most important. How can macroeconomic policy be used to pursue these goals? **Monetary policy**, which involves policies that affect bank lending, interest rates, and financial capital markets, is conducted by a nation's central bank. For the United States, this is the Federal Reserve. **Fiscal policy**, which involves government spending and taxes, is determined by a nation's legislative body. For the United States, this is the Congress and the executive branch, which originates the federal budget. These are the main tools the government has to work with. Americans tend to expect that government can fix whatever economic problems we encounter, but to what extent is that expectation realistic? These are just some of the issues that will be explored in the macroeconomic chapters of this book.

KEY CONCEPTS AND SUMMARY

Microeconomics and macroeconomics are two different perspectives on the economy. The microeconomic perspective focuses on parts of the economy: individuals, firms, and industries. The macroeconomic perspective looks at the economy as a whole, focusing on goals like growth in the standard of living, unemployment, and inflation. Macroeconomics has two types of policies for pursuing these goals: monetary policy and fiscal policy.

SELF-CHECK QUESTIONS

What would be another example of a “system” in the real world that could serve as a metaphor for micro and macroeconomics?

REVIEW QUESTIONS

1. What is the difference between microeconomics and macroeconomics?
2. What are examples of individual economic agents?
3. What are the three main goals of macroeconomics?

CRITICAL THINKING QUESTIONS

1. A balanced federal budget and a balance of trade are considered secondary goals of macroeconomics, while growth in the standard of living (for example) is considered a primary goal. Why do you think that is so?
2. Macroeconomics is an aggregate of what happens at the microeconomic level. Would it be possible for what happens at the macro level to differ from how economic agents would react to some stimulus at the micro level? *Hint:* Think about the behavior of crowds.

GLOSSARY

fiscal policy economic policies that involve government spending and taxes

macroeconomics the branch of economics that focuses on broad issues such as growth, unemployment, inflation, and trade balance.

microeconomics the branch of economics that focuses on actions of particular agents within the economy, like households, workers, and business firms

monetary policy policy that involves altering the level of interest rates, the availability of credit in the economy, and the extent of borrowing

SOLUTIONS

Answers to Self-Check Questions

There are many physical systems that would work, for example, the study of planets (micro) in the solar system (macro), or solar systems (micro) in the galaxy (macro).

1.3 HOW ECONOMISTS USE THEORIES AND MODELS TO UNDERSTAND ECONOMIC ISSUES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Interpret a circular flow diagram
- Explain the importance of economic theories and models
- Describe goods and services markets and labor markets

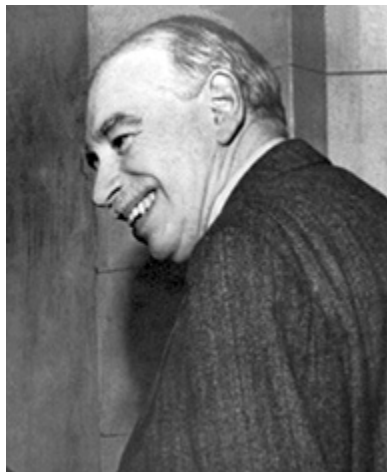


Figure 1. John Maynard Keynes.

One of the most influential economists in modern times was John Maynard Keynes. (Credit: Wikimedia Commons)

John Maynard Keynes (1883–1946), one of the greatest economists of the twentieth century, pointed out that economics is not just a subject area but also a way of thinking. Keynes, shown in Figure 1, famously wrote in the introduction to a fellow economist’s book: “[Economics] is a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which helps its possessor to draw correct conclusions.” In other words, economics teaches you how to think, not what to think.

Watch this video about John Maynard Keynes and his influence on economics.



Economists see the world through a different lens than anthropologists, biologists, classicists, or practitioners of any other discipline. They analyze issues and problems with economic theories that are based on particular assumptions about human behavior, that are different than the assumptions an anthropologist or psychologist might use. A **theory** is a simplified representation of how two or more variables interact with each other. The purpose of a theory is to take a complex, real-world issue and simplify it down to its essentials. If done well, this enables the analyst to understand the issue and any problems around it. A good theory is simple enough to be understood, while complex enough to capture the key features of the object or situation being studied.

Sometimes economists use the term **model** instead of theory. Strictly speaking, a theory is a more abstract representation, while a model is more applied or empirical representation. Models are used to test theories, but for this course we will use the terms interchangeably.

For example, an architect who is planning a major office building will often build a physical model that sits on a tabletop to show how the entire city block will look after the new building is constructed. Companies often build models of their new products, which are more rough and unfinished than the final product will be, but can still demonstrate how the new product will work.

A good model to start with in economics is the **circular flow diagram**, which is shown in Figure 2. It pictures the economy as consisting of two groups—households and firms—that interact in two markets: the **goods and services market** in which firms sell and households buy and the **labor market** in which households sell labor to business firms or other employees.

Of course, in the real world, there are many different markets for goods and services and markets for many different types of labor. The circular flow diagram simplifies this to make the picture easier to grasp. In the diagram, firms produce goods and services, which they sell to households in return for revenues. This is shown in the outer circle, and represents the two sides of the product market (for example, the market for goods and services) in which households demand and firms supply. Households sell their labor as workers to firms in return for wages, salaries and benefits. This is shown in the inner circle and represents the two sides of the labor market in which households supply and firms demand.

This version of the circular flow model is stripped down to the essentials, but it has enough features to explain how the product and labor markets work in the economy. We could easily add details to this basic model if we wanted to introduce more real-world elements, like financial markets, governments, and interactions with the rest of the globe (imports and exports).

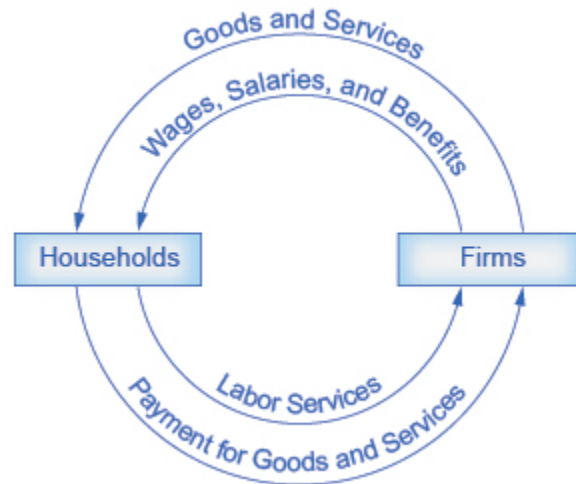


Figure 2. The Circular Flow Diagram. The circular flow diagram shows how households and firms interact in the goods and services market, and in the labor market. The direction of the arrows shows that in the goods and services market, households receive goods and services and pay firms for them. In the labor market, households provide labor and receive payment from firms through wages, salaries, and benefits.

Economists carry a set of theories in their heads like a carpenter carries around a toolkit. When they see an economic issue or problem, they go through the theories they know to see if they can find one that fits. Then they use the theory to derive insights about the issue or problem. In economics, theories are expressed as diagrams, graphs, or even as mathematical equations. (Do not worry. In this course, we will mostly use graphs.) Economists do not figure out the answer to the problem first and then draw the graph to illustrate. Rather, they use the graph of the theory to help them figure out the answer. Although at the introductory level, you can sometimes figure out the right answer without applying a model, if you keep studying economics, before too long you will run into issues and problems that you will need to graph to solve. Both micro and macroeconomics are explained in terms of theories and models. The most well-known theories are probably those of supply and demand, but you will learn a number of others.

KEY CONCEPTS AND SUMMARY

Economists analyze problems differently than do other disciplinary experts. The main tools economists use are economic theories or models. A theory is not an illustration of the answer to a problem. Rather, a theory is a tool for determining the answer.

SELF-CHECK QUESTIONS

1. Suppose we extend the circular flow model to add imports and exports. Copy the circular flow diagram

onto a sheet of paper and then add a foreign country as a third agent. Draw a rough sketch of the flows of imports, exports, and the payments for each on your diagram.

2. What is an example of a problem in the world today, not mentioned in the chapter, that has an economic dimension?

REVIEW QUESTIONS

1. How did John Maynard Keynes define economics?
2. Are households primarily buyers or sellers in the goods and services market? In the labor market?
3. Are firms primarily buyers or sellers in the goods and services market? In the labor market?

CRITICAL THINKING QUESTIONS

1. Why is it unfair or meaningless to criticize a theory as “unrealistic?”
2. Suppose, as an economist, you are asked to analyze an issue unlike anything you have ever done before. Also, suppose you do not have a specific model for analyzing that issue. What should you do? *Hint: What would a carpenter do in a similar situation?*

GLOSSARY

circular flow diagram a diagram that views the economy as consisting of households and firms interacting in a goods and services market and a labor market

goods and services market a market in which firms are sellers of what they produce and households are buyers

labor market the market in which households sell their labor as workers to business firms or other employers

model see theory

theory a representation of an object or situation that is simplified while including enough of the key features to help us understand the object or situation

SOLUTIONS

Answers to Self-Check Questions

1. Draw a box outside the original circular flow to represent the foreign country. Draw an arrow from the foreign country to firms, to represent imports. Draw an arrow in the reverse direction representing payments for imports. Draw an arrow from firms to the foreign country to represent exports. Draw an arrow in the reverse direction to represent payments for imports.
2. There are many such problems. Consider the AIDS epidemic. Why are so few AIDS patients in Africa and

Southeast Asia treated with the same drugs that are effective in the United States and Europe? It is because neither those patients nor the countries in which they live have the resources to purchase the same drugs.

1.4 HOW ECONOMIES CAN BE ORGANIZED: AN OVERVIEW OF ECONOMIC SYSTEMS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Contrast traditional economies, command economies, and market economies
- Explain gross domestic product (GDP)
- Assess the importance and effects of globalization

Think about what a complex system a modern economy is. It includes all production of goods and services, all buying and selling, all employment. The economic life of every individual is interrelated, at least to a small extent, with the economic lives of thousands or even millions of other individuals. Who organizes and coordinates this system? Who insures that, for example, the number of televisions a society provides is the same as the amount it needs and wants? Who insures that the right number of employees work in the electronics industry? Who insures that televisions are produced in the best way possible? How does it all get done?

There are at least three ways societies have found to organize an economy. The first is the **traditional economy**, which is the oldest economic system and can be found in parts of Asia, Africa, and South America. Traditional economies organize their economic affairs the way they have always done (i.e., tradition). Occupations stay in the family. Most families are farmers who grow the crops they have always grown using traditional methods. What you produce is what you get to consume. Because things are driven by tradition, there is little economic progress or development.

Command economies are very different. In a **command economy**, economic effort is devoted to goals passed down from a ruler or ruling class. Ancient Egypt was a good example: a large part of economic life was devoted to building pyramids, like those shown in Figure 1, for the pharaohs. Medieval manor life is another example: the lord provided the land for growing crops and protection in the event of war. In return, vassals provided labor and soldiers to do the lord's bidding. In the last century, communism emphasized command economies.

In a command economy, the government decides what goods and services will be produced and what prices will be charged for them. The government decides what methods of production will be used and how much workers will be paid. Many necessities like healthcare and education are provided for free. Currently, Cuba and North Korea have command economies.



Figure 1. A Command Economy. Ancient Egypt was an example of a command economy. (Credit: Jay Bergesen/Flickr Creative Commons)



Figure 2. A Market Economy. Nothing says “market” more than The New York Stock Exchange. (Credit: Erik Drost/Flickr Creative Commons)

Although command economies have a very centralized structure for economic decisions, market economies have a very decentralized structure. A **market** is an institution that brings together buyers and sellers of goods or services, who may be either individuals or businesses. The New York Stock Exchange, shown in Figure 2, is a prime example of market in which buyers and sellers are brought together. In a **market economy**, decision-making is decentralized. Market economies are based on **private enterprise**: the means of production (resources and businesses) are owned and operated by private individuals or groups of private individuals. Businesses supply goods and services based on demand. (In a command economy, by contrast, resources and businesses are owned by the government.) What goods and services are supplied depends on what is demanded. A person’s income is based on his or her ability to convert resources (especially labor) into something that society val-

ues. The more society values the person's output, the higher the income (think Lady Gaga or LeBron James). In this scenario, economic decisions are determined by market forces, not governments.

Most economies in the real world are mixed; they combine elements of command and market (and even traditional) systems. The U.S. economy is positioned toward the market-oriented end of the spectrum. Many countries in Europe and Latin America, while primarily market-oriented, have a greater degree of government involvement in economic decisions than does the U.S. economy. China and Russia, while they are closer to having a market-oriented system now than several decades ago, remain closer to the command economy end of the spectrum. A rich resource of information about countries and their economies can be found on the Heritage Foundation's website, as the following Clear It Up feature discusses.

WHAT COUNTRIES ARE CONSIDERED ECONOMICALLY FREE?

Who is in control of economic decisions? Are people free to do what they want and to work where they want? Are businesses free to produce when they want and what they choose, and to hire and fire as they wish? Are banks free to choose who will receive loans? Or does the government control these kinds of choices? Each year, researchers at the Heritage Foundation and the *Wall Street Journal* look at 50 different categories of economic freedom for countries around the world. They give each nation a score based on the extent of economic freedom in each category.

The 2015 Heritage Foundation's Index of Economic Freedom report ranked 178 countries around the world: some examples of the most free and the least free countries are listed in Table 1. Several countries were not ranked because of extreme instability that made judgments about economic freedom impossible. These countries include Afghanistan, Iraq, Syria, and Somalia.

The assigned rankings are inevitably based on estimates, yet even these rough measures can be useful for discerning trends. In 2015, 101 of the 178 included countries shifted toward greater economic freedom, although 77 of the countries shifted toward less economic freedom. In recent decades, the overall trend has been a *higher level of economic freedom around the world*.

Most Economic Freedom	Least Economic Freedom
1. Hong Kong	167. Timor-Leste
2. Singapore	168. Democratic Republic of Congo
3. New Zealand	169. Argentina
4. Australia	170. Republic of Congo
5. Switzerland	171. Iran
6. Canada	172. Turkmenistan
7. Chile	173. Equatorial Guinea
8. Estonia	174. Eritrea
9. Ireland	175. Zimbabwe
10. Mauritius	176. Venezuela
11. Denmark	177. Cuba
12. United States	178. North Korea

Table 1. Economic Freedoms, 2015 (Source: The Heritage Foundation, 2015 Index of Economic Freedom, Country Rankings, <http://www.heritage.org/index/ranking>)

REGULATIONS: THE RULES OF THE GAME

Markets and government regulations are always entangled. There is no such thing as an absolutely free market. Regulations always define the “rules of the game” in the economy. Economies that are primarily market-oriented have fewer regulations—ideally just enough to maintain an even playing field for participants. At a minimum, these laws govern matters like safeguarding private property against theft, protecting people from violence, enforcing legal contracts, preventing fraud, and collecting taxes. Conversely, even the most command-oriented economies operate using markets. How else would buying and selling occur? But the decisions of what will be produced and what prices will be charged are heavily regulated. Heavily regulated economies often have **underground economies**, which are markets where the buyers and sellers make transactions without the government’s approval.

The question of how to organize economic institutions is typically not a black-or-white choice between all market or all government, but instead involves a balancing act over the appropriate combination of market freedom and government rules.



Figure 3. Globalization. Cargo ships are one mode of transportation for shipping goods in the global economy. (Credit: Raul Valdez/Flickr Creative Commons)

THE RISE OF GLOBALIZATION

Recent decades have seen a trend toward **globalization**, which is the expanding cultural, political, and economic connections between people around the world. One measure of this is the increased buying and selling of goods, services, and assets across national borders—in other words, international trade and financial capital flows.

Globalization has occurred for a number of reasons. Improvements in shipping, as illustrated by the container ship shown in Figure 3, and air cargo have driven down transportation costs. Innovations in computing and telecommunications have made it easier and cheaper to manage long-distance economic connections of production and sales. Many valuable products and services in the modern economy can take the form of information—for example: computer software; financial advice; travel planning; music, books and movies; and blueprints for designing a building. These products and many

others can be transported over telephones and computer networks at ever-lower costs. Finally, international agreements and treaties between countries have encouraged greater trade.

Table 2 presents one measure of globalization. It shows the percentage of domestic economic production that was exported for a selection of countries from 2010 to 2013, according to an entity known as The World Bank. **Exports** are the goods and services that are produced domestically and sold abroad. **Imports** are the goods and services that are produced abroad and then sold domestically. The size of total production in an economy is measured by the **gross domestic product (GDP)**. Thus, the ratio of exports divided by GDP measures what share of a country's total economic production is sold in other countries.

Country	2010	2011	2012	2013
Higher Income Countries				
United States	12.4	13.6	13.6	13.5
Belgium	76.2	81.4	82.2	82.8
Canada	29.1	30.7	30.0	30.1
France	26.0	27.8	28.1	28.3
Middle Income Countries				
Brazil	10.9	11.9	12.6	12.6
Mexico	29.9	31.2	32.6	31.7
South Korea	49.4	55.7	56.3	53.9
Lower Income Countries				
Chad	36.8	38.9	36.9	32.2
China	29.4	28.5	27.3	26.4
India	22.0	23.9	24.0	24.8
Nigeria	25.3	31.3	31.4	18.0

Table 2. The Extent of Globalization (exports/GDP) (Source: <http://databank.worldbank.org/data/>)

In recent decades, the export/GDP ratio has generally risen, both worldwide and for the U.S. economy. Interestingly, the share of U.S. exports in proportion to the U.S. economy is well below the global average, in part because large economies like the United States can contain more of the division of labor inside their national borders. However, smaller economies like Belgium, Korea, and Canada need to trade across their borders with other countries to take full advantage of division of labor, specialization, and economies of scale. In this sense, the enormous U.S. economy is less affected by globalization than most other countries.

Table 2 also shows that many medium and low income countries around the world, like Mexico and China, have also experienced a surge of globalization in recent decades. If an astronaut in orbit could put on special glasses that make all economic transactions visible as brightly colored lines and look down at Earth, the astronaut would see the planet covered with connections.

So, hopefully, you now have an idea of what economics is about. Before you move to any other chapter of study, be sure to read the very important appendix to this chapter called The Use of Mathematics in Principles of Economics. It is essential that you learn more about how to read and use models in economics.

DECISIONS ... DECISIONS IN THE SOCIAL MEDIA AGE

The world we live in today provides nearly instant access to a wealth of information. Consider that as recently as the late 1970s, the Farmer's Almanac, along with the Weather Bureau of the U.S. Department of Agriculture, were the primary sources American farmers used to determine when to plant and harvest their crops. Today, farmers are more likely to access, online, weather forecasts from the National Oceanic and Atmospheric Administration or watch the Weather Channel. After all, knowing the upcoming forecast could drive when to harvest crops. Consequently, knowing the upcoming weather could change the amount of crop harvested.

Some relatively new information forums, such as Facebook, are rapidly changing how information is distributed; hence, influencing decision making. In 2014, the Pew Research Center reported that 71% of online adults use Facebook. Facebook post topics range from the National Basketball Association, to celebrity singers and performers, to farmers.

Information helps us make decisions. Decisions as simple as what to wear today to how many reporters should be sent to cover a crash. Each of these decisions is an economic decision. After all, resources are scarce. If ten reporters are sent to cover an accident, they are not available to cover other stories or complete other tasks. Information provides the knowledge needed to make the best possible decisions on how to utilize scarce resources. Welcome to the world of economics!

KEY CONCEPTS AND SUMMARY

Societies can be organized as traditional, command, or market-oriented economies. Most societies are a mix. The last few decades have seen globalization evolve as a result of growth in commercial and financial networks that cross national borders, making businesses and workers from different economies increasingly interdependent.

SELF-CHECK QUESTIONS

1. The chapter defines *private enterprise* as a characteristic of market-oriented economies. What would *public enterprise* be? *Hint:* It is a characteristic of command economies.
2. Why might Belgium, France, Italy, and Sweden have a higher export to GDP ratio than the United States?

REVIEW QUESTIONS

1. What are the three ways that societies can organize themselves economically?
2. What is globalization? How do you think it might have affected the economy over the past decade?

CRITICAL THINKING QUESTIONS

1. Why do you think that most modern countries' economies are a mix of command and market types?
2. Can you think of ways that globalization has helped you economically? Can you think of ways that it has not?

REFERENCES

The Heritage Foundation. 2015. "2015 Index of Economic Freedom." Accessed March 11, 2015. <http://www.heritage.org/index/ranking>.

Garling, Caleb. "S.F. plane crash: Reporting, emotions on social media," *The San Francisco Chronicle*. July 7, 2013. <http://www.sfgate.com/news/article/S-F-plane-crash-Reporting-emotions-on-social-4651639.php>.

Irvine, Jessica. "Social Networking Sites are Factories of Modern Ideas." *The Sydney Morning Herald*. November 25, 2011. <http://www.smh.com.au/federal-politics/society-and-culture/social-networking-sites-are-factories-of-modern-ideas-20111124-1nwy3.html#ixzz2YZhPYeME>.

Pew Research Center. 2015. "Social Networking Fact Sheet." Accessed March 11, 2015. <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/>.

The World Bank Group. 2015. "World Data Bank." Accessed March 30, 2014. <http://databank.world-bank.org/data/>.

GLOSSARY

command economy an economy where economic decisions are passed down from government authority and where resources are owned by the government

exports products (goods and services) made domestically and sold abroad

globalization the trend in which buying and selling in markets have increasingly crossed national borders

gross domestic product (GDP) measure of the size of total production in an economy

imports products (goods and services) made abroad and then sold domestically

market interaction between potential buyers and sellers; a combination of demand and supply

market economy an economy where economic decisions are decentralized, resources are owned by private individuals, and businesses supply goods and services based on demand

private enterprise system where the means of production (resources and businesses) are owned and operated by private individuals or groups of private individuals

traditional economy typically an agricultural economy where things are done the same as they have always been done

underground economy a market where the buyers and sellers make transactions in violation of one or more government regulations

SOLUTIONS

Answers to Self-Check Questions

1. Public enterprise means the factors of production (resources and businesses) are owned and operated by the government.
2. The United States is a large country economically speaking, so it has less need to trade internationally than the other countries mentioned. (This is the same reason that France and Italy have lower ratios than

Belgium or Sweden.) One additional reason is that each of the other countries is a member of the European Union, where trade between members occurs without barriers to trade, like tariffs and quotas.

CHAPTER 2. CHOICE IN A WORLD OF SCARCITY

INTRODUCTION TO CHOICE IN A WORLD OF SCARCITY



Figure 1. Choices and Tradeoffs. In general, the higher the degree, the higher the salary. So why aren't more people pursuing higher degrees? The short answer: choices and tradeoffs. (Credit: modification of work by "Jim, the Photographer"/Flickr Creative Commons)

CHOICES ... TO WHAT DEGREE?

In 2015, the median income for workers who hold master's degrees varies from males to females. The average of the two is \$2,951 weekly. Multiply this average by 52 weeks, and you get an average salary of \$153,452. Compare that to the median weekly earnings for a full-time worker over 25 with no higher than a bachelor's degree: \$1,224 weekly and \$63,648 a year. What about those with no higher than a high school diploma in 2015? They earn just \$664 weekly and \$34,528 over 12 months. In other words, says the Bureau of Labor Statistics (BLS), earning a bachelor's degree boosted salaries 54% over what you would have earned if you had stopped your education after high school. A master's degree yields a salary almost double that of a high school diploma.

Given these statistics, we might expect a lot of people to choose to go to college and at least earn a bachelor's degree. Assuming that people want to improve their material well-being, it seems like they would make those choices that give them the greatest opportunity to consume goods and services. As it turns out, the analysis is not nearly as simple as this. In fact, in 2014, the BLS reported that while almost 88% of the population in the United States had a high school diploma, only 33.6% of 25–65 year olds had bachelor's degrees, and only 7.4% of 25–65 year olds in 2014 had earned a master's.

This brings us to the subject of this chapter: why people make the choices they make and how economists go about explaining those choices.

CHAPTER OBJECTIVES

Introduction to Choice in a World of Scarcity

In this chapter, you will learn about:

- How Individuals Make Choices Based on Their Budget Constraint
- The Production Possibilities Frontier and Social Choices
- Confronting Objections to the Economic Approach

You will learn quickly when you examine the relationship between economics and scarcity that choices involve tradeoffs. Every choice has a cost.

In 1968, the Rolling Stones recorded “You Can’t Always Get What You Want.” Economists chuckled, because they had been singing a similar tune for decades. English economist Lionel Robbins (1898–1984), in his *Essay on the Nature and Significance of Economic Science* in 1932, described not always getting what you want in this way:

The time at our disposal is limited. There are only twenty-four hours in the day. We have to choose between the different uses to which they may be put. ... Everywhere we turn, if we choose one thing we must relinquish others which, in different circumstances, we would wish not to have relinquished. Scarcity of means to satisfy given ends is an almost ubiquitous condition of human nature.

Because people live in a world of scarcity, they cannot have all the time, money, possessions, and experiences they wish. Neither can society.

This chapter will continue our discussion of scarcity and the economic way of thinking by first introducing three critical concepts: opportunity cost, marginal decision making, and diminishing returns. Later, it will consider whether the economic way of thinking accurately describes either how choices *are* made or how they *should* be made.

2.1 HOW INDIVIDUALS MAKE CHOICES BASED ON THEIR BUDGET CONSTRAINT

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Calculate and graph budgets constraints
- Explain opportunity sets and opportunity costs
- Evaluate the law of diminishing marginal utility
- Explain how marginal analysis and utility influence choices

Consider the typical consumer's budget problem. Consumers have a limited amount of income to spend on the things they need and want. Suppose Alphonso has \$10 in spending money each week that he can allocate between bus tickets for getting to work and the burgers that he eats for lunch. Burgers cost \$2 each, and bus tickets are 50 cents each. Figure 1 shows Alphonso's **budget constraint**, that is, the outer boundary of his **opportunity set**. The opportunity set identifies all the opportunities for spending within his budget. The budget constraint indicates all the combinations of burgers and bus tickets Alphonso can afford when he exhausts his budget, given the prices of the two goods. (There are actually many different kinds of budget constraints. You will learn more about them in the chapter on Consumer Choices.)

The vertical axis in the figure shows burger purchases and the horizontal axis shows bus ticket purchases. If Alphonso spends all his money on burgers, he can afford five per week. ($\$10 \text{ per week} / \$2 \text{ per burger} = 5 \text{ burgers per week}$.) But if he does this, he will not be able to afford any bus tickets. This choice (zero bus tickets and five burgers) is shown by point A in the figure. Alternatively, if Alphonso spends all his money on bus tickets, he can afford 20 per week. ($\$10 \text{ per week} / \$0.50 \text{ per bus ticket} = 20 \text{ bus tickets per week}$.) Then, however, he will not be able to afford any burgers. This alternative choice (20 bus tickets and zero burgers) is shown by point F.

If Alphonso is like most people, he will choose some combination that includes both bus tickets and burgers. That is, he will choose some combination on the budget constraint that connects points A and F. Every point on (or inside) the constraint shows a combination of burgers and bus tickets that Alphonso can afford. Any point outside the constraint is not affordable, because it would cost more money than Alphonso has in his budget.

The budget constraint clearly shows the tradeoff Alphonso faces in choosing between burgers and bus tickets. Suppose he is currently at point D, where he can afford 12 bus tickets and two burgers. What

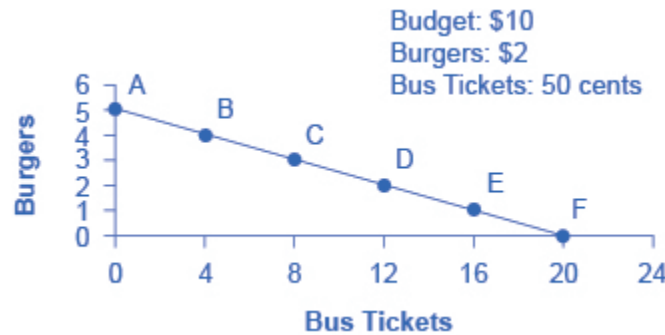


Figure 1. The Budget Constraint: Alphonso's Consumption Choice Opportunity Frontier. Each point on the budget constraint represents a combination of burgers and bus tickets whose total cost adds up to Alphonso's budget of \$10. The slope of the budget constraint is determined by the relative price of burgers and bus tickets. All along the budget set, giving up one burger means gaining four bus tickets.

would it cost Alphonso for one more burger? It would be natural to answer \$2, but that's not the way economists think. Instead they ask, how many bus tickets would Alphonso have to give up to get one more burger, while staying within his budget? The answer is four bus tickets. That is the true cost to Alphonso of one more burger.

THE CONCEPT OF OPPORTUNITY COST

Economists use the term **opportunity cost** to indicate what must be given up to obtain something that is desired. The idea behind opportunity cost is that the cost of one item is the lost opportunity to do or consume something else; in short, opportunity cost is the value of the next best alternative. For Alphonso, the opportunity cost of a burger is the four bus tickets he would have to give up. He would decide whether or not to choose the burger depending on whether the value of the burger exceeds the value of the forgone alternative—in this case, bus tickets. Since people must choose, they inevitably face tradeoffs in which they have to give up things they desire to get other things they desire more.

View this website for an example of opportunity cost—paying someone else to wait in line for you.



A fundamental principle of economics is that every choice has an opportunity cost. If you sleep

through your economics class (not recommended, by the way), the opportunity cost is the learning you miss from not attending class. If you spend your income on video games, you cannot spend it on movies. If you choose to marry one person, you give up the opportunity to marry anyone else. In short, opportunity cost is all around us and part of human existence.

The following Work It Out feature shows a step-by-step analysis of a budget constraint calculation. Read through it to understand another important concept—slope—that is further explained in the appendix The Use of Mathematics in Principles of Economics.

UNDERSTANDING BUDGET CONSTRAINTS

Budget constraints are easy to understand if you apply a little math. The appendix The Use of Mathematics in Principles of Economics explains all the math you are likely to need in this book. So if math is not your strength, you might want to take a look at the appendix.

Step 1: The equation for any budget constraint is:

$$\text{Budget} = P_1 \times Q_1 + P_2 \times Q_2$$

where P and Q are the price and quantity of items purchased and Budget is the amount of income one has to spend.

Step 2. Apply the budget constraint equation to the scenario. In Alphonso's case, this works out to be:

$$\text{Budget} = P_1 \times Q_1 + P_2 \times Q_2$$

$$\text{\$10 budget} = \text{\$2 per burger} \times \text{quantity of burgers} + \text{\$0.50 per bus ticket} \times \text{quantity of bus tickets}$$

$$\text{\$10} = \text{\$2} \times Q_{\text{burgers}} + \text{\$0.50} \times Q_{\text{bus tickets}}$$

Step 3. Using a little algebra, we can turn this into the familiar equation of a line:

$$y = b + mx$$

For Alphonso, this is:

$$\text{\$10} = \text{\$2} \times Q_{\text{burgers}} + \text{\$0.50} \times Q_{\text{bus tickets}}$$

Step 4. Simplify the equation. Begin by multiplying both sides of the equation by 2:

$$2 \times \text{\$10} = 2 \times \text{\$2} \times Q_{\text{burgers}} + 2 \times \text{\$0.50} \times Q_{\text{bus tickets}}$$

$$\text{\$20} = \text{\$4} \times Q_{\text{burgers}} + \text{\$1} \times Q_{\text{bus tickets}}$$

Step 5. Subtract one bus ticket from both sides:

$$\text{\$20} - Q_{\text{bus tickets}} = \text{\$4} \times Q_{\text{burgers}}$$

Divide each side by 4 to yield the answer:

$$5 - 0.25 \times Q_{\text{bus tickets}} = Q_{\text{burgers}}$$

or

$$Q_{\text{burgers}} = 5 - 0.25 \times Q_{\text{bus tickets}}$$

Step 6. Notice that this equation fits the budget constraint in Figure 1. The vertical intercept is 5 and the slope is -0.25 , just as the equation says. If you plug 20 bus tickets into the equation, you get 0 burgers. If you plug other numbers of bus tickets into the equation, you get the results shown in Table 1, which are the points on Alphonso's budget constraint.

Point	Quantity of Burgers (at \$2)	Quantity of Bus Tickets (at 50 cents)
A	5	0
B	4	4
C	3	8
D	2	12
E	1	16
F	0	20

Table 1.

Step 7. Notice that the slope of a budget constraint always shows the opportunity cost of the good which is on the horizontal axis. For Alphonso, the slope is -0.25 , indicating that for every four bus tickets he buys, Alphonso must give up 1 burger.

There are two important observations here. First, the algebraic sign of the slope is negative, which means that the only way to get more of one good is to give up some of the other. Second, the slope is defined as the price of bus tickets (whatever is on the horizontal axis in the graph) divided by the price of burgers (whatever is on the vertical axis), in this case $\$0.50/\$2 = 0.25$. So if you want to determine the opportunity cost quickly, just divide the two prices.

IDENTIFYING OPPORTUNITY COST

In many cases, it is reasonable to refer to the opportunity cost as the **price**. If your cousin buys a new bicycle for \$300, then \$300 measures the amount of “other consumption” that he has given up. For practical purposes, there may be no special need to identify the specific alternative product or products that could have been bought with that \$300, but sometimes the price as measured in dollars may not accurately capture the true opportunity cost. This problem can loom especially large when costs of time are involved.

For example, consider a boss who decides that all employees will attend a two-day retreat to “build team spirit.” The out-of-pocket monetary cost of the event may involve hiring an outside consulting firm to run the retreat, as well as room and board for all participants. But an opportunity cost exists as well: during the two days of the retreat, none of the employees are doing any other work.

Attending college is another case where the opportunity cost exceeds the monetary cost. The out-of-pocket costs of attending college include tuition, books, room and board, and other expenses. But in addition, during the hours that you are attending class and studying, it is impossible to work at a paying job. Thus, college imposes both an out-of-pocket cost and an opportunity cost of lost earnings.

WHAT IS THE OPPORTUNITY COST ASSOCIATED WITH INCREASED AIRPORT SECURITY MEASURES?

After the terrorist plane hijackings on September 11, 2001, many steps were proposed to improve air travel safety. For example, the federal government could provide armed “sky marshals” who would travel inconspicuously with the rest of the passengers. The cost of having a sky marshal on every flight would be roughly \$3 billion per year. Retrofitting all U.S. planes with reinforced cockpit doors to make it harder for terrorists to take over the plane would have a price tag of \$450 million. Buying more sophisticated security equipment for airports, like three-dimensional baggage scanners and cameras linked to face recognition software, could cost another \$2 billion.

But the single biggest cost of greater airline security does not involve spending money. It is the opportunity cost of addi-

tional waiting time at the airport. According to the United States Department of Transportation (DOT), more than 800 million passengers took plane trips in the United States in 2012. Since the 9/11 hijackings, security screening has become more intensive, and consequently, the procedure takes longer than in the past. Say that, on average, each air passenger spends an extra 30 minutes in the airport per trip. Economists commonly place a value on time to convert an opportunity cost in time into a monetary figure. Because many air travelers are relatively high-paid business people, conservative estimates set the average price of time for air travelers at \$20 per hour. By these back-of-the-envelope calculations, the opportunity cost of delays in airports could be as much as $800 \text{ million} \times 0.5 \text{ hours} \times \$20/\text{hour}$, or \$8 billion per year. Clearly, the opportunity costs of waiting time can be just as important as costs that involve direct spending.

In some cases, realizing the opportunity cost can alter behavior. Imagine, for example, that you spend \$8 on lunch every day at work. You may know perfectly well that bringing a lunch from home would cost only \$3 a day, so the opportunity cost of buying lunch at the restaurant is \$5 each day (that is, the \$8 buying lunch costs minus the \$3 your lunch from home would cost). \$5 each day does not seem to be that much. However, if you project what that adds up to in a year—250 days a year \times \$5 per day equals \$1,250, the cost, perhaps, of a decent vacation. If the opportunity cost is described as “a nice vacation” instead of “\$5 a day,” you might make different choices.

MARGINAL DECISION-MAKING AND DIMINISHING MARGINAL UTILITY

The budget constraint framework helps to emphasize that most choices in the real world are not about getting all of one thing or all of another; that is, they are not about choosing either the point at one end of the budget constraint or else the point all the way at the other end. Instead, most choices involve **marginal analysis**, which means comparing the benefits and costs of choosing a little more or a little less of a good.

People desire goods and services for the satisfaction or **utility** those goods and services provide. Utility, as we will see in the chapter on Consumer Choices, is subjective but that does not make it less real. Economists typically assume that the more of some good one consumes (for example, slices of pizza), the more utility one obtains. At the same time, the utility a person receives from consuming the first unit of a good is typically more than the utility received from consuming the fifth or the tenth unit of that same good. When Alphonso chooses between burgers and bus tickets, for example, the first few bus rides that he chooses might provide him with a great deal of utility—perhaps they help him get to a job interview or a doctor’s appointment. But later bus rides might provide much less utility—they may only serve to kill time on a rainy day. Similarly, the first burger that Alphonso chooses to buy may be on a day when he missed breakfast and is ravenously hungry. However, if Alphonso has a burger every single day, the last few burgers may taste pretty boring. The general pattern that consumption of the first few units of any good tends to bring a higher level of utility to a person than consumption of later units is a common pattern. Economists refer to this pattern as the **law of diminishing marginal utility**, which means that as a person receives more of a good, the additional (or marginal) utility from each additional unit of the good declines. In other words, the first slice of pizza brings more satisfaction than the sixth.

The law of diminishing marginal utility explains why people and societies rarely make all-or-nothing choices. You would not say, “My favorite food is ice cream, so I will eat nothing but ice cream from now on.” Instead, even if you get a very high level of utility from your favorite food, if you ate it exclusively, the additional or marginal utility from those last few servings would not be very high. Similarly, most workers do not say: “I enjoy leisure, so I’ll never work.” Instead, workers recognize that

even though some leisure is very nice, a combination of all leisure and no income is not so attractive. The budget constraint framework suggests that when people make choices in a world of scarcity, they will use marginal analysis and think about whether they would prefer a little more or a little less.

SUNK COSTS

In the budget constraint framework, all decisions involve what will happen next: that is, what quantities of goods will you consume, how many hours will you work, or how much will you save. These decisions do not look back to past choices. Thus, the budget constraint framework assumes that **sunk costs**, which are costs that were incurred in the past and cannot be recovered, should not affect the current decision.

Consider the case of Selena, who pays \$8 to see a movie, but after watching the film for 30 minutes, she knows that it is truly terrible. Should she stay and watch the rest of the movie because she paid for the ticket, or should she leave? The money she spent is a sunk cost, and unless the theater manager is feeling kindly, Selena will not get a refund. But staying in the movie still means paying an opportunity cost in time. Her choice is whether to spend the next 90 minutes suffering through a cinematic disaster or to do something—anything—else. The lesson of sunk costs is to forget about the money and time that is irretrievably gone and instead to focus on the marginal costs and benefits of current and future options.

For people and firms alike, dealing with sunk costs can be frustrating. It often means admitting an earlier error in judgment. Many firms, for example, find it hard to give up on a new product that is doing poorly because they spent so much money in creating and launching the product. But the lesson of sunk costs is to ignore them and make decisions based on what will happen in the future.

FROM A MODEL WITH TWO GOODS TO ONE OF MANY GOODS

The budget constraint diagram containing just two goods, like most models used in this book, is not realistic. After all, in a modern economy people choose from thousands of goods. However, thinking about a model with many goods is a straightforward extension of what we discussed here. Instead of drawing just one budget constraint, showing the tradeoff between two goods, you can draw multiple budget constraints, showing the possible tradeoffs between many different pairs of goods. Or in more advanced classes in economics, you would use mathematical equations that include many possible goods and services that can be purchased, together with their quantities and prices, and show how the total spending on all goods and services is limited to the overall budget available. The graph with two goods that was presented here clearly illustrates that every choice has an opportunity cost, which is the point that does carry over to the real world.

KEY CONCEPTS AND SUMMARY

Economists see the real world as one of scarcity: that is, a world in which people's desires exceed what is possible. As a result, economic behavior involves tradeoffs in which individuals, firms, and society must give up something that they desire to obtain things that they desire more. Individuals face the tradeoff of what quantities of goods and services to consume. The budget constraint, which is the frontier of the opportunity set, illustrates the range of choices available. The slope of the budget constraint is determined by the relative price of the choices. Choices beyond the budget constraint are not affordable.

Opportunity cost measures cost by what is given up in exchange. Sometimes opportunity cost can be measured in money, but it is often useful to consider time as well, or to measure it in terms of the actual resources that must be given up.

Most economic decisions and tradeoffs are not all-or-nothing. Instead, they involve marginal analysis, which means they are about decisions on the margin, involving a little more or a little less. The law of diminishing marginal utility points out that as a person receives more of something—whether it is a specific good or another resource—the additional marginal gains tend to become smaller. Because sunk costs occurred in the past and cannot be recovered, they should be disregarded in making current decisions.

SELF-CHECK QUESTIONS

Suppose Alphonso's town raised the price of bus tickets to \$1 per trip (while the price of burgers stayed at \$2 and his budget remained \$10 per week.) Draw Alphonso's new budget constraint. What happens to the opportunity cost of bus tickets?

REVIEW QUESTIONS

1. Explain why scarcity leads to tradeoffs.
2. Explain why individuals make choices that are directly on the budget constraint, rather than inside the budget constraint or outside it.

CRITICAL THINKING QUESTIONS

Suppose Alphonso's town raises the price of bus tickets from \$0.50 to \$1 and the price of burgers rises from \$2 to \$4. Why is the opportunity cost of bus tickets unchanged? Suppose Alphonso's weekly spending money increases from \$10 to \$20. How is his budget constraint affected from all three changes? Explain.

PROBLEMS

Use this information to answer the following 4 questions: Marie has a weekly budget of \$24, which she likes to spend on magazines and pies.

1. If the price of a magazine is \$4 each, what is the maximum number of magazines she could buy in a week?
2. If the price of a pie is \$12, what is the maximum number of pies she could buy in a week?
3. Draw Marie's budget constraint with pies on the horizontal axis and magazines on the vertical axis. What is the slope of the budget constraint?
4. What is Marie's opportunity cost of purchasing a pie?

REFERENCES

Bureau of Labor Statistics, U.S. Department of Labor. 2015. "Median Weekly Earnings by Educational Attainment in 2014." Accessed March 27, 2015. <http://www.bls.gov/opub/ted/2015/median-weekly-earnings-by-education-gender-race-and-ethnicity-in-2014.htm>.

Robbins, Lionel. *An Essay on the Nature and Significance of Economic Science*. London: Macmillan. 1932.

United States Department of Transportation. "Total Passengers on U.S Airlines and Foreign Airlines U.S. Flights Increased 1.3% in 2012 from 2011." Accessed October 2013. http://www.rita.dot.gov/bts/press_releases/bts016_13

GLOSSARY

budget constraint all possible consumption combinations of goods that someone can afford, given the prices of goods, when all income is spent; the boundary of the opportunity set

law of diminishing marginal utility as we consume more of a good or service, the utility we get from additional units of the good or service tend to become smaller than what we received from earlier units

marginal analysis examination of decisions on the margin, meaning a little more or a little less from the status quo

opportunity cost measures cost by what is given up in exchange; opportunity cost measures the value of the forgone alternative

opportunity set all possible combinations of consumption that someone can afford given the prices of goods and the individual's income

sunk costs costs that are made in the past and cannot be recovered

utility satisfaction, usefulness, or value one obtains from consuming goods and services

SOLUTIONS

Answers to Self-Check Questions

The opportunity cost of bus tickets is the number of burgers that must be given up to obtain one more bus ticket. Originally, when the price of bus tickets was 50 cents per trip, this opportunity cost was $0.50/2 = .25$ burgers. The reason for this is that at the original prices, one burger (\$2) costs the same as four bus tickets (\$0.50), so the opportunity cost of a burger is four bus tickets, and the opportunity cost of a bus ticket is .25 (the inverse of the opportunity cost of a burger). With the new, higher price of bus tickets, the opportunity cost rises to $\$1/\2 or 0.50. You can see this graphically since the slope of the new budget constraint is flatter than the original one. If Alphonso spends all of his budget on burgers, the higher price of bus tickets has no impact so the horizontal intercept of the budget constraint is the same. If he spends all of his budget on bus tickets, he can now afford only half as many, so the vertical intercept is half as much. In short, the budget constraint rotates clockwise around the horizontal intercept, flattening as it goes and the opportunity cost of bus tickets increases.

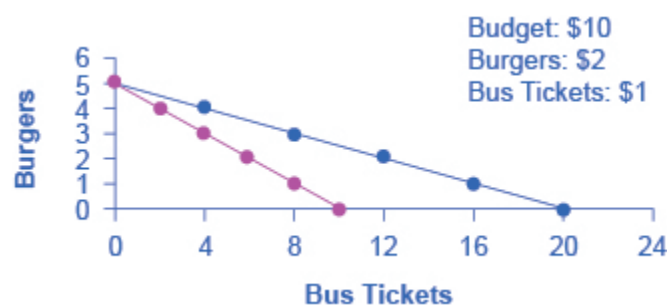


Figure 2.

2.2 THE PRODUCTION POSSIBILITIES FRONTIER AND SOCIAL CHOICES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Interpret production possibilities frontier graphs
- Contrast a budget constraint and a production possibilities frontier
- Explain the relationship between a production possibilities frontier and the law of diminishing returns
- Contrast productive efficiency and allocative efficiency
- Define comparative advantage

Just as individuals cannot have everything they want and must instead make choices, society as a whole cannot have everything it might want, either. This section of the chapter will explain the constraints faced by society, using a model called the **production possibilities frontier (PPF)**. There are more similarities than differences between individual choice and social choice. As you read this section, focus on the similarities.

Because society has limited resources (e.g., labor, land, capital, raw materials) at any point in time, there is a limit to the quantities of goods and services it can produce. Suppose a society desires two products, healthcare and education. This situation is illustrated by the production possibilities frontier in Figure 1.

In Figure 1, healthcare is shown on the vertical axis and education is shown on the horizontal axis. If the society were to allocate all of its resources to healthcare, it could produce at point A. But it would not have any resources to produce education. If it were to allocate all of its resources to education, it could produce at point F. Alternatively, the society could choose to produce any combination of healthcare and education shown on the production possibilities frontier. In effect, the production possibilities frontier plays the same role for society as the budget constraint plays for Alphonso. Society can choose any combination of the two goods on or inside the PPF. But it does not have enough resources to produce outside the PPF.

Most important, the production possibilities frontier clearly shows the tradeoff between healthcare and education. Suppose society has chosen to operate at point B, and it is considering producing more education. Because the PPF is downward sloping from left to right, the only way society can obtain more education is by giving up some healthcare. That is the tradeoff society faces. Suppose it consid-

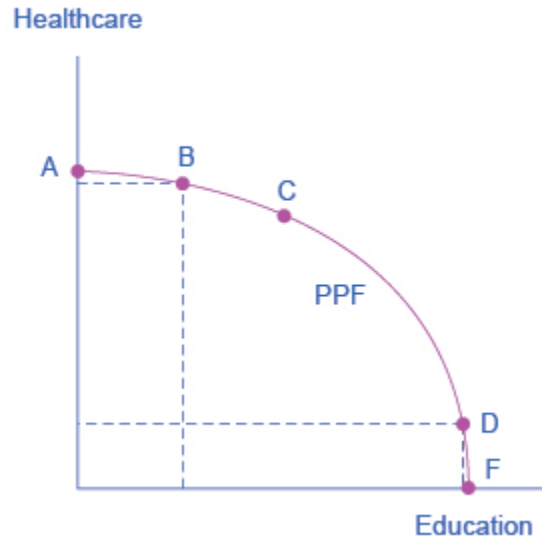


Figure 1. A Healthcare vs. Education Production Possibilities Frontier. This production possibilities frontier shows a tradeoff between devoting social resources to healthcare and devoting them to education. At A all resources go to healthcare and at B, most go to healthcare. At D most resources go to education, and at F, all go to education.

ers moving from point B to point C. What would the opportunity cost be for the additional education? The opportunity cost would be the healthcare society has to give up. Just as with Alphonso's budget constraint, the opportunity cost is shown by the **slope** of the production possibilities frontier. By now you might be saying, "Hey, this PPF is sounding like the budget constraint." If so, read the following Clear It Up feature.

WHAT'S THE DIFFERENCE BETWEEN A BUDGET CONSTRAINT AND A PPF?

There are two major differences between a budget constraint and a production possibilities frontier. The first is the fact that the budget constraint is a straight line. This is because its slope is given by the relative prices of the two goods. In contrast, the PPF has a curved shape because of the law of the diminishing returns. The second is the absence of specific numbers on the axes of the PPF. There are no specific numbers because we do not know the exact amount of resources this imaginary economy has, nor do we know how many resources it takes to produce healthcare and how many resources it takes to produce education. If this were a real world example, that data would be available. An additional reason for the lack of numbers is that there is no single way to measure levels of education and healthcare. However, when you think of improvements in education, you can think of accomplishments like more years of school completed, fewer high-school dropouts, and higher scores on standardized tests. When you think of improvements in healthcare, you can think of longer life expectancies, lower levels of infant mortality, and fewer outbreaks of disease.

Whether or not we have specific numbers, conceptually we can measure the opportunity cost of additional education as society moves from point B to point C on the PPF. The additional education is measured by the horizontal distance between B and C. The foregone healthcare is given by the vertical distance between B and C. The slope of the PPF between B and C is (approximately) the vertical distance (the "rise") over the horizontal distance (the "run"). This is the opportunity cost of the additional education.

THE SHAPE OF THE PPF AND THE LAW OF DIMINISHING RETURNS

The budget constraints presented earlier in this chapter, showing individual choices about what quantities of goods to consume, were all straight lines. The reason for these straight lines was that the slope of the budget constraint was determined by relative prices of the two goods in the **consumption budget constraint**. However, the production possibilities frontier for healthcare and education was drawn as a curved line. Why does the PPF have a different shape?

To understand why the PPF is curved, start by considering point A at the top left-hand side of the PPF. At point A, all available resources are devoted to healthcare and none are left for education. This situation would be extreme and even ridiculous. For example, children are seeing a doctor every day, whether they are sick or not, but not attending school. People are having cosmetic surgery on every part of their bodies, but no high school or college education exists. Now imagine that some of these resources are diverted from healthcare to education, so that the economy is at point B instead of point A. Diverting some resources away from A to B causes relatively little reduction in health because the last few marginal dollars going into healthcare services are not producing much additional gain in health. However, putting those marginal dollars into education, which is completely without resources at point A, can produce relatively large gains. For this reason, the shape of the PPF from A to B is relatively flat, representing a relatively small drop-off in health and a relatively large gain in education.

Now consider the other end, at the lower right, of the production possibilities frontier. Imagine that society starts at choice D, which is devoting nearly all resources to education and very few to healthcare, and moves to point F, which is devoting *all* spending to education and none to healthcare. For the sake of concreteness, you can imagine that in the movement from D to F, the last few doctors must become high school science teachers, the last few nurses must become school librarians rather than dispensers of vaccinations, and the last few emergency rooms are turned into kindergartens. The gains to education from adding these last few resources to education are very small. However, the opportunity cost lost to health will be fairly large, and thus the slope of the PPF between D and F is steep, showing a large drop in health for only a small gain in education.

The lesson is not that society is likely to make an extreme choice like devoting no resources to education at point A or no resources to health at point F. Instead, the lesson is that the gains from committing additional marginal resources to education depend on how much is already being spent. If on the one hand, very few resources are currently committed to education, then an increase in resources used can bring relatively large gains. On the other hand, if a large number of resources are already committed to education, then committing additional resources will bring relatively smaller gains.

This pattern is common enough that it has been given a name: the **law of diminishing returns**, which holds that as additional increments of resources are added to a certain purpose, the marginal benefit from those additional increments will decline. When government spends a certain amount more on reducing crime, for example, the original gains in reducing crime could be relatively large. But additional increases typically cause relatively smaller reductions in crime, and paying for enough police and security to reduce crime to nothing at all would be tremendously expensive.

The curvature of the production possibilities frontier shows that as additional resources are added to education, moving from left to right along the horizontal axis, the original gains are fairly large, but gradually diminish. Similarly, as additional resources are added to healthcare, moving from bottom to top on the vertical axis, the original gains are fairly large, but again gradually diminish. In this way,

the law of diminishing returns produces the outward-bending shape of the production possibilities frontier.

PRODUCTIVE EFFICIENCY AND ALLOCATIVE EFFICIENCY

The study of economics does not presume to tell a society what choice it should make along its production possibilities frontier. In a market-oriented economy with a democratic government, the choice will involve a mixture of decisions by individuals, firms, and government. However, economics can point out that some choices are unambiguously better than others. This observation is based on the concept of efficiency. In everyday usage, efficiency refers to lack of waste. An inefficient machine operates at high **cost**, while an efficient machine operates at lower cost, because it is not wasting energy or materials. An inefficient organization operates with long delays and high costs, while an efficient organization meets schedules, is focused, and performs within budget.

The production possibilities frontier can illustrate two kinds of efficiency: productive efficiency and allocative efficiency. Figure 2 illustrates these ideas using a production possibilities frontier between healthcare and education.

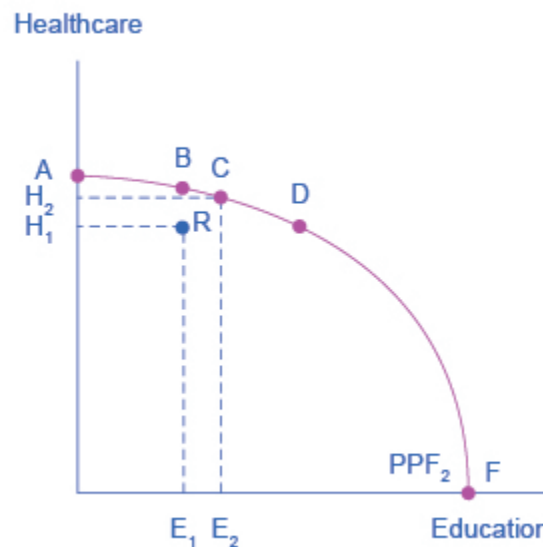


Figure 2. Productive and Allocative Efficiency. Productive efficiency means it is impossible to produce more of one good without decreasing the quantity that is produced of another good.

Thus, all choices along a given PPF like B, C, and D display productive efficiency, but R does not. Allocative efficiency means that the particular mix of goods being produced—that is, the specific choice along the production possibilities frontier—represents the allocation that society most desires.

Productive efficiency means that, given the available inputs and technology, it is impossible to produce more of one good without decreasing the quantity that is produced of another good. All choices on the PPF in Figure 2, including A, B, C, D, and F, display productive efficiency. As a firm moves from any one of these choices to any other, either healthcare increases and education decreases or vice

versa. However, any choice inside the production possibilities frontier is productively inefficient and wasteful because it is possible to produce more of one good, the other good, or some combination of both goods.

For example, point R is productively inefficient because it is possible at choice C to have more of both goods: education on the horizontal axis is higher at point C than point R (E_2 is greater than E_1), and healthcare on the vertical axis is also higher at point C than point R (H_2 is greater than H_1).

The particular mix of goods and services being produced—that is, the specific combination of healthcare and education chosen along the production possibilities frontier—can be shown as a ray (line) from the origin to a specific point on the PPF. Output mixes that had more healthcare (and less education) would have a steeper ray, while those with more education (and less healthcare) would have a flatter ray.

Allocative efficiency means that the particular mix of goods a society produces represents the combination that society most desires. How to determine what a society desires can be a controversial question, and is usually discussed in political science, sociology, and philosophy classes as well as in economics. At its most basic, allocative efficiency means producers supply the quantity of each product that consumers demand. Only one of the productively efficient choices will be the allocatively efficient choice for society as a whole.

WHY SOCIETY MUST CHOOSE

Every economy faces two situations in which it may be able to expand consumption of all goods. In the first case, a society may discover that it has been using its resources inefficiently, in which case by improving efficiency and producing on the production possibilities frontier, it can have more of all goods (or at least more of some and less of none). In the second case, as resources grow over a period of years (e.g., more labor and more capital), the economy grows. As it does, the production possibilities frontier for a society will tend to shift outward and society will be able to afford more of all goods.

But improvements in productive efficiency take time to discover and implement, and economic growth happens only gradually. So, a society must choose between tradeoffs in the present. For government, this process often involves trying to identify where additional spending could do the most good and where reductions in spending would do the least harm. At the individual and **firm** level, the **market economy** coordinates a process in which firms seek to produce goods and services in the quantity, quality, and price that people want. But for both the government and the market economy in the short term, increases in production of one good typically mean offsetting decreases somewhere else in the economy.

THE PPF AND COMPARATIVE ADVANTAGE

While every society must choose how much of each good it should produce, it does not need to produce every single good it consumes. Often how much of a good a country decides to produce depends on how expensive it is to produce it versus buying it from a different country. As we saw earlier, the curvature of a country's PPF gives us information about the tradeoff between devoting resources to producing one good versus another. In particular, its slope gives the opportunity cost of producing one more unit of the good in the x-axis in terms of the other good (in the y-axis). Countries tend to

have different opportunity costs of producing a specific good, either because of different climates, geography, technology or skills.

Suppose two countries, the US and Brazil, need to decide how much they will produce of two crops: sugar cane and wheat. Due to its climatic conditions, Brazil can produce a lot of sugar cane per acre but not much wheat. Conversely, the U.S. can produce a lot of wheat per acre, but not much sugar cane. Clearly, Brazil has a lower opportunity cost of producing sugar cane (in terms of wheat) than the U.S. The reverse is also true; the U.S. has a lower opportunity cost of producing wheat than Brazil. This can be illustrated by the PPFs of the two countries in Figure 3.

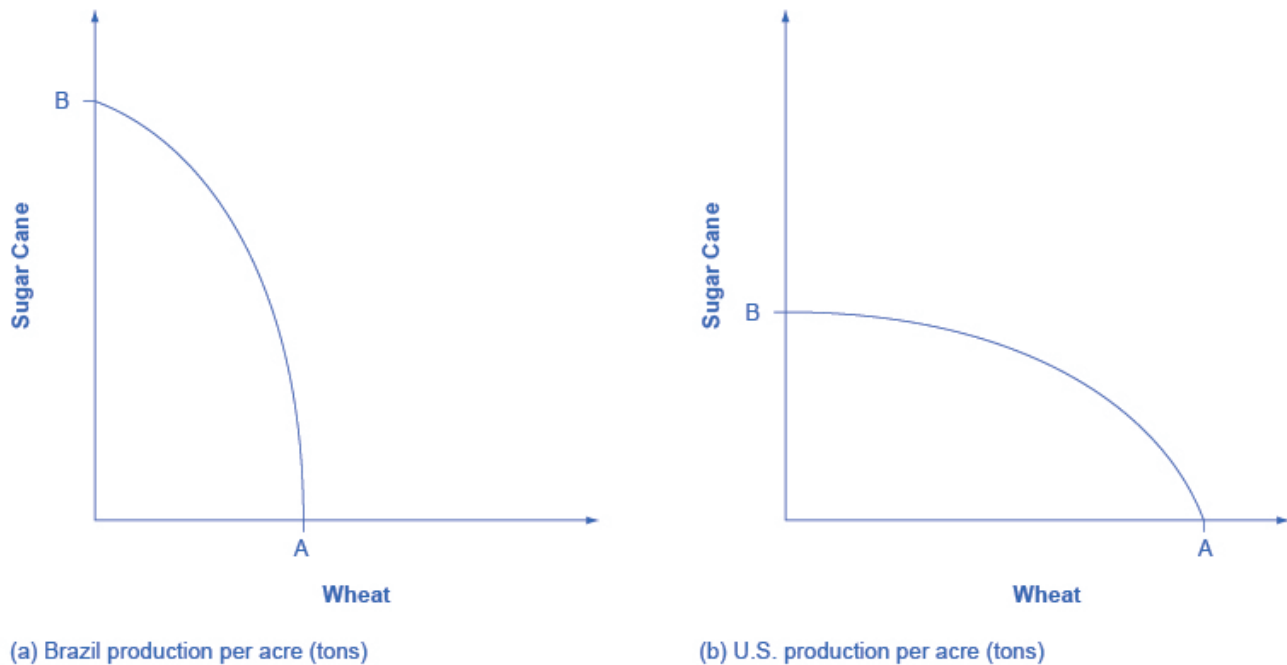


Figure 3. Production Possibility Frontier for the U.S. and Brazil. The U.S. PPF is flatter than the Brazil PPF implying that the opportunity cost of wheat in term of sugar cane is lower in the U.S. than in Brazil. Conversely, the opportunity cost of sugar cane is lower in Brazil. The U.S. has comparative advantage in wheat and Brazil has comparative advantage in sugar cane.

When a country can produce a good at a lower opportunity cost than another country, we say that this country has a **comparative advantage** in that good. In our example, Brazil has a comparative advantage in sugar cane and the U.S. has a comparative advantage in wheat. One can easily see this with a simple observation of the extreme production points in the PPFs of the two countries. If Brazil devoted all of its resources to producing wheat, it would be producing at point A. If however it had devoted all of its resources to producing sugar cane instead, it would be producing a much larger amount, at point B. By moving from point A to point B Brazil would give up a relatively small quantity in wheat production to obtain a large production in sugar cane. The opposite is true for the U.S. If the U.S. moved from point A to B and produced only sugar cane, this would result in a large opportunity cost in terms of foregone wheat production.

The slope of the PPF gives the opportunity cost of producing an additional unit of wheat. While the slope is not constant throughout the PPFs, it is quite apparent that the PPF in Brazil is much steeper than in the U.S., and therefore the opportunity cost of wheat generally higher in Brazil. In the chapter

on International Trade you will learn that countries' differences in comparative advantage determine which goods they will choose to produce and trade. When countries engage in trade, they specialize in the production of the goods that they have comparative advantage in, and trade part of that production for goods they do not have comparative advantage in. With trade, goods are produced where the opportunity cost is lowest, so total production increases, benefiting both trading parties.

KEY CONCEPTS AND SUMMARY

A production possibilities frontier defines the set of choices society faces for the combinations of goods and services it can produce given the resources available. The shape of the PPF is typically curved outward, rather than straight. Choices outside the PPF are unattainable and choices inside the PPF are wasteful. Over time, a growing economy will tend to shift the PPF outwards.

The law of diminishing returns holds that as increments of additional resources are devoted to producing something, the marginal increase in output will become smaller and smaller. All choices along a production possibilities frontier display productive efficiency; that is, it is impossible to use society's resources to produce more of one good without decreasing production of the other good. The specific choice along a production possibilities frontier that reflects the mix of goods society prefers is the choice with allocative efficiency. The curvature of the PPF is likely to differ by country, which results in different countries having comparative advantage in different goods. Total production can increase if countries specialize in the goods they have comparative advantage in and trade some of their production for the remaining goods.

SELF-CHECK QUESTIONS

1. Return to the example in Figure 2. Suppose there is an improvement in medical technology that enables more healthcare to be provided with the same amount of resources. How would this affect the production possibilities curve and, in particular, how would it affect the opportunity cost of education?
2. Could a nation be producing in a way that is allocatively efficient, but productively inefficient?
3. What are the similarities between a consumer's budget constraint and society's production possibilities frontier, not just graphically but analytically?

REVIEW QUESTIONS

1. What is comparative advantage?
2. What does a production possibilities frontier illustrate?
3. Why is a production possibilities frontier typically drawn as a curve, rather than a straight line?
4. Explain why societies cannot make a choice above their production possibilities frontier and should not make a choice below it.
5. What are diminishing marginal returns?
6. What is productive efficiency? Allocative efficiency?

CRITICAL THINKING QUESTIONS

1. During the Second World War, Germany's factories were decimated. It also suffered many human casualties, both soldiers and civilians. How did the war affect Germany's production possibilities curve?
2. It is clear that productive inefficiency is a waste since resources are being used in a way that produces less goods and services than a nation is capable of. Why is allocative inefficiency also wasteful?

GLOSSARY

allocative efficiency when the mix of goods being produced represents the mix that society most desires

comparative advantage when a country can produce a good at a lower cost in terms of other goods; or, when a country has a lower opportunity cost of production

law of diminishing returns as additional increments of resources are added to producing a good or service, the marginal benefit from those additional increments will decline

production possibilities frontier (PPF) a diagram that shows the productively efficient combinations of two products that an economy can produce given the resources it has available.

productive efficiency when it is impossible to produce more of one good (or service) without decreasing the quantity produced of another good (or service)

SOLUTIONS

Answers to Self-Check Questions

1. Because of the improvement in technology, the vertical intercept of the PPF would be at a higher level of healthcare. In other words, the PPF would rotate clockwise around the horizontal intercept. This would make the PPF steeper, corresponding to an increase in the opportunity cost of education, since resources devoted to education would now mean forgoing a greater quantity of healthcare.
2. No. Allocative efficiency requires productive efficiency, because it pertains to choices along the production possibilities frontier.
3. Both the budget constraint and the PPF show the constraint that each operates under. Both show a tradeoff between having more of one good but less of the other. Both show the opportunity cost graphically as the slope of the constraint (budget or PPF).

2.3 CONFRONTING OBJECTIONS TO THE ECONOMIC APPROACH

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Analyze arguments against economic approaches to decision-making
- Interpret a tradeoff diagram
- Contrast normative statements and positive statements

It is one thing to understand the economic approach to decision-making and another thing to feel comfortable applying it. The sources of discomfort typically fall into two categories: that people do not act in the way that fits the economic way of thinking, and that even if people did act that way, they should try not to. Let's consider these arguments in turn.

FIRST OBJECTION: PEOPLE, FIRMS, AND SOCIETY DO NOT ACT LIKE THIS

The economic approach to decision-making seems to require more information than most individuals possess and more careful decision-making than most individuals actually display. After all, do you or any of your friends draw a budget constraint and mutter to yourself about maximizing utility before you head to the shopping mall? Do members of the U.S. Congress contemplate production possibilities frontiers before they vote on the annual budget? The messy ways in which people and societies operate somehow doesn't look much like neat budget constraints or smoothly curving production possibilities frontiers.

However, the economics approach can be a useful way to analyze and understand the tradeoffs of economic decisions even so. To appreciate this point, imagine for a moment that you are playing basketball, dribbling to the right, and throwing a bounce-pass to the left to a teammate who is running toward the basket. A physicist or engineer could work out the correct speed and trajectory for the pass, given the different movements involved and the weight and bounciness of the ball. But when you are playing basketball, you do not perform any of these calculations. You just pass the ball, and if you are a good player, you will do so with high accuracy.

Someone might argue: "The scientist's formula of the bounce-pass requires a far greater knowledge of physics and far more specific information about speeds of movement and weights than the basketball player actually has, so it must be an unrealistic description of how basketball passes are actually made." This reaction would be wrongheaded. The fact that a good player can throw the ball accurately

because of practice and skill, without making a physics calculation, does not mean that the physics calculation is wrong.

Similarly, from an economic point of view, someone who goes shopping for groceries every week has a great deal of practice with how to purchase the combination of goods that will provide that person with utility, even if the shopper does not phrase decisions in terms of a budget constraint. Government institutions may work imperfectly and slowly, but in general, a democratic form of government feels pressure from voters and social institutions to make the choices that are most widely preferred by people in that society. So, when thinking about the economic actions of groups of people, firms, and society, it is reasonable, as a first approximation, to analyze them with the tools of economic analysis. For more on this, read about behavioral economics in the chapter on Consumer Choices.

SECOND OBJECTION: PEOPLE, FIRMS, AND SOCIETY SHOULD NOT ACT THIS WAY

The economics approach portrays people as self-interested. For some critics of this approach, even if self-interest is an accurate description of how people behave, these behaviors are not moral. Instead, the critics argue that people should be taught to care more deeply about others. Economists offer several answers to these concerns.

First, economics is not a form of moral instruction. Rather, it seeks to describe economic behavior as it actually exists. Philosophers draw a distinction between **positive statements**, which describe the world as it is, and **normative statements**, which describe how the world should be. For example, an economist could analyze a proposed subway system in a certain city. If the expected benefits exceed the costs, he concludes that the project is worth doing—an example of positive analysis. Another economist argues for extended unemployment compensation during the Great Depression because a rich country like the United States should take care of its less fortunate citizens—an example of normative analysis.

Even if the line between positive and normative statements is not always crystal clear, economic analysis does try to remain rooted in the study of the actual people who inhabit the actual economy. Fortunately however, the assumption that individuals are purely self-interested is a simplification about human nature. In fact, we need to look no further than to **Adam Smith**, the very father of modern economics to find evidence of this. The opening sentence of his book, *The Theory of Moral Sentiments*, puts it very clearly: “How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.” Clearly, individuals are both self-interested and altruistic.

Second, self-interested behavior and profit-seeking can be labeled with other names, such as personal choice and freedom. The ability to make personal choices about buying, working, and saving is an important personal freedom. Some people may choose high-pressure, high-paying jobs so that they can earn and spend a lot of money on themselves. Others may earn a lot of money and give it to charity or spend it on their friends and family. Others may devote themselves to a career that can require a great deal of time, energy, and expertise but does not offer high financial rewards, like being an elementary school teacher or a social worker. Still others may choose a job that does not take lots of their time or provide a high level of income, but still leaves time for family, friends, and contemplation. Some people may prefer to work for a large company; others might want to start their own business. People’s freedom to make their own economic choices has a moral value worth respecting.

IS A DIAGRAM BY ANY OTHER NAME THE SAME?

When you study economics, you may feel buried under an avalanche of diagrams: diagrams in the text, diagrams in the lectures, diagrams in the problems, and diagrams on exams. Your goal should be to recognize the common underlying logic and pattern of the diagrams, not to memorize each of the individual diagrams.

This chapter uses only one basic diagram, although it is presented with different sets of labels. The consumption budget constraint and the production possibilities frontier for society, as a whole, are the same basic diagram. Figure 1 shows an individual budget constraint and a production possibilities frontier for two goods, Good 1 and Good 2. The tradeoff diagram always illustrates three basic themes: scarcity, tradeoffs, and economic efficiency.

The first theme is **scarcity**. It is not feasible to have unlimited amounts of both goods. But even if the budget constraint or a PPF shifts, scarcity remains—just at a different level. The second theme is **tradeoffs**. As depicted in the budget constraint or the production possibilities frontier, it is necessary to give up some of one good to gain more of the other good. The details of this tradeoff vary. In a budget constraint, the tradeoff is determined by the relative prices of the goods: that is, the relative price of two goods in the consumption choice budget constraint. These tradeoffs appear as a straight line. However, the tradeoffs in many production possibilities frontiers are represented by a curved line because the law of diminishing returns holds that as resources are added to an area, the marginal gains tend to diminish. Regardless of the specific shape, tradeoffs remain.

The third theme is **economic efficiency**, or getting the most benefit from scarce resources. All choices on the production possibilities frontier show productive efficiency because in such cases, there is no way to increase the quantity of one good without decreasing the quantity of the other. Similarly, when an individual makes a choice along a budget constraint, there is no way to increase the quantity of one good without decreasing the quantity of the other. The choice on a production possibilities set that is socially preferred, or the choice on an individual's budget constraint that is personally preferred, will display allocative efficiency.

The basic budget constraint/production possibilities frontier diagram will recur throughout this book. Some examples include using these tradeoff diagrams to analyze trade, labor supply versus leisure, saving versus consumption, environmental protection and economic output, equality of incomes and economic output, and the macroeconomic tradeoff between consumption and investment. Do not be confused by the different labels. The budget constraint/production possibilities frontier diagram is always just a tool for thinking carefully about scarcity, tradeoffs, and efficiency in a particular situation.

Third, self-interested behavior can lead to positive social results. For example, when people work hard to make a living, they create economic output. Consumers who are looking for the best deals will encourage businesses to offer goods and services that meet their needs. Adam Smith, writing in *The Wealth of Nations*, christened this property the **invisible hand**. In describing how consumers and producers interact in a market economy, Smith wrote:

Every individual...generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. By preferring the support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain. And he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention...By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.

The metaphor of the invisible hand suggests the remarkable possibility that broader social good can emerge from selfish individual actions.

Fourth, even people who focus on their own self-interest in the economic part of their life often set

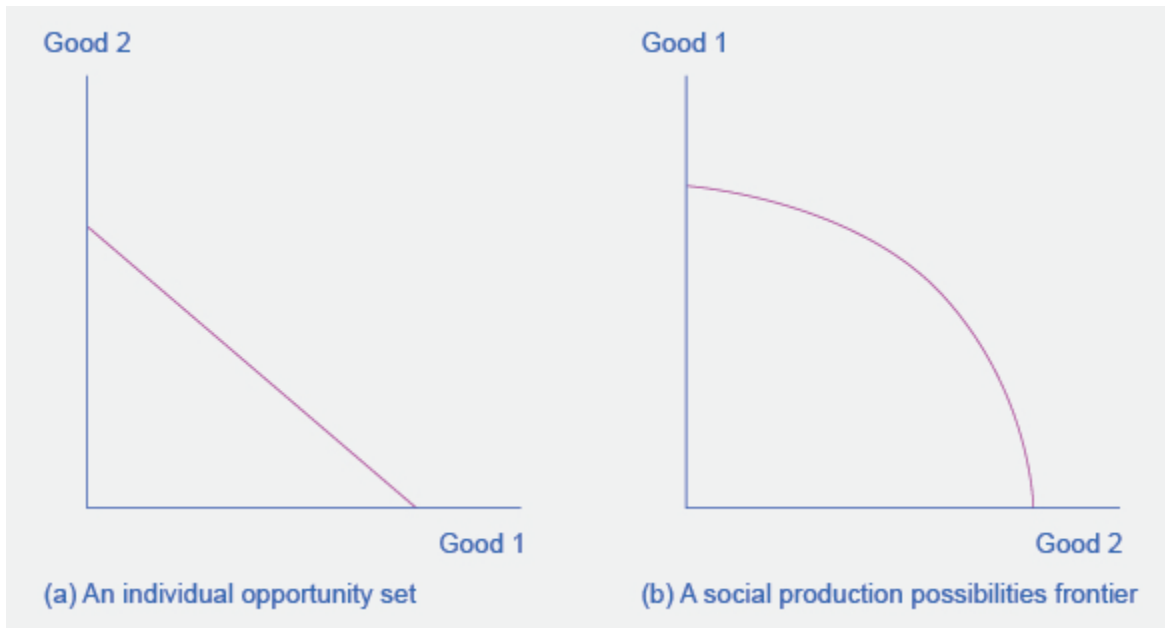


Figure 1. The Tradeoff Diagram. Both the individual opportunity set (or budget constraint) and the social production possibilities frontier show the constraints under which individual consumers and society as a whole operate. Both diagrams show the tradeoff in choosing more of one good at the cost of less of the other.

aside their own narrow self-interest in other parts of life. For example, you might focus on your own self-interest when asking your employer for a raise or negotiating to buy a car. But then you might turn around and focus on other people when you volunteer to read stories at the local library, help a friend move to a new apartment, or donate money to a charity. Self-interest is a reasonable starting point for analyzing many economic decisions, without needing to imply that people never do anything that is not in their own immediate self-interest.

CHOICES ... TO WHAT DEGREE?

What have we learned? We know that scarcity impacts all the choices we make. So, an economist might argue that people do not go on to get bachelor's degrees or master's degrees because they do not have the resources to make those choices or because their incomes are too low and/or the price of these degrees is too high. A bachelor's degree or a master's degree may not be available in their opportunity set.

The price of these degrees may be too high not only because the actual price, college tuition (and perhaps room and board), is too high. An economist might also say that for many people, the full opportunity cost of a bachelor's degree or a master's degree is too high. For these people, they are unwilling or unable to make the tradeoff of giving up years of working, and earning an income, to earn a degree.

Finally, the statistics introduced at the start of the chapter reveal information about **intertemporal choices**. An economist might say that people choose not to get a college degree because they may have to borrow money to go to college, and the interest they have to pay on that loan in the future will affect their decisions today. Also, it could be that some people have a preference for current consumption over future consumption, so they choose to work now at a lower salary and consume now, rather than putting that consumption off until after they graduate college.

KEY CONCEPTS AND SUMMARY

The economic way of thinking provides a useful approach to understanding human behavior. Economists make the careful distinction between positive statements, which describe the world as it is, and normative statements, which describe how the world should be. Even when economics analyzes the gains and losses from various events or policies, and thus draws normative conclusions about how the world should be, the analysis of economics is rooted in a positive analysis of how people, firms, and governments actually behave, not how they should behave.

SELF-CHECK QUESTIONS

1. Individuals may not act in the rational, calculating way described by the economic model of decision making, measuring utility and costs at the margin, but can you make a case that they behave approximately that way?
2. Would an op-ed piece in a newspaper urging the adoption of a particular economic policy be considered a positive or normative statement?
3. Would a research study on the effects of soft drink consumption on children's cognitive development be considered a positive or normative statement?

REVIEW QUESTIONS

1. What is the difference between a positive and a normative statement?
2. Is the economic model of decision-making intended as a literal description of how individuals, firms, and the governments actually make decisions?
3. What are four responses to the claim that people should not behave in the way described in this chapter?

CRITICAL THINKING QUESTIONS

1. What assumptions about the economy must be true for the invisible hand to work? To what extent are those assumptions valid in the real world?
2. Do economists have any particular expertise at making normative arguments? In other words, they have expertise at making positive statements (i.e., what *will* happen) about some economic policy, for example, but do they have special expertise to judge whether or not the policy *should* be undertaken?

REFERENCES

Smith, Adam. "Of Restraints upon the Importation from Foreign Countries." In *The Wealth of Nations*. London: Methuen & Co., 1904, first pub 1776), I.V. 2.9.

Smith, Adam. "Of the Propriety of Action." In *The Theory of Moral Sentiments*. London: A. Millar, 1759, 1.

GLOSSARY

invisible hand idea that self-interested behavior by individuals can lead to positive social outcomes

normative statement statement which describes how the world should be

positive statement statement which describes the world as it is

SOLUTIONS

Answers to Self-Check Questions

1. When individuals compare cost per unit in the grocery store, or characteristics of one product versus another, they are behaving approximately like the model describes.
2. Since an op-ed makes a case for what should be, it is considered normative.
3. Assuming that the study is not taking an explicit position about whether soft drink consumption is good or bad, but just reporting the science, it would be considered positive.

CHAPTER 3: DEFINING ECONOMICS: A PLURALISTIC APPROACH

Defining Economics: A Pluralistic Approach

DEFINING ECONOMICS: A PLURALISTIC APPROACH

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Multiple Definitions of Economics
- Diversity of Paradigms in Economics
- A Critical Examination of the Orthodox (Neoclassical) Definition of Economics
- The Presentation of an Alternative Definition

3.1 THE IMPORTANCE OF DEFINITIONS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Discuss how definitions provide theoretical direction.

The presentation of an alternative definition may appear generic, even unimportant. A definition often washes over a reader, leaving the reader influenced, but without any obvious new insight as to the intended meaning or implications of a word. Worse, without proper guidance, a reader may interpret a definition differently than what is intended by the author, or in the case of economics, different from how the discipline is actually employing the definition. Definitions are, however, important, potentially very important. A definition has the potential to provide clarity. At their best, definitions act like a compass, providing a lost reader with several potential directions from which to proceed.

3.2 MULTIPLE PERSPECTIVES REQUIRE MULTIPLE DEFINITIONS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define the term paradigm.

The importance of clarity and directionality becomes heightened when a discipline has more than one potential definition. The discipline of economics, similar to the other social sciences, does not ascribe to only one definition. The definition of economics that is utilized by an economist will often depend on the paradigmatic perspective of the economist.

Paradigm – A school of thought. It represents the boundary of a discipline, framing the types of questions and phenomena that will be analyzed, and the approach or method of study a theorist will employ.

The discipline of economics has more than one paradigm. Not surprisingly, the discipline of economics also has more than one definition.

3.3 A BRIEF SYNOPSIS OF DIFFERENT ECONOMIC PERSPECTIVES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define the term ideology.
- Describe the relationship between ideology and paradigms
- Identify some of the differences between competing economic paradigms.

Multiple theoretical perspectives tend to create multiple ideas as to what economic questions, phenomena, and methods are most appropriate. With paradigmatic differences come ideological differences.

Ideology – The ideas, beliefs, subjective values, and prevailing world views that a person holds. One’s ideology influences the way in which a person perceives what they believe is correct or incorrect about their social circumstance or the larger world in which they live.

Broadly speaking, the discipline of economics is occupied by theorists who tend to fall within one of three categories.

1. Conservative economists
2. Progressive economists
3. Radical economists

In terms of beliefs, while there is some crossover, the differences between each of the ideological perspectives tends to be overwhelming.

Conservative Economists Believe...

- That markets works well on their own without outside interference.
- Because markets work well, a “Laissez-Faire” (Hands Off) view of markets is embraced.
- Given the belief in well-functioning markets, it is also believed, overwhelmingly, that markets should be responsible for how societies allocate resources
- Limited government involvement in markets is best because extensive government involvement is disruptive to market performance.

Progressive Economists Believe...

- Markets should generally be responsible for the allocation of resources
- Free markets generate inequality
- Free markets are often unstable
- Government intervention is necessary

Radical Economists Believe...

- Market economies are classed based economic systems.
- Class is determined by one's position in the economy: owner, laborer, etc.
- Market economies empower the wealthy, typically owners of assets.
- Market economies exploit labor
- Free market capitalism is unstable
- Capitalism needs to be replaced

COMPETING PARADIGMATIC VIEWS

How might different economists address Different Microeconomic Issues?

1. Consider the issue of Negative *Externalities* – *When a third party suffers a loss from the actions of others. For example, consider the massive social costs associated with carbon emissions and climate change.*
 - **Conservative Economists:** Argue that the solution to market failure associated with externalities is to assign property rights. If the party causing or impacted by the externality has property rights, then markets and accurate prices emerge.
 - **Progressive Economists:** Argue that the government is needed to assist in correcting for externalities. Taxes and/or regulations can be instituted that require producers of externalities to account for the social costs they have generated.
 - **Radical Economists:** Argue that externalities are pervasive in market economies. Assigning property rights encourages producers of externalities to produce more negative externalities in an effort to extort payments from those afflicted. Also it is impossible to tax and/or regulate all externalities. An alternative to the market, as an allocative institution, is needed.

How might different economists address structural issues in economics?

1. Consider the issue of *Preferred Economic Institutional Structure(s)*,
 - **Conservative Economists:** Argue that the free market is the ideal institution for allocating resources to their best (socially optimal) outcomes. Advocate that government should be limited to institutions useful toward the defense of private property and the enforcement of economic contracts.

- **Progressive Economists:** Argue that as a result of market failure such as environmental pollution and significant unemployment, institutions, typically government institutions, are necessary in order to correct market failure and return markets to a more efficient allocation of resources.
- **Radical Economists:** Argue that government institutions are an incomplete and ineffective mechanism for addressing market failure. Advocate for the creation of alternative institutions, such as worker cooperatives, that are designed to challenge and potentially supplant institutions favorable to the current operation of the economic system.

How might different economists address economic development issues?

1. Consider the issue of **Income Inequality**,

- **Conservative Economists:** Argue that income inequality arises from productivity differences that emerge between and among individuals. Efforts to reduce inequality should be avoided as they distort incentives and market allocation.
- **Progressive Economists:** Argue that income inequality results from social circumstances. Children born into higher income (higher wealth) families have more opportunities available to garner both the resources and skills necessary for higher incomes. Hence, these economists favor government programs that promote equality of opportunity.
- **Radical Economists:** Argue that income inequality is a byproduct of a class-based economic system. Also inequality is exacerbated by social circumstances. Generally, they would support government programs that promote equality of opportunity, however, also argue that those programs cannot eliminate inequality in a system built upon inequality.

How might different economists address world economic issues?

1. Consider the issue of **Global Poverty**,

- **Conservative Economists:** Argue that free trade between nations will drive specialization and the expansion of material production, thus gradually reducing global poverty.
- **Progressive Economists:** Argue for a degree of government intervention in international trade. Suggest trade via such policies as infant industry, fair labor standards, and environmental protections. Trade is an economic development tool.
- **Radical Economists:** Argue that international trade worsens class tensions. Low wages in developing countries cause wages to decline in developed countries. Poverty is a part of capitalism, free trade and/or managed trade cannot end global poverty.

3.4 DECONSTRUCTING THE ORTHODOX DEFINITION OF ECONOMICS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Discuss the orthodox definition of economics.
- Understand the organization of the orthodox definition of economics.
- Describe the paradigmatic direction of orthodox economics.

Sometimes, in order to better understand how something works it is useful to break apart, or deconstruct (de-engineer), the object of examination. During periods of war, oppositional armies are known to commandeer their opponent's weaponry in order to de-engineer, or deconstruct, the equipment. The United States Army Air Forces Operation LUSTY (Luftwaffe Secret Technology) sought to capture German technology both during and after World War Two. By taking apart a piece of equipment piece by piece the examiner gains insight into both how the equipment operates as well as its weaknesses. In times of war, stealing good ideas and developing better ways to destroy an opponent's weapons can be invaluable. The object of deconstruction does not have to be physical or mechanical for the process to be useful; deconstructing an idea is also a valuable process for enriching understanding. Regarding the orthodox definition of economics, as it is the object of both examination and critique, this same type of deconstruction will also be useful.

In the first chapter of this text an example of the orthodox definition of economics was provided to the reader. The definition provided, similar to variations on the orthodox definition of economics in most textbooks, owes its origins to a famous early 20th century economist, Lionel Robbins. Robbins famously defined economics as, "the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses." Contemporary orthodox economic textbooks tend to utilize a generic variation on Robbins' definition, often stipulating something akin to "Economics is the study of the allocation of limited resources over unlimited wants." In order to examine the appropriateness, or lack thereof, of the orthodox definition of economics, a closer examination of the definition is necessary.

One approach toward deconstructing the orthodox definition of economics is to define, specifically within the context of the orthodox economic meaning of the words, the other terms within the definition. Given the organization of the orthodox definition, with wants being conditioned by the availability of resources, one possible starting point for deconstruction is the orthodox definition of the term resources.

As should be noted from chapter two of this text, the orthodox economic definition of resources emphasizes that resources are those things that are needed for production. While other resources may be relevant, orthodox economics nearly exclusively focuses on *Land, Labor, and Capital* as the essential resources necessary for production. When more resources are present, more production is possible and when fewer resources are present, less production is possible. Obviously, if resources are unlimited, then production would be unlimited and all wants, and presumably needs, could be met. Of course, as noted by the definition, resources are assumed to be limited, otherwise referred to as scarce.

The terms “limited” or “scarce” are fairly self-explanatory. By definition, if something is scarce or limited, then there is only so much of the thing in question. In this case, orthodox economists are assuming that Land, Labor, and Capital are in some way constrained. In many ways this is a sensible assumption as the planet we reside upon has only so much available land, human population, albeit significant, is limited to 7.4 billion people, and capital, consisting of tools and equipment, are limited because only so many machines are available at any one time. Given limited or scarce resources, the ability for wants to be met must be limited.

Limits to wants are an important consideration. Given the orthodox definition of economics, the term wants could, presumably, represent any and all things people could want. However, because wants are deemed to be conditioned by resources and what resources are capable of producing, it stands to reason that what orthodox economics means by its use of the term “wants” are those things that are produced by resources. Since resources are factors of production, land, labor, and capital, then “wants” must be those things that are produced by land, labor, and capital—essentially products (goods and services).

But why do people want products? According to orthodox economics people want products because people garner satisfaction or utility (happiness) from their consumption of products. Given that wants are considered to be unlimited, which is to say endless or infinite, then, by logical continuation, people must have an endless desire for utility or happiness. Importantly, considering that wants must be the byproduct of the products produced by resources, orthodox economics is arguing that people have an endless desire for happiness and that happiness is derived, apparently exclusively, from the products that people acquire. Additionally, consistent with the circular logic of orthodox economics, if human beings have unlimited wants, then resources must be scarce. All wants can never be met, rendering resources on a finite planet scarce, if wants are endless.

In summary, given the above set of definitions the following if-then statement provides a summary of the belief of orthodox economists. If wants are unlimited and resources are used to produce the things that people want, those things being products for consumption, then economists assume people have an insatiable desire to acquire products (aka. inherently greedy). Additionally, if, presumably, products bring happiness and people have an endless desire for products, then it must be that people have an endless desire to be happy.

3.5 A CRITICAL EXAMINATION OF THE ORTHODOX DEFINITION OF ECONOMICS AND ITS RESULTANT IMPACTS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Discuss critiques of the orthodox definition of economics.
- Understand the differences between facts and beliefs.
- Describe the differences between needs and wants.
- Understand the interconnection between facts and values.
- Compare the differences between scarcity and surplus.
- Compare the differences between individual and social approach to economics.
- Identify the relationship between unlimited wants and environmental degradation.

Chapter one presented the orthodox case for utilizing the limited resource, unlimited want definition. For economists outside of the orthodox paradigm, however, the orthodox story contains troubling elements. While potentially insightful in certain instances, the orthodox definition is also potentially misleading in terms of clarifying what it is that is actually happening in the economy. What follows are, from a heterodox (alternative) economic perspective, four critiques of the orthodox definition of economics and its implications.

ARE UNLIMITED WANTS (AND THEREFORE SCARCE RESOURCES) A FACT OF LIFE OR A BELIEF?

A fact is something immutable, it just is, whereas a belief is something that a person or people think, whether there is evidence to support their contention or not. An atheist or a theist cannot prove their beliefs, God may or may not exist, but there is no way to empirically prove either view. Atheism and theism are beliefs, not facts. In terms of belief-based versus fact-based analyses in the realm of economics, noted 20th century political economist Karl Polanyi criticized the orthodox (neoclassical) scarcity story as scarcity assumed within the mind of the orthodox economist. In other words, scarcity is a belief of the orthodox economist, not a fact.

Outside of orthodox economics, such as within heterodox (alternative) economic circles as well as the other social science disciplines, the issue of scarcity as induced by unlimited wants has been extensively evaluated and re-evaluated. For many theorists, the concept of unlimited/endless want of material items is simply not an immutable fact present in all times and places among all human beings. Among anthropologists, for example, there is a rich history of ethnographic studies docu-

menting ways in which small-scale, non-market oriented economies appear to behave differently than orthodox economists predict. One such anthropologist, Marshall Sahlins, explicitly notes that people residing in hunter-gatherer societies, while viewed as deeply impoverished by the standards of people residing in advanced industrial economies, are actually quite affluent because “all the people’s wants are easily satisfied.” For Sahlins and other anthropologists, the hunter-gatherer’s affluence is not one of vast material abundance, but rather a happiness shaped by being contented with having basic needs and wants met. By having little and demanding little, a successful hunt can yield great want satisfaction.

Borrowing from psychology, there is also a sub-discipline of economics known as Happiness Economics. Utilizing survey data, theorists pursuing the study of happiness have repeatedly found that people that focus on the goals of monetary gain and material benefits tend to be less happy than people who are less focused or not focused on those goals. It seems that for every endless maximizer of products, there are many more people that are apt to be content and satisfied with far fewer material items, particularly if fewer material items mean a greater level of interpersonal satisfaction like love, the respect of peers, affection, family fulfillment, and relationship fulfillment. Apparently, these and other equally non-material but important psychological factors are more important to the pursuit of happiness than is monetary wealth.

CAN FACTS AND VALUES REALLY BE SEPARATED?

Utilizing beliefs, ideas, or subjective values is not, however, an unacceptable practice for most social scientists. Orthodox economics insists that facts and values are to remain separate, otherwise known as positive economics. The orthodox fact/value dichotomy is, however, misleading. Nearly all theorists apply some sort of pre-analytical value to their analysis. The very use of assumptions implies that some kind of value-based decision making process was utilized by the theorist. The irony in this instance is that orthodox economics insists that it is a fact-based science when, in *fact*, it utilizes value based (normative) propositions.

Within their very definition, orthodox economists employ subjective values, such as the desire to maximize acquisition (an assumption, or belief, not a fact), to frame their vision of the discipline of economics. Once again, the use of subjective values on their own is not the problem or even a problem (contrary to the orthodox perspective that economics should be a positive science), but it becomes a problem when a definition states something as “is” as opposed to “believed.” Compare the following phrases, “the study of economics is” and “one belief is that economics studies.” The first phrase leads the reader to believe that economists deal in absolute facts while the second phrase implies that some economists may have settled on an accepted set of ideas, but those ideas remain up for debate. An honest presentation of the orthodox definition of economics would note its subjective, normative, non-positive nature.

SCARCITY OR SURPLUS?

There is another fundamental question that Polanyi’s critique alludes to: does scarcity as perceived by the mind actually mean that there is a scarcity of resources? For Polanyi, the orthodox definition of economics is circular, it has neither a beginning nor an end, just continuous scarcity because of limitless want. After all, if human beings have unlimited wants, then resources have to be scarce, there is simply no way for all wants to be met if wants are endless. Close inspection of the orthodox definition of economics makes it clear that the picture of the homeless man on the bench presented in

chapter one is not actually a result of scarcity, at least not as scarcity is defined within the orthodox tradition.

Again, recall that the orthodox definition of economics describes scarcity as resulting from endless “**want**.” Endless want does not mean that there are not enough resources available to fulfill all of humanities basic **needs**. Nor does endless “want” mean that many wants cannot be met given available resources. Obviously the man on the park bench may have some very legitimate wants (certainly needs, like adequate shelter), but is he on the bench because there are not enough resources available to make sure his basic needs, such as a home, or other wants, are met? At least in terms of available housing, the answer to this question is an overwhelming no. According to the U.S. Department of Housing and Urban Development on any given night in January of 2015, 564,708 people were homeless. At the same time, the number of vacant residential properties in the United States numbered approximately 1.4 million.

The homeless man’s lack of housing is not the byproduct of scarcity, at least not by the definition of scarcity that many people are accustomed. Rather, the homeless man on the bench is the byproduct of either misallocated resources or a maldistribution of production.

Homelessness (or house-less-ness) is not the only instance in which the orthodox economic scarcity story breaks down. There are other well documented instances. One particularly disturbing instance relates to malnourishment and general starvation. According to the World Hunger Education Service (WHES) around the world there is enough food produced on an annualized basis to ensure that every person is adequately nourished. In fact, the sheer amount of calories per person generated by global food production has risen since the early 1960s through the mid-2000s, with the amount of per capita calories produced being enough to adequately feed an adult in both developed and developing economies. At the same time, “The United Nations Food and Agriculture Organization estimates that about 795 million people of the 7.3 billion people in the world, or one in nine, were suffering from chronic undernourishment in 2014-2016.”

<http://www.fao.org/3/a-i4674e.pdf>

Again, this instance of basic needs not being met may be perceived as a situation of scarcity. However, as reported by the WHES, poverty is the overwhelming driver of global malnourishment. While poverty implies a scarcity of income, possibly due to a scarcity of employment possibilities, the actual product that resolves malnourishment, food is produced in enough abundance to adequately feed everyone. Once again, basic human needs are going unmet and this has nothing at all to do with the scarcity of resources.

An interesting outcome of reorienting the definition of economics away from the scarcity story, when evaluated on what is actually available, is not that resources are endless or that all wants can necessarily be met, but that it is apparent that there are often more than enough resources to meet all of humanity’s material needs as well as many, many, many of humanities wants. Reoriented this way the story of economics becomes a story of surplus and abundance, rather than shortfall and hardship. A much less dismal, and likely more realistic, science indeed!

NEEDS VERSUS WANTS

A distinction must be made between needs and wants, as they are decidedly not the same.

Need – Reflects something that is necessary for a human being to sustain basic physical and psychological health.

One fairly well known measure defining needs is Maslow's hierarchy of needs. In Maslow's hierarchy basic physiological needs, like food, water, air, shelter, safety, must be met first, followed by psychological needs, such as love and affection, with self-actualizing needs being met last. Importantly, while Maslow's hierarchy provides a framework, it must also be understood that the concept of need evolves depending on the social circumstances and technological make up of a given economy. In the absence of basic physiological needs being met, a person's life will suffer to the degree that in many instances their life may be shortened. For example, homeless people have shorter life spans, 42 to 52 years, than non-homeless people, 78 years.

A want is different from a need.

Want – is something that is desired, but also not necessary for an individual to live and function in a society.

For example, while basic shelter with appropriate utilities such as gas, electric, water, sanitation, and basic cooking, cleaning, and food storage appliances will generally fulfill basic needs, additional products such as dishwashers, large screen televisions, hot tubs, and many, many other products are wants.

To further distinguish between needs and wants, consider the following example. In some communities the existing transportation infrastructure is nearly completely and exclusively automobile dependent. In order for a person to have a reasonable expectation of functioning, physically, socially, and economically in this socioeconomic environment, a person will need access to an automobile. In this case, a functional automobile is, in fact, a need, but a high end SUV is not a need, it is a want.

Based on the orthodox definition of economics it is clear that the orthodoxy is interested in wants rather than needs. The absence of needs from the orthodox definition is a curiosity. There are several possible reasons why needs are not explicitly referenced in the orthodox definition. Two possible reasons will be addressed here. First, it is possible that orthodox economists do not differentiate between needs and wants, thus all needs really are just wants. Second, orthodox economists may be assuming that needs are already being met. If basic needs are met, then it is not only possible but reasonable to study what people want.

In critique, the orthodox merger of needs and wants seems to produce potentially misleading insights. While one can certainly make the argument that all people want their needs to be met, such that a need is nothing more than a type of want, basic needs must be met or the study of additional wants becomes useless as there can be no other wants. Additionally, even if it is assumed that needs are being met, this does not actually mean that needs are being met in the world where people really live. If needs are not being met, can orthodox economics be trusted to study the situation and then prescribe a solution that might ease human suffering? The answer to this question is unclear as the orthodox definition of economics does not provide direction toward the study of needs.

INDIVIDUAL PERSPECTIVE OR SOCIAL PERSPECTIVE?

As a discipline, economics is categorized as a social science. The social sciences, in general, seek to develop understandings about human interactions on a social level. For example, an anthropologist

may be interested in understanding the role of culture as it influences human behavior. Culture can be represented as social norms, rituals, and/or customs that different groups of people, mostly collectively, participate in performing. Another example of a social science is political science. A political scientist may be interested in examining political institutions and legal structures. In order to better understand why individuals or larger social groups of people exhibit certain behaviors, an understanding of how a society structures its laws, the rewards or penalties people encounter as a result of socially derived rules, is essential.

For the orthodox economist, however, the social element of economics is secondary to that of the individual. Again, the orthodox definition of economics gives the orthodox economist direction. Recall, the focus of the orthodox economist is on want fulfillment with wants being satisfied by goods and services produced by resources. The reason products are wanted is because products, presumably, provide the recipient with happiness, or utility. Since people are presumed to have endless wants, people are also presumed to have an endless desire for utility. In many ways orthodox economics is really just the study of the maximization of utility. But how is utility measured? Well, it's not. Utility is subjective, it is in the eye of the beholder. As such, the study of utility maximization is the study of individual choices. After all, the only way to know what makes a person happy, or bringing the most utility is to evaluate the consumption choices that people make. As such, the focus of orthodox economics is that of individual choice making.

For some in the orthodox economic community, the focus on the individual is a conscious choice, purposeful and known, and often times affectionately referred to as “**methodological individualism**.” A methodology is an approach for addressing concerns that a discipline believes are essential. Methodological individualism is a method anchored to examining the individual, placing front and center the issue most essential to that of the orthodox economist: individual decision-making. But if the individual is of utmost importance, then the role of social is much less clear within orthodox economics. This is not to say that the larger society is ignored by orthodox economists. For orthodox economics the social does emerge but, consistent with their dedication to studying individual choice, it is as a byproduct of individual maximizing behavior. Once individuals pursue their own self-interest, the orthodox economist is tasked with simply summing or aggregating the individual behavior and the result is the larger social outcome.

For those outside of the orthodox economic tradition, many important considerations are seemingly lost by the direction of orthodox economic analysis. Take for example the question of how broader societal norms as a function of social institutions may have influenced the behavior of individuals. By assuming away the social and focusing exclusively on the individual, the orthodox economist will be apt to ignore such issues as how social institutions may have influenced individual behavior such that it is now not prone to maximizing individual acquisition. In other words, let's say people really don't just look out for themselves all of the time, seeking to acquire as many products for themselves as they can, but instead, people behave such that they are concerned for the well-being of others, maybe because the norms of their society encourage a less than self-centered set of behaviors. Non-acquisitive, non-self-interested individual behavior tends to be outside of the scope of orthodox economics, but it is not outside the scope of actual human economic interactions. For many non-orthodox economists, the absence of a clear social perspective limits the understandings available to the practitioner of methodological individualism.

1. *UTILITY AS EVERYTHING, UTILITY AS NOTHING*

Some orthodox economists have objected to the idea that they are assuming that people are greedy, only wanting products or those things that can get products, namely resources or money as money is what is used to perform purchases. One orthodox economic argument that has emerged claims that the orthodox economic model can be extended to include the story of “**enlightened self-interest**.” Enlightened self-interest is the idea that it can be in one’s self-interest to act to benefit others. In some regards enlightened self-interest can be interpreted as someone serving society such that through their service society is better-off and, subsequently, as a member of society, the individual then benefits.

A common example of enlightened self-interest is charitable giving. Charitable giving is the opposite of individual acquisition as the person giving to charity is presumably giving away something of personal value for themselves so that someone else can benefit. Orthodox economists argue that the reason a person engages in charity, or another type of enlightened self-interest, is because they gain utility from knowing that they have done something that is perceived as “good” for society.

There are two problems that become evident when the orthodox enlightened self-interest argument is compared to the orthodox definition of economics. First, enlightened self-interest is inconsistent with the basic premise described by the orthodox definition of economics. To once again reiterate, the orthodox definition of economics is stipulating that people’s unlimited wants are fulfilled by resources and resources are utilized to produce products, thus want fulfillment is synonymous with product acquisition. Returning to the charitable giving example, it is clear that charitable giving is not product acquisition but rather product dissipation. Thus orthodox economics cannot, without contradicting itself, apply its definition of economics to a story of enlightened self-interest. Another definition of economics would be necessary to account for human behavioral actions associated with what is perceived as enlightened self-interest, or really any human action that is not associated with the acquisition of products.

Some thoughtful orthodox economists have recognized the contradiction between their definition of economics and the idea of enlightened self-interest. As a result, some orthodox economists have recast their position arguing that utility can be derived from anything, not just material products. In other words, for the orthodox economist all human actions and behaviors are driven around utility maximization and thus the study of economics is really the study of utility maximization. When recast in this light, however, the second problem appears. If all actions provide a person with utility, and all people are utility maximizers, then any action can be said to be utility maximizing, otherwise the person would not engage in the action. Based on this construction, the orthodox economist would be reduced to observing what people do and then proclaiming that it must be in their best interest. As a critical thinking exercise an important question now emerges: when exactly does orthodox economics become a social science? Saying that people do as they do because otherwise they wouldn’t do it does not represent science, it has very little empirical function. It is also not focused on the social as it does not account for any social factors that may influence or disrupt individual behavior. And it provides almost no predictive possibilities other than identifying what might happen in the future based on what has happened in the past. Once again, confronted with such limitations many economists suggest that another definition of economics may be more substantive and useful.

BREAKOUT BOX – WHAT ABOUT BEHAVIORAL ECONOMICS?

Because people frequently engage in behaviors that are not considered rational, that is to say individually acquisitive, orthodox economics has been targeted for criticism that their models do not accurately depict the real world. In response a sub-field of orthodox economics, behavioral economics, has developed that seeks to utilize experiments and psychology to better understand actual human behavior. Behavioral economics tends to start from the position that otherwise-rational, optimizing people will deviate from standard orthodox economic actions because people can be influenced by non-economic factors, such as religious beliefs or cultural norms

A common approach to identifying the factors that condition individual choice-making is to run experiments that test human response to incentives. A controlled experiment provides researchers with empirical, data-driven evidence as to how and why people may not respond to opportunities for individual gain in the ways that orthodox economics predicts.

For many outside of orthodox economics, the development of behavioral economics is a step in right direction for the larger science of economics. Having evidence and being able to explain a wider range of phenomena rather than simply stipulating that people do as they do otherwise they wouldn't do it, appears to move the discipline of economics toward more realistic insights. At the same time, critiques of orthodox economics are also apt to note that behavioral economics still does not go far enough with its approach. Behavioral economics emphasizes an analysis that begins with the standard orthodox isolated maximizing individual and then searches for deviations from this behavior. By the very nature of their focus, behavioral economics under-emphasizes social considerations. As such, data may be derived that draws out the idea that an individual is making choices that are influenced by social institutions, but absent the study of those social institutions, the theorist is possibly still in the dark as to why the social institutions elicit the response that they do on the part of the individual. The result is that behavioral economics broadens the scope of economic inquiry but, for those outside of orthodox economics, it is still too narrow of an approach.

LIMITLESS VERSUS LIMITS: THE IMPLICATIONS FOR THE HEALTH OF THE PLANETARY ENVIRONMENT

The final critique presented here may be the most important in terms of its implications for the future of humanity. The Earth is a finite biosphere. There is simply only so much atmosphere, land, water, minerals, ores, and other natural resources within the bounds of the planet. In this way, the Earth really is representative of a circumstance of scarcity. Certainly improved efficiency is capable of extending the productive capability of the available natural resources, but this does not change the fact that there is a limit to how much is actually available.

Regardless, for many orthodox economists, the principle of endless growth is unquestionably desirable and it is downright sacrilegious to doubt the merits of endless economic growth. Consider the orthodox definition of economics and its emphasis on the desirability of endless happiness and that happiness being derived from the endless consumption of products. Given the premise of unlimited wants, the need for endless production is obvious and, by continuation, endless production is a synonym for endless, exponential, economic growth. In the simplest sense, more products produced means that the material well-being of society is improved. For orthodox economists, the improved material well-being of society cannot be understated. For example, it is not uncommon for orthodox economists to argue that given enough growth, social maladies such as poverty can be alleviated and,

in time, eliminated. Of course, the orthodox economic perspective leaves aside incredibly important considerations such as whether or not new material products are equitably distributed. But, as orthodox economists would note, an unequal distribution of products is a separate problem from inadequate production and overt want, so the existence of inequality does not undermine the importance of economic growth.

The issue of endless growth in a finite biosphere presents orthodox economics with interesting conundrums. If resources are finite, how is endless growth possible? Again, orthodox economics does provide an answer in form of the advancement of knowledge and technology that allow for improvements in productivity and efficiencies creating opportunities for endless growth. But what if endless growth then yields waste that pollutes the natural environment making human life on the planet ever more difficult to sustain? The answer to this question is more difficult and the failure to answer is more perilous. Even if knowledge and technological advancement can improve productive efficiencies, absent the elimination of waste, exponential economic growth will add waste to the land, sea, and air. Just as lily pads can overrun a pond, with exponential economic growth and its associated waste, the effluent byproducts of industrial production can inundate and overwhelm a finite biosphere.

The mass annihilation of the Plain's Bison in the 19th century is a grave depiction of the ability of humanity to inflict harm upon the natural environment.



Figure 2. Endless economic growth necessitates endless production. At the expense of rich, biodiverse, rainforest human beings burn and clear cut the forest in order expand agricultural capacity. (Credit: Jami Dwyer, Wikimedia, Public Domain)



Figure 1. A large pile of bison skulls c. 1870s. (Wikimedia, Public Domain)

It is clear that human economic activity nearly inevitably leads to the creation of waste byproducts, pollution. Not only is the Earth's biosphere limited in terms of what it can provide in the way of natural resources toward productive output, it is also limited in terms of how much waste it is capable of absorbing on land, in the sea, and in the air. For example, the Stockholm Resilience Centre has identified nine planetary boundaries (environmental thresholds) that must not be eclipsed if human life is to safely exist. The nine planetary boundaries are stratospheric ozone depletion, biodiversity loss, chemical pollution, climate change, ocean acidification, freshwater consumption, land system change, nitrogen and phosphorous flows to the biosphere and oceans, and atmospheric aerosol loading. Of these nine boundaries, three in particular, climate change, ocean acidification, and stratospheric ozone depletion, are of such consequence that if the threshold is surpassed, then the planet's basic ecology will likely be fundamentally destabilized. In the case of climate change and ocean acidification, the two issues are caused by human economic activity and are inherently interrelated. Climate change is caused by carbon dioxide emissions and ocean acidification is triggered when the oceans absorb and process carbon dioxide. Strikingly, at least in the case of climate change, the proposed boundary for atmospheric carbon dioxide has been surpassed as 350 parts per million (ppm) is the boundary while actual carbon dioxide levels now exceed 400 ppm. Clearly, human economic

activity, and the associated environmental destruction of the Earth's biosphere as a result of economic growth, does present a powerful challenge to the importance of economic growth.

Given the above-stated considerations, going forward humanity must consider the possibility of orienting its economic activity within the context of limits. The direction of orthodox economics, as guided by its definition, is not, however, oriented toward accepting limits. In the face of the abundance of scientific evidence pointing toward human factors as the cause of a de-stabilizing natural environment, the orthodox definition of economics appears to point economic analysis in a direction counter-productive to health and well-being of life on this planet. As a result, it stands to reason that another definition of economics, one less rigid and one-dimensional, may be of use.

ENVIRONMENTAL RESOURCE ECONOMICS

Orthodox economics is not entirely oblivious to environmental concerns. Nearly a century ago orthodox economic thinkers such as A. C. Pigou recognized the potential for a concept known as **externalities**. Discussed in chapter 18: Environmental Protection and Negative Externalities, negative externalities represent the idea that sometimes economic activities create unaccounted for costs that impact third parties, or those people who are not directly involved in the economic activity. Any failure to properly account for a negative externality means that the market system has failed to efficiently allocate resources. For Pigou and orthodox economics, the solution to negative externalities is the development of institutional mechanisms that assign penalties to those that create negative externalities.

A common orthodox example of an institutional mechanism designed to address negative externalities is the government being granted the regulatory task of environmental protection. From the standpoint of orthodox economics, pollution and environmental degradation represent a negative externality, an unaccounted for cost on the larger society. As a regulator, the government can utilize any number of regulatory tools to try and mitigate pollution. For example, one choice preferred by many, but not all, orthodox economists, is the use of taxation as a way to increase costs to polluters. Overall, orthodox environmental resource economics places the environment as sub-ordinate to the economy, and argues that marginal analysis should be utilized to measure the economic costs and benefits of environmental changes.

As will be later illustrated in chapter 18: Environmental Protection and Negative Externalities, by invoking a tax on polluters in a market economy, costs to producers rise. As costs rise the producer will produce less of the product the production of which is creating pollution. Additionally, with less production of the product, the supply of the product will decline generating an increase in price of the product. The outcome is three-fold. First, with less production comes less pollution. Second, with less production the product now sells at a higher price, with the price capturing more of the negative externality and reflecting something closer to the true cost of the product. Third, the tax revenue generated from the tax has the potential to be used to compensate the victims of the negative externality. On the face of it, it appears as though the taxation solution to an externality is providing a final outcome that is close to ideal.

Critics of the orthodox economic externality resolution argument have their doubts about the efficacy of the orthodox approach. The following questions represent a small sample of possible concerns. For starters, it is possible to conceive of an economy with a seemingly endless variety of negative externalities, does this mean that all externalities will be taxed? Failure to properly tax negative exter-

nalities means that the economy is rife with inefficiencies. Additionally, how large of a government bureaucratic structure would be needed to adequately address all externalities? Might the solution be worse than the cure? Also, even if all products to be taxed could be identified, how can a regulator be certain to precisely calculate the size of the tax in order to correct for the costs of the negative externality.

Of course, even if all of the questions above could be sorted out and resolved, there are still basic fundamental issues and contradictions, associated with the orthodox definition of economics, which the orthodox argument fails to address. For example, all efforts to address negative environmental externalities call for the reduction of waste and some kind of constraint on economic activity and growth. Any check on growth clearly and obviously runs afoul of the basic tenet of orthodox economics, as is clear in their definition, of endless want of products and the associated desirability to try and meet those wants. Also, any reduction in production will also act to reduce consumption, presumably limiting consumers' abilities to increase their utility by acquiring products. The primary conclusion being that orthodox economics' own solution for externalities runs counter to the foundations of their definition of economics.

Some critics of the standard orthodox environmental resource economics arguments, through their application and advocacy for **ecological economics**, have sought to re-orient the relationship between the economy and the environment. Whereas orthodox economics views the natural environment as an external factor that may be impacted by economic activity, ecological economists place the economy in a sub-ordinate role to the natural environment. Ecological economics generally tends to support the notion that the natural environment must be nurtured and protected, as, without it, an economy and human life cannot function. As such ecological economics has a more cautious perspective regarding economic activity and how it impacts the natural environment. As a result, standard orthodox economic concepts such as endless economic growth and want fulfillment are of secondary importance to long term sustained living made possible through the protection of the natural environment. As such, ecological economics operates outside of the orthodox definition of economics.

GLOSSARY

need

something that is necessary for a human being to sustain basic physical and psychological health

want

something that is desired, but also not necessary for an individual to live and function in a society

negative externalities

situations in which economic activities create unaccounted for costs that impact third parties, or those people who are not directly involved in the economic activity

ecological economics

a heterodox school of economic thought that situates the economy as a subordinate structure within the natural environment

3.6 AN ALTERNATIVE APPROACH TO DEFINING ECONOMICS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe an alternative definition of economics.
- Understand the basis for framing an alternative definition of economics.

The preceding presentation centered on deconstructing and then challenging the orthodox definition of economics. Of central concern throughout is the way in which the orthodox definition of economics potentially handcuffs the economist into viewing economic activity through a very narrow and specific lens. Given the problems associated with the orthodox definition of economics, a concerned student may be apt to ask, what does an alternative definition of economics look like? The answer to this question is that there are many possible alternative definitional variations.

The structure of an alternative definition of economics can take many forms, although the overall essence of alternative definitions is frequently similar. With regard to structure, some alternative definitions are apt to simplify economic activity to its most basic form. For example, economics can be the study of how human beings must work together, and with nature, to produce those things that fulfill the material needs and wants of society. In other instances alternative definitions of economics seek to reduce the definition to the essential characteristics of an economy. For example, economics can be defined as the study of how human beings organize production, distribution, and consumption. Still in other instances alternative definitions of economics focus on interdisciplinary facets of the structure of economic decision making. For instance, economics can be defined as the study of how cultural norms, social institutions, political structures, and general decision making processes influence human behavior toward economic ends.

Importantly, each of the possible definitional structures stipulated above appear to avoid the pitfalls that plague the orthodox definition of economics. First, none of the above alternative definitional possibilities assumes scarcity. Rather, each definition is open to the possibility of differing forms of economic organization and a diversity of ways to meet societal needs and wants. Second, none of the above alternative definitional structures make unprovable assumptions about individual human behavior and motivation. Rather, the social elements of the organization of economic activity are emphasized, opening the door to analyzing how individuals respond to social conditions. Third, none of the above alternative definitional structures predicates endless want and the need and desirability for endless economic growth. Rather, each is open to the possibility of alternative economic structures that can be coordinated to operate within the limits of the natural biosphere.

Going forward, within the context of the alternative perspectives developed within this textbook, and in addition to the above definitional structures, the following definition of economics is suggested.

Economics – Is the study of social provisioning, in which an understanding of the development of political economies is rooted in social, political, natural, and cultural processes.

CHAPTER 4. DEMAND AND SUPPLY

INTRODUCTION TO DEMAND AND SUPPLY



Figure 1. Farmer's Market. Organic vegetables and fruits that are grown and sold within a specific geographical region should, in theory, cost less than conventional produce because the transportation costs are less. That is not, however, usually the case. (Credit: modification of work by Natalie Maynor/Flickr Creative Commons)

WHY CAN WE NOT GET ENOUGH OF ORGANIC?

Organic food is increasingly popular, not just in the United States, but worldwide. At one time, consumers had to go to specialty stores or farmer's markets to find organic produce. Now it is available in most grocery stores. In short, organic is part of the mainstream.

Ever wonder why organic food costs more than conventional food? Why, say, does an organic Fuji apple cost \$1.99 a pound, while its conventional counterpart costs \$1.49 a pound? The same price relationship is true for just about every organic product on the market. If many organic foods are locally grown, would they not take less time to get to market and therefore be cheaper? What are the forces that keep those prices from coming down? Turns out those forces have a lot to do with this chapter's topic: demand and supply.

CHAPTER OBJECTIVES

Introduction to Demand and Supply

In this chapter, you will learn about:

- Demand, Supply, and Equilibrium in Markets for Goods and Services
- Shifts in Demand and Supply for Goods and Services
- Changes in Equilibrium Price and Quantity: The Four-Step Process
- Price Ceilings and Price Floors

An auction bidder pays thousands of dollars for a dress Whitney Houston wore. A collector spends a small fortune for a few drawings by John Lennon. People usually react to purchases like these in two ways: their jaw drops because they think these are high prices to pay for such goods or they think these are rare, desirable items and the amount paid seems right.

Visit this website to read a list of bizarre items that have been purchased for their ties to celebrities. These examples represent an interesting facet of demand and supply.



When economists talk about prices, they are less interested in making judgments than in gaining a practical understanding of what determines prices and why prices change. Consider a price most of us contend with weekly: that of a gallon of gas. Why was the average price of gasoline in the United States \$3.71 per gallon in June 2014? Why did the price for gasoline fall sharply to \$2.07 per gallon by January 2015? To explain these price movements, economists focus on the determinants of what gasoline buyers are willing to pay and what gasoline sellers are willing to accept.

As it turns out, the price of gasoline in June of any given year is nearly always higher than the price in January of that same year; over recent decades, gasoline prices in midsummer have averaged about 10 cents per gallon more than their midwinter low. The likely reason is that people drive more in the summer, and are also willing to pay more for gas, but that does not explain how steeply gas prices fell. Other factors were at work during those six months, such as increases in supply and decreases in the demand for crude oil.

This chapter introduces the economic model of demand and supply—one of the most powerful models in all of economics. The discussion here begins by examining how demand and supply determine

the price and the quantity sold in markets for goods and services, and how changes in demand and supply lead to changes in prices and quantities.

4.1 DEMAND, SUPPLY, AND EQUILIBRIUM IN MARKETS FOR GOODS AND SERVICES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain demand, quantity demanded, and the law of demand
- Identify a demand curve and a supply curve
- Explain supply, quantity supply, and the law of supply
- Explain equilibrium, equilibrium price, and equilibrium quantity

First let's first focus on what economists mean by demand, what they mean by supply, and then how demand and supply interact in a market.

DEMAND FOR GOODS AND SERVICES

Economists use the term **demand** to refer to the amount of some good or service consumers are willing and able to purchase at each price. Demand is based on needs and wants—a consumer may be able to differentiate between a need and a want, but from an economist's perspective they are the same thing. Demand is also based on ability to pay. If you cannot pay for it, you have no effective demand.

What a buyer pays for a unit of the specific good or service is called **price**. The total number of units purchased at that price is called the **quantity demanded**. A rise in price of a good or service almost always decreases the quantity demanded of that good or service. Conversely, a fall in price will increase the quantity demanded. When the price of a gallon of gasoline goes up, for example, people look for ways to reduce their consumption by combining several errands, commuting by carpool or mass transit, or taking weekend or vacation trips closer to home. Economists call this inverse relationship between price and quantity demanded the **law of demand**. The law of demand assumes that all other variables that affect demand (to be explained in the next module) are held constant.

An example from the market for gasoline can be shown in the form of a table or a graph. A table that shows the quantity demanded at each price, such as Table 1, is called a **demand schedule**. Price in this case is measured in dollars per gallon of gasoline. The quantity demanded is measured in millions of gallons over some time period (for example, per day or per year) and over some geographic area (like a state or a country). A **demand curve** shows the relationship between price and quantity demanded on a graph like Figure 1, with quantity on the horizontal axis and the price per gallon on the vertical axis. (Note that this is an exception to the normal rule in mathematics that the independent variable

(x) goes on the horizontal axis and the dependent variable (y) goes on the vertical. Economics is not math.)

The demand schedule shown by Table 1 and the demand curve shown by the graph in Figure 1 are two ways of describing the same relationship between price and quantity demanded.

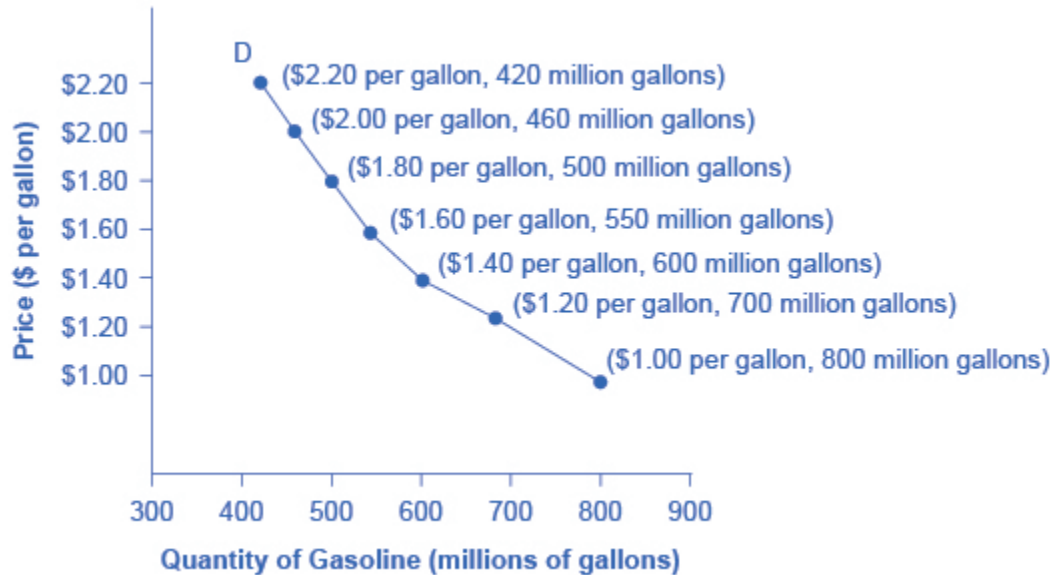


Figure 1. A Demand Curve for Gasoline. The demand schedule shows that as price rises, quantity demanded decreases, and vice versa. These points are then graphed, and the line connecting them is the demand curve (D). The downward slope of the demand curve again illustrates the law of demand—the inverse relationship between prices and quantity demanded.

Price (per gallon)	Quantity Demanded (millions of gallons)
\$1.00	800
\$1.20	700
\$1.40	600
\$1.60	550
\$1.80	500
\$2.00	460
\$2.20	420

Table 1. Price and Quantity Demanded of Gasoline

Demand curves will appear somewhat different for each product. They may appear relatively steep or flat, or they may be straight or curved. Nearly all demand curves share the fundamental similarity that they slope down from left to right. So demand curves embody the law of demand: As the price increases, the quantity demanded decreases, and conversely, as the price decreases, the quantity demanded increases.

Confused about these different types of demand? Read the next Clear It Up feature.

IS DEMAND THE SAME AS QUANTITY DEMANDED?

In economic terminology, demand is not the same as quantity demanded. When economists talk about demand, they mean the relationship between a range of prices and the quantities demanded at those prices, as illustrated by a demand curve or a demand schedule. When economists talk about quantity demanded, they mean only a certain point on the demand curve, or one quantity on the demand schedule. In short, demand refers to the curve and quantity demanded refers to the (specific) point on the curve.

SUPPLY OF GOODS AND SERVICES

When economists talk about **supply**, they mean the amount of some good or service a producer is willing to supply at each price. Price is what the producer receives for selling one unit of a **good** or **service**. A rise in price almost always leads to an increase in the **quantity supplied** of that good or service, while a fall in price will decrease the quantity supplied. When the price of gasoline rises, for example, it encourages profit-seeking firms to take several actions: expand exploration for oil reserves; drill for more oil; invest in more pipelines and oil tankers to bring the oil to plants where it can be refined into gasoline; build new oil refineries; purchase additional pipelines and trucks to ship the gasoline to gas stations; and open more gas stations or keep existing gas stations open longer hours. Economists call this positive relationship between price and quantity supplied—that a higher price leads to a higher quantity supplied and a lower price leads to a lower quantity supplied—the **law of supply**. The law of supply assumes that all other variables that affect supply (to be explained in the next module) are held constant.

Still unsure about the different types of supply? See the following Clear It Up feature.

IS SUPPLY THE SAME AS QUANTITY SUPPLIED?

In economic terminology, supply is not the same as quantity supplied. When economists refer to supply, they mean the relationship between a range of prices and the quantities supplied at those prices, a relationship that can be illustrated with a supply curve or a supply schedule. When economists refer to quantity supplied, they mean only a certain point on the supply curve, or one quantity on the supply schedule. In short, supply refers to the curve and quantity supplied refers to the (specific) point on the curve.

Figure 2 illustrates the law of supply, again using the market for gasoline as an example. Like demand, supply can be illustrated using a table or a graph. A **supply schedule** is a table, like Table 2, that shows the quantity supplied at a range of different prices. Again, price is measured in dollars per gallon of gasoline and quantity supplied is measured in millions of gallons. A **supply curve** is a graphic illustration of the relationship between price, shown on the vertical axis, and quantity, shown on the horizontal axis. The supply schedule and the supply curve are just two different ways of showing the same information. Notice that the horizontal and vertical axes on the graph for the supply curve are the same as for the demand curve.

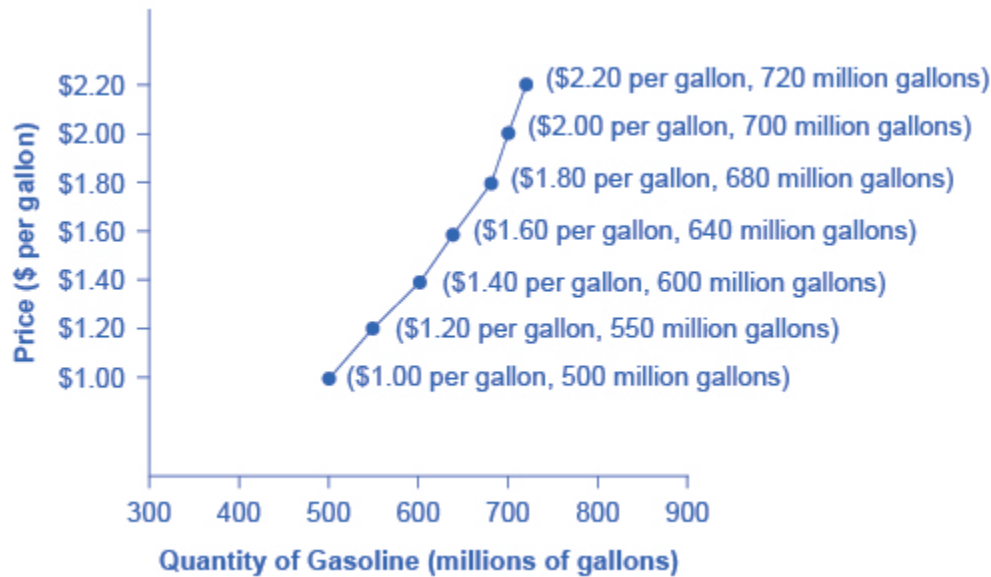


Figure 2. A Supply Curve for Gasoline. The supply schedule is the table that shows quantity supplied of gasoline at each price. As price rises, quantity supplied also increases, and vice versa.

The supply curve (S) is created by graphing the points from the supply schedule and then connecting them. The upward slope of the supply curve illustrates the law of supply—that a higher price leads to a higher quantity supplied, and vice versa.

Price (per gallon)	Quantity Supplied (millions of gallons)
\$1.00	500
\$1.20	550
\$1.40	600
\$1.60	640
\$1.80	680
\$2.00	700
\$2.20	720

Table 2. Price and Supply of Gasoline

The shape of supply curves will vary somewhat according to the product: steeper, flatter, straighter, or curved. Nearly all supply curves, however, share a basic similarity: they slope up from left to right and illustrate the law of supply: as the price rises, say, from \$1.00 per gallon to \$2.20 per gallon, the quantity supplied increases from 500 gallons to 720 gallons. Conversely, as the price falls, the quantity supplied decreases.

EQUILIBRIUM—WHERE DEMAND AND SUPPLY INTERSECT

Because the graphs for demand and supply curves both have price on the vertical axis and quantity on the horizontal axis, the demand curve and supply curve for a particular good or service can appear on the same graph. Together, demand and supply determine the price and the quantity that will be bought and sold in a market.

Figure 3 illustrates the interaction of demand and supply in the market for gasoline. The demand curve (D) is identical to Figure 1. The supply curve (S) is identical to Figure 2. Table 3 contains the same information in tabular form.

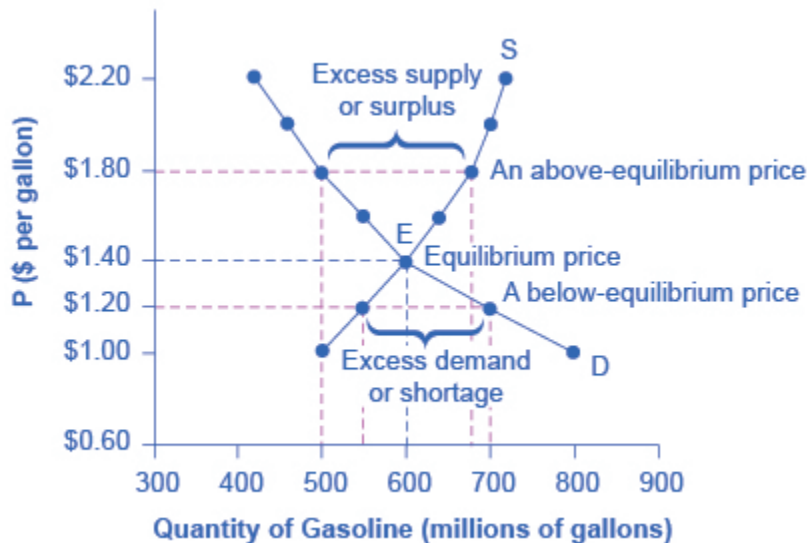


Figure 3. Demand and Supply for Gasoline. The demand curve (D) and the supply curve (S) intersect at the equilibrium point E, with a price of \$1.40 and a quantity of 600. The equilibrium is the only price where quantity demanded is equal to quantity supplied. At a price above equilibrium like \$1.80, quantity supplied exceeds the quantity demanded, so there is excess supply. At a price below equilibrium such as \$1.20, quantity demanded exceeds quantity supplied, so there is excess demand.

Price (per gallon)	Quantity demanded (millions of gallons)	Quantity supplied (millions of gallons)
\$1.00	800	500
\$1.20	700	550
\$1.40	600	600
\$1.60	550	640
\$1.80	500	680
\$2.00	460	700
\$2.20	420	720

Table 3. Price, Quantity Demanded, and Quantity Supplied

Remember this: When two lines on a diagram cross, this intersection usually means something. The point where the supply curve (S) and the demand curve (D) cross, designated by point E in Figure 3, is called the **equilibrium**. The **equilibrium price** is the only price where the plans of consumers and the plans of producers agree—that is, where the amount of the product consumers want to buy (quantity demanded) is equal to the amount producers want to sell (quantity supplied). This common quantity is called the **equilibrium quantity**. At any other price, the quantity demanded does not equal the quantity supplied, so the market is not in equilibrium at that price.

In Figure 3, the equilibrium price is \$1.40 per gallon of gasoline and the equilibrium quantity is 600 million gallons. If you had only the demand and supply schedules, and not the graph, you could find the equilibrium by looking for the price level on the tables where the quantity demanded and the quantity supplied are equal.

The word “equilibrium” means “balance.” If a market is at its equilibrium price and quantity, then it has no reason to move away from that point. However, if a market is not at equilibrium, then economic pressures arise to move the market toward the equilibrium price and the equilibrium quantity.

Imagine, for example, that the price of a gallon of gasoline was above the equilibrium price—that is, instead of \$1.40 per gallon, the price is \$1.80 per gallon. This above-equilibrium price is illustrated by the dashed horizontal line at the price of \$1.80 in Figure 3. At this higher price, the quantity demanded drops from 600 to 500. This decline in quantity reflects how consumers react to the higher price by finding ways to use less gasoline.

Moreover, at this higher price of \$1.80, the quantity of gasoline supplied rises from the 600 to 680, as the higher price makes it more profitable for gasoline producers to expand their output. Now, consider how quantity demanded and quantity supplied are related at this above-equilibrium price. Quantity demanded has fallen to 500 gallons, while quantity supplied has risen to 680 gallons. In fact, at any above-equilibrium price, the quantity supplied exceeds the quantity demanded. We call this an **excess supply** or a **surplus**.

With a surplus, gasoline accumulates at gas stations, in tanker trucks, in pipelines, and at oil refineries. This accumulation puts pressure on gasoline sellers. If a surplus remains unsold, those firms involved in making and selling gasoline are not receiving enough cash to pay their workers and to cover their expenses. In this situation, some producers and sellers will want to cut prices, because it is better to sell at a lower price than not to sell at all. Once some sellers start cutting prices, others will follow to avoid losing sales. These price reductions in turn will stimulate a higher quantity demanded. So, if the price is above the equilibrium level, incentives built into the structure of demand and supply will create pressures for the price to fall toward the equilibrium.

Now suppose that the price is below its equilibrium level at \$1.20 per gallon, as the dashed horizontal line at this price in Figure 3 shows. At this lower price, the quantity demanded increases from 600 to 700 as drivers take longer trips, spend more minutes warming up the car in the driveway in winter-time, stop sharing rides to work, and buy larger cars that get fewer miles to the gallon. However, the below-equilibrium price reduces gasoline producers’ incentives to produce and sell gasoline, and the quantity supplied falls from 600 to 550.

When the price is below equilibrium, there is **excess demand**, or a **shortage**—that is, at the given price the quantity demanded, which has been stimulated by the lower price, now exceeds the quantity supplied, which had been depressed by the lower price. In this situation, eager gasoline buyers mob the gas stations, only to find many stations running short of fuel. Oil companies and gas stations recognize that they have an opportunity to make higher profits by selling what gasoline they have at a higher price. As a result, the price rises toward the equilibrium level. Read Demand, Supply, and Efficiency for more discussion on the importance of the demand and supply model.

KEY CONCEPTS AND SUMMARY

A demand schedule is a table that shows the quantity demanded at different prices in the market. A

demand curve shows the relationship between quantity demanded and price in a given market on a graph. The law of demand states that a higher price typically leads to a lower quantity demanded.

A supply schedule is a table that shows the quantity supplied at different prices in the market. A supply curve shows the relationship between quantity supplied and price on a graph. The law of supply says that a higher price typically leads to a higher quantity supplied.

The equilibrium price and equilibrium quantity occur where the supply and demand curves cross. The equilibrium occurs where the quantity demanded is equal to the quantity supplied. If the price is below the equilibrium level, then the quantity demanded will exceed the quantity supplied. Excess demand or a shortage will exist. If the price is above the equilibrium level, then the quantity supplied will exceed the quantity demanded. Excess supply or a surplus will exist. In either case, economic pressures will push the price toward the equilibrium level.

SELF-CHECK QUESTIONS

Review Figure 3. Suppose the price of gasoline is \$1.60 per gallon. Is the quantity demanded higher or lower than at the equilibrium price of \$1.40 per gallon? And what about the quantity supplied? Is there a shortage or a surplus in the market? If so, of how much?

REVIEW QUESTIONS

1. What determines the level of prices in a market?
2. What does a downward-sloping demand curve mean about how buyers in a market will react to a higher price?
3. Will demand curves have the same exact shape in all markets? If not, how will they differ?
4. Will supply curves have the same shape in all markets? If not, how will they differ?
5. What is the relationship between quantity demanded and quantity supplied at equilibrium? What is the relationship when there is a shortage? What is the relationship when there is a surplus?
6. How can you locate the equilibrium point on a demand and supply graph?
7. If the price is above the equilibrium level, would you predict a surplus or a shortage? If the price is below the equilibrium level, would you predict a surplus or a shortage? Why?
8. When the price is above the equilibrium, explain how market forces move the market price to equilibrium. Do the same when the price is below the equilibrium.
9. What is the difference between the demand and the quantity demanded of a product, say milk? Explain in words and show the difference on a graph with a demand curve for milk.
10. What is the difference between the supply and the quantity supplied of a product, say milk? Explain in words and show the difference on a graph with the supply curve for milk.

CRITICAL THINKING QUESTIONS

1. Review Figure 3. Suppose the government decided that, since gasoline is a necessity, its price should be legally capped at \$1.30 per gallon. What do you anticipate would be the outcome in the gasoline market?
2. Explain why the following statement is false: "In the goods market, no buyer would be willing to pay more than the equilibrium price."
3. Explain why the following statement is false: "In the goods market, no seller would be willing to sell for less than the equilibrium price."

PROBLEMS

Review Figure 3 again. Suppose the price of gasoline is \$1.00. Will the quantity demanded be lower or higher than at the equilibrium price of \$1.40 per gallon? Will the quantity supplied be lower or higher? Is there a shortage or a surplus in the market? If so, of how much?

REFERENCES

- Costanza, Robert, and Lisa Wainger. "No Accounting For Nature: How Conventional Economics Distorts the Value of Things." *The Washington Post*. September 2, 1990.
- European Commission: Agriculture and Rural Development. 2013. "Overview of the CAP Reform: 2014-2024." Accessed April 13, 205. <http://ec.europa.eu/agriculture/cap-post-2013/>.
- Radford, R. A. "The Economic Organisation of a P.O.W. Camp." *Economica*. no. 48 (1945): 189-201. <http://www.jstor.org/stable/2550133>.

GLOSSARY

- demand curve** a graphic representation of the relationship between price and quantity demanded of a certain good or service, with quantity on the horizontal axis and the price on the vertical axis
- demand schedule** a table that shows a range of prices for a certain good or service and the quantity demanded at each price
- demand** the relationship between price and the quantity demanded of a certain good or service
- equilibrium price** the price where quantity demanded is equal to quantity supplied
- equilibrium quantity** the quantity at which quantity demanded and quantity supplied are equal for a certain price level
- equilibrium** the situation where quantity demanded is equal to the quantity supplied; the combination of price and quantity where there is no economic pressure from surpluses or shortages that would cause price or quantity to change
- excess demand** at the existing price, the quantity demanded exceeds the quantity supplied; also called a shortage
- excess supply** at the existing price, quantity supplied exceeds the quantity demanded; also called a surplus
- law of demand** the common relationship that a higher price leads to a lower quantity demanded of a

certain good or service and a lower price leads to a higher quantity demanded, while all other variables are held constant

law of supply the common relationship that a higher price leads to a greater quantity supplied and a lower price leads to a lower quantity supplied, while all other variables are held constant

price what a buyer pays for a unit of the specific good or service

quantity demanded the total number of units of a good or service consumers are willing to purchase at a given price

quantity supplied the total number of units of a good or service producers are willing to sell at a given price

shortage at the existing price, the quantity demanded exceeds the quantity supplied; also called excess demand

supply curve a line that shows the relationship between price and quantity supplied on a graph, with quantity supplied on the horizontal axis and price on the vertical axis

supply schedule a table that shows a range of prices for a good or service and the quantity supplied at each price

supply the relationship between price and the quantity supplied of a certain good or service

surplus at the existing price, quantity supplied exceeds the quantity demanded; also called excess supply

SOLUTIONS

Answers to Self-Check Questions

Since \$1.60 per gallon is above the equilibrium price, the quantity demanded would be lower at 550 gallons and the quantity supplied would be higher at 640 gallons. (These results are due to the laws of demand and supply, respectively.)

The outcome of lower Q_d and higher Q_s would be a surplus in the gasoline market of $640 - 550 = 90$ gallons.

4.2 SHIFTS IN DEMAND AND SUPPLY FOR GOODS AND SERVICES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify factors that affect demand
- Graph demand curves and demand shifts
- Identify factors that affect supply
- Graph supply curves and supply shifts

The previous module explored how **price** affects the quantity demanded and the quantity supplied. The result was the demand curve and the supply curve. Price, however, is not the only thing that influences demand. Nor is it the only thing that influences supply. For example, how is demand for vegetarian food affected if, say, health concerns cause more consumers to avoid eating meat? Or how is the supply of diamonds affected if diamond producers discover several new diamond mines? What are the major factors, in addition to the price, that influence demand or supply?

Visit this website to read a brief note on how marketing strategies can influence supply and demand of products.



WHAT FACTORS AFFECT DEMAND?

We defined demand as the amount of some product a consumer is willing and able to purchase at each price. That suggests at least two factors in addition to price that affect demand. Willingness to purchase suggests a desire, based on what economists call tastes and preferences. If you neither need nor want something, you will not buy it. Ability to purchase suggests that income is important. Professors are usually able to afford better housing and transportation than students, because they have

more income. Prices of related goods can affect demand also. If you need a new car, the price of a Honda may affect your demand for a Ford. Finally, the size or composition of the population can affect demand. The more children a family has, the greater their demand for clothing. The more driving-age children a family has, the greater their demand for car insurance, and the less for diapers and baby formula.

These factors matter both for demand by an individual and demand by the market as a whole. Exactly how do these various factors affect demand, and how do we show the effects graphically? To answer those questions, we need the *ceteris paribus* assumption.

THE CETERIS PARIBUS ASSUMPTION

A **demand curve** or a **supply curve** is a relationship between two, and only two, variables: quantity on the horizontal axis and price on the vertical axis. The assumption behind a demand curve or a supply curve is that no relevant economic factors, other than the product's price, are changing. Economists call this assumption **ceteris paribus**, a Latin phrase meaning "other things being equal." Any given demand or supply curve is based on the *ceteris paribus* assumption that all else is held equal. A demand curve or a supply curve is a relationship between two, and only two, variables when all other variables are kept constant. If all else is not held equal, then the laws of supply and demand will not necessarily hold, as the following Clear It Up feature shows.

WHEN DOES CETERIS PARIBUS APPLY?

Ceteris paribus is typically applied when we look at how changes in price affect demand or supply, but *ceteris paribus* can be applied more generally. In the real world, demand and supply depend on more factors than just price. For example, a consumer's demand depends on income and a producer's supply depends on the cost of producing the product. How can we analyze the effect on demand or supply if multiple factors are changing at the same time—say price rises and income falls? The answer is that we examine the changes one at a time, assuming the other factors are held constant.

For example, we can say that an increase in the price reduces the amount consumers will buy (assuming income, and anything else that affects demand, is unchanged). Additionally, a decrease in income reduces the amount consumers can afford to buy (assuming price, and anything else that affects demand, is unchanged). This is what the *ceteris paribus* assumption really means. In this particular case, after we analyze each factor separately, we can combine the results. The amount consumers buy falls for two reasons: first because of the higher price and second because of the lower income.

HOW DOES INCOME AFFECT DEMAND?

Let's use income as an example of how factors other than price affect demand. Figure 1 shows the initial demand for automobiles as D_0 . At point Q, for example, if the price is \$20,000 per car, the quantity of cars demanded is 18 million. D_0 also shows how the quantity of cars demanded would change as a result of a higher or lower price. For example, if the price of a car rose to \$22,000, the quantity demanded would decrease to 17 million, at point R.

The original demand curve D_0 , like every demand curve, is based on the *ceteris paribus* assumption that no other economically relevant factors change. Now imagine that the economy expands in a way that raises the incomes of many people, making cars more affordable. How will this affect demand? How can we show this graphically?

Return to Figure 1. The price of cars is still \$20,000, but with higher incomes, the quantity demanded has now increased to 20 million cars, shown at point S. As a result of the higher income levels, the demand curve shifts to the right to the new demand curve D_1 , indicating an increase in demand. Table 4 shows clearly that this increased demand would occur at every price, not just the original one.

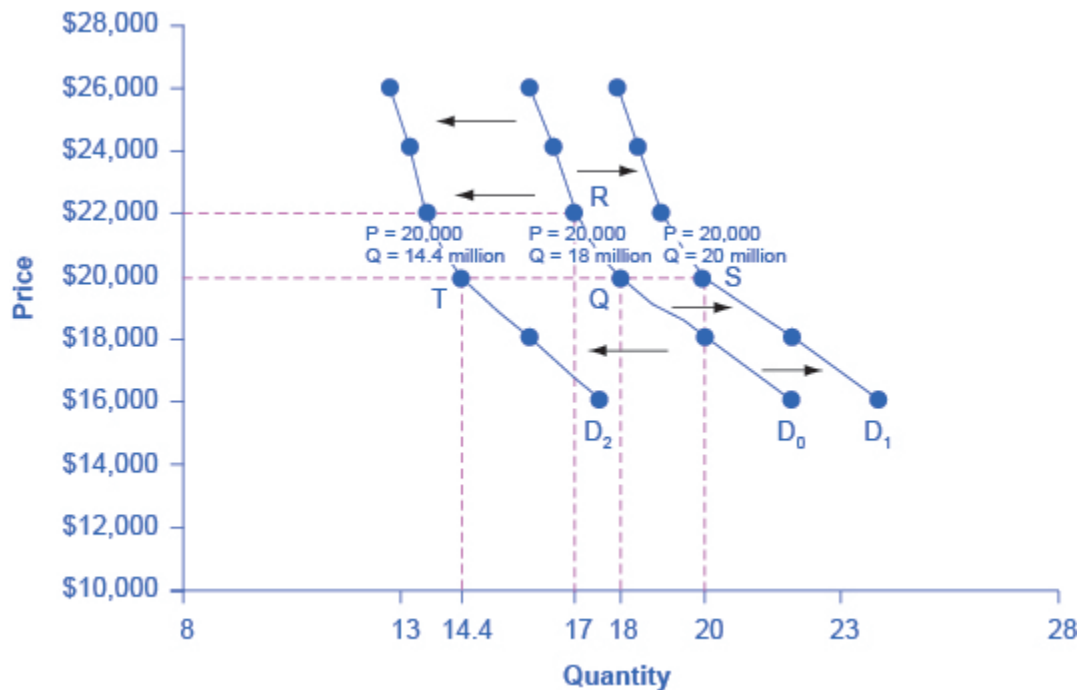


Figure 1. Shifts in Demand: A Car Example. Increased demand means that at every given price, the quantity demanded is higher, so that the demand curve shifts to the right from D_0 to D_1 . Decreased demand means that at every given price, the quantity demanded is lower, so that the demand curve shifts to the left from D_0 to D_2 .

Price	Decrease to D_2	Original Quantity Demanded D_0	Increase to D_1
\$16,000	17.6 million	22.0 million	24.0 million
\$18,000	16.0 million	20.0 million	22.0 million
\$20,000	14.4 million	18.0 million	20.0 million
\$22,000	13.6 million	17.0 million	19.0 million
\$24,000	13.2 million	16.5 million	18.5 million
\$26,000	12.8 million	16.0 million	18.0 million

Table 4. Price and Demand Shifts: A Car Example

Now, imagine that the economy slows down so that many people lose their jobs or work fewer hours, reducing their incomes. In this case, the decrease in income would lead to a lower quantity of cars demanded at every given price, and the original demand curve D_0 would shift left to D_2 . The shift from D_0 to D_2 represents such a decrease in demand: At any given price level, the quantity demanded is now lower. In this example, a price of \$20,000 means 18 million cars sold along the original demand curve, but only 14.4 million sold after demand fell.

When a demand curve shifts, it does not mean that the quantity demanded by every individual buyer

changes by the same amount. In this example, not everyone would have higher or lower income and not everyone would buy or not buy an additional car. Instead, a shift in a demand curve captures an pattern for the market as a whole.

In the previous section, we argued that higher income causes greater demand at every price. This is true for most goods and services. For some—luxury cars, vacations in Europe, and fine jewelry—the effect of a rise in income can be especially pronounced. A product whose demand rises when income rises, and vice versa, is called a **normal good**. A few exceptions to this pattern do exist. As incomes rise, many people will buy fewer generic brand groceries and more name brand groceries. They are less likely to buy used cars and more likely to buy new cars. They will be less likely to rent an apartment and more likely to own a home, and so on. A product whose demand falls when income rises, and vice versa, is called an **inferior good**. In other words, when income increases, the demand curve shifts to the left.

OTHER FACTORS THAT SHIFT DEMAND CURVES

Income is not the only factor that causes a shift in demand. Other things that change demand include tastes and preferences, the composition or size of the population, the prices of related goods, and even expectations. A change in any one of the underlying factors that determine what quantity people are willing to buy at a given price will cause a shift in demand. Graphically, the new demand curve lies either to the right (an increase) or to the left (a decrease) of the original demand curve. Let's look at these factors.

Changing Tastes or Preferences

From 1980 to 2014, the per-person consumption of chicken by Americans rose from 48 pounds per year to 85 pounds per year, and consumption of beef fell from 77 pounds per year to 54 pounds per year, according to the U.S. Department of Agriculture (USDA). Changes like these are largely due to movements in taste, which change the quantity of a good demanded at every price: that is, they shift the demand curve for that good, rightward for chicken and leftward for beef.

Changes in the Composition of the Population

The proportion of elderly citizens in the United States population is rising. It rose from 9.8% in 1970 to 12.6% in 2000, and will be a projected (by the **U.S. Census Bureau**) 20% of the population by 2030. A society with relatively more children, like the United States in the 1960s, will have greater demand for goods and services like tricycles and day care facilities. A society with relatively more elderly persons, as the United States is projected to have by 2030, has a higher demand for nursing homes and hearing aids. Similarly, changes in the size of the population can affect the demand for housing and many other goods. Each of these changes in demand will be shown as a shift in the demand curve.

The demand for a product can also be affected by changes in the prices of related goods such as substitutes or complements. A **substitute** is a good or service that can be used in place of another good or service. As electronic books, like this one, become more available, you would expect to see a decrease in demand for traditional printed books. A lower price for a substitute decreases demand for the other product. For example, in recent years as the price of tablet computers has fallen, the quantity demanded has increased (because of the law of demand). Since people are purchasing tablets, there has been a decrease in demand for laptops, which can be shown graphically as a leftward shift in the demand curve for laptops. A higher price for a substitute good has the reverse effect.

Other goods are **complements** for each other, meaning that the goods are often used together, because consumption of one good tends to enhance consumption of the other. Examples include breakfast cereal and milk; notebooks and pens or pencils, golf balls and golf clubs; gasoline and sport utility vehicles; and the five-way combination of bacon, lettuce, tomato, mayonnaise, and bread. If the price of golf clubs rises, since the quantity demanded of golf clubs falls (because of the law of demand), demand for a complement good like golf balls decreases, too. Similarly, a higher price for skis would shift the demand curve for a complement good like ski resort trips to the left, while a lower price for a complement has the reverse effect.

Changes in Expectations about Future Prices or Other Factors that Affect Demand

While it is clear that the price of a good affects the quantity demanded, it is also true that expectations about the future price (or expectations about tastes and preferences, income, and so on) can affect demand. For example, if people hear that a hurricane is coming, they may rush to the store to buy flashlight batteries and bottled water. If people learn that the price of a good like coffee is likely to rise in the future, they may head for the store to stock up on coffee now. These changes in demand are shown as shifts in the curve. Therefore, a **shift in demand** happens when a change in some economic factor (other than price) causes a different quantity to be demanded at every price. The following Work It Out feature shows how this happens.

SHIFT IN DEMAND

A shift in demand means that at any price (and at every price), the quantity demanded will be different than it was before. Following is an example of a shift in demand due to an income increase.

Step 1. Draw the graph of a demand curve for a normal good like pizza. Pick a price (like P_0). Identify the corresponding Q_0 . An example is shown in Figure 2.

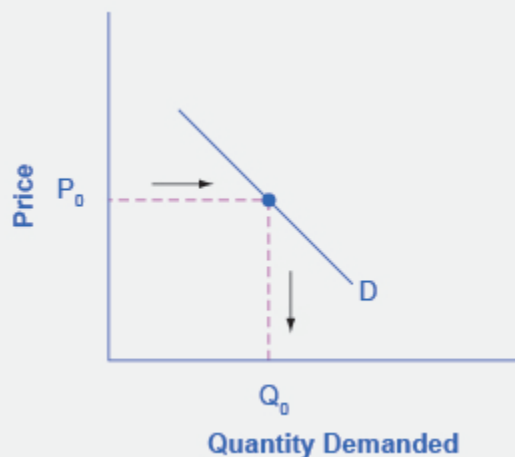


Figure 2. Demand Curve. The demand curve can be used to identify how much consumers would buy at any given price.

Step 2. Suppose income increases. As a result of the change, are consumers going to buy more or less pizza? The answer is more. Draw a dotted horizontal line from the chosen price, through the original quantity demanded, to the new point with the new Q_1 . Draw a dotted vertical line down to the horizontal axis and label the new Q_1 . An example is provided in Figure 3.

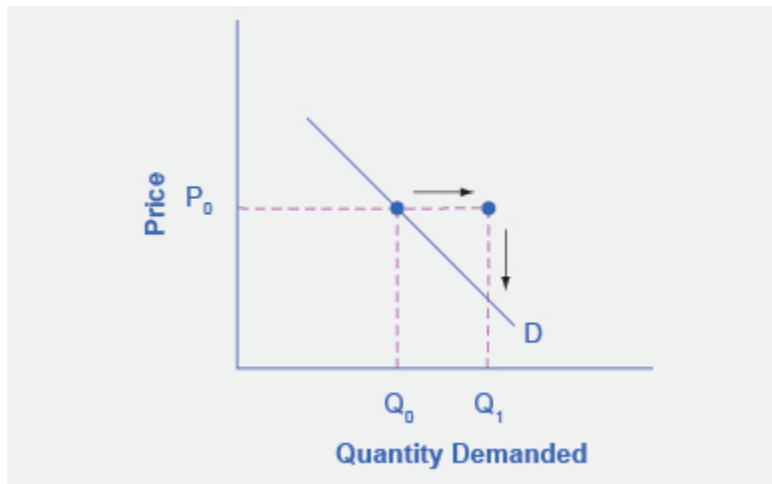


Figure 3. Demand Curve with Income Increase. With an increase in income, consumers will purchase larger quantities, pushing demand to the right.

Step 3. Now, shift the curve through the new point. You will see that an increase in income causes an upward (or rightward) shift in the demand curve, so that at any price the quantities demanded will be higher, as shown in Figure 4.

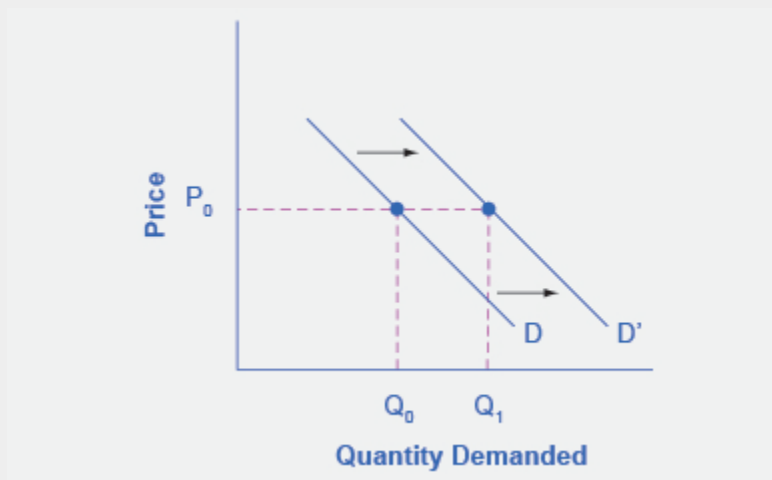


Figure 4. Demand Curve Shifted Right. With an increase in income, consumers will purchase larger quantities, pushing demand to the right, and causing the demand curve to shift right.

SUMMING UP FACTORS THAT CHANGE DEMAND

Six factors that can shift demand curves are summarized in Figure 5. The direction of the arrows indicates whether the demand curve shifts represent an increase in demand or a decrease in demand. Notice that a change in the price of the good or service itself is not listed among the factors that can shift a demand curve. A change in the price of a good or service causes a movement along a specific demand curve, and it typically leads to some change in the quantity demanded, but it does not shift the demand curve.

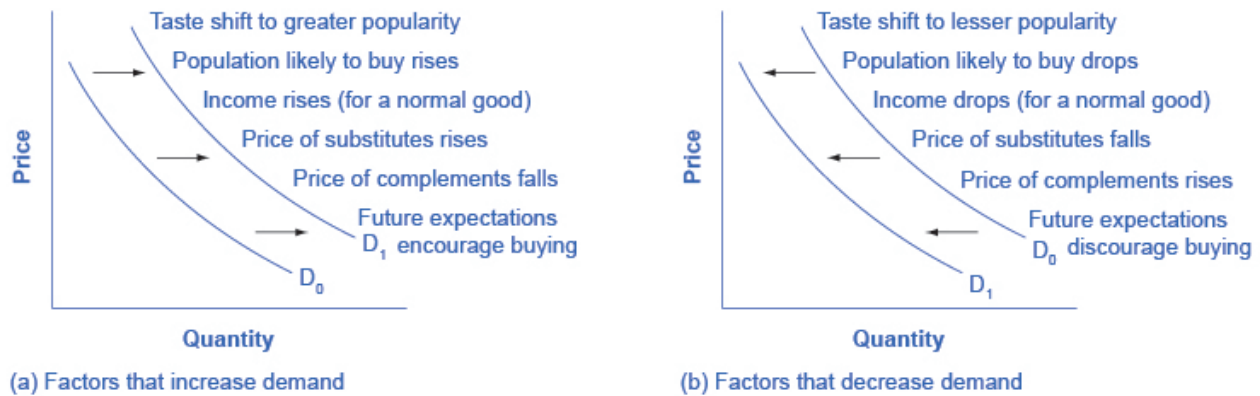


Figure 5. Factors That Shift Demand Curves. (a) A list of factors that can cause an increase in demand from D_0 to D_1 . (b) The same factors, if their direction is reversed, can cause a decrease in demand from D_0 to D_1 .

When a demand curve shifts, it will then intersect with a given supply curve at a different equilibrium price and quantity. We are, however, getting ahead of our story. Before discussing how changes in demand can affect equilibrium price and quantity, we first need to discuss shifts in supply curves.

HOW PRODUCTION COSTS AFFECT SUPPLY

A supply curve shows how quantity supplied will change as the price rises and falls, assuming *ceteris paribus* so that no other economically relevant factors are changing. If other factors relevant to supply do change, then the entire supply curve will shift. Just as a shift in demand is represented by a change in the quantity demanded at every price, a **shift in supply** means a change in the quantity supplied at every price.

In thinking about the factors that affect supply, remember what motivates firms: profits, which are the difference between revenues and costs. Goods and services are produced using combinations of labor, materials, and machinery, or what we call **inputs** or **factors of production**. If a firm faces lower costs of production, while the prices for the good or service the firm produces remain unchanged, a firm's profits go up. When a firm's profits increase, it is more motivated to produce output, since the more it produces the more profit it will earn. So, when costs of production fall, a firm will tend to supply a larger quantity at any given price for its output. This can be shown by the supply curve shifting to the right.

Take, for example, a messenger company that delivers packages around a city. The company may find that buying gasoline is one of its main costs. If the price of gasoline falls, then the company will find it can deliver messages more cheaply than before. Since lower costs correspond to higher profits, the messenger company may now supply more of its services at any given price. For example, given the lower gasoline prices, the company can now serve a greater area, and increase its supply.

Conversely, if a firm faces higher costs of production, then it will earn lower profits at any given selling price for its products. As a result, a higher cost of production typically causes a firm to supply a smaller quantity at any given price. In this case, the supply curve shifts to the left.

Consider the supply for cars, shown by curve S_0 in Figure 6. Point J indicates that if the price is \$20,000, the quantity supplied will be 18 million cars. If the price rises to \$22,000 per car, *ceteris*

paribus, the quantity supplied will rise to 20 million cars, as point K on the S_0 curve shows. The same information can be shown in table form, as in Table 5.

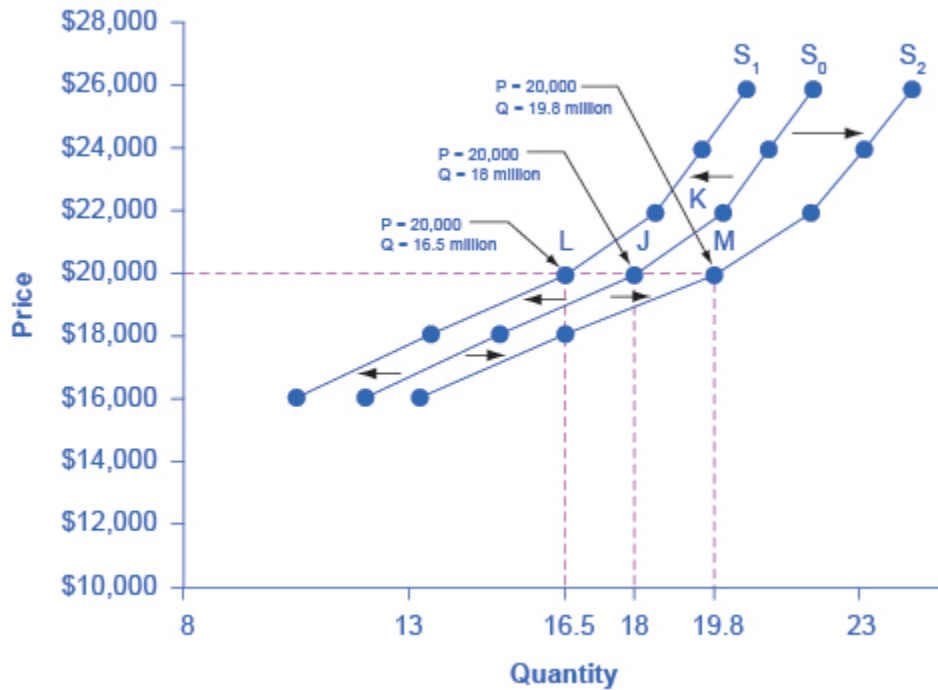


Figure 6. Shifts in Supply: A Car Example. Decreased supply means that at every given price, the quantity supplied is lower, so that the supply curve shifts to the left, from S_0 to S_1 . Increased supply means that at every given price, the quantity supplied is higher, so that the supply curve shifts to the right, from S_0 to S_2 .

Price	Decrease to S_1	Original Quantity Supplied S_0	Increase to S_2
\$16,000	10.5 million	12.0 million	13.2 million
\$18,000	13.5 million	15.0 million	16.5 million
\$20,000	16.5 million	18.0 million	19.8 million
\$22,000	18.5 million	20.0 million	22.0 million
\$24,000	19.5 million	21.0 million	23.1 million
\$26,000	20.5 million	22.0 million	24.2 million

Table 5. Price and Shifts in Supply: A Car Example

Now, imagine that the price of steel, an important ingredient in manufacturing cars, rises, so that producing a car has become more expensive. At any given price for selling cars, car manufacturers will react by supplying a lower quantity. This can be shown graphically as a leftward shift of supply, from S_0 to S_1 , which indicates that at any given price, the quantity supplied decreases. In this example, at a price of \$20,000, the quantity supplied decreases from 18 million on the original supply curve (S_0) to 16.5 million on the supply curve S_1 , which is labeled as point L.

Conversely, if the price of steel decreases, producing a car becomes less expensive. At any given price for selling cars, car manufacturers can now expect to earn higher profits, so they will supply a higher

quantity. The shift of supply to the right, from S_0 to S_2 , means that at all prices, the quantity supplied has increased. In this example, at a price of \$20,000, the quantity supplied increases from 18 million on the original supply curve (S_0) to 19.8 million on the supply curve S_2 , which is labeled M.

OTHER FACTORS THAT AFFECT SUPPLY

In the example above, we saw that changes in the prices of inputs in the production process will affect the cost of production and thus the supply. Several other things affect the cost of production, too, such as changes in weather or other natural conditions, new technologies for production, and some government policies.

The cost of production for many agricultural products will be affected by changes in natural conditions. For example, in 2014 the Manchurian Plain in Northeastern China, which produces most of the country's wheat, corn, and soybeans, experienced its most severe drought in 50 years. A drought decreases the supply of agricultural products, which means that at any given price, a lower quantity will be supplied; conversely, especially good weather would shift the supply curve to the right.

When a **firm** discovers a new technology that allows the firm to produce at a lower cost, the supply curve will shift to the right, as well. For instance, in the 1960s a major scientific effort nicknamed the Green Revolution focused on breeding improved seeds for basic crops like wheat and rice. By the early 1990s, more than two-thirds of the wheat and rice in low-income countries around the world was grown with these Green Revolution seeds—and the harvest was twice as high per acre. A technological improvement that reduces costs of production will shift supply to the right, so that a greater quantity will be produced at any given price.

Government policies can affect the cost of production and the supply curve through taxes, regulations, and subsidies. For example, the U.S. government imposes a tax on alcoholic beverages that collects about \$8 billion per year from producers. Taxes are treated as costs by businesses. Higher costs decrease supply for the reasons discussed above. Other examples of policy that can affect cost are the wide array of government regulations that require firms to spend money to provide a cleaner environment or a safer workplace; complying with regulations increases costs.

A government subsidy, on the other hand, is the opposite of a tax. A subsidy occurs when the government pays a firm directly or reduces the firm's taxes if the firm carries out certain actions. From the firm's perspective, taxes or regulations are an additional cost of production that shifts supply to the left, leading the firm to produce a lower quantity at every given price. Government subsidies reduce the cost of production and increase supply at every given price, shifting supply to the right. The following Work It Out feature shows how this shift happens.

SHIFT IN SUPPLY

We know that a supply curve shows the minimum price a firm will accept to produce a given quantity of output. What happens to the supply curve when the cost of production goes up? Following is an example of a shift in supply due to a production cost increase.

Step 1. Draw a graph of a supply curve for pizza. Pick a quantity (like Q_0). If you draw a vertical line up from Q_0 to the supply curve, you will see the price the firm chooses. An example is shown in Figure 7.

Step 2. Why did the firm choose that price and not some other? One way to think about this is that the price is composed

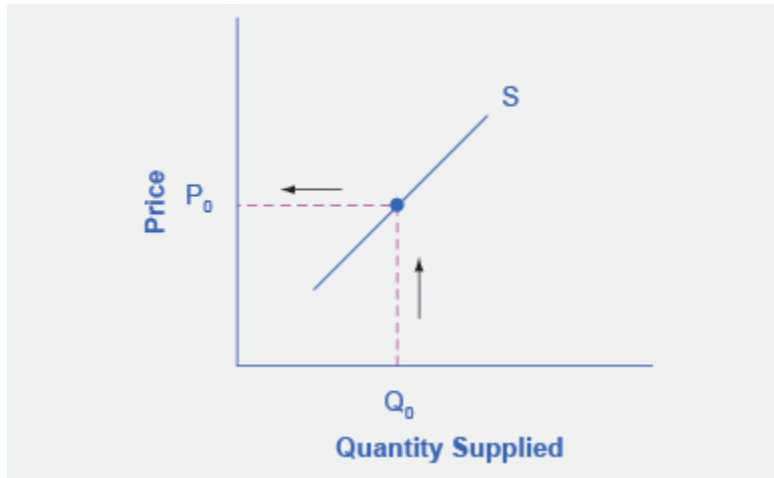


Figure 7. Supply Curve. The supply curve can be used to show the minimum price a firm will accept to produce a given quantity of output.

of two parts. The first part is the average cost of production, in this case, the cost of the pizza ingredients (dough, sauce, cheese, pepperoni, and so on), the cost of the pizza oven, the rent on the shop, and the wages of the workers. The second part is the firm's desired profit, which is determined, among other factors, by the profit margins in that particular business. If you add these two parts together, you get the price the firm wishes to charge. The quantity Q_0 and associated price P_0 give you one point on the firm's supply curve, as shown in Figure 8.

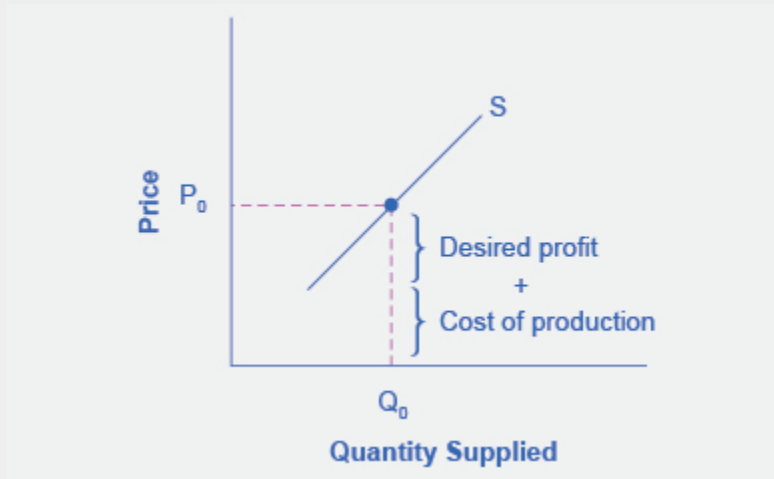


Figure 8. Setting Prices. The cost of production and the desired profit equal the price a firm will set for a product.

Step 3. Now, suppose that the cost of production goes up. Perhaps cheese has become more expensive by \$0.75 per pizza. If that is true, the firm will want to raise its price by the amount of the increase in cost (\$0.75). Draw this point on the supply curve directly above the initial point on the curve, but \$0.75 higher, as shown in Figure 9.

Step 4. Shift the supply curve through this point. You will see that an increase in cost causes an upward (or a leftward) shift of the supply curve so that at any price, the quantities supplied will be smaller, as shown in Figure 10.

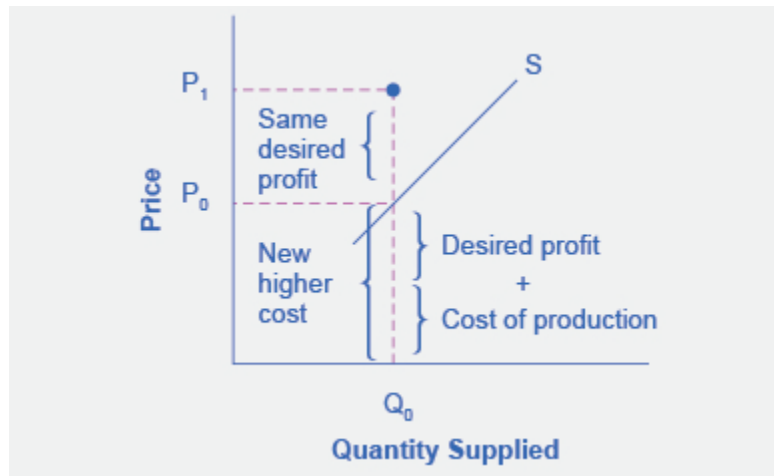


Figure 9. Increasing Costs Leads to Increasing Price. Because the cost of production and the desired profit equal the price a firm will set for a product, if the cost of production increases, the price for the product will also need to increase.

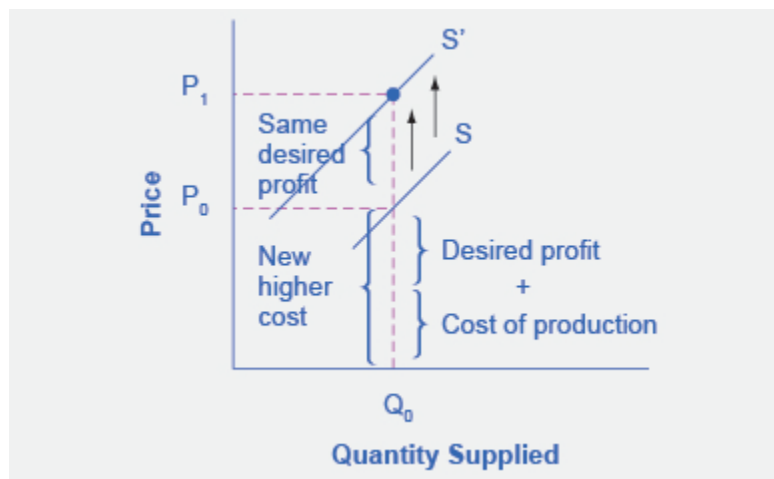


Figure 10. Supply Curve Shifts. When the cost of production increases, the supply curve shifts upwardly to a new price level.

SUMMING UP FACTORS THAT CHANGE SUPPLY

Changes in the cost of inputs, natural disasters, new technologies, and the impact of government decisions all affect the cost of production. In turn, these factors affect how much firms are willing to supply at any given price.

Figure 11 summarizes factors that change the supply of goods and services. Notice that a change in the price of the product itself is not among the factors that shift the supply curve. Although a change in price of a good or service typically causes a change in quantity supplied or a movement along the supply curve for that specific good or service, it does not cause the supply curve itself to shift.

Because demand and supply curves appear on a two-dimensional diagram with only price and quantity on the axes, an unwary visitor to the land of economics might be fooled into believing that eco-

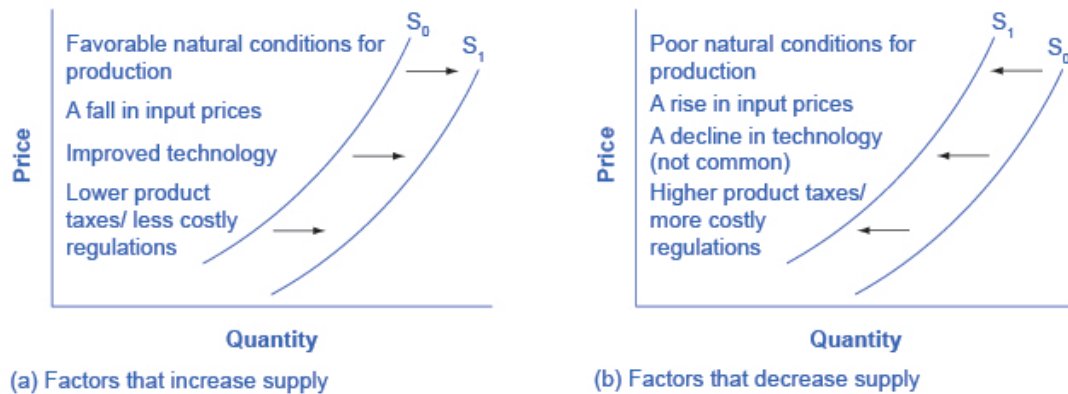


Figure 11. Factors That Shift Supply Curves. (a) A list of factors that can cause an increase in supply from S_0 to S_1 . (b) The same factors, if their direction is reversed, can cause a decrease in supply from S_0 to S_1 .

nomics is about only four topics: demand, supply, price, and quantity. However, demand and supply are really “umbrella” concepts: demand covers all the factors that affect demand, and supply covers all the factors that affect supply. Factors other than price that affect demand and supply are included by using shifts in the demand or the supply curve. In this way, the two-dimensional demand and supply model becomes a powerful tool for analyzing a wide range of economic circumstances.

KEY CONCEPTS AND SUMMARY

Economists often use the *ceteris paribus* or “other things being equal” assumption: while examining the economic impact of one event, all other factors remain unchanged for the purpose of the analysis. Factors that can shift the demand curve for goods and services, causing a different quantity to be demanded at any given price, include changes in tastes, population, income, prices of substitute or complement goods, and expectations about future conditions and prices. Factors that can shift the supply curve for goods and services, causing a different quantity to be supplied at any given price, include input prices, natural conditions, changes in technology, and government taxes, regulations, or subsidies.

SELF-CHECK QUESTIONS

- Why do economists use the *ceteris paribus* assumption?
- In an analysis of the market for paint, an economist discovers the facts listed below. State whether each of these changes will affect supply or demand, and in what direction.
 - There have recently been some important cost-saving inventions in the technology for making paint.
 - Paint is lasting longer, so that property owners need not repaint as often.
 - Because of severe hailstorms, many people need to repaint now.
 - The hailstorms damaged several factories that make paint, forcing them to close down for several months.
- Many changes are affecting the market for oil. Predict how each of the following events will affect the

equilibrium price and quantity in the market for oil. In each case, state how the event will affect the supply and demand diagram. Create a sketch of the diagram if necessary.

- a. Cars are becoming more fuel efficient, and therefore get more miles to the gallon.
- b. The winter is exceptionally cold.
- c. A major discovery of new oil is made off the coast of Norway.
- d. The economies of some major oil-using nations, like Japan, slow down.
- e. A war in the Middle East disrupts oil-pumping schedules.
- f. Landlords install additional insulation in buildings.
- g. The price of solar energy falls dramatically.
- h. Chemical companies invent a new, popular kind of plastic made from oil.

REVIEW QUESTIONS

1. When analyzing a market, how do economists deal with the problem that many factors that affect the market are changing at the same time?
2. Name some factors that can cause a shift in the demand curve in markets for goods and services.
3. Name some factors that can cause a shift in the supply curve in markets for goods and services.

CRITICAL THINKING QUESTIONS

1. Consider the demand for hamburgers. If the price of a substitute good (for example, hot dogs) increases and the price of a complement good (for example, hamburger buns) increases, can you tell for sure what will happen to the demand for hamburgers? Why or why not? Illustrate your answer with a graph.
2. How do you suppose the demographics of an aging population of “Baby Boomers” in the United States will affect the demand for milk? Justify your answer.
3. We know that a change in the price of a product causes a movement along the demand curve. Suppose consumers believe that prices will be rising in the future. How will that affect demand for the product in the present? Can you show this graphically?
4. Suppose there is a soda tax to curb obesity. What should a reduction in the soda tax do to the supply of sodas and to the equilibrium price and quantity? Can you show this graphically? *Hint:* assume that the soda tax is collected from the sellers

PROBLEMS

1. Table 6 shows information on the demand and supply for bicycles, where the quantities of bicycles are measured in thousands.

Price	Qd	Qs
\$120	50	36
\$150	40	40
\$180	32	48
\$210	28	56
\$240	24	70

Table 6. Demand and Supply for Bicycles

1. What is the quantity demanded and the quantity supplied at a price of \$210?
 2. At what price is the quantity supplied equal to 48,000?
 3. Graph the demand and supply curve for bicycles. How can you determine the equilibrium price and quantity from the graph? How can you determine the equilibrium price and quantity from the table? What are the equilibrium price and equilibrium quantity?
 4. If the price was \$120, what would the quantities demanded and supplied be? Would a shortage or surplus exist? If so, how large would the shortage or surplus be?
2. The computer market in recent years has seen many more computers sell at much lower prices. What shift in demand or supply is most likely to explain this outcome? Sketch a demand and supply diagram and explain your reasoning for each.
1. A rise in demand
 2. A fall in demand
 3. A rise in supply
 4. A fall in supply

REFERENCES

Landsburg, Steven E. *The Armchair Economist: Economics and Everyday Life*. New York: The Free Press. 2012. specifically Section IV: How Markets Work.

National Chicken Council. 2015. "Per Capita Consumption of Poultry and Livestock, 1965 to Estimated 2015, in Pounds." Accessed April 13, 2015. <http://www.nationalchickencouncil.org/about-the-industry/statistics/per-capita-consumption-of-poultry-and-livestock-1965-to-estimated-2012-in-pounds/>.

Wessel, David. "Saudi Arabia Fears \$40-a-Barrel Oil, Too." *The Wall Street Journal*. May 27, 2004, p. 42. <http://online.wsj.com/news/articles/SB108561000087822300>.

GLOSSARY

ceteris paribus other things being equal

complements goods that are often used together so that consumption of one good tends to enhance consumption of the other

factors of production the combination of labor, materials, and machinery that is used to produce goods and services; also called inputs

inferior good a good in which the quantity demanded falls as income rises, and in which quantity demanded rises and income falls

inputs the combination of labor, materials, and machinery that is used to produce goods and services; also called factors of production

normal good a good in which the quantity demanded rises as income rises, and in which quantity demanded falls as income falls

shift in demand when a change in some economic factor (other than price) causes a different quantity to be demanded at every price

shift in supply when a change in some economic factor (other than price) causes a different quantity to be supplied at every price

substitute a good that can replace another to some extent, so that greater consumption of one good can mean less of the other

SOLUTIONS

Answers to Self-Check Questions

1. To make it easier to analyze complex problems. *Ceteris paribus* allows you to look at the effect of one factor at a time on what it is you are trying to analyze. When you have analyzed all the factors individually, you add the results together to get the final answer.
2.
 1. An improvement in technology that reduces the cost of production will cause an increase in supply. Alternatively, you can think of this as a reduction in price necessary for firms to supply any quantity. Either way, this can be shown as a rightward (or downward) shift in the supply curve.
 2. An improvement in product quality is treated as an increase in tastes or preferences, meaning consumers demand more paint at any price level, so demand increases or shifts to the right. If this seems counterintuitive, note that demand in the future for the longer-lasting paint will fall, since consumers are essentially shifting demand from the future to the present.
 3. An increase in need causes an increase in demand or a rightward shift in the demand curve.
 4. Factory damage means that firms are unable to supply as much in the present. Technically, this is an increase in the cost of production. Either way you look at it, the supply curve shifts to the left.
3.
 1. More fuel-efficient cars means there is less need for gasoline. This causes a leftward shift in the demand for gasoline and thus oil. Since the demand curve is shifting down the supply curve, the equilibrium price and quantity both fall.
 2. Cold weather increases the need for heating oil. This causes a rightward shift in the demand for heating oil and thus oil. Since the demand curve is shifting up the supply curve, the equilibrium price and quantity both rise.
 3. A discovery of new oil will make oil more abundant. This can be shown as a rightward shift in the supply curve, which will cause a decrease in the equilibrium price along with an increase in the equilibrium quantity. (The supply curve shifts down the demand curve so price and quantity follow the law of demand. If price goes down, then the quantity goes up.)
 4. When an economy slows down, it produces less output and demands less input, including energy, which is used in the production of virtually everything. A decrease in demand for energy will be reflected as a decrease in the demand for oil, or a leftward shift in demand for oil. Since the

demand curve is shifting down the supply curve, both the equilibrium price and quantity of oil will fall.

5. Disruption of oil pumping will reduce the supply of oil. This leftward shift in the supply curve will show a movement up the demand curve, resulting in an increase in the equilibrium price of oil and a decrease in the equilibrium quantity.
6. Increased insulation will decrease the demand for heating. This leftward shift in the demand for oil causes a movement down the supply curve, resulting in a decrease in the equilibrium price and quantity of oil.
7. Solar energy is a substitute for oil-based energy. So if solar energy becomes cheaper, the demand for oil will decrease as consumers switch from oil to solar. The decrease in demand for oil will be shown as a leftward shift in the demand curve. As the demand curve shifts down the supply curve, both equilibrium price and quantity for oil will fall.
8. A new, popular kind of plastic will increase the demand for oil. The increase in demand will be shown as a rightward shift in demand, raising the equilibrium price and quantity of oil.

4.3 CHANGES IN EQUILIBRIUM PRICE AND QUANTITY: THE FOUR-STEP PROCESS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify equilibrium price and quantity through the four-step process
- Graph equilibrium price and quantity
- Contrast shifts of demand or supply and movements along a demand or supply curve
- Graph demand and supply curves, including equilibrium price and quantity, based on real-world examples

Let's begin this discussion with a single economic event. It might be an event that affects demand, like a change in income, population, tastes, prices of substitutes or complements, or expectations about future prices. It might be an event that affects supply, like a change in natural conditions, input prices, or technology, or government policies that affect production. How does this economic event affect equilibrium price and quantity? We will analyze this question using a four-step process.

Step 1. Draw a demand and supply model before the economic change took place. To establish the model requires four standard pieces of information: The law of demand, which tells us the slope of the demand curve; the law of supply, which gives us the slope of the supply curve; the shift variables for demand; and the shift variables for supply. From this model, find the initial equilibrium values for price and quantity.

Step 2. Decide whether the economic change being analyzed affects demand or supply. In other words, does the event refer to something in the list of demand factors or supply factors?

Step 3. Decide whether the effect on demand or supply causes the curve to shift to the right or to the left, and sketch the new demand or supply curve on the diagram. In other words, does the event increase or decrease the amount consumers want to buy or producers want to sell?

Step 4. Identify the new equilibrium and then compare the original equilibrium price and quantity to the new equilibrium price and quantity.

Let's consider one example that involves a shift in supply and one that involves a shift in demand. Then we will consider an example where both supply and demand shift.

GOOD WEATHER FOR SALMON FISHING

In the summer of 2000, weather conditions were excellent for commercial salmon fishing off the California coast. Heavy rains meant higher than normal levels of water in the rivers, which helps the salmon to breed. Slightly cooler ocean temperatures stimulated the growth of plankton, the microscopic organisms at the bottom of the ocean food chain, providing everything in the ocean with a hearty food supply. The ocean stayed calm during fishing season, so commercial fishing operations did not lose many days to bad weather. How did these climate conditions affect the quantity and price of salmon? Figure 1 illustrates the four-step approach, which is explained below, to work through this problem. Table 7 provides the information to work the problem as well.

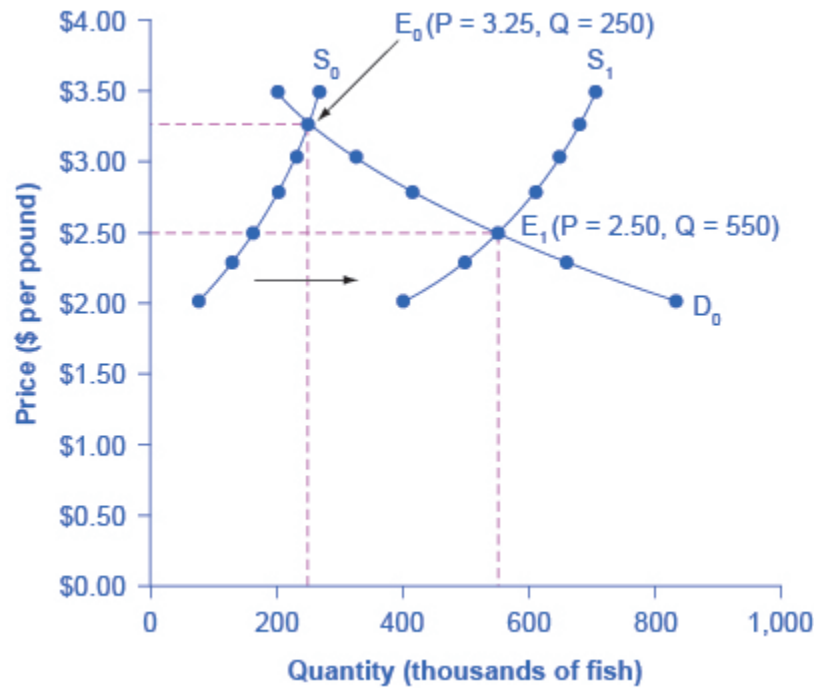


Figure 1. Good Weather for Salmon Fishing: The Four-Step Process. Unusually good weather leads to changes in the price and quantity of salmon.

Price per Pound	Quantity Supplied in 1999	Quantity Supplied in 2000	Quantity Demanded
\$2.00	80	400	840
\$2.25	120	480	680
\$2.50	160	550	550
\$2.75	200	600	450
\$3.00	230	640	350
\$3.25	250	670	250
\$3.50	270	700	200

Table 7. Salmon Fishing

Step 1. Draw a demand and supply model to illustrate the market for salmon in the year before the good weather conditions began. The demand curve D_0 and the supply curve S_0 show that the original

equilibrium price is \$3.25 per pound and the original equilibrium quantity is 250,000 fish. (This price per pound is what commercial buyers pay at the fishing docks; what consumers pay at the grocery is higher.)

Step 2. Did the economic event affect supply or demand? Good weather is an example of a natural condition that affects supply.

Step 3. Was the effect on supply an increase or a decrease? Good weather is a change in natural conditions that increases the quantity supplied at any given price. The supply curve shifts to the right, moving from the original supply curve S_0 to the new supply curve S_1 , which is shown in both the table and the figure.

Step 4. Compare the new equilibrium price and quantity to the original equilibrium. At the new equilibrium E_1 , the equilibrium price falls from \$3.25 to \$2.50, but the equilibrium quantity increases from 250,000 to 550,000 salmon. Notice that the equilibrium quantity demanded increased, even though the demand curve did not move.

In short, good weather conditions increased supply of the California commercial salmon. The result was a higher equilibrium quantity of salmon bought and sold in the market at a lower price.

NEWSPAPERS AND THE INTERNET

According to the **Pew Research Center for People and the Press**, more and more people, especially younger people, are getting their news from online and digital sources. The majority of U.S. adults now own smartphones or tablets, and most of those Americans say they use them in part to get the news. From 2004 to 2012, the share of Americans who reported getting their news from digital sources increased from 24% to 39%. How has this affected consumption of print news media, and radio and television news? Figure 2 and the text below illustrates using the four-step analysis to answer this question.

Step 1. Develop a demand and supply model to think about what the market looked like before the event. The demand curve D_0 and the supply curve S_0 show the original relationships. In this case, the analysis is performed without specific numbers on the price and quantity axis.

Step 2. Did the change described affect supply or demand? A change in tastes, from traditional news sources (print, radio, and television) to digital sources, caused a change in demand for the former.

Step 3. Was the effect on demand positive or negative? A shift to digital news sources will tend to mean a lower quantity demanded of traditional news sources at every given price, causing the demand curve for print and other traditional news sources to shift to the left, from D_0 to D_1 .

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium (E_1) occurs at a lower quantity and a lower price than the original equilibrium (E_0).

The decline in print news reading predates 2004. Print newspaper circulation peaked in 1973 and has declined since then due to competition from television and radio news. In 1991, 55% of Americans indicated they got their news from print sources, while only 29% did so in 2012. Radio news has followed a similar path in recent decades, with the share of Americans getting their news from radio

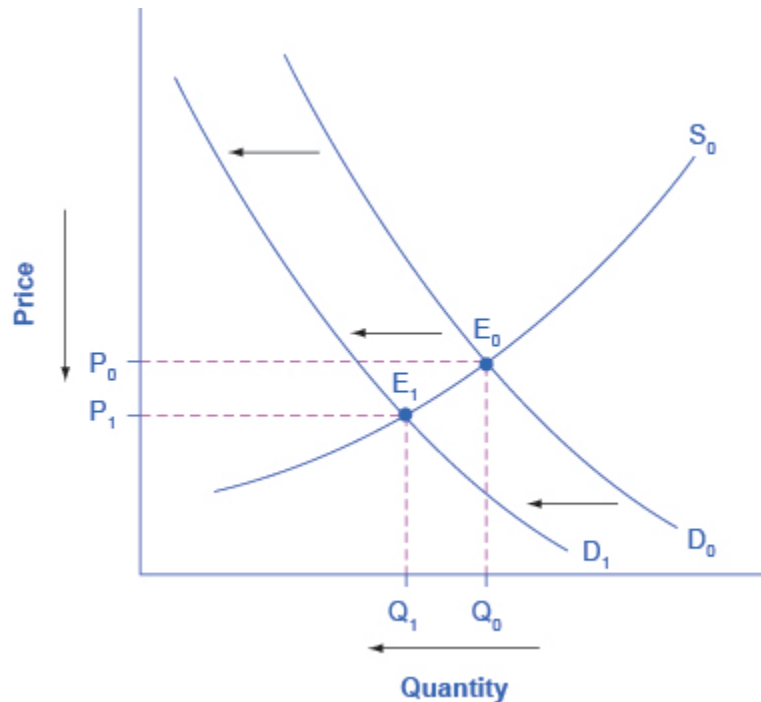


Figure 2. The Print News Market: A Four-Step Analysis. A change in tastes from print news sources to digital sources results in a leftward shift in demand for the former. The result is a decrease in both equilibrium price and quantity.

declining from 54% in 1991 to 33% in 2012. Television news has held its own over the last 15 years, with a market share staying in the mid to upper fifties. What does this suggest for the future, given that two-thirds of Americans under 30 years old say they do not get their news from television at all?

THE INTERCONNECTIONS AND SPEED OF ADJUSTMENT IN REAL MARKETS

In the real world, many factors that affect demand and supply can change all at once. For example, the demand for cars might increase because of rising incomes and population, and it might decrease because of rising gasoline prices (a complementary good). Likewise, the supply of cars might increase because of innovative new technologies that reduce the cost of car production, and it might decrease as a result of new government regulations requiring the installation of costly pollution-control technology.

Moreover, rising incomes and population or changes in gasoline prices will affect many markets, not just cars. How can an economist sort out all these interconnected events? The answer lies in the *ceteris paribus* assumption. Look at how each economic event affects each market, one event at a time, holding all else constant. Then combine the analyses to see the net effect.

A COMBINED EXAMPLE

The U.S. Postal Service is facing difficult challenges. Compensation for postal workers tends to increase most years due to cost-of-living increases. At the same time, more and more people are using email, text, and other digital message forms such as Facebook and Twitter to communicate with

friends and others. What does this suggest about the continued viability of the Postal Service? Figure 3 and the text below illustrates using the four-step analysis to answer this question.

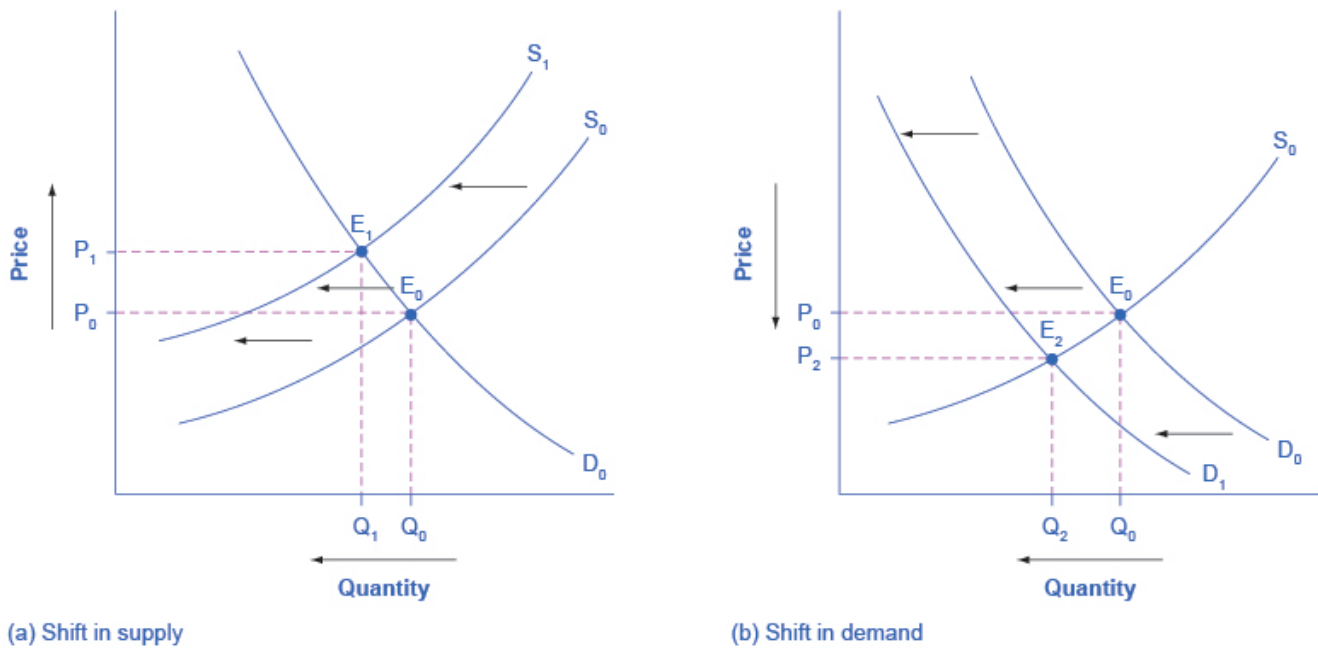


Figure 3. Higher Compensation for Postal Workers: A Four-Step Analysis. (a) Higher labor compensation causes a leftward shift in the supply curve, a decrease in the equilibrium quantity, and an increase in the equilibrium price. (b) A change in tastes away from Postal Services causes a leftward shift in the demand curve, a decrease in the equilibrium quantity, and a decrease in the equilibrium price.

Since this problem involves two disturbances, we need two four-step analyses, the first to analyze the effects of higher compensation for postal workers, the second to analyze the effects of many people switching from “snailmail” to email and other digital messages.

Figure 3 (a) shows the shift in supply discussed in the following steps.

Step 1. Draw a demand and supply model to illustrate what the market for the U.S. Postal Service looked like before this scenario starts. The demand curve D_0 and the supply curve S_0 show the original relationships.

Step 2. Did the change described affect supply or demand? Labor compensation is a cost of production. A change in production costs caused a change in supply for the Postal Service.

Step 3. Was the effect on supply positive or negative? Higher labor compensation leads to a lower quantity supplied of postal services at every given price, causing the supply curve for postal services to shift to the left, from S_0 to S_1 .

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium (E_1) occurs at a lower quantity and a higher price than the original equilibrium (E_0).

Figure 3 (b) shows the shift in demand discussed in the following steps.

Step 1. Draw a demand and supply model to illustrate what the market for U.S. Postal Services looked like before this scenario starts. The demand curve D_0 and the supply curve S_0 show the original relationships. Note that this diagram is independent from the diagram in panel (a).

Step 2. Did the change described affect supply or demand? A change in tastes away from snailmail toward digital messages will cause a change in demand for the Postal Service.

Step 3. Was the effect on supply positive or negative? Higher labor compensation leads to a lower quantity supplied of postal services at every given price, causing the supply curve for postal services to shift to the left, from D_0 to D_1 .

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium (E_2) occurs at a lower quantity and a lower price than the original equilibrium (E_0).

The final step in a scenario where both supply and demand shift is to combine the two individual analyses to determine what happens to the equilibrium quantity and price. Graphically, we superimpose the previous two diagrams one on top of the other, as in Figure 4.

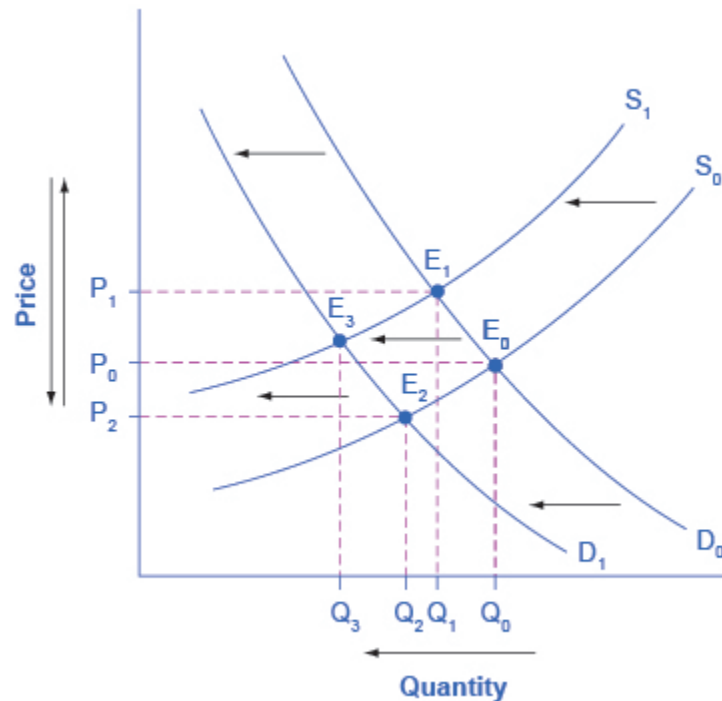


Figure 4. Combined Effect of Decreased Demand and Decreased Supply. Supply and demand shifts cause changes in equilibrium price and quantity.

Following are the results:

Effect on Quantity: The effect of higher labor compensation on Postal Services because it raises the cost of production is to decrease the equilibrium quantity. The effect of a change in tastes away from snailmail is to decrease the equilibrium quantity. Since both shifts are to the left, the overall impact is

a decrease in the equilibrium quantity of Postal Services (Q_3). This is easy to see graphically, since Q_3 is to the left of Q_0 .

Effect on Price: The overall effect on price is more complicated. The effect of higher labor compensation on Postal Services, because it raises the cost of production, is to increase the equilibrium price. The effect of a change in tastes away from snailmail is to decrease the equilibrium price. Since the two effects are in opposite directions, unless we know the magnitudes of the two effects, the overall effect is unclear. This is not unusual. When both curves shift, typically we can determine the overall effect on price or on quantity, but not on both. In this case, we determined the overall effect on the equilibrium quantity, but not on the equilibrium price. In other cases, it might be the opposite.

The next Clear It Up feature focuses on the difference between shifts of supply or demand and movements along a curve.

WHAT IS THE DIFFERENCE BETWEEN SHIFTS OF DEMAND OR SUPPLY VERSUS MOVEMENTS ALONG A DEMAND OR SUPPLY CURVE?

One common mistake in applying the demand and supply framework is to confuse the shift of a demand or a supply curve with movement along a demand or supply curve. As an example, consider a problem that asks whether a drought will increase or decrease the equilibrium quantity and equilibrium price of wheat. Lee, a student in an introductory economics class, might reason:

“Well, it is clear that a drought reduces supply, so I will shift back the supply curve, as in the shift from the original supply curve S_0 to S_1 shown on the diagram (called Shift 1). So the equilibrium moves from E_0 to E_1 , the **equilibrium quantity** is lower and the equilibrium price is higher. Then, a higher price makes farmers more likely to supply the good, so the supply curve shifts right, as shown by the shift from S_1 to S_2 , on the diagram (shown as Shift 2), so that the equilibrium now moves from E_1 to E_2 . The higher price, however, also reduces demand and so causes demand to shift back, like the shift from the original demand curve, D_0 to D_1 on the diagram (labeled Shift 3), and the equilibrium moves from E_2 to E_3 .”

At about this point, Lee suspects that this answer is headed down the wrong path. Think about what might be wrong with Lee’s logic, and then read the answer that follows.

Answer: Lee’s first step is correct: that is, a drought shifts back the supply curve of wheat and leads to a prediction of a lower equilibrium quantity and a higher equilibrium price. This corresponds to a movement along the original demand curve (D_0), from E_0 to E_1 . The rest of Lee’s argument is wrong, because it mixes up shifts in supply with quantity supplied, and shifts in demand with quantity demanded. A higher or lower price never shifts the supply curve, as suggested by the shift in supply from S_1 to S_2 . Instead, a price change leads to a movement along a given supply curve. Similarly, a higher or lower price never shifts a demand curve, as suggested in the shift from D_0 to D_1 . Instead, a price change leads to a movement along a given demand curve. Remember, a change in the price of a good never causes the demand or supply curve for that good to shift.

Think carefully about the timeline of events: What happens first, what happens next? What is cause, what is effect? If you keep the order right, you are more likely to get the analysis correct.

In the four-step analysis of how economic events affect equilibrium price and quantity, the movement from the old to the new equilibrium seems immediate. As a practical matter, however, prices and quantities often do not zoom straight to equilibrium. More realistically, when an economic event causes demand or supply to shift, prices and quantities set off in the general direction of equilibrium. Indeed, even as they are moving toward one new equilibrium, prices are often then pushed by another change in demand or supply toward another equilibrium.

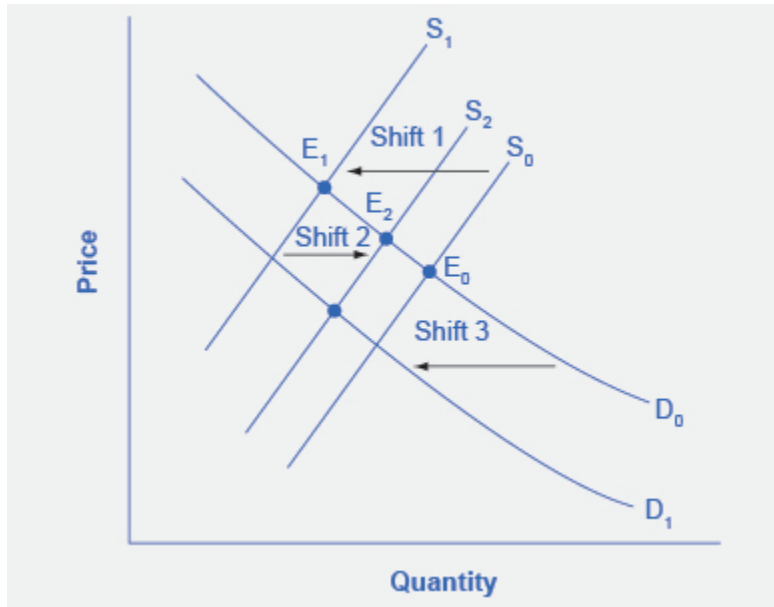


Figure 5. Shifts of Demand or Supply versus Movements along a Demand or Supply Curve. A shift in one curve never causes a shift in the other curve. Rather, a shift in one curve causes a movement along the second curve.

KEY CONCEPTS AND SUMMARY

When using the supply and demand framework to think about how an event will affect the equilibrium price and quantity, proceed through four steps: (1) sketch a supply and demand diagram to think about what the market looked like before the event; (2) decide whether the event will affect supply or demand; (3) decide whether the effect on supply or demand is negative or positive, and draw the appropriate shifted supply or demand curve; (4) compare the new equilibrium price and quantity to the original ones.

SELF-CHECK QUESTIONS

1. Let's think about the market for air travel. From August 2014 to January 2015, the price of jet fuel decreased roughly 47%. Using the four-step analysis, how do you think this fuel price decrease affected the equilibrium price and quantity of air travel?
2. A tariff is a tax on imported goods. Suppose the U.S. government cuts the tariff on imported flat screen televisions. Using the four-step analysis, how do you think the tariff reduction will affect the equilibrium price and quantity of flat screen TVs?

REVIEW QUESTIONS

1. How does one analyze a market where both demand and supply shift?
2. What causes a movement along the demand curve? What causes a movement along the supply curve?

CRITICAL THINKING QUESTIONS

1. Use the four-step process to analyze the impact of the advent of the iPod (or other portable digital music players) on the equilibrium price and quantity of the Sony Walkman (or other portable audio cassette players).
2. Use the four-step process to analyze the impact of a reduction in tariffs on imports of iPods on the equilibrium price and quantity of Sony Walkman-type products.
3. Suppose both of these events took place at the same time. Combine your analyses of the impacts of the iPod and the tariff reduction to determine the likely impact on the equilibrium price and quantity of Sony Walkman-type products. Show your answer graphically.

PROBLEMS

1. Demand and supply in the market for cheddar cheese is illustrated in Table 8. Graph the data and find the equilibrium. Next, create a table showing the change in quantity demanded or quantity supplied, and a graph of the new equilibrium, in each of the following situations:
 - a. The price of milk, a key input for cheese production, rises, so that the supply decreases by 80 pounds at every price.
 - b. A new study says that eating cheese is good for your health, so that demand increases by 20% at every price.

Price per Pound	Qd	Qs
\$3.00	750	540
\$3.20	700	600
\$3.40	650	650
\$3.60	620	700
\$3.80	600	720
\$4.00	590	730

Table 8. Demand and supply in the market for cheddar cheese

2. Supply and demand for movie tickets in a city are shown in Table 9. Graph demand and supply and identify the equilibrium. Then calculate in a table and graph the effect of the following two changes.
 - a. Three new nightclubs open. They offer decent bands and have no cover charge, but make their money by selling food and drink. As a result, demand for movie tickets falls by six units at every price.
 - b. The city eliminates a tax that it had been placing on all local entertainment businesses. The result is that the quantity supplied of movies at any given price increases by 10%.

Price per Pound	Qd	Qs
\$5.00	26	16
\$6.00	24	18
\$7.00	22	20
\$8.00	21	21
\$9.00	20	22

Table 9. Demand and supply for movie tickets

REFERENCES

Pew Research Center. "Pew Research: Center for the People & the Press." <http://www.people-press.org/>.

SOLUTIONS

Answers to Self-Check Questions

- Step 1. Draw the graph with the initial supply and demand curves. Label the initial equilibrium price and quantity.

Step 2. Did the economic event affect supply or demand? Jet fuel is a cost of producing air travel, so an increase in jet fuel price affects supply.

Step 3. An increase in the price of jet fuel caused a decrease in the cost of air travel. We show this as a downward or rightward shift in supply.

Step 4. A rightward shift in supply causes a movement down the demand curve, lowering the equilibrium price of air travel and increasing the equilibrium quantity.
- Step 1. Draw the graph with the initial supply and demand curves. Label the initial equilibrium price and quantity.

Step 2. Did the economic event affect supply or demand? A tariff is treated like a cost of production, so this affects supply.

Step 3. A tariff reduction is equivalent to a decrease in the cost of production, which we can show as a rightward (or downward) shift in supply.

Step 4. A rightward shift in supply causes a movement down the demand curve, lowering the equilibrium price and raising the equilibrium quantity.

4.4 PRICE CEILINGS AND PRICE FLOORS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain price controls, price ceilings, and price floors
- Analyze demand and supply as a social adjustment mechanism

Controversy sometimes surrounds the prices and quantities established by demand and supply, especially for products that are considered necessities. In some cases, discontent over prices turns into public pressure on politicians, who may then pass legislation to prevent a certain price from climbing “too high” or falling “too low.”

The demand and supply model shows how people and firms will react to the incentives provided by these laws to control prices, in ways that will often lead to undesirable consequences. Alternative policy tools can often achieve the desired goals of price control laws, while avoiding at least some of their costs and tradeoffs.

PRICE CEILINGS

Laws that government enacts to regulate prices are called **Price controls**. Price controls come in two flavors. A **price ceiling** keeps a price from rising above a certain level (the “ceiling”), while a **price floor** keeps a price from falling below a certain level (the “floor”). This section uses the demand and supply framework to analyze price ceilings. The next section discusses price floors.

In many markets for goods and services, demanders outnumber suppliers. Consumers, who are also potential voters, sometimes unite behind a political proposal to hold down a certain price. In some cities, such as Albany, renters have pressed political leaders to pass rent control laws, a price ceiling that usually works by stating that rents can be raised by only a certain maximum percentage each year.

Rent control becomes a politically hot topic when rents begin to rise rapidly. Everyone needs an affordable place to live. Perhaps a change in tastes makes a certain suburb or town a more popular place to live. Perhaps locally-based businesses expand, bringing higher incomes and more people into the area. Changes of this sort can cause a change in the demand for rental housing, as Figure 1 illustrates. The original equilibrium (E_0) lies at the intersection of supply curve S_0 and demand curve D_0 , corresponding to an equilibrium price of \$500 and an equilibrium quantity of 15,000 units of rental housing. The effect of greater income or a change in tastes is to shift the demand curve for rental

housing to the right, as shown by the data in Table 10 and the shift from D_0 to D_1 on the graph. In this market, at the new equilibrium E_1 , the price of a rental unit would rise to \$600 and the equilibrium quantity would increase to 17,000 units.

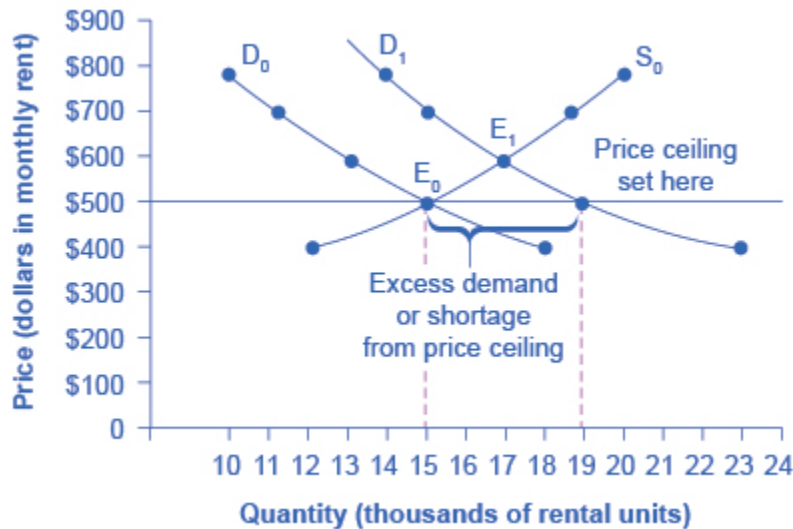


Figure 1. A Price Ceiling Example—Rent Control. The original intersection of demand and supply occurs at E_0 . If demand shifts from D_0 to D_1 , the new equilibrium would be at E_1 —unless a price ceiling prevents the price from rising. If the price is not permitted to rise, the quantity supplied remains at 15,000. However, after the change in demand, the quantity demanded rises to 19,000, resulting in a shortage.

Price	Original Quantity Supplied	Original Quantity Demanded	New Quantity Demanded
\$400	12,000	18,000	23,000
\$500	15,000	15,000	19,000
\$600	17,000	13,000	17,000
\$700	19,000	11,000	15,000
\$800	20,000	10,000	14,000

Table 10. Rent Control

Suppose that a rent control law is passed to keep the price at the original equilibrium of \$500 for a typical apartment. In Figure 1, the horizontal line at the price of \$500 shows the legally fixed maximum price set by the rent control law. However, the underlying forces that shifted the demand curve to the right are still there. At that price (\$500), the quantity supplied remains at the same 15,000 rental units, but the quantity demanded is 19,000 rental units. In other words, the quantity demanded exceeds the quantity supplied, so there is a shortage of rental housing. One of the ironies of price ceilings is that while the price ceiling was intended to help renters, there are actually fewer apartments rented out under the price ceiling (15,000 rental units) than would be the case at the market rent of \$600 (17,000 rental units).

Price ceilings do not simply benefit renters at the expense of landlords. Rather, some renters (or

potential renters) lose their housing as landlords convert apartments to co-ops and condos. Even when the housing remains in the rental market, landlords tend to spend less on maintenance and on essentials like heating, cooling, hot water, and lighting. The first rule of economics is you do not get something for nothing—everything has an opportunity cost. So if renters get “cheaper” housing than the market requires, they tend to also end up with lower quality housing.

Price ceilings have been proposed for other products. For example, price ceilings to limit what producers can charge have been proposed in recent years for prescription drugs, doctor and hospital fees, the charges made by some automatic teller bank machines, and auto insurance rates. Price ceilings are enacted in an attempt to keep prices low for those who demand the product. But when the market price is not allowed to rise to the equilibrium level, quantity demanded exceeds quantity supplied, and thus a shortage occurs. Those who manage to purchase the product at the lower price given by the price ceiling will benefit, but sellers of the product will suffer, along with those who are not able to purchase the product at all. Quality is also likely to deteriorate.

PRICE FLOORS

A price floor is the lowest legal price that can be paid in markets for goods and services, labor, or financial capital. Perhaps the best-known example of a price floor is the minimum wage, which is based on the normative view that someone working full time ought to be able to afford a basic standard of living. The federal minimum wage at the end of 2014 was \$7.25 per hour, which yields an income for a single person slightly higher than the poverty line. As the cost of living rises over time, the Congress periodically raises the federal minimum wage.

Price floors are sometimes called “price supports,” because they support a price by preventing it from falling below a certain level. Around the world, many countries have passed laws to create agricultural price supports. Farm prices and thus farm incomes fluctuate, sometimes widely. So even if, on average, farm incomes are adequate, some years they can be quite low. The purpose of price supports is to prevent these swings.

The most common way price supports work is that the government enters the market and buys up the product, adding to demand to keep prices higher than they otherwise would be. According to the Common Agricultural Policy reform passed in 2013, the **European Union (EU)** will spend about 60 billion euros per year, or 67 billion dollars per year, or roughly 38% of the EU budget, on price supports for Europe’s farmers from 2014 to 2020.

Figure 2 illustrates the effects of a government program that assures a price above the equilibrium by focusing on the market for wheat in Europe. In the absence of government intervention, the price would adjust so that the quantity supplied would equal the quantity demanded at the equilibrium point E_0 , with price P_0 and quantity Q_0 . However, policies to keep prices high for farmers keeps the price above what would have been the market equilibrium level—the price P_f shown by the dashed horizontal line in the diagram. The result is a quantity supplied in excess of the quantity demanded (Q_d). When quantity supplied exceeds quantity demanded, a surplus exists.

The high-income areas of the world, including the United States, Europe, and Japan, are estimated to spend roughly \$1 billion per day in supporting their farmers. If the government is willing to purchase the excess supply (or to provide payments for others to purchase it), then farmers will benefit from the price floor, but taxpayers and consumers of food will pay the costs. Numerous proposals have been

offered for reducing farm subsidies. In many countries, however, political support for subsidies for farmers remains strong. Either because this is viewed by the population as supporting the traditional rural way of life or because of the lobbying power of the agro-business industry.

For more detail on the effects price ceilings and floors have on demand and supply, see the following Clear It Up feature.

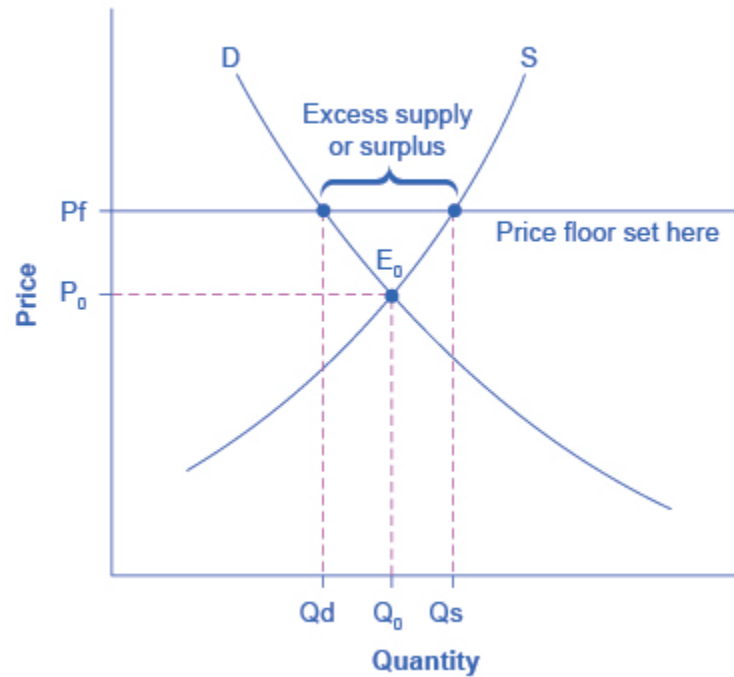


Figure 2. European Wheat Prices: A Price Floor Example. The intersection of demand (D) and supply (S) would be at the equilibrium point E_0 . However, a price floor set at P_f holds the price above E_0 and prevents it from falling. The result of the price floor is that the quantity supplied Q_s exceeds the quantity demanded Q_d . There is excess supply, also called a surplus.

DO PRICE CEILINGS AND FLOORS CHANGE DEMAND OR SUPPLY?

Neither price ceilings nor price floors cause demand or supply to change. They simply set a price that limits what can be legally charged in the market. Remember, changes in price do not cause demand or supply to change. Price ceilings and price floors can cause a different choice of quantity demanded along a demand curve, but they do not move the demand curve. Price controls can cause a different choice of quantity supplied along a supply curve, but they do not shift the supply curve.

KEY CONCEPTS AND SUMMARY

Price ceilings prevent a price from rising above a certain level. When a price ceiling is set below the equilibrium price, quantity demanded will exceed quantity supplied, and excess demand or shortages will result. Price floors prevent a price from falling below a certain level. When a price floor is set

above the equilibrium price, quantity supplied will exceed quantity demanded, and excess supply or surpluses will result. Price floors and price ceilings often lead to unintended consequences.

SELF-CHECK QUESTIONS

1. What is the effect of a price ceiling on the quantity demanded of the product? What is the effect of a price ceiling on the quantity supplied? Why exactly does a price ceiling cause a shortage?
2. Does a price ceiling change the equilibrium price?
3. What would be the impact of imposing a price floor below the equilibrium price?

REVIEW QUESTIONS

1. Does a price ceiling attempt to make a price higher or lower?
2. How does a price ceiling set below the equilibrium level affect quantity demanded and quantity supplied?
3. Does a price floor attempt to make a price higher or lower?
4. How does a price floor set above the equilibrium level affect quantity demanded and quantity supplied?

CRITICAL THINKING QUESTIONS

1. Most government policy decisions have winners and losers. What are the effects of raising the minimum wage? It is more complex than simply producers lose and workers gain. Who are the winners and who are the losers, and what exactly do they win and lose? To what extent does the policy change achieve its goals?
2. Agricultural price supports result in governments holding large inventories of agricultural products. Why do you think the government cannot simply give the products away to poor people?
3. Can you propose a policy that would induce the market to supply more rental housing units?

PROBLEMS

A low-income country decides to set a price ceiling on bread so it can make sure that bread is affordable to the poor. The conditions of demand and supply are given in Table 11. What are the equilibrium price and equilibrium quantity before the price ceiling? What will the excess demand or the shortage (that is, quantity demanded minus quantity supplied) be if the price ceiling is set at \$2.40? At \$2.00? At \$3.60?

Price	Qd	Qs
\$1.60	9,000	5,000
\$2.00	8,500	5,500
\$2.40	8,000	6,400
\$2.80	7,500	7,500
\$3.20	7,000	9,000
\$3.60	6,500	11,000
\$4.00	6,000	15,000

Table 11. Demand and Supply of Bread in Low-Income Country

GLOSSARY

price ceiling a legal maximum price

price control government laws to regulate prices instead of letting market forces determine prices

price floor a legal minimum price

total surplus see social surplus

SOLUTIONS

Answers to Self-Check Questions

1. A price ceiling (which is below the equilibrium price) will cause the quantity demanded to rise and the quantity supplied to fall. This is why a price ceiling creates a shortage.
2. A price ceiling is just a legal restriction. Equilibrium is an economic condition. People may or may not obey the price ceiling, so the actual price may be at or above the price ceiling, but the price ceiling does not change the equilibrium price.
3. A price ceiling is a legal maximum price, but a price floor is a legal minimum price and, consequently, it would leave room for the price to rise to its equilibrium level. In other words, a price floor below equilibrium will not be binding and will have no effect.

4.5 DEMAND, SUPPLY, AND EFFICIENCY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Contrast consumer surplus, producer surplus, and social surplus
- Explain why price floors and price ceilings can be inefficient
- Analyze demand and supply as a social adjustment mechanism

The familiar **demand and supply diagram** holds within it the concept of economic efficiency. One typical way that economists define **efficiency** is when it is impossible to improve the situation of one party without imposing a cost on another. Conversely, if a situation is inefficient, it becomes possible to benefit at least one party without imposing costs on others.

Efficiency in the demand and supply model has the same basic meaning: The economy is getting as much benefit as possible from its scarce resources and all the possible gains from trade have been achieved. In other words, the optimal amount of each good and service is being produced and consumed.

CONSUMER SURPLUS, PRODUCER SURPLUS, SOCIAL SURPLUS

Consider a market for tablet computers, as shown in Figure 1. The equilibrium price is \$80 and the equilibrium quantity is 28 million. To see the benefits to consumers, look at the segment of the demand curve above the **equilibrium** point and to the left. This portion of the demand curve shows that at least some demanders would have been willing to pay more than \$80 for a tablet.

For example, point J shows that if the price was \$90, 20 million tablets would be sold. Those consumers who would have been willing to pay \$90 for a tablet based on the utility they expect to receive from it, but who were able to pay the equilibrium price of \$80, clearly received a benefit beyond what they had to pay for. Remember, the demand curve traces consumers' willingness to pay for different quantities. The amount that individuals would have been willing to pay, minus the amount that they actually paid, is called **consumer surplus**. Consumer surplus is the area labeled F—that is, the area above the market price and below the demand curve.

The supply curve shows the quantity that firms are willing to supply at each price. For example, point K in Figure 1 illustrates that, at \$45, firms would still have been willing to supply a quantity of 14 million. Those producers who would have been willing to supply the tablets at \$45, but who were

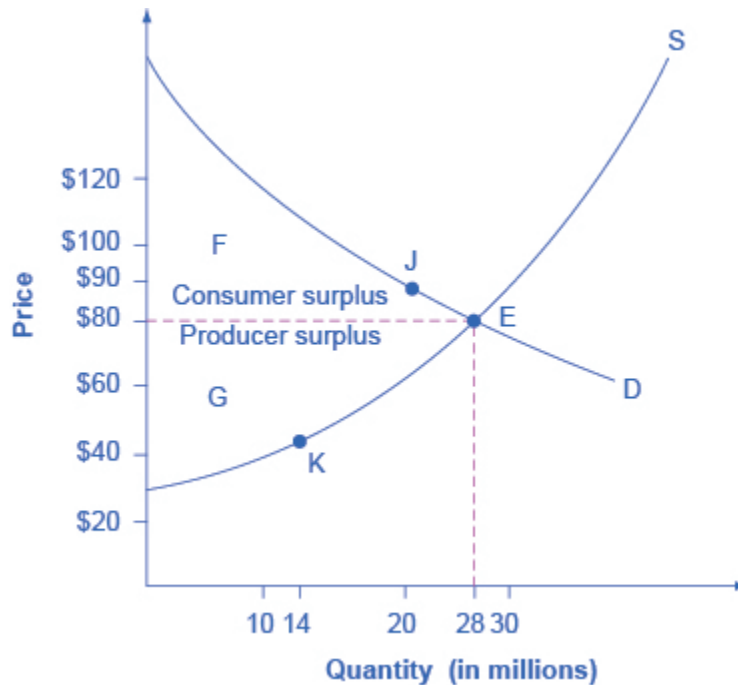


Figure 1. Consumer and Producer Surplus. The somewhat triangular area labeled by F shows the area of consumer surplus, which shows that the equilibrium price in the market was less than what many of the consumers were willing to pay. Point J on the demand curve shows that, even at the price of \$90, consumers would have been willing to purchase a quantity of 20 million. The somewhat triangular area labeled by G shows the area of producer surplus, which shows that the equilibrium price received in the market was more than what many of the producers were willing to accept for their products. For example, point K on the supply curve shows that at a price of \$45, firms would have been willing to supply a quantity of 14 million.

instead able to charge the equilibrium price of \$80, clearly received an extra benefit beyond what they required to supply the product. The amount that a seller is paid for a good minus the seller's actual cost is called **producer surplus**. In Figure 1, producer surplus is the area labeled G—that is, the area between the market price and the segment of the supply curve below the equilibrium.

The sum of consumer surplus and producer surplus is **social surplus**, also referred to as **economic surplus** or **total surplus**. In Figure 1, social surplus would be shown as the area F + G. Social surplus is larger at equilibrium quantity and price than it would be at any other quantity. This demonstrates the economic efficiency of the market equilibrium. In addition, at the efficient level of output, it is impossible to produce greater consumer surplus without reducing producer surplus, and it is impossible to produce greater producer surplus without reducing consumer surplus.

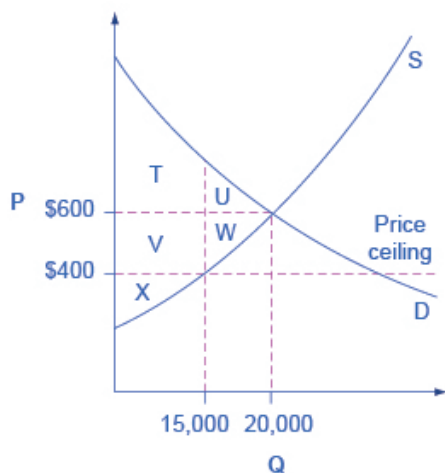
INEFFICIENCY OF PRICE FLOORS AND PRICE CEILINGS

The imposition of a price floor or a price ceiling will prevent a market from adjusting to its equilibrium price and quantity, and thus will create an inefficient outcome. But there is an additional twist here. Along with creating inefficiency, price floors and ceilings will also transfer some consumer surplus to producers, or some producer surplus to consumers.

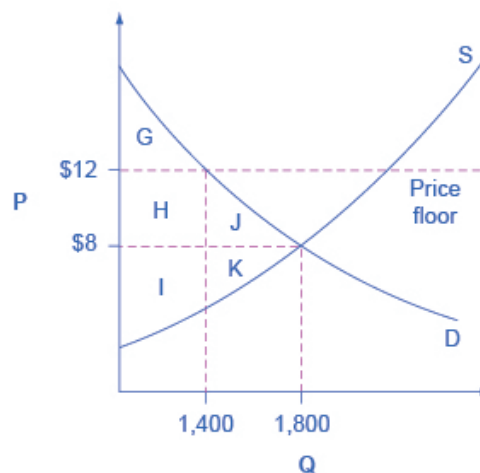
Imagine that several firms develop a promising but expensive new drug for treating back pain. If this therapy is left to the market, the equilibrium price will be \$600 per month and 20,000 people will use the drug, as shown in Figure 2 (a). The original level of consumer surplus is $T + U$ and producer surplus is $V + W + X$. However, the government decides to impose a price ceiling of \$400 to make the drug more affordable. At this price ceiling, firms in the market now produce only 15,000.

As a result, two changes occur. First, an inefficient outcome occurs and the total surplus of society is reduced. The loss in social surplus that occurs when the economy produces at an inefficient quantity is called **deadweight loss**. In a very real sense, it is like money thrown away that benefits no one. In Figure 2 (a), the deadweight loss is the area $U + W$. When deadweight loss exists, it is possible for both consumer and producer surplus to be higher, in this case because the **price control** is blocking some suppliers and demanders from transactions they would both be willing to make.

A second change from the **price ceiling** is that some of the producer surplus is transferred to consumers. After the price ceiling is imposed, the new consumer surplus is $T + V$, while the new producer surplus is X . In other words, the price ceiling transfers the area of surplus (V) from producers to consumers. Note that the gain to consumers is less than the loss to producers, which is just another way of seeing the deadweight loss.



(a) Reduced social surplus from a price ceiling



(b) Reduced social surplus from a price floor

Figure 2. (a) Efficiency and Price Floors and Ceilings. The original equilibrium price is \$600 with a quantity of 20,000. Consumer surplus is $T + U$, and producer surplus is $V + W + X$. A price ceiling is imposed at \$400, so firms in the market now produce only a quantity of 15,000. As a result, the new consumer surplus is $T + V$, while the new producer surplus is X . (b) The original equilibrium is \$8 at a quantity of 1,800. Consumer surplus is $G + H + J$, and producer surplus is $I + K$. A price floor is imposed at \$12, which means that quantity demanded falls to 1,400. As a result, the new consumer surplus is G , and the new producer surplus is $H + I$.

Figure 2 (b) shows a price floor example using a string of struggling movie theaters, all in the same city. The current equilibrium is \$8 per movie ticket, with 1,800 people attending movies. The original consumer surplus is $G + H + J$, and producer surplus is $I + K$. The city government is worried that movie theaters will go out of business, reducing the entertainment options available to citizens, so it decides to impose a price floor of \$12 per ticket. As a result, the quantity demanded of movie tickets falls to 1,400. The new consumer surplus is G , and the new producer surplus is $H + I$. In effect, the **price floor** causes the area H to be transferred from consumer to producer surplus, but also causes a deadweight loss of $J + K$.

This analysis shows that a price ceiling, like a law establishing rent controls, will transfer some producer surplus to consumers—which helps to explain why consumers often favor them. Conversely, a price floor like a guarantee that farmers will receive a certain price for their crops will transfer some consumer surplus to producers, which explains why producers often favor them. However, both price floors and price ceilings block some transactions that buyers and sellers would have been willing to make, and creates deadweight loss. Removing such barriers, so that prices and quantities can adjust to their equilibrium level, will increase the economy's social surplus.

DEMAND AND SUPPLY AS A SOCIAL ADJUSTMENT MECHANISM

The demand and supply model emphasizes that prices are not set only by demand or only by supply, but by the interaction between the two. In 1890, the famous economist **Alfred Marshall** wrote that asking whether supply or demand determined a price was like arguing “whether it is the upper or the under blade of a pair of scissors that cuts a piece of paper.” The answer is that both blades of the demand and supply scissors are always involved.

The adjustments of equilibrium price and quantity in a market-oriented economy often occur without much government direction or oversight. If the coffee crop in Brazil suffers a terrible frost, then the supply curve of coffee shifts to the left and the price of coffee rises. Some people—call them the coffee addicts—continue to drink coffee and pay the higher price. Others switch to tea or soft drinks. No government commission is needed to figure out how to adjust coffee prices, which companies will be allowed to process the remaining supply, which supermarkets in which cities will get how much coffee to sell, or which consumers will ultimately be allowed to drink the brew. Such adjustments in response to price changes happen all the time in a market economy, often so smoothly and rapidly that we barely notice them.

Think for a moment of all the seasonal foods that are available and inexpensive at certain times of the year, like fresh corn in midsummer, but more expensive at other times of the year. People alter their diets and restaurants alter their menus in response to these fluctuations in prices without fuss or fanfare. For both the U.S. economy and the world economy as a whole, markets—that is, demand and supply—are the primary social mechanism for answering the basic questions about what is produced, how it is produced, and for whom it is produced.

WHY CAN WE NOT GET ENOUGH OF ORGANIC?

Organic food is grown without synthetic pesticides, chemical fertilizers or genetically modified seeds. In recent decades, the demand for organic products has increased dramatically. The Organic Trade Association reported sales increased from \$1 billion in 1990 to \$35.1 billion in 2013, more than 90% of which were sales of food products.

Why, then, are organic foods more expensive than their conventional counterparts? The answer is a clear application of the theories of supply and demand. As people have learned more about the harmful effects of chemical fertilizers, growth hormones, pesticides and the like from large-scale factory farming, our tastes and preferences for safer, organic foods have increased. This change in tastes has been reinforced by increases in income, which allow people to purchase pricier products, and has made organic foods more mainstream. This has led to an increased demand for organic foods. Graphically, the demand curve has shifted right, and we have moved up the supply curve as producers have responded to the higher prices by supplying a greater quantity.

In addition to the movement along the supply curve, we have also had an increase in the number of farmers converting to organic farming over time. This is represented by a shift to the right of the supply curve. Since both demand and supply have shifted to the right, the resulting equilibrium quantity of organic foods is definitely higher, but the price will only fall when the increase in supply is larger than the increase in demand. We may need more time before we see lower prices in organic foods. Since the production costs of these foods may remain higher than conventional farming, because organic fertilizers and pest management techniques are more expensive, they may never fully catch up with the lower prices of non-organic foods.

As a final, specific example: The Environmental Working Group's "Dirty Dozen" list of fruits and vegetables, which test high for pesticide residue even after washing, was released in April 2013. The inclusion of strawberries on the list has led to an increase in demand for organic strawberries, resulting in both a higher equilibrium price and quantity of sales.

KEY CONCEPTS AND SUMMARY

Consumer surplus is the gap between the price that consumers are willing to pay, based on their preferences, and the market equilibrium price. Producer surplus is the gap between the price for which producers are willing to sell a product, based on their costs, and the market equilibrium price. Social surplus is the sum of consumer surplus and producer surplus. Total surplus is larger at the equilibrium quantity and price than it will be at any other quantity and price. Deadweight loss is loss in total surplus that occurs when the economy produces at an inefficient quantity.

SELF-CHECK QUESTIONS

1. Does a price ceiling increase or decrease the number of transactions in a market? Why? What about a price floor?
2. If a price floor benefits producers, why does a price floor reduce social surplus?

REVIEW QUESTIONS

1. What is consumer surplus? How is it illustrated on a demand and supply diagram?
2. What is producer surplus? How is it illustrated on a demand and supply diagram?
3. What is total surplus? How is it illustrated on a demand and supply diagram?
4. What is the relationship between total surplus and economic efficiency?
5. What is deadweight loss?

CRITICAL THINKING QUESTIONS

1. What term would an economist use to describe what happens when a shopper gets a “good deal” on a product?
2. Explain why voluntary transactions improve social welfare.
3. Why would a free market never operate at a quantity greater than the equilibrium quantity? *Hint: What would be required for a transaction to occur at that quantity?*

GLOSSARY

consumer surplus the extra benefit consumers receive from buying a good or service, measured by what the individuals would have been willing to pay minus the amount that they actually paid

deadweight loss the loss in social surplus that occurs when a market produces an inefficient quantity

economic surplus see social surplus

producer surplus the extra benefit producers receive from selling a good or service, measured by the price the producer actually received minus the price the producer would have been willing to accept

social surplus the sum of consumer surplus and producer surplus

SOLUTIONS

Answers to Self-Check Questions

1. Assuming that people obey the price ceiling, the market price will be below equilibrium, which means that Q_d will be more than Q_s . Buyers can only buy what is offered for sale, so the number of transactions will fall to Q_s . This is easy to see graphically. By analogous reasoning, with a price floor the market price will be above the equilibrium price, so Q_d will be less than Q_s . Since the limit on transactions here is demand, the number of transactions will fall to Q_d . Note that because both price floors and price ceilings reduce the number of transactions, social surplus is less.
2. Because the losses to consumers are greater than the benefits to producers, so the net effect is negative. Since the lost consumer surplus is greater than the additional producer surplus, social surplus falls.

CHAPTER 5. LABOR AND FINANCIAL MARKETS

INTRODUCTION TO LABOR AND FINANCIAL MARKETS



Figure 1. People often think of demand and supply in relation to goods, but labor markets, such as the nursing profession, can also apply to this analysis. (Credit: modification of work by “Fotos GOVBA”/Flickr Creative Commons)

BABY BOOMERS COME OF AGE

The Census Bureau reports that as of 2013, 20% of the U.S. population was over 60 years old, which means that almost 63 million people are reaching an age when they will need increased medical care.

The baby boomer population, the group born between 1946 and 1964, is comprised of approximately 74 million people who have just reached retirement age. As this population grows older, they will be faced with common healthcare issues such as heart conditions, arthritis, and Alzheimer’s that may require hospitalization, long-term, or at-home nursing care. Aging baby boomers and advances in life-saving and life-extending technologies will increase the demand for healthcare and nursing. Additionally, the Affordable Care Act, which expands access to healthcare for millions of Americans, will further increase the demand.

According to the Bureau of Labor Statistics, registered nursing jobs are expected to increase by 19% between 2012 and

2022. The median annual wage of \$67,930 (in 2012) is also expected to increase. The BLS forecasts that 526,000 new nurses will be needed by 2022. One concern is the low rate of enrollment in nursing programs to help meet the growing demand. According to the American Association of Colleges of Nursing (AACN), enrollment in 2011 increased by only 5.1% due to a shortage of nursing educators and teaching facilities.

These data tell us, as economists, that the market for healthcare professionals, and nurses in particular, will face several challenges. Our study of supply and demand will help us to analyze what might happen in the labor market for nursing and other healthcare professionals, as discussed in the second half of this case at the end of the chapter.

CHAPTER OBJECTIVES

Introduction to Labor and Financial Markets

In this chapter, you will learn about:

- Demand and Supply at Work in Labor Markets
- Demand and Supply in Financial Markets
- The Market System as an Efficient Mechanism for Information

The theories of supply and demand do not apply just to markets for goods. They apply to any market, even markets for labor and financial services. Labor markets are markets for employees or jobs. Financial services markets are markets for saving or borrowing.

When we think about demand and supply curves in goods and services markets, it is easy to picture who the demanders and suppliers are: businesses produce the products and households buy them. Who are the demanders and suppliers in labor and financial service markets? In labor markets job seekers (individuals) are the suppliers of labor, while firms and other employers who hire labor are the demanders for labor. In financial markets, any individual or firm who saves contributes to the supply of money, and any who borrows (person, firm, or government) contributes to the demand for money.

As a college student, you most likely participate in both labor and financial markets. Employment is a fact of life for most college students: In 2011, says the BLS, 52% of undergraduates worked part time and another 20% worked full time. Most college students are also heavily involved in financial markets, primarily as borrowers. Among full-time students, about half take out a loan to help finance their education each year, and those loans average about \$6,000 per year. Many students also borrow for other expenses, like purchasing a car. As this chapter will illustrate, we can analyze labor markets and financial markets with the same tools we use to analyze demand and supply in the goods markets.

5.1 DEMAND AND SUPPLY AT WORK IN LABOR MARKETS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Predict shifts in the demand and supply curves of the labor market
- Explain the impact of new technology on the demand and supply curves of the labor market
- Explain price floors in the labor market such as minimum wage or a living wage

Markets for labor have demand and supply curves, just like markets for goods. The law of demand applies in labor markets this way: A higher **salary** or **wage**—that is, a higher price in the labor market—leads to a decrease in the quantity of labor demanded by employers, while a lower salary or wage leads to an increase in the quantity of labor demanded. The law of supply functions in labor markets, too: A higher price for labor leads to a higher quantity of labor supplied; a lower price leads to a lower quantity supplied.

EQUILIBRIUM IN THE LABOR MARKET

In 2013, about 34,000 registered nurses worked in the Minneapolis-St. Paul-Bloomington, Minnesota-Wisconsin metropolitan area, according to the BLS. They worked for a variety of employers: hospitals, doctors' offices, schools, health clinics, and nursing homes. Figure 1 illustrates how demand and supply determine equilibrium in this labor market. The demand and supply schedules in Table 1 list the quantity supplied and quantity demanded of nurses at different salaries.

Annual Salary	Quantity Demanded	Quantity Supplied
\$55,000	45,000	20,000
\$60,000	40,000	27,000
\$65,000	37,000	31,000
\$70,000	34,000	34,000
\$75,000	33,000	38,000
\$80,000	32,000	41,000

Table 1. Demand and Supply of Nurses in Minneapolis-St. Paul-Bloomington

The horizontal axis shows the quantity of nurses hired. In this example, labor is measured by number of workers, but another common way to measure the quantity of labor is by the number of hours

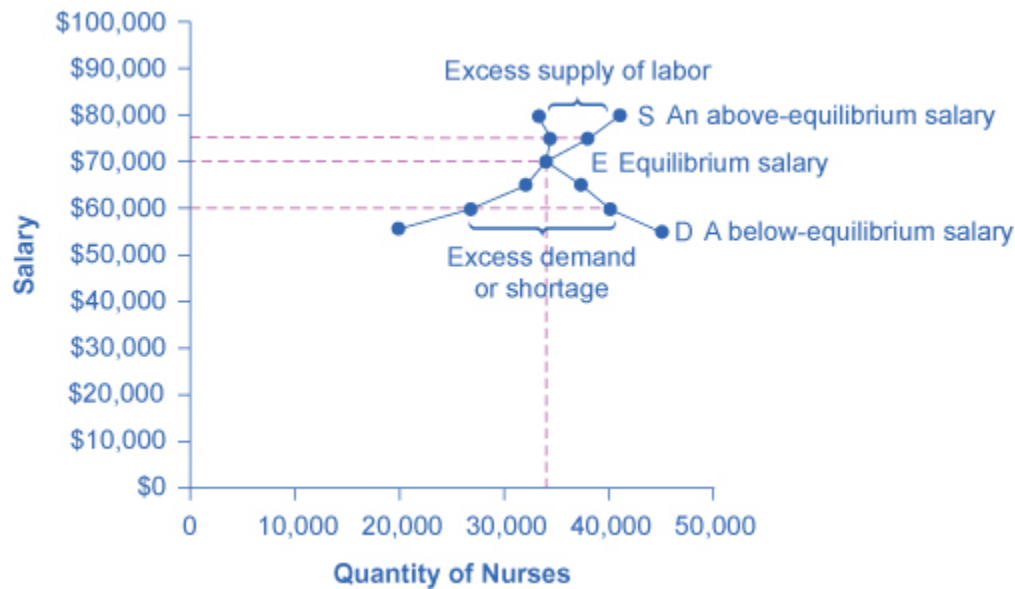


Figure 1. Labor Market Example: Demand and Supply for Nurses in Minneapolis-St. Paul-Bloomington. The demand curve (D) of those employers who want to hire nurses intersects with the supply curve (S) of those who are qualified and willing to work as nurses at the equilibrium point (E). The equilibrium salary is \$70,000 and the equilibrium quantity is 34,000 nurses. At an above-equilibrium salary of \$75,000, quantity supplied increases to 38,000, but the quantity of nurses demanded at the higher pay declines to 33,000. At this above-equilibrium salary, an excess supply or surplus of nurses would exist. At a below-equilibrium salary of \$60,000, quantity supplied declines to 27,000, while the quantity demanded at the lower wage increases to 40,000 nurses. At this below-equilibrium salary, excess demand or a surplus exists.

worked. The vertical axis shows the price for nurses' labor—that is, how much they are paid. In the real world, this "price" would be total labor compensation: salary plus benefits. It is not obvious, but benefits are a significant part (as high as 30 percent) of labor compensation. In this example, the price of labor is measured by salary on an annual basis, although in other cases the price of labor could be measured by monthly or weekly pay, or even the wage paid per hour. As the salary for nurses rises, the quantity demanded will fall. Some hospitals and nursing homes may cut back on the number of nurses they hire, or they may lay off some of their existing nurses, rather than pay them higher salaries. Employers who face higher nurses' salaries may also try to replace some nursing functions by investing in physical equipment, like computer monitoring and diagnostic systems to monitor patients, or by using lower-paid health care aides to reduce the number of nurses they need.

As the salary for nurses rises, the quantity supplied will rise. If nurses' salaries in Minneapolis-St. Paul-Bloomington are higher than in other cities, more nurses will move to Minneapolis-St. Paul-Bloomington to find jobs, more people will be willing to train as nurses, and those currently trained as nurses will be more likely to pursue nursing as a full-time job. In other words, there will be more nurses looking for jobs in the area.

At **equilibrium**, the quantity supplied and the quantity demanded are equal. Thus, every employer who wants to hire a nurse at this equilibrium wage can find a willing worker, and every nurse who

wants to work at this equilibrium salary can find a job. In Figure 1, the supply curve (S) and demand curve (D) intersect at the equilibrium point (E). The equilibrium quantity of nurses in the Minneapolis-St. Paul-Bloomington area is 34,000, and the equilibrium salary is \$70,000 per year. This example simplifies the nursing market by focusing on the “average” nurse. In reality, of course, the market for nurses is actually made up of many smaller markets, like markets for nurses with varying degrees of experience and credentials. Many markets contain closely related products that differ in quality; for instance, even a simple product like gasoline comes in regular, premium, and super-premium, each with a different price. Even in such cases, discussing the average price of gasoline, like the average salary for nurses, can still be useful because it reflects what is happening in most of the submarkets.

When the price of labor is not at the equilibrium, economic incentives tend to move salaries toward the equilibrium. For example, if salaries for nurses in Minneapolis-St. Paul-Bloomington were above the equilibrium at \$75,000 per year, then 38,000 people want to work as nurses, but employers want to hire only 33,000 nurses. At that above-equilibrium salary, excess supply or a surplus results. In a situation of excess supply in the **labor market**, with many applicants for every job opening, employers will have an incentive to offer lower wages than they otherwise would have. Nurses’ salary will move down toward equilibrium.

In contrast, if the salary is below the equilibrium at, say, \$60,000 per year, then a situation of excess demand or a shortage arises. In this case, employers encouraged by the relatively lower wage want to hire 40,000 nurses, but only 27,000 individuals want to work as nurses at that salary in Minneapolis-St. Paul-Bloomington. In response to the shortage, some employers will offer higher pay to attract the nurses. Other employers will have to match the higher pay to keep their own employees. The higher salaries will encourage more nurses to train or work in Minneapolis-St. Paul-Bloomington. Again, price and quantity in the labor market will move toward equilibrium.

SHIFTS IN LABOR DEMAND

The demand curve for labor shows the quantity of labor employers wish to hire at any given salary or wage rate, under the *ceteris paribus* assumption. A change in the wage or salary will result in a change in the quantity demanded of labor. If the wage rate increases, employers will want to hire fewer employees. The quantity of labor demanded will decrease, and there will be a movement upward along the demand curve. If the wages and salaries decrease, employers are more likely to hire a greater number of workers. The quantity of labor demanded will increase, resulting in a downward movement along the demand curve.

Shifts in the demand curve for labor occur for many reasons. One key reason is that the demand for labor is based on the demand for the good or service that is being produced. For example, the more new automobiles consumers demand, the greater the number of workers automakers will need to hire. Therefore the demand for labor is called a “derived demand.” Here are some examples of derived demand for labor:

- The demand for chefs is dependent on the demand for restaurant meals.
- The demand for pharmacists is dependent on the demand for prescription drugs.
- The demand for attorneys is dependent on the demand for legal services.

As the demand for the goods and services increases, the demand for labor will increase, or shift to

the right, to meet employers' production requirements. As the demand for the goods and services decreases, the demand for labor will decrease, or shift to the left. Table 2 shows that in addition to the derived demand for labor, demand can also increase or decrease (shift) in response to several factors.

Factors	Results
Demand for Output	When the demand for the good produced (output) increases, both the output price and profitability increase. As a result, producers demand more labor to ramp up production.
Education and Training	A well-trained and educated workforce causes an increase in the demand for that labor by employers. Increased levels of productivity within the workforce will cause the demand for labor to shift to the right. If the workforce is not well-trained or educated, employers will not hire from within that labor pool, since they will need to spend a significant amount of time and money training that workforce. Demand for such will shift to the left.
Technology	Technology changes can act as either substitutes for or complements to labor. When technology acts as a substitute, it replaces the need for the number of workers an employer needs to hire. For example, word processing decreased the number of typists needed in the workplace. This shifted the demand curve for typists left. An increase in the availability of certain technologies may increase the demand for labor. Technology that acts as a complement to labor will increase the demand for certain types of labor, resulting in a rightward shift of the demand curve. For example, the increased use of word processing and other software has increased the demand for information technology professionals who can resolve software and hardware issues related to a firm's network. More and better technology will increase demand for skilled workers who know how to use technology to enhance workplace productivity. Those workers who do not adapt to changes in technology will experience a decrease in demand.
Number of Companies	An increase in the number of companies producing a given product will increase the demand for labor resulting in a shift to the right. A decrease in the number of companies producing a given product will decrease the demand for labor resulting in a shift to the left.
Government Regulations	Complying with government regulations can increase or decrease the demand for labor at any given wage. In the healthcare industry, government rules may require that nurses be hired to carry out certain medical procedures. This will increase the demand for nurses. Less-trained healthcare workers would be prohibited from carrying out these procedures, and the demand for these workers will shift to the left.
Price and Availability of Other Inputs	Labor is not the only input into the production process. For example, a salesperson at a call center needs a telephone and a computer terminal to enter data and record sales. The demand for salespersons at the call center will increase if the number of telephones and computer terminals available increases. This will cause a rightward shift of the demand curve. As the amount of inputs increases, the demand for labor will increase. If the terminal or the telephones malfunction, then the demand for that labor force will decrease. As the quantity of other inputs decreases, the demand for labor will decrease. Similarly, if prices of other inputs fall, production will become more profitable and suppliers will demand more labor to increase production. The opposite is also true. Higher input prices lower demand for labor

Table 2. Factors That Can Shift Demand

Click here to read more about "Trends and Challenges for Work in the 21st Century."



SHIFTS IN LABOR SUPPLY

The supply of labor is upward-sloping and adheres to the law of supply: The higher the price, the

greater the quantity supplied and the lower the price, the less quantity supplied. The supply curve models the tradeoff between supplying labor into the market or using time in leisure activities at every given price level. The higher the wage, the more labor is willing to work and forego leisure activities. Table 3 lists some of the factors that will cause the supply to increase or decrease.

Factors	Results
Number of Workers	An increased number of workers will cause the supply curve to shift to the right. An increased number of workers can be due to several factors, such as immigration, increasing population, an aging population, and changing demographics. Policies that encourage immigration will increase the supply of labor, and vice versa. Population grows when birth rates exceed death rates; this eventually increases supply of labor when the former reach working age. An aging and therefore retiring population will decrease the supply of labor. Another example of changing demographics is more women working outside of the home, which increases the supply of labor.
Required Education	The more required education, the lower the supply. There is a lower supply of PhD mathematicians than of high school mathematics teachers; there is a lower supply of cardiologists than of primary care physicians; and there is a lower supply of physicians than of nurses.
Government Policies	Government policies can also affect the supply of labor for jobs. On the one hand, the government may support rules that set high qualifications for certain jobs: academic training, certificates or licenses, or experience. When these qualifications are made tougher, the number of qualified workers will decrease at any given wage. On the other hand, the government may also subsidize training or even reduce the required level of qualifications. For example, government might offer subsidies for nursing schools or nursing students. Such provisions would shift the supply curve of nurses to the right. In addition, government policies that change the relative desirability of working versus not working also affect the labor supply. These include unemployment benefits, maternity leave, child care benefits and welfare policy. For example, child care benefits may increase the labor supply of working mothers. Long term unemployment benefits may discourage job searching for unemployed workers. All these policies must therefore be carefully designed to minimize any negative labor supply effects.

Table 3. Factors that Can Shift Supply

A change in salary will lead to a movement along labor demand or labor supply curves, but it will not shift those curves. However, other events like those outlined here will cause either the demand or the supply of labor to shift, and thus will move the labor market to a new equilibrium salary and quantity.

TECHNOLOGY AND WAGE INEQUALITY: THE FOUR-STEP PROCESS

Economic events can change the equilibrium salary (or wage) and quantity of labor. Consider how the wave of new information technologies, like computer and telecommunications networks, has affected low-skill and high-skill workers in the U.S. economy. From the perspective of employers who demand labor, these new technologies are often a substitute for low-skill laborers like file clerks who used to keep file cabinets full of paper records of transactions. However, the same new technologies are a complement to high-skill workers like managers, who benefit from the technological advances by being able to monitor more information, communicate more easily, and juggle a wider array of responsibilities. So, how will the new technologies affect the wages of high-skill and low-skill workers? For this question, the four-step process of analyzing how shifts in supply or demand affect a market (introduced in Demand and Supply) works in this way:

Step 1. What did the markets for low-skill labor and high-skill labor look like before the arrival of the new technologies? In Figure 2 (a) and Figure 2 (b), S_0 is the original supply curve for labor and D_0 is the original demand curve for labor in each market. In each graph, the original point of equilibrium, E_0 , occurs at the price W_0 and the quantity Q_0 .

Step 2. Does the new technology affect the supply of labor from households or the demand for labor

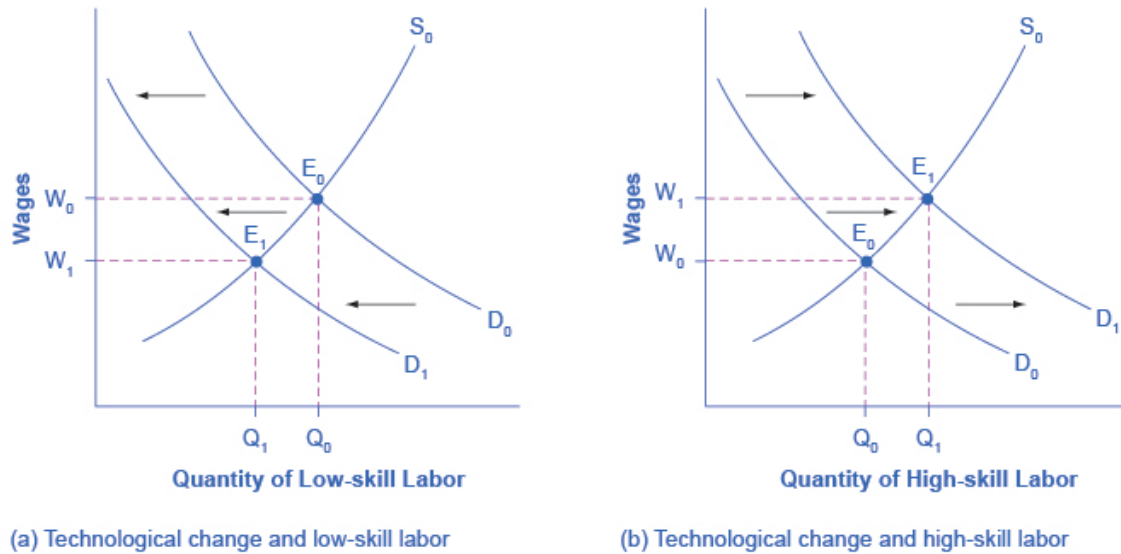


Figure 2. Technology and Wages: Applying Demand and Supply (a) The demand for low-skill labor shifts to the left when technology can do the job previously done by these workers. (b) New technologies can also increase the demand for high-skill labor in fields such as information technology and network administration.

from firms? The technology change described here affects demand for labor by firms that hire workers.

Step 3. Will the new technology increase or decrease demand? Based on the description earlier, as the substitute for low-skill labor becomes available, demand for low-skill labor will shift to the left, from D_0 to D_1 . As the technology complement for high-skill labor becomes cheaper, demand for high-skill labor will shift to the right, from D_0 to D_1 .

Step 4. The new equilibrium for low-skill labor, shown as point E_1 with price W_1 and quantity Q_1 , has a lower wage and quantity hired than the original equilibrium, E_0 . The new equilibrium for high-skill labor, shown as point E_1 with price W_1 and quantity Q_1 , has a higher wage and quantity hired than the original equilibrium (E_0).

So, the demand and supply model predicts that the new computer and communications technologies will raise the pay of high-skill workers but reduce the pay of low-skill workers. Indeed, from the 1970s to the mid-2000s, the wage gap widened between high-skill and low-skill labor. According to the National Center for Education Statistics, in 1980, for example, a college graduate earned about 30% more than a high school graduate with comparable job experience, but by 2012, a college graduate earned about 60% more than an otherwise comparable high school graduate. Many economists believe that the trend toward greater wage inequality across the U.S. economy was primarily caused by the new technologies.

Visit this website to read about ten tech skills that have lost relevance in today's workforce.



PRICE FLOORS IN THE LABOR MARKET: LIVING WAGES AND MINIMUM WAGES

In contrast to goods and services markets, price ceilings are rare in labor markets, because rules that prevent people from earning income are not politically popular. There is one exception: sometimes limits are proposed on the high incomes of top business executives.

The labor market, however, presents some prominent examples of price floors, which are often used as an attempt to increase the wages of low-paid workers. The U.S. government sets a **minimum wage**, a price floor that makes it illegal for an employer to pay employees less than a certain hourly rate. In mid-2009, the U.S. minimum wage was raised to \$7.25 per hour. Local political movements in a number of U.S. cities have pushed for a higher minimum wage, which they call a **living wage**. Promoters of living wage laws maintain that the minimum wage is too low to ensure a reasonable standard of living. They base this conclusion on the calculation that, if you work 40 hours a week at a minimum wage of \$7.25 per hour for 50 weeks a year, your annual income is \$14,500, which is less than the official U.S. government definition of what it means for a family to be in poverty. (A family with two adults earning minimum wage and two young children will find it more cost efficient for one parent to provide childcare while the other works for income. So the family income would be \$14,500, which is significantly lower than the federal poverty line for a family of four, which was \$23,850 in 2014.)

Supporters of the living wage argue that full-time workers should be assured a high enough wage so that they can afford the essentials of life: food, clothing, shelter, and healthcare. Since Baltimore passed the first living wage law in 1994, several dozen cities enacted similar laws in the late 1990s and the 2000s. The living wage ordinances do not apply to all employers, but they have specified that all employees of the city or employees of firms that are hired by the city be paid at least a certain wage that is usually a few dollars per hour above the U.S. minimum wage.

Figure 3 illustrates the situation of a city considering a living wage law. For simplicity, we assume that there is no federal minimum wage. The wage appears on the vertical axis, because the wage is the price in the labor market. Before the passage of the living wage law, the equilibrium wage is \$10 per hour and the city hires 1,200 workers at this wage. However, a group of concerned citizens persuades the city council to enact a living wage law requiring employers to pay no less than \$12 per hour. In response to the higher wage, 1,600 workers look for jobs with the city. At this higher wage, the city, as an employer, is willing to hire only 700 workers. At the price floor, the quantity supplied exceeds the quantity demanded, and a surplus of labor exists in this market. For workers who continue to have a job at a higher salary, life has improved. For those who were willing to work at the old wage rate but lost their jobs with the wage increase, life has not improved. Table 4 shows the differences in supply and demand at different wages.

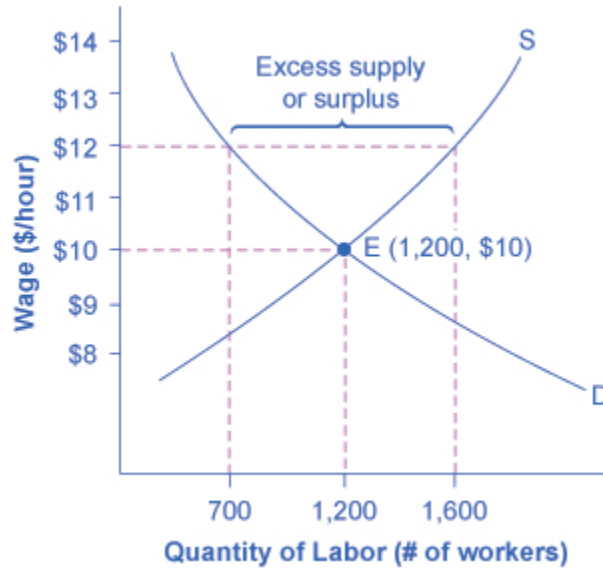


Figure 3. A Living Wage: Example of a Price Floor The original equilibrium in this labor market is a wage of \$10/hour and a quantity of 1,200 workers, shown at point E. Imposing a wage floor at \$12/hour leads to an excess supply of labor. At that wage, the quantity of labor supplied is 1,600 and the quantity of labor demanded is only 700.

Wage	Quantity Labor Demanded	Quantity Labor Supplied
\$8/hr	1,900	500
\$9/hr	1,500	900
\$10/hr	1,200	1,200
\$11/hr	900	1,400
\$12/hr	700	1,600
\$13/hr	500	1,800
\$14/hr	400	1,900

Table 4. Living Wage: Example of a Price Floor

THE MINIMUM WAGE AS AN EXAMPLE OF A PRICE FLOOR

The U.S. minimum wage is a price floor that is set either very close to the equilibrium wage or even slightly below it. About 1% of American workers are actually paid the minimum wage. In other words, the vast majority of the U.S. labor force has its wages determined in the labor market, not as a result of the government price floor. But for workers with low skills and little experience, like those without a high school diploma or teenagers, the minimum wage is quite important. In many cities, the federal minimum wage is apparently below the market price for unskilled labor, because employers offer more than the minimum wage to checkout clerks and other low-skill workers without any government prodding.

Economists have attempted to estimate how much the minimum wage reduces the quantity

demand of low-skill labor. A typical result of such studies is that a 10% increase in the minimum wage would decrease the hiring of unskilled workers by 1 to 2%, which seems a relatively small reduction. In fact, some studies have even found no effect of a higher minimum wage on employment at certain times and places—although these studies are controversial.

Let's suppose that the minimum wage lies just slightly *below* the equilibrium wage level. Wages could fluctuate according to market forces above this price floor, but they would not be allowed to move beneath the floor. In this situation, the price floor minimum wage is said to be *nonbinding*—that is, the price floor is not determining the market outcome. Even if the minimum wage moves just a little higher, it will still have no effect on the quantity of employment in the economy, as long as it remains below the equilibrium wage. Even if the minimum wage is increased by enough so that it rises slightly above the equilibrium wage and becomes binding, there will be only a small excess supply gap between the quantity demanded and quantity supplied.

These insights help to explain why U.S. minimum wage laws have historically had only a small impact on employment. Since the minimum wage has typically been set close to the equilibrium wage for low-skill labor and sometimes even below it, it has not had a large effect in creating an excess supply of labor. However, if the minimum wage were increased dramatically—say, if it were doubled to match the living wages that some U.S. cities have considered—then its impact on reducing the quantity demanded of employment would be far greater. The following Clear It Up feature describes in greater detail some of the arguments for and against changes to minimum wage.

WHAT'S THE HARM IN RAISING THE MINIMUM WAGE?

Because of the law of demand, a higher required wage will reduce the amount of low-skill employment either in terms of employees or in terms of work hours. Although there is controversy over the numbers, let's say for the sake of the argument that a 10% rise in the minimum wage will reduce the employment of low-skill workers by 2%. Does this outcome mean that raising the minimum wage by 10% is bad public policy? Not necessarily.

If 98% of those receiving the minimum wage have a pay increase of 10%, but 2% of those receiving the minimum wage lose their jobs, are the gains for society as a whole greater than the losses? The answer is not clear, because job losses, even for a small group, may cause more pain than modest income gains for others. For one thing, we need to consider which minimum wage workers are losing their jobs. If the 2% of minimum wage workers who lose their jobs are struggling to support families, that is one thing. If those who lose their job are high school students picking up spending money over summer vacation, that is something else.

Another complexity is that many minimum wage workers do not work full-time for an entire year. Imagine a minimum wage worker who holds different part-time jobs for a few months at a time, with bouts of unemployment in between. The worker in this situation receives the 10% raise in the minimum wage when working, but also ends up working 2% fewer hours during the year because the higher minimum wage reduces how much employers want people to work. Overall, this worker's income would rise because the 10% pay raise would more than offset the 2% fewer hours worked.

Of course, these arguments do not prove that raising the minimum wage is necessarily a good idea either. There may well be other, better public policy options for helping low-wage workers. (The Poverty and Economic Inequality chapter discusses some possibilities.) The lesson from this maze of minimum wage arguments is that complex social problems rarely have simple answers. Even those who agree on how a proposed economic policy affects quantity demanded and quantity supplied may still disagree on whether the policy is a good idea.

KEY CONCEPTS AND SUMMARY

In the labor market, households are on the supply side of the market and firms are on the demand side. In the market for financial capital, households and firms can be on either side of the market: they are suppliers of financial capital when they save or make financial investments, and demanders of financial capital when they borrow or receive financial investments.

In the demand and supply analysis of labor markets, the price can be measured by the annual salary or hourly wage received. The quantity of labor can be measured in various ways, like number of workers or the number of hours worked.

Factors that can shift the demand curve for labor include: a change in the quantity demanded of the product that the labor produces; a change in the production process that uses more or less labor; and a change in government policy that affects the quantity of labor that firms wish to hire at a given wage. Demand can also increase or decrease (shift) in response to: workers' level of education and training, technology, the number of companies, and availability and price of other inputs.

The main factors that can shift the supply curve for labor are: how desirable a job appears to workers relative to the alternatives, government policy that either restricts or encourages the quantity of workers trained for the job, the number of workers in the economy, and required education.

SELF-CHECK QUESTIONS

1. In the labor market, what causes a movement along the demand curve? What causes a shift in the demand curve?
2. In the labor market, what causes a movement along the supply curve? What causes a shift in the supply curve?
3. Why is a living wage considered a price floor? Does imposing a living wage have the same outcome as a minimum wage?

REVIEW QUESTIONS

1. What is the "price" commonly called in the labor market?
2. Are households demanders or suppliers in the goods market? Are firms demanders or suppliers in the goods market? What about the labor market and the financial market?
3. Name some factors that can cause a shift in the demand curve in labor markets.
4. Name some factors that can cause a shift in the supply curve in labor markets.

CRITICAL THINKING QUESTIONS

1. Other than the demand for labor, what would be another example of a "derived demand?"

2. Suppose that a 5% increase in the minimum wage causes a 5% reduction in employment. How would this affect employers and how would it affect workers? In your opinion, would this be a good policy?
3. What assumption is made for a minimum wage to be a nonbinding price floor? What assumption is made for a living wage price floor to be binding?

PROBLEMS

1. Identify each of the following as involving either demand or supply. Draw a circular flow diagram and label the flows A through F. (Some choices can be on both sides of the goods market.)
 - a. Households in the labor market
 - b. Firms in the goods market
 - c. Firms in the financial market
 - d. Households in the goods market
 - e. Firms in the labor market
 - f. Households in the financial market
2. Predict how each of the following events will raise or lower the equilibrium wage and quantity of coal miners in West Virginia. In each case, sketch a demand and supply diagram to illustrate your answer.
 - a. The price of oil rises.
 - b. New coal-mining equipment is invented that is cheap and requires few workers to run.
 - c. Several major companies that do not mine coal open factories in West Virginia, offering a lot of well-paid jobs.
 - d. Government imposes costly new regulations to make coal-mining a safer job.

REFERENCES

American Community Survey. 2012. "School Enrollment and Work Status: 2011." Accessed April 13, 2015. <http://www.census.gov/prod/2013pubs/acsbr11-14.pdf>.

National Center for Educational Statistics. "Digest of Education Statistics." (2008 and 2010). Accessed December 11, 2013. nces.ed.gov.

GLOSSARY

minimum wage a price floor that makes it illegal for an employer to pay employees less than a certain hourly rate

SOLUTIONS

Answers to Self-Check Questions

1. Changes in the wage rate (the price of labor) cause a movement along the demand curve. A change in anything else that affects demand for labor (e.g., changes in output, changes in the production process that use more or less labor, government regulation) causes a shift in the demand curve.
2. Changes in the wage rate (the price of labor) cause a movement along the supply curve. A change in anything else that affects supply of labor (e.g., changes in how desirable the job is perceived to be, government policy to promote training in the field) causes a shift in the supply curve.
3. Since a living wage is a suggested minimum wage, it acts like a price floor (assuming, of course, that it is followed). If the living wage is binding, it will cause an excess supply of labor at that wage rate.

5.2 DEMAND AND SUPPLY IN FINANCIAL MARKETS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify the demanders and suppliers in a financial market.
- Explain how interest rates can affect supply and demand
- Analyze the economic effects of U.S. debt in terms of domestic financial markets
- Explain the role of price ceilings and usury laws in the U.S.

United States' households, institutions, and domestic businesses saved almost \$1.9 trillion in 2013. Where did that savings go and what was it used for? Some of the savings ended up in banks, which in turn loaned the money to individuals or businesses that wanted to borrow money. Some was invested in private companies or loaned to government agencies that wanted to borrow money to raise funds for purposes like building roads or mass transit. Some firms reinvested their savings in their own businesses.

In this section, we will determine how the demand and supply model links those who wish to supply **financial capital** (i.e., savings) with those who demand financial capital (i.e., borrowing). Those who save money (or make financial investments, which is the same thing), whether individuals or businesses, are on the supply side of the financial market. Those who borrow money are on the demand side of the financial market. For a more detailed treatment of the different kinds of financial investments like bank accounts, stocks and bonds, see the Financial Markets chapter.

WHO DEMANDS AND WHO SUPPLIES IN FINANCIAL MARKETS?

In any market, the price is what suppliers receive and what demanders pay. In financial markets, those who supply financial capital through saving expect to receive a rate of return, while those who demand financial capital by receiving funds expect to pay a rate of return. This rate of return can come in a variety of forms, depending on the type of investment.

The simplest example of a rate of return is the **interest rate**. For example, when you supply money into a savings account at a bank, you receive interest on your deposit. The interest paid to you as a percent of your deposits is the interest rate. Similarly, if you demand a loan to buy a car or a computer, you will need to pay interest on the money you borrow.

Let's consider the market for borrowing money with credit cards. In 2014, almost 200 million Amer-

icans were cardholders. Credit cards allow you to borrow money from the card's issuer, and pay back the borrowed amount plus interest, though most allow you a period of time in which you can repay the loan without paying interest. A typical credit card interest rate ranges from 12% to 18% per year. In 2014, Americans had about \$793 billion outstanding in credit card debts. About half of U.S. families with credit cards report that they almost always pay the full balance on time, but one-quarter of U.S. families with credit cards say that they “hardly ever” pay off the card in full. In fact, in 2014, 56% of consumers carried an unpaid balance in the last 12 months. Let's say that, on average, the annual interest rate for credit card borrowing is 15% per year. So, Americans pay tens of billions of dollars every year in interest on their credit cards—plus basic fees for the credit card or fees for late payments.

Figure 1 illustrates demand and supply in the financial market for credit cards. The horizontal axis of the financial market shows the quantity of money that is loaned or borrowed in this market. The vertical or price axis shows the rate of return, which in the case of credit card borrowing can be measured with an interest rate. Table 5 shows the quantity of financial capital that consumers demand at various interest rates and the quantity that credit card firms (often banks) are willing to supply.

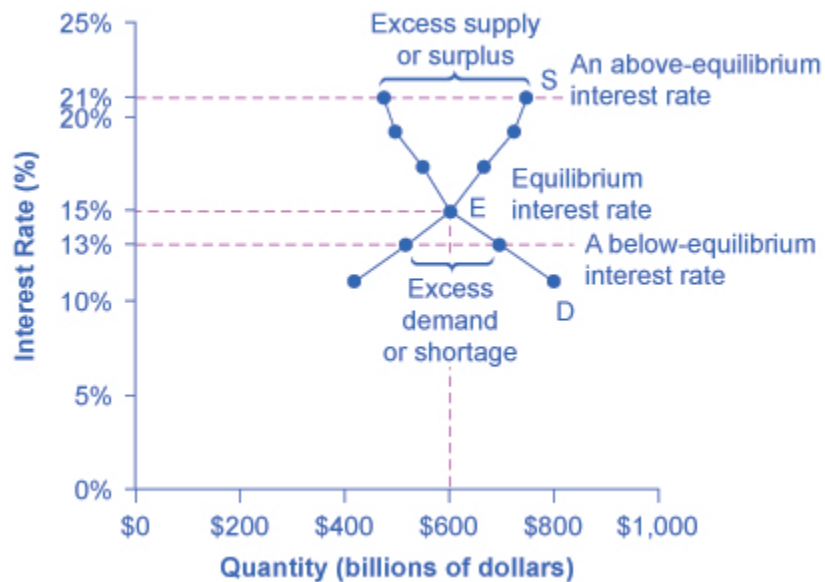


Figure 1. Demand and Supply for Borrowing Money with Credit Cards. In this market for credit card borrowing, the demand curve (D) for borrowing financial capital intersects the supply curve (S) for lending financial capital at equilibrium E. At the equilibrium, the interest rate (the “price” in this market) is 15% and the quantity of financial capital being loaned and borrowed is \$600 billion. The equilibrium price is where the quantity demanded and the quantity supplied are equal. At an above-equilibrium interest rate like 21%, the quantity of financial capital supplied would increase to \$750 billion, but the quantity demanded would decrease to \$480 billion. At a below-equilibrium interest rate like 13%, the quantity of financial capital demanded would increase to \$700 billion, but the quantity of financial capital supplied would decrease to \$510 billion.

Interest Rate (%)	Quantity of Financial Capital Demanded (Borrowing) (\$ billions)	Quantity of Financial Capital Supplied (Lending) (\$ billions)
11	\$800	\$420
13	\$700	\$510
15	\$600	\$600
17	\$550	\$660
19	\$500	\$720
21	\$480	\$750

Table 5. Demand and Supply for Borrowing Money with Credit Cards

The laws of demand and supply continue to apply in the financial markets. According to the **law of demand**, a higher rate of return (that is, a higher price) will decrease the quantity demanded. As the interest rate rises, consumers will reduce the quantity that they borrow. According to the law of supply, a higher price increases the quantity supplied. Consequently, as the interest rate paid on credit card borrowing rises, more firms will be eager to issue credit cards and to encourage customers to use them. Conversely, if the interest rate on credit cards falls, the quantity of financial capital supplied in the credit card market will decrease and the quantity demanded will fall.

EQUILIBRIUM IN FINANCIAL MARKETS

In the financial market for credit cards shown in Figure 1, the supply curve (S) and the demand curve (D) cross at the equilibrium point (E). The equilibrium occurs at an interest rate of 15%, where the quantity of funds demanded and the quantity supplied are equal at an equilibrium quantity of \$600 billion.

If the interest rate (remember, this measures the “price” in the financial market) is above the equilibrium level, then an excess supply, or a surplus, of financial capital will arise in this market. For example, at an interest rate of 21%, the quantity of funds supplied increases to \$750 billion, while the quantity demanded decreases to \$480 billion. At this above-equilibrium interest rate, firms are eager to supply loans to credit card borrowers, but relatively few people or businesses wish to borrow. As a result, some credit card firms will lower the interest rates (or other fees) they charge to attract more business. This strategy will push the interest rate down toward the equilibrium level.

If the interest rate is below the equilibrium, then excess demand or a shortage of funds occurs in this market. At an interest rate of 13%, the quantity of funds credit card borrowers demand increases to \$700 billion; but the quantity credit card firms are willing to supply is only \$510 billion. In this situation, credit card firms will perceive that they are overloaded with eager borrowers and conclude that they have an opportunity to raise interest rates or fees. The interest rate will face economic pressures to creep up toward the equilibrium level.

SHIFTS IN DEMAND AND SUPPLY IN FINANCIAL MARKETS

Those who supply financial capital face two broad decisions: how much to save, and how to divide up their savings among different forms of financial investments. We will discuss each of these in turn.

Participants in financial markets must decide when they prefer to consume goods: now or in the future. Economists call this **intertemporal decision making** because it involves decisions across

time. Unlike a decision about what to buy from the grocery store, decisions about investment or saving are made across a period of time, sometimes a long period.

Most workers save for retirement because their income in the present is greater than their needs, while the opposite will be true once they retire. So they save today and supply financial markets. If their income increases, they save more. If their perceived situation in the future changes, they change the amount of their saving. For example, there is some evidence that Social Security, the program that workers pay into in order to qualify for government checks after retirement, has tended to reduce the quantity of financial capital that workers save. If this is true, Social Security has shifted the supply of financial capital at any interest rate to the left.

By contrast, many college students need money today when their income is low (or nonexistent) to pay their college expenses. As a result, they borrow today and demand from financial markets. Once they graduate and become employed, they will pay back the loans. Individuals borrow money to purchase homes or cars. A business seeks financial investment so that it has the funds to build a factory or invest in a research and development project that will not pay off for five years, ten years, or even more. So when consumers and businesses have greater confidence that they will be able to repay in the future, the quantity demanded of financial capital at any given interest rate will shift to the right.

For example, in the technology boom of the late 1990s, many businesses became extremely confident that investments in new technology would have a high rate of return, and their demand for financial capital shifted to the right. Conversely, during the Great Recession of 2008 and 2009, their demand for financial capital at any given interest rate shifted to the left.

To this point, we have been looking at saving in total. Now let us consider what affects saving in different types of financial investments. In deciding between different forms of financial investments, suppliers of financial capital will have to consider the rates of return and the risks involved. Rate of return is a positive attribute of investments, but risk is a negative. If Investment A becomes more risky, or the return diminishes, then savers will shift their funds to Investment B—and the supply curve of financial capital for Investment A will shift back to the left while the supply curve of capital for Investment B shifts to the right.

THE UNITED STATES AS A GLOBAL BORROWER

In the global economy, trillions of dollars of financial investment cross national borders every year. In the early 2000s, financial investors from foreign countries were investing several hundred billion dollars per year more in the U.S. economy than U.S. financial investors were investing abroad. The following Work It Out deals with one of the macroeconomic concerns for the U.S. economy in recent years.

THE EFFECT OF GROWING U.S. DEBT

Imagine that the U.S. economy became viewed as a less desirable place for foreign investors to put their money because of fears about the growth of the U.S. public debt. Using the four-step process for analyzing how changes in supply and demand affect equilibrium outcomes, how would increased U.S. public debt affect the equilibrium price and quantity for capital in U.S. financial markets?

Step 1. Draw a diagram showing demand and supply for financial capital that represents the original scenario in which for-

eign investors are pouring money into the U.S. economy. Figure 2 shows a demand curve, D , and a supply curve, S , where the supply of capital includes the funds arriving from foreign investors. The original equilibrium E_0 occurs at interest rate R_0 and quantity of financial investment Q_0 .

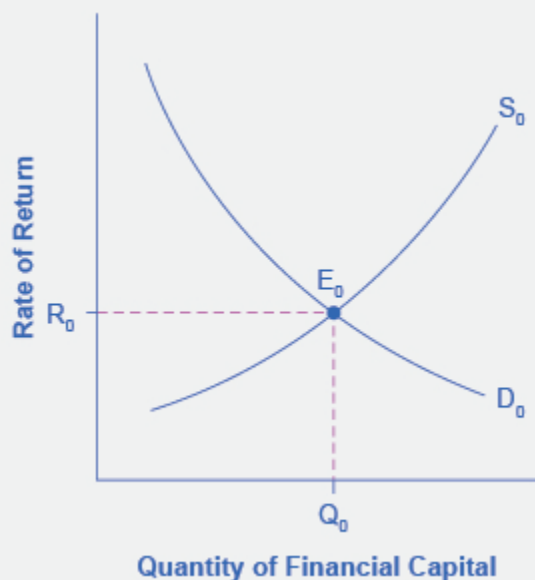


Figure 2. The United States as a Global Borrower Before U.S. Debt Uncertainty.

The graph shows the demand for financial capital from and supply of financial capital into the U.S. financial markets by the foreign sector before the increase in uncertainty regarding U.S. public debt. The original equilibrium (E_0) occurs at an equilibrium rate of return (R_0) and the equilibrium quantity is at Q_0 .

Step 2. Will the diminished confidence in the U.S. economy as a place to invest affect demand or supply of financial capital? Yes, it will affect supply. Many foreign investors look to the U.S. financial markets to store their money in safe financial vehicles with low risk and stable returns. As the U.S. debt increases, debt servicing will increase—that is, more current income will be used to pay the interest rate on past debt. Increasing U.S. debt also means that businesses may have to pay higher interest rates to borrow money, because business is now competing with the government for financial resources.

Step 3. Will supply increase or decrease? When the enthusiasm of foreign investors' for investing their money in the U.S. economy diminishes, the supply of financial capital shifts to the left. Figure 3 shows the supply curve shift from S_0 to S_1 .

Step 4. Thus, foreign investors' diminished enthusiasm leads to a new equilibrium, E_1 , which occurs at the higher interest rate, R_1 , and the lower quantity of financial investment, Q_1 .

The economy has experienced an enormous inflow of foreign capital. According to the U.S. Bureau of Economic Analysis, by the third quarter of 2014, U.S. investors had accumulated \$24.6 trillion of foreign assets, but foreign investors owned a total of \$30.8 trillion of U.S. assets. If foreign investors were to pull their money out of the U.S. economy and invest elsewhere in the world, the result could be a significantly lower quantity of financial investment in the United States, available only at a higher interest rate. This reduced inflow of foreign financial investment could impose hardship on U.S. consumers and firms interested in borrowing.

In a modern, developed economy, financial capital often moves invisibly through electronic transfers

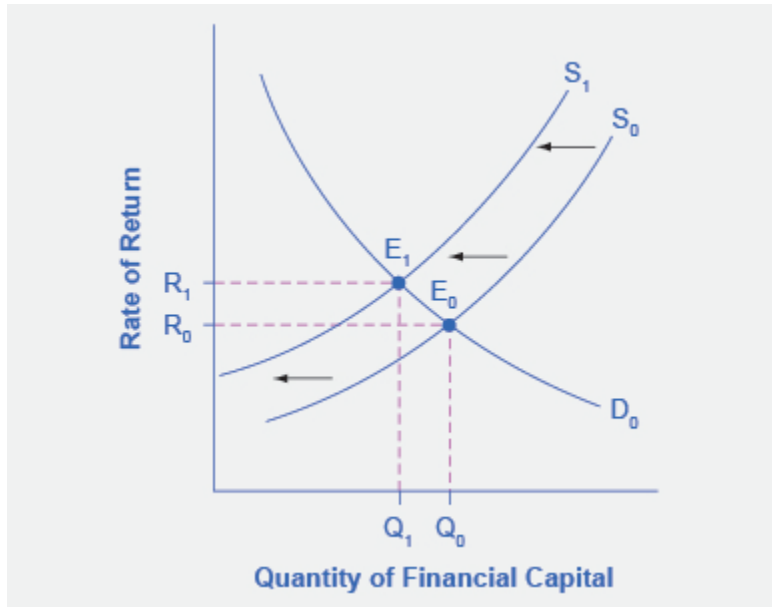


Figure 3. The United States as a Global Borrower Before and After U.S. Debt Uncertainty. The graph shows the demand for financial capital and supply of financial capital into the U.S. financial markets by the foreign sector before and after the increase in uncertainty regarding U.S. public debt. The original equilibrium (E_0) occurs at an equilibrium rate of return (R_0) and the equilibrium quantity is at Q_0 .

between one bank account and another. Yet these flows of funds can be analyzed with the same tools of demand and supply as markets for goods or labor.

PRICE CEILINGS IN FINANCIAL MARKETS: USURY LAWS

As we noted earlier, about 200 million Americans own credit cards, and their interest payments and fees total tens of billions of dollars each year. It is little wonder that political pressures sometimes arise for setting limits on the interest rates or fees that credit card companies charge. The firms that issue credit cards, including banks, oil companies, phone companies, and retail stores, respond that the higher interest rates are necessary to cover the losses created by those who borrow on their credit cards and who do not repay on time or at all. These companies also point out that cardholders can avoid paying interest if they pay their bills on time.

Consider the credit card market as illustrated in Figure 4. In this financial market, the vertical axis shows the interest rate (which is the price in the financial market). Demanders in the credit card market are households and businesses; suppliers are the companies that issue credit cards. This figure does not use specific numbers, which would be hypothetical in any case, but instead focuses on the underlying economic relationships. Imagine a law imposes a price ceiling that holds the interest rate charged on credit cards at the rate R_c , which lies below the interest rate R_0 that would otherwise have prevailed in the market. The price ceiling is shown by the horizontal dashed line in Figure 4. The demand and supply model predicts that at the lower price ceiling interest rate, the quantity demanded of credit card debt will increase from its original level of Q_0 to Q_d ; however, the quantity supplied of credit card debt will decrease from the original Q_0 to Q_s . At the price ceiling (R_c), quantity demanded

will exceed quantity supplied. Consequently, a number of people who want to have credit cards and are willing to pay the prevailing interest rate will find that companies are unwilling to issue cards to them. The result will be a credit shortage.

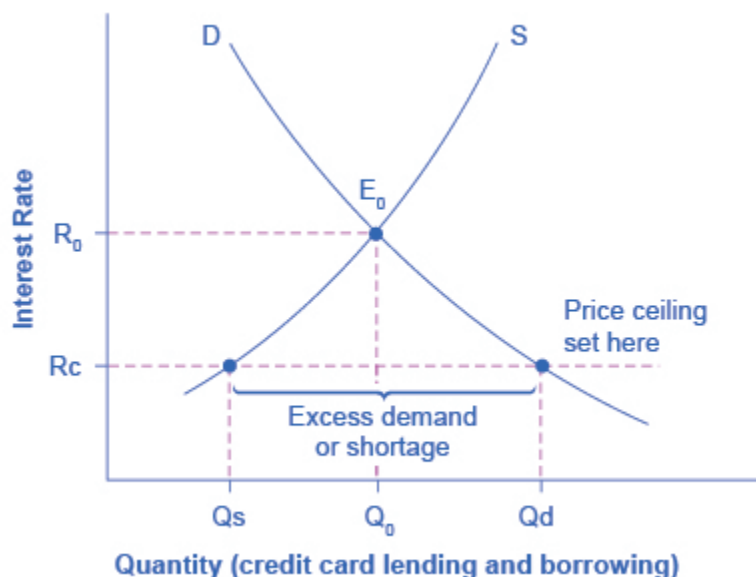


Figure 4. Credit Card Interest Rates: Another Price Ceiling Example. The original intersection of demand D and supply S occurs at equilibrium E_0 . However, a price ceiling is set at the interest rate R_c , below the equilibrium interest rate R_0 , and so the interest rate cannot adjust upward to the equilibrium. At the price ceiling, the quantity demanded, Q_d , exceeds the quantity supplied, Q_s . There is excess demand, also called a shortage.

Many states do have **usury laws**, which impose an upper limit on the interest rate that lenders can charge. However, in many cases these upper limits are well above the market interest rate. For example, if the interest rate is not allowed to rise above 30% per year, it can still fluctuate below that level according to market forces. A price ceiling that is set at a relatively high level is nonbinding, and it will have no practical effect unless the equilibrium price soars high enough to exceed the price ceiling.

KEY CONCEPTS AND SUMMARY

In the demand and supply analysis of financial markets, the “price” is the rate of return or the interest rate received. The quantity is measured by the money that flows from those who supply financial capital to those who demand it.

Two factors can shift the supply of financial capital to a certain investment: if people want to alter their existing levels of consumption, and if the riskiness or return on one investment changes relative to other investments. Factors that can shift demand for capital include business confidence and consumer confidence in the future—since financial investments received in the present are typically repaid in the future.

SELF-CHECK QUESTIONS

1. In the financial market, what causes a movement along the demand curve? What causes a shift in the demand curve?
2. In the financial market, what causes a movement along the supply curve? What causes a shift in the supply curve?
3. If a usury law limits interest rates to no more than 35%, what would the likely impact be on the amount of loans made and interest rates paid?
4. Which of the following changes in the financial market will lead to a decline in interest rates:
 - a. a rise in demand
 - b. a fall in demand
 - c. a rise in supply
 - d. a fall in supply
5. Which of the following changes in the financial market will lead to an increase in the quantity of loans made and received:
 - a. a rise in demand
 - b. a fall in demand
 - c. a rise in supply
 - d. a fall in supply

REVIEW QUESTIONS

1. How is equilibrium defined in financial markets?
2. What would be a sign of a shortage in financial markets?
3. Would usury laws help or hinder resolution of a shortage in financial markets?

CRITICAL THINKING QUESTIONS

1. Suppose the U.S. economy began to grow more rapidly than other countries in the world. What would be the likely impact on U.S. financial markets as part of the global economy?
2. If the government imposed a federal interest rate ceiling of 20% on all loans, who would gain and who would lose?

PROBLEMS

1. Predict how each of the following economic changes will affect the equilibrium price and quantity in the financial market for home loans. Sketch a demand and supply diagram to support your answers.
 - a. The number of people at the most common ages for home-buying increases.
 - b. People gain confidence that the economy is growing and that their jobs are secure.
 - c. Banks that have made home loans find that a larger number of people than they expected are not repaying those loans.
 - d. Because of a threat of a war, people become uncertain about their economic future.
 - e. The overall level of saving in the economy diminishes.
 - f. The federal government changes its bank regulations in a way that makes it cheaper and easier for banks to make home loans.
2. Table 6 shows the amount of savings and borrowing in a market for loans to purchase homes, measured in millions of dollars, at various interest rates. What is the equilibrium interest rate and quantity in the capital financial market? How can you tell? Now, imagine that because of a shift in the perceptions of foreign investors, the supply curve shifts so that there will be \$10 million less supplied at every interest rate. Calculate the new equilibrium interest rate and quantity, and explain why the direction of the interest rate shift makes intuitive sense.

Interest Rate	Qs	Qd
5%	130	170
6%	135	150
7%	140	140
8%	145	135
9%	150	125
10%	155	110

Table 6. Loans to Purchase Homes at Various Interest Rates

REFERENCES

CreditCards.com. 2013. <http://www.creditcards.com/credit-card-news/credit-card-industry-facts-personal-debt-statistics-1276.php>.

GLOSSARY

interest rate the “price” of borrowing in the financial market; a rate of return on an investment

usury laws laws that impose an upper limit on the interest rate that lenders can charge

SOLUTIONS

Answers to Self-Check Questions

1. Changes in the interest rate (i.e., the price of financial capital) cause a movement along the demand curve. A change in anything else (non-price variable) that affects demand for financial capital (e.g., changes in confidence about the future, changes in needs for borrowing) would shift the demand curve.
2. Changes in the interest rate (i.e., the price of financial capital) cause a movement along the supply curve. A change in anything else that affects the supply of financial capital (a non-price variable) such as income or future needs would shift the supply curve.
3. If market interest rates stay in their normal range, an interest rate limit of 35% would not be binding. If the equilibrium interest rate rose above 35%, the interest rate would be capped at that rate, and the quantity of loans would be lower than the equilibrium quantity, causing a shortage of loans.
4. b and c will lead to a fall in interest rates. At a lower demand, lenders will not be able to charge as much, and with more available lenders, competition for borrowers will drive rates down.
5. a and c will increase the quantity of loans. More people who want to borrow will result in more loans being given, as will more people who want to lend.

5.3 THE MARKET SYSTEM AS AN EFFICIENT MECHANISM FOR INFORMATION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Apply demand and supply models to analyze prices and quantities
- Explain the effects of price controls on the equilibrium of prices and quantities

Prices exist in markets for goods and services, for labor, and for financial capital. In all of these markets, prices serve as a remarkable social mechanism for collecting, combining, and transmitting information that is relevant to the market—namely, the relationship between demand and supply—and then serving as messengers to convey that information to buyers and sellers. In a market-oriented economy, no government agency or guiding intelligence oversees the set of responses and interconnections that result from a change in price. Instead, each consumer reacts according to that person's preferences and budget set, and each profit-seeking producer reacts to the impact on its expected profits. The following Clear It Up feature examines the **demand and supply models**.

WHY ARE DEMAND AND SUPPLY CURVES IMPORTANT?

The demand and supply model is the second fundamental diagram for this course. (The opportunity set model introduced in the Choice in a World of Scarcity chapter was the first.) Just as it would be foolish to try to learn the arithmetic of long division by memorizing every possible combination of numbers that can be divided by each other, it would be foolish to try to memorize every specific example of demand and supply in this chapter, this textbook, or this course. Demand and supply is not primarily a list of examples; it is a model to analyze prices and quantities. Even though demand and supply diagrams have many labels, they are fundamentally the same in their logic. Your goal should be to understand the underlying model so you can use it to analyze *any* market.

Figure 1 displays a generic demand and supply curve. The horizontal axis shows the different measures of quantity: a quantity of a good or service, or a quantity of labor for a given job, or a quantity of financial capital. The vertical axis shows a measure of price: the price of a good or service, the wage in the labor market, or the rate of return (like the interest rate) in the financial market.

The demand and supply model can explain the existing levels of prices, wages, and rates of return. To carry out such an analysis, think about the quantity that will be demanded at each price and the quantity that will be supplied at each price—that is, think about the shape of the demand and supply curves—and how these forces will combine to produce equilibrium.

Demand and supply can also be used to explain how economic events will cause changes in prices, wages, and rates of return. There are only four possibilities: the change in any single event may cause the demand curve to shift right or to shift left; or it may cause the supply curve to shift right or to shift left. The key to analyzing the effect of an economic event on equilibrium prices and quantities is to determine which of these four possibilities occurred. The way to do this correctly is to think back to the list of factors that shift the demand and supply curves. Note that if more than one variable is changing at the same time, the overall impact will depend on the degree of the shifts; when there are multiple variables, economists isolate each change and analyze it independently.



Figure 1. Demand and Supply Curves. The figure displays a generic demand and supply curve. The horizontal axis shows the different measures of quantity: a quantity of a good or service, a quantity of labor for a given job, or a quantity of financial capital. The vertical axis shows a measure of price: the price of a good or service, the wage in the labor market, or the rate of return (like the interest rate) in the financial market. The demand and supply curves can be used to explain how economic events will cause changes in prices, wages, and rates of return.

An increase in the price of some product signals consumers that there is a shortage and the product should perhaps be economized on. For example, if you are thinking about taking a plane trip to Hawaii, but the ticket turns out to be expensive during the week you intend to go, you might consider other weeks when the ticket might be cheaper. The price could be high because you were planning to travel during a holiday when demand for traveling is high. Or, maybe the cost of an input like jet fuel increased or the airline has raised the price temporarily to see how many people are willing to pay it. Perhaps all of these factors are present at the same time. You do not need to analyze the market and break down the price change into its underlying factors. You just have to look at the price of a ticket and decide whether and when to fly.

In the same way, price changes provide useful information to producers. Imagine the situation of a farmer who grows oats and learns that the price of oats has risen. The higher price could be due to an

increase in demand caused by a new scientific study proclaiming that eating oats is especially healthful. Or perhaps the price of a substitute grain, like corn, has risen, and people have responded by buying more oats. But the oat farmer does not need to know the details. The farmer only needs to know that the price of oats has risen and that it will be profitable to expand production as a result.

The actions of individual consumers and producers as they react to prices overlap and interlock in markets for goods, labor, and financial capital. A change in any single market is transmitted through these multiple interconnections to other markets. The vision of the role of flexible prices helping markets to reach equilibrium and linking different markets together helps to explain why **price controls** can be so counterproductive. Price controls are government laws that serve to regulate prices rather than allow the various markets to determine prices. There is an old proverb: “Don’t kill the messenger.” In ancient times, messengers carried information between distant cities and kingdoms. When they brought bad news, there was an emotional impulse to kill the messenger. But killing the messenger did not kill the bad news. Moreover, killing the messenger had an undesirable side effect: Other messengers would refuse to bring news to that city or kingdom, depriving its citizens of vital information.

Those who seek price controls are trying to kill the messenger—or at least to stifle an unwelcome message that prices are bringing about the equilibrium level of price and quantity. But price controls do nothing to affect the underlying forces of demand and supply, and this can have serious repercussions. During China’s “Great Leap Forward” in the late 1950s, food prices were kept artificially low, with the result that 30 to 40 million people died of starvation because the low prices depressed farm production. Changes in demand and supply will continue to reveal themselves through consumers’ and producers’ behavior. Immobilizing the price messenger through price controls will deprive everyone in the economy of critical information. Without this information, it becomes difficult for everyone—buyers and sellers alike—to react in a flexible and appropriate manner as changes occur throughout the economy.

BABY BOOMERS COME OF AGE

The theory of supply and demand can explain what happens in the labor markets and suggests that the demand for nurses will increase as healthcare needs of baby boomers increase, as Figure 2 shows. The impact of that increase will result in an average salary higher than the \$67,930 earned in 2012 referenced in the first part of this case. The new equilibrium (E_1) will be at the new equilibrium price (P_{e1}). Equilibrium quantity will also increase from Q_{e0} to Q_{e1} .

Suppose that as the demand for nurses increases, the supply shrinks due to an increasing number of nurses entering retirement and increases in the tuition of nursing degrees. The impact of a decreasing supply of nurses is captured by the leftward shift of the supply curve in Figure 3. The shifts in the two curves result in higher salaries for nurses, but the overall impact in the quantity of nurses is uncertain, as it depends on the relative shifts of supply and demand.

While we do not know if the number of nurses will increase or decrease relative to their initial employment, we know they will have higher salaries. The situation of the labor market for nurses described in the beginning of the chapter is different from this example, because instead of a shrinking supply, we had the supply growing at a lower rate than the growth in demand. Since both curves were shifting to the right, we would have an unequivocal increase in the quantity of nurses. And because the shift in the demand curve was larger than the one in the supply, we would expect higher wages as a result.

KEY CONCEPTS AND SUMMARY

The market price system provides a highly efficient mechanism for disseminating information about

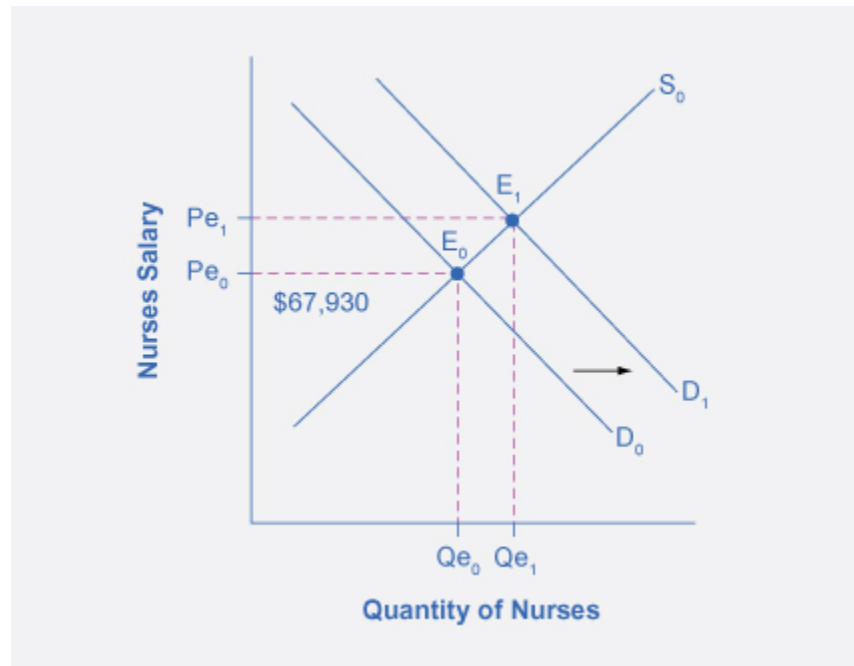


Figure 2. Impact of Increasing Demand for Nurses 2012-2022. In 2012, the median salary for nurses was \$67,930. As demand for services increases, the demand curve shifts to the right (from D_0 to D_1) and the equilibrium quantity of nurses increases from Q_{e_0} to Q_{e_1} .

The equilibrium salary increases from P_{e_0} to P_{e_1} .

relative scarcities of goods, services, labor, and financial capital. Market participants do not need to know why prices have changed, only that the changes require them to revisit previous decisions they made about supply and demand. Price controls hide information about the true scarcity of products and thereby cause misallocation of resources.

SELF-CHECK QUESTIONS

1. Identify the most accurate statement. A price floor will have the largest effect if it is set:

- a. substantially above the equilibrium price
- b. slightly above the equilibrium price
- c. slightly below the equilibrium price
- d. substantially below the equilibrium price

Sketch all four of these possibilities on a demand and supply diagram to illustrate your answer.

2. A price ceiling will have the largest effect:

- a. substantially below the equilibrium price
- b. slightly below the equilibrium price
- c. substantially above the equilibrium price
- d. slightly above the equilibrium price

Sketch all four of these possibilities on a demand and supply diagram to illustrate your answer.

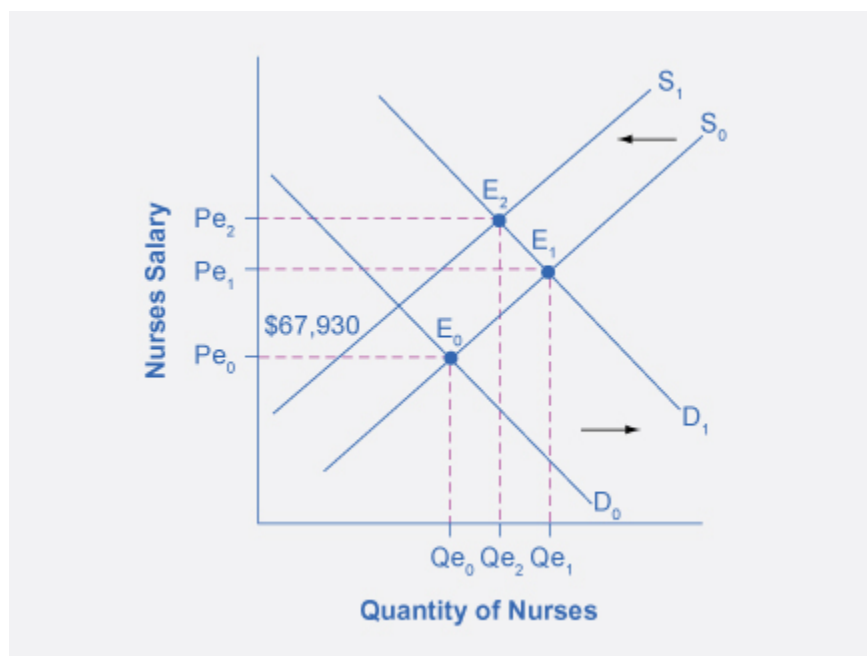


Figure 3. Impact of Decreasing Supply of Nurses between 2012 and 2022. Initially, salaries increase as demand for nursing increases to Pe_1 . When demand increases, so too does the equilibrium quantity, from Qe_0 to Qe_1 . The decrease in the supply of nurses due to nurses retiring from the workforce and fewer nursing graduates (*ceteris paribus*), causes a leftward shift of the supply curve resulting in even higher salaries for nurses, at Pe_2 , but an uncertain outcome for the equilibrium quantity of nurses, which in this representation is less than Qe_1 , but more than the initial Qe_0 .

3. Select the correct answer. A price floor will usually shift:

- a. demand
- b. supply
- c. both
- d. neither

Illustrate your answer with a diagram.

4. Select the correct answer. A price ceiling will usually shift:

- a. demand
- b. supply
- c. both
- d. neither

REVIEW QUESTIONS

Whether the product market or the labor market, what happens to the equilibrium price and quantity for each of the four possibilities: increase in demand, decrease in demand, increase in supply, and decrease in supply.

CRITICAL THINKING QUESTIONS

1. Why are the factors that shift the demand for a product different from the factors that shift the demand for labor? Why are the factors that shift the supply of a product different from those that shift the supply of labor?
2. During a discussion several years ago on building a pipeline to Alaska to carry natural gas, the U.S. Senate passed a bill stipulating that there should be a guaranteed minimum price for the natural gas that would be carried through the pipeline. The thinking behind the bill was that if private firms had a guaranteed price for their natural gas, they would be more willing to drill for gas and to pay to build the pipeline.
 - a. Using the demand and supply framework, predict the effects of this price floor on the price, quantity demanded, and quantity supplied.
 - b. With the enactment of this price floor for natural gas, what are some of the likely unintended consequences in the market?
 - c. Suggest some policies other than the price floor that the government can pursue if it wishes to encourage drilling for natural gas and for a new pipeline in Alaska.

PROBLEMS

1. Imagine that to preserve the traditional way of life in small fishing villages, a government decides to impose a price floor that will guarantee all fishermen a certain price for their catch.
 - a. Using the demand and supply framework, predict the effects on the price, quantity demanded, and quantity supplied.
 - b. With the enactment of this price floor for fish, what are some of the likely unintended consequences in the market?
 - c. Suggest some policies other than the price floor to make it possible for small fishing villages to continue.
2. What happens to the price and the quantity bought and sold in the cocoa market if countries producing cocoa experience a drought and a new study is released demonstrating the health benefits of cocoa? Illustrate your answer with a demand and supply graph.

SOLUTIONS

Answers to Self-Check Questions

1. A price floor prevents a price from falling below a certain level, but has no effect on prices above that level. It will have its biggest effect in creating excess supply (as measured by the entire area inside the dotted lines on the graph, from D to S) if it is substantially above the equilibrium price. This is illustrated in the following figure.

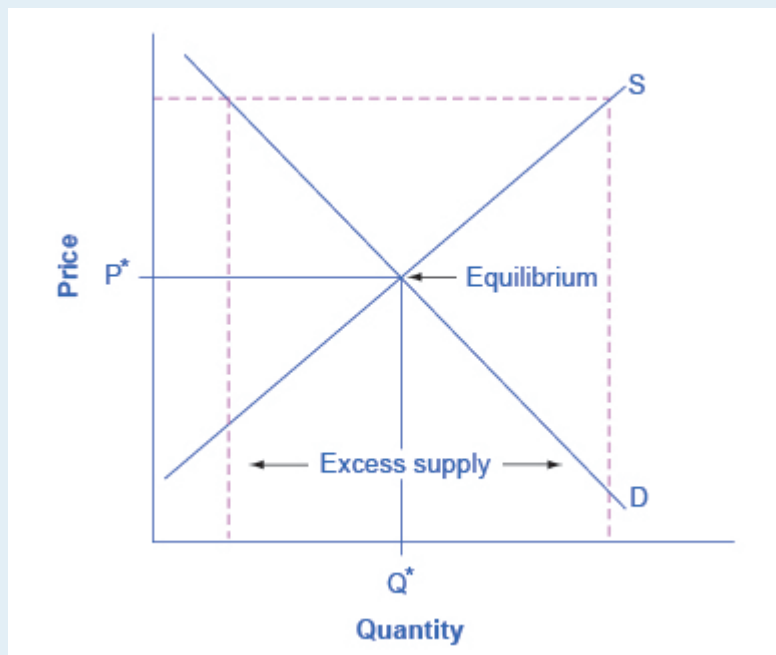


Figure 4.

It will have a lesser effect if it is slightly above the equilibrium price. This is illustrated in the next figure.

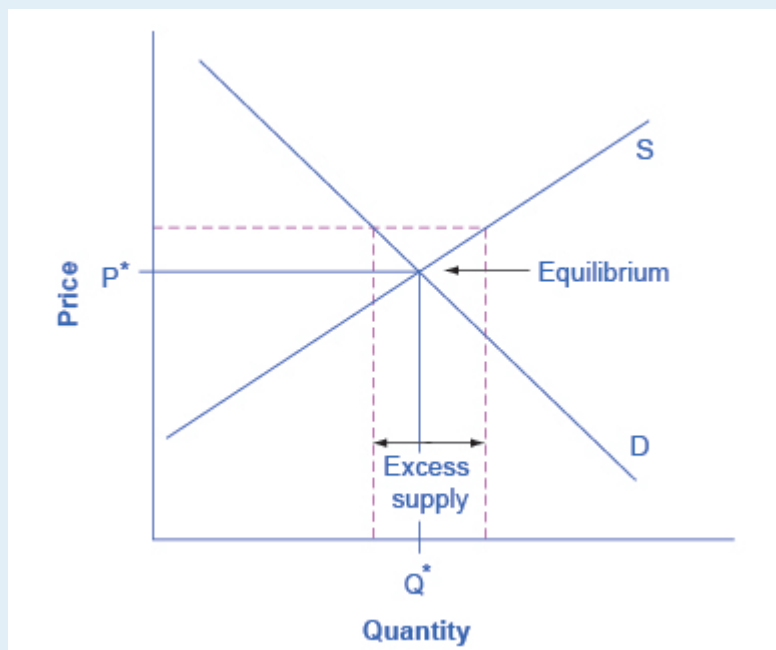


Figure 5.

It will have no effect if it is set either slightly or substantially below the equilibrium price, since an equilibrium price above a price floor will not be affected by that price floor. The following figure illustrates these situations.

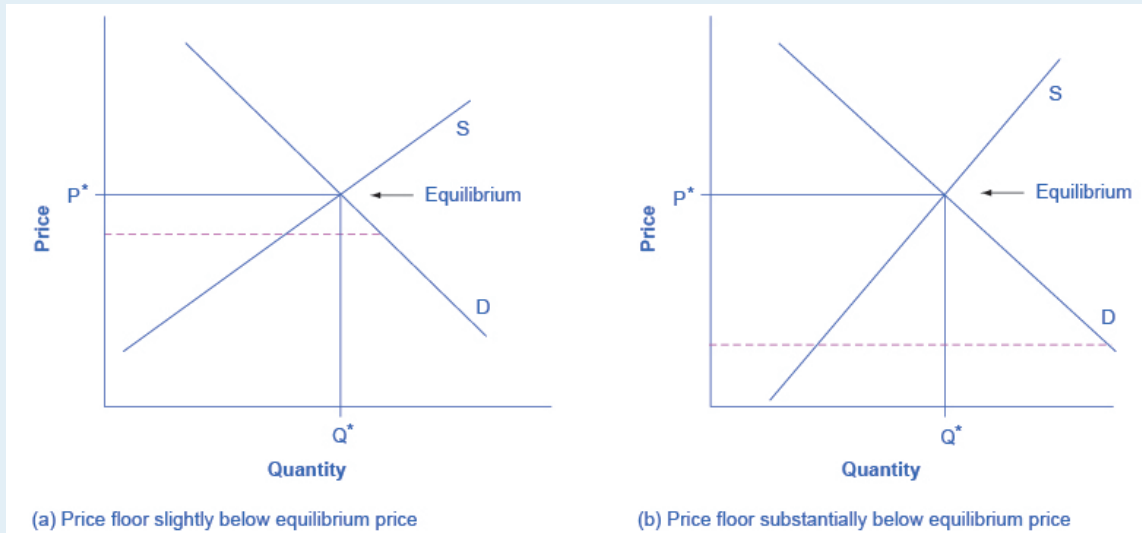


Figure 6.

2. A price ceiling prevents a price from rising above a certain level, but has no effect on prices below that level. It will have its biggest effect in creating excess demand if it is substantially below the equilibrium price. The following figure illustrates these situations.

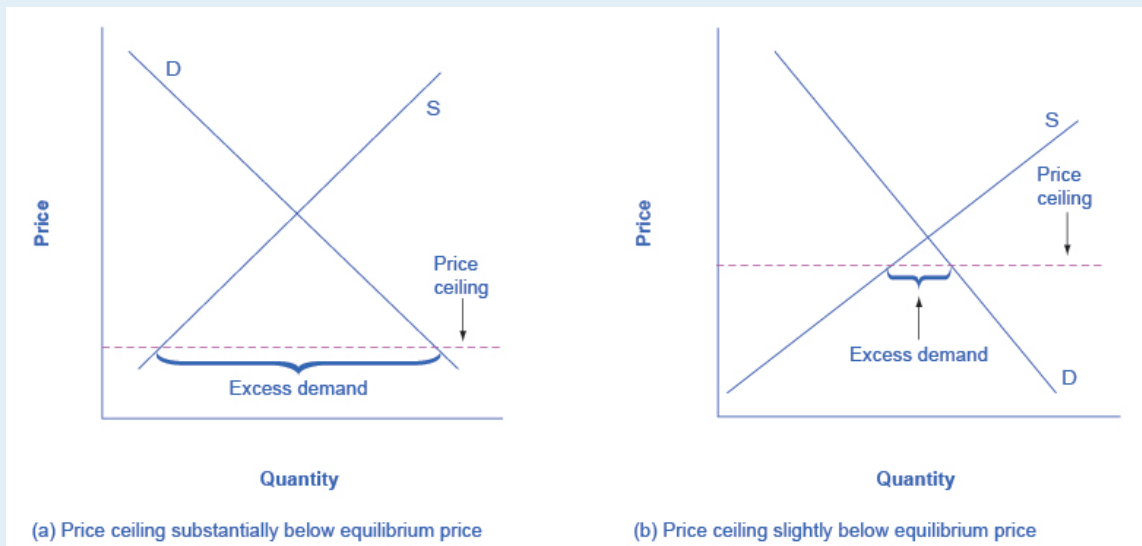
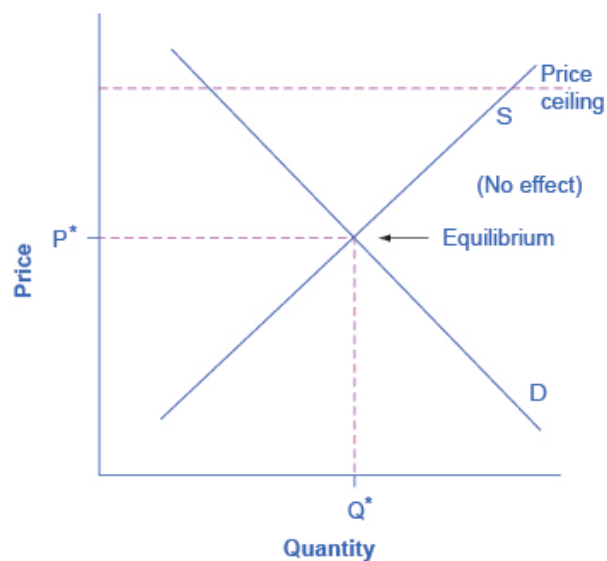


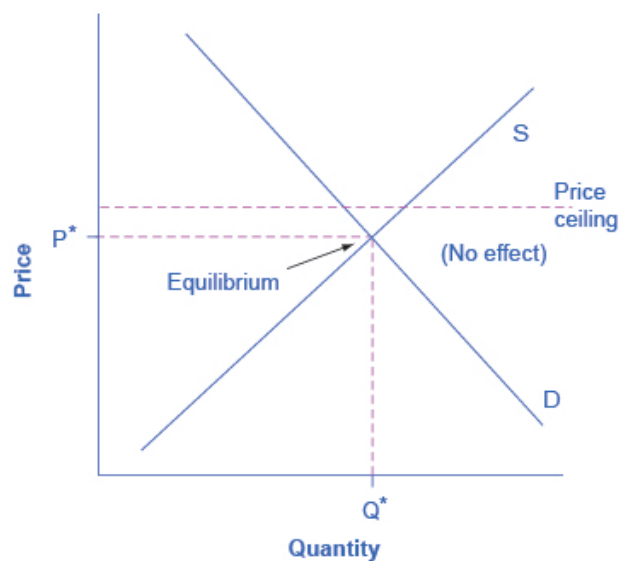
Figure 7.

When the price ceiling is set substantially or slightly above the equilibrium price, it will have no effect on creating excess demand. The following figure illustrates these situations.

3. Neither. A shift in demand or supply means that at every price, either a greater or a lower quantity is demanded or supplied. A price floor does not shift a demand curve or a supply curve. However, if the price



(a) Price ceiling substantially above equilibrium price



(b) Price ceiling slightly above equilibrium price

Figure 8.

floor is set above the equilibrium, it will cause the quantity supplied on the supply curve to be greater than the quantity demanded on the demand curve, leading to excess supply.

4. Neither. A shift in demand or supply means that at every price, either a greater or a lower quantity is demanded or supplied. A price ceiling does not shift a demand curve or a supply curve. However, if the price ceiling is set below the equilibrium, it will cause the quantity demanded on the demand curve to be greater than the quantity supplied on the supply curve, leading to excess demand.

CHAPTER 6. ELASTICITY

INTRODUCTION TO ELASTICITY

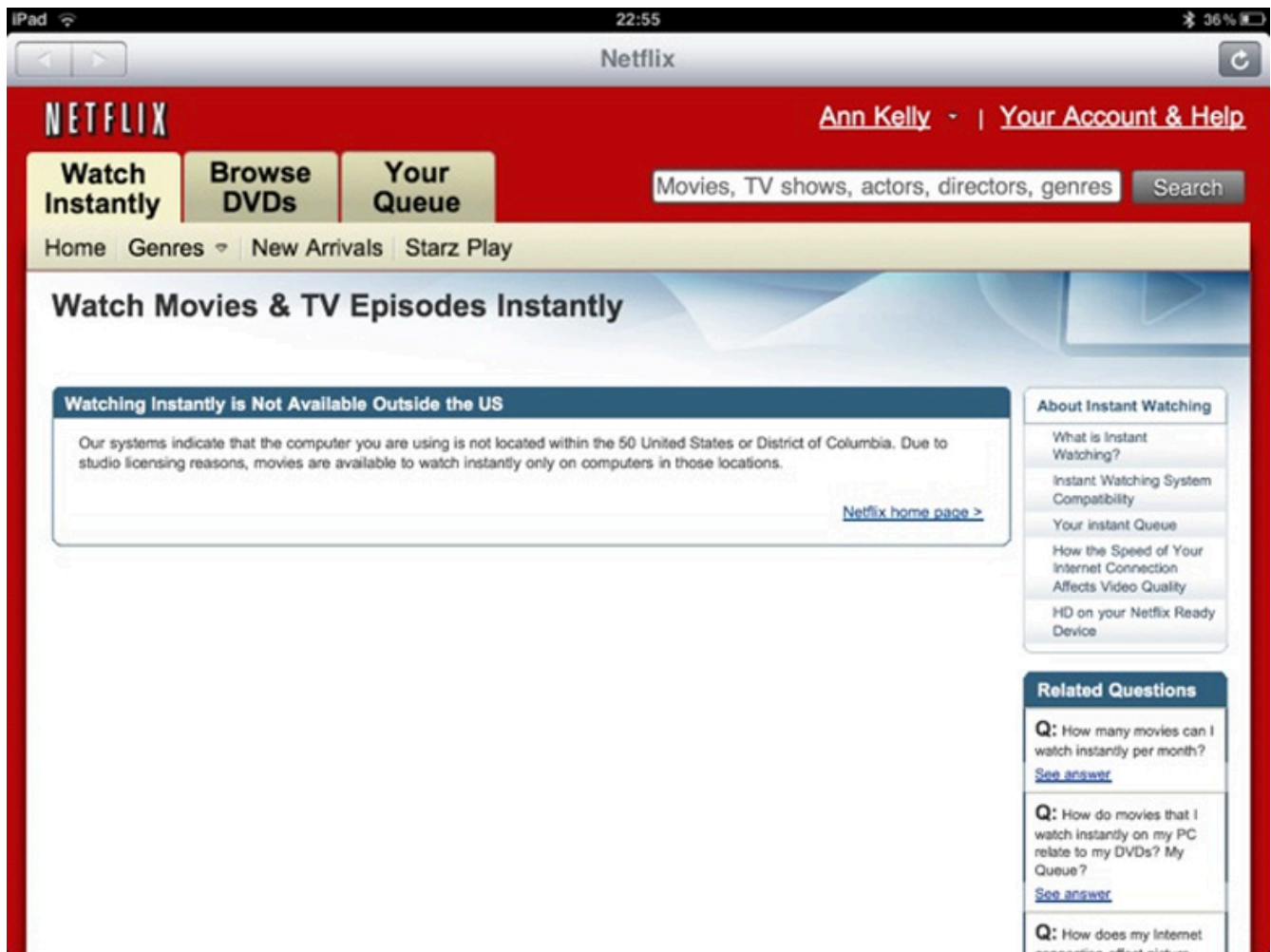


Figure 1. Netflix On-Demand Media. Netflix, Inc. is an American provider of on-demand Internet streaming media to many countries around the world, including the United States, and of flat rate DVD-by-mail in the United States. (Credit: modification of work by Traci Lawson/Flickr Creative Commons)

THAT WILL BE HOW MUCH?

Imagine going to your favorite coffee shop and having the waiter inform you the pricing has changed. Instead of \$3 for a cup of coffee, you will now be charged \$2 for coffee, \$1 for creamer, and \$1 for your choice of sweetener. If you pay your usual \$3 for a cup of coffee, you must choose between creamer and sweetener. If you want both, you now face an extra

charge of \$1. Sound absurd? Well, that is the situation Netflix customers found themselves in—a 60% price hike to retain the same service in 2011.

In early 2011, Netflix consumers paid about \$10 a month for a package consisting of streaming video and DVD rentals. In July 2011, the company announced a packaging change. Customers wishing to retain both streaming video and DVD rental would be charged \$15.98 per month, a price increase of about 60%. In 2014, Netflix also raised its streaming video subscription price from \$7.99 to \$8.99 per month for new U.S. customers. The company also changed its policy of 4K streaming content from \$9.00 to \$12.00 per month that year.

How would customers of the 18-year-old firm react? Would they abandon Netflix? Would the ease of access to other venues make a difference in how consumers responded to the Netflix price change? The answers to those questions will be explored in this chapter: the change in quantity with respect to a change in price, a concept economists call elasticity.

CHAPTER OBJECTIVES

Introduction to Elasticity

In this chapter, you will learn about:

- Price Elasticity of Demand and Price Elasticity of Supply
- Polar Cases of Elasticity and Constant Elasticity
- Elasticity and Pricing
- Elasticity in Areas Other Than Price

Anyone who has studied economics knows the law of demand: a higher price will lead to a lower quantity demanded. What you may not know is how much lower the quantity demanded will be. Similarly, the law of supply shows that a higher price will lead to a higher quantity supplied. The question is: How much higher? This chapter will explain how to answer these questions and why they are critically important in the real world.

To find answers to these questions, we need to understand the concept of elasticity. **Elasticity** is an economics concept that measures responsiveness of one variable to changes in another variable. Suppose you drop two items from a second-floor balcony. The first item is a tennis ball. The second item is a brick. Which will bounce higher? Obviously, the tennis ball. We would say that the tennis ball has greater elasticity.

Consider an economic example. Cigarette taxes are an example of a “sin tax,” a tax on something that is bad for you, like alcohol. Cigarettes are taxed at the state and national levels. State taxes range from a low of 17 cents per pack in Missouri to \$4.35 per pack in New York. The average state cigarette tax is \$1.51 per pack. The 2014 federal tax rate on cigarettes was \$1.01 per pack, but in 2015 the Obama Administration proposed raising the federal tax nearly a dollar to \$1.95 per pack. The key question is: How much would cigarette purchases decline?

Taxes on cigarettes serve two purposes: to raise tax revenue for government and to discourage consumption of cigarettes. However, if a higher cigarette tax discourages consumption by quite a lot, meaning a greatly reduced quantity of cigarettes is sold, then the cigarette tax on each pack will not raise much revenue for the government. Alternatively, a higher cigarette tax that does not discourage

consumption by much will actually raise more tax revenue for the government. Thus, when a government agency tries to calculate the effects of altering its cigarette tax, it must analyze how much the tax affects the quantity of cigarettes consumed. This issue reaches beyond governments and taxes; every firm faces a similar issue. Every time a firm considers raising the price that it charges, it must consider how much a price increase will reduce the quantity demanded of what it sells. Conversely, when a firm puts its products on sale, it must expect (or hope) that the lower price will lead to a significantly higher quantity demanded.

6.1 PRICE ELASTICITY OF DEMAND AND PRICE ELASTICITY OF SUPPLY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Calculate the price elasticity of demand
- Calculate the price elasticity of supply

Both the demand and supply curve show the relationship between price and the number of units demanded or supplied. **Price elasticity** is the ratio between the percentage change in the quantity demanded (Qd) or supplied (Qs) and the corresponding percent change in price. The **price elasticity of demand** is the percentage change in the quantity *demanded* of a good or service divided by the percentage change in the price. The **price elasticity of supply** is the percentage change in quantity *supplied* divided by the percentage change in price.

Elasticities can be usefully divided into three broad categories: elastic, inelastic, and unitary. An **elastic demand** or **elastic supply** is one in which the elasticity is greater than one, indicating a high responsiveness to changes in price. Elasticities that are less than one indicate low responsiveness to price changes and correspond to **inelastic demand** or **inelastic supply**. **Unitary elasticities** indicate proportional responsiveness of either demand or supply, as summarized in Table 1.

If:	Then:	And It Is Called:
$\% \text{ change in quantity} > \% \text{ change in price}$	$\frac{\% \text{ change in quantity}}{\% \text{ change in price}} > 1$	Elastic
$\% \text{ change in quantity} = \% \text{ change in price}$	$\frac{\% \text{ change in quantity}}{\% \text{ change in price}} = 1$	Unitary
$\% \text{ change in quantity} < \% \text{ change in price}$	$\frac{\% \text{ change in quantity}}{\% \text{ change in price}} < 1$	Inelastic

Table 1. Elastic, Inelastic, and Unitary: Three Cases of Elasticity

Before we get into the nitty gritty of elasticity, enjoy this article on elasticity and ticket prices at the Super Bowl.



To calculate elasticity, instead of using simple percentage changes in quantity and price, economists use the average percent change in both quantity and price. This is called the Midpoint Method for Elasticity, and is represented in the following equations:

$$\% \text{ change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1)/2} \times 100$$

$$\% \text{ change in price} = \frac{P_2 - P_1}{(P_2 + P_1)/2} \times 100$$

The advantage of the is **Midpoint Method** is that one obtains the same elasticity between two price points whether there is a price increase or decrease. This is because the formula uses the same base for both cases.

CALCULATING PRICE ELASTICITY OF DEMAND

Let's calculate the elasticity between points A and B and between points G and H shown in Figure 1.

First, apply the formula to calculate the elasticity as price decreases from \$70 at point B to \$60 at point A:

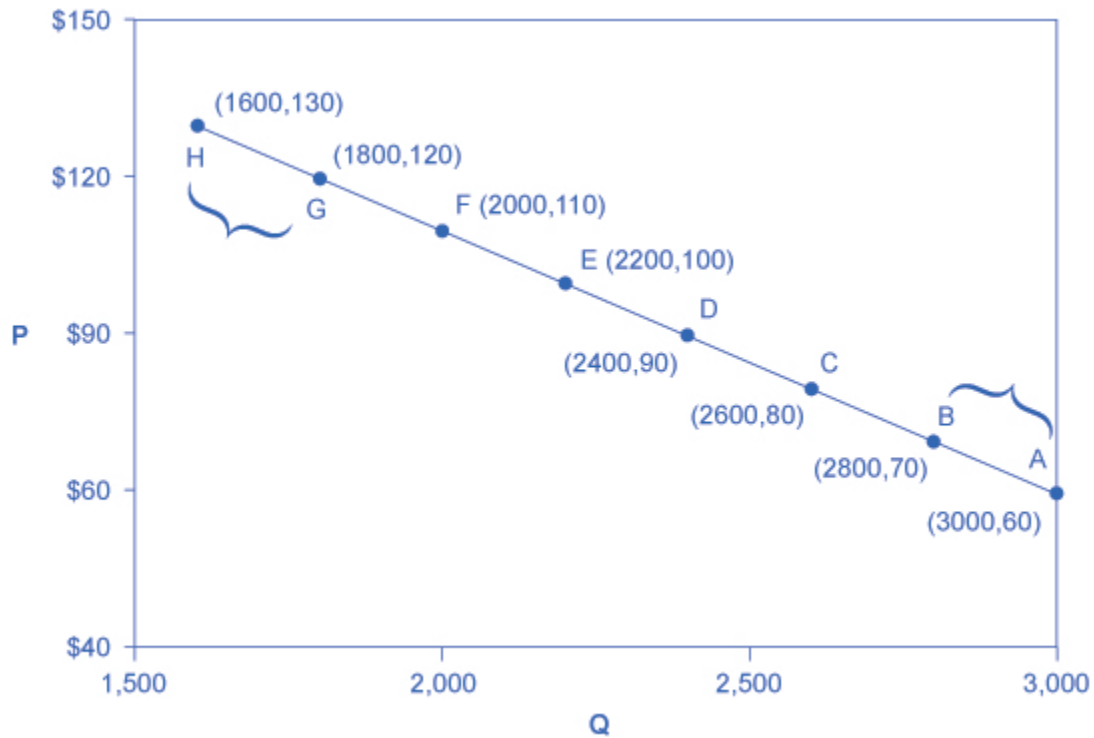


Figure 1. Calculating the Price Elasticity of Demand. The price elasticity of demand is calculated as the percentage change in quantity divided by the percentage change in price.

$$\% \text{ change in quantity} = \frac{3,000 - 2,800}{(3,000 + 2,800)/2} \times 100$$

$$= \frac{200}{2,900} \times 100$$

$$= 6.9$$

$$\% \text{ change in price} = \frac{60 - 70}{(60 + 70)/2} \times 100$$

$$= \frac{-10}{65} \times 100$$

$$= -15.4$$

$$\text{Price Elasticity of Demand} = \frac{6.9\%}{-15.4\%}$$

$$= 0.45$$

$$\frac{6.9\%}{-15.4\%}$$

Therefore, the elasticity of demand between these two points is -15.4% which is 0.45, an amount smaller than one, showing that the demand is inelastic in this interval. Price elasticities of demand are *always* negative since price and quantity demanded always move in opposite directions (on the demand curve). By convention, we always talk about elasticities as positive numbers. So mathematically, we take the absolute value of the result. We will ignore this detail from now on, while remembering to interpret elasticities as positive numbers.

This means that, along the demand curve between point B and A, if the price changes by 1%, the quantity demanded will change by 0.45%. A change in the price will result in a smaller percentage change in the quantity demanded. For example, a 10% *increase* in the price will result in only a 4.5% *decrease* in quantity demanded. A 10% *decrease* in the price will result in only a 4.5% *increase* in the quantity demanded. Price elasticities of demand are negative numbers indicating that the demand curve is downward sloping, but are read as absolute values. The following Work It Out feature will walk you through calculating the price elasticity of demand.

FINDING THE PRICE ELASTICITY OF DEMAND

Calculate the price elasticity of demand using the data in Figure 1 for an increase in price from G to H. Has the elasticity increased or decreased?

Step 1. We know that:

$$\text{Price Elasticity of Demand} = \frac{\% \text{ change in quantity}}{\% \text{ change in price}}$$

Step 2. From the **Midpoint Formula** we know that:

$$\% \text{ change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1)/2} \times 100$$

$$\% \text{ change in price} = \frac{P_2 - P_1}{(P_2 + P_1)/2} \times 100$$

Step 3. So we can use the values provided in the figure in each equation:

$$\% \text{ change in quantity} \quad \frac{1,600-1,800}{(1,600+1,800)/2} \times 100$$

$$\frac{-200}{1,700} \times 100$$

$$-11.76$$

$$\% \text{ change in price} \quad \frac{130-120}{(130+120)/2} \times 100$$

$$\frac{10}{125} \times 100$$

$$8.0$$

Step 4. Then, those values can be used to determine the price elasticity of demand:

$$\text{Price Elasticity of Demand} \quad \frac{\% \text{ change in quantity}}{\% \text{ change in price}}$$

$$\frac{-11.76}{8}$$

$$1.47$$

Therefore, the elasticity of demand from G to H 1.47. The magnitude of the elasticity has increased (in absolute value) as we moved up along the **demand curve** from points A to B. Recall that the elasticity between these two points was 0.45. Demand was inelastic between points A and B and elastic between points G and H. This shows us that price elasticity of demand changes at different points along a **straight-line demand curve**.

CALCULATING THE PRICE ELASTICITY OF SUPPLY

Assume that an apartment rents for \$650 per month and at that price 10,000 units are rented as shown in Figure 2. When the price increases to \$700 per month, 13,000 units are supplied into the market. By what percentage does apartment supply increase? What is the price sensitivity?

Using the **Midpoint Method**,

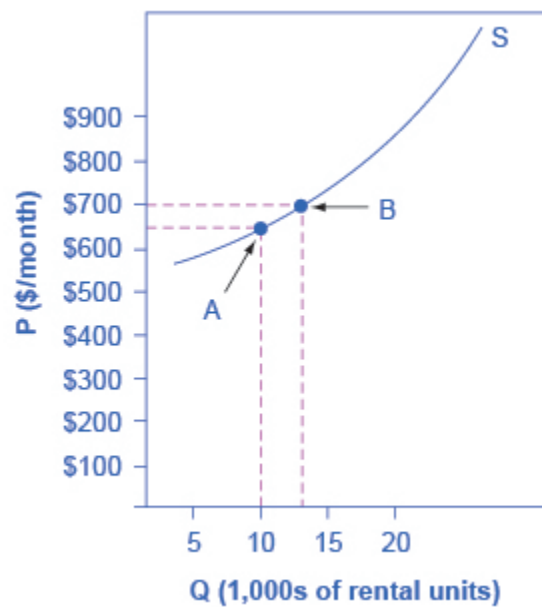


Figure 2. Price Elasticity of Supply. The price elasticity of supply is calculated as the percentage change in quantity divided by the percentage change in price.

$$\% \text{ change in quantity} = \frac{13,000 - 10,000}{(13,000 + 10,000)/2} \times 100$$

$$\frac{3,000}{11,500} \times 100$$

$$26.1$$

$$\% \text{ change in price} = \frac{\$700 - \$650}{(\$700 + \$650)/2} \times 100$$

$$\frac{50}{675} \times 100$$

$$7.4$$

$$\text{Price Elasticity of Demand} = \frac{26.1\%}{7.4\%}$$

$$3.53$$

Again, as with the elasticity of demand, the elasticity of supply is not followed by any units. Elasticity is a ratio of one percentage change to another percentage change—nothing more—and is read as an absolute value. In this case, a 1% rise in price causes an increase in quantity supplied of 3.5%. The greater than one elasticity of supply means that the percentage change in quantity supplied will be

greater than a one percent price change. If you're starting to wonder if the concept of slope fits into this calculation, read the following Clear It Up box.

IS THE ELASTICITY THE SLOPE?

It is a common mistake to confuse the slope of either the supply or demand curve with its elasticity. The slope is the rate of change in units along the curve, or the rise/run (change in y over the change in x). For example, in Figure 1, each point shown on the demand curve, price drops by \$10 and the number of units demanded increases by 200. So the slope is $-10/200$ along the entire demand curve and does not change. The price elasticity, however, changes along the curve. Elasticity between points A and B was 0.45 and increased to 1.47 between points G and H. Elasticity is the *percentage* change, which is a different calculation from the slope and has a different meaning.

When we are at the upper end of a demand curve, where price is high and the quantity demanded is low, a small change in the quantity demanded, even in, say, one unit, is pretty big in percentage terms. A change in price of, say, a dollar, is going to be much less important in percentage terms than it would have been at the bottom of the demand curve. Likewise, at the bottom of the demand curve, that one unit change when the quantity demanded is high will be small as a percentage.

So, at one end of the demand curve, where we have a large percentage change in quantity demanded over a small percentage change in price, the elasticity value would be high, or demand would be relatively elastic. Even with the same change in the price and the same change in the quantity demanded, at the other end of the demand curve the quantity is much higher, and the price is much lower, so the percentage change in quantity demanded is smaller and the percentage change in price is much higher. That means at the bottom of the curve we'd have a small numerator over a large denominator, so the elasticity measure would be much lower, or inelastic.

As we move along the demand curve, the values for quantity and price go up or down, depending on which way we are moving, so the percentages for, say, a \$1 difference in price or a one unit difference in quantity, will change as well, which means the ratios of those percentages will change.

KEY CONCEPTS AND SUMMARY

Price elasticity measures the responsiveness of the quantity demanded or supplied of a good to a change in its price. It is computed as the percentage change in quantity demanded (or supplied) divided by the percentage change in price. Elasticity can be described as elastic (or very responsive), unit elastic, or inelastic (not very responsive). Elastic demand or supply curves indicate that quantity demanded or supplied respond to price changes in a greater than proportional manner. An inelastic demand or supply curve is one where a given percentage change in price will cause a smaller percentage change in quantity demanded or supplied. A unitary elasticity means that a given percentage change in price leads to an equal percentage change in quantity demanded or supplied.

SELF-CHECK QUESTIONS

1. From the data shown in Table 2 about demand for smart phones, calculate the price elasticity of demand from: point B to point C, point D to point E, and point G to point H. Classify the elasticity at each point as elastic, inelastic, or unit elastic.

Points	P	Q
A	60	3,000
B	70	2,800
C	80	2,600
D	90	2,400
E	100	2,200
F	110	2,000
G	120	1,800
H	130	1,600

Table 2.

2. From the data shown in Table 3 about supply of alarm clocks, calculate the price elasticity of supply from: point J to point K, point L to point M, and point N to point P. Classify the elasticity at each point as elastic, inelastic, or unit elastic.

Point	Price	Quantity Supplied
J	\$8	50
K	\$9	70
L	\$10	80
M	\$11	88
N	\$12	95
P	\$13	100

Table 3.

REVIEW QUESTIONS

1. What is the formula for calculating elasticity?
2. What is the price elasticity of demand? Can you explain it in your own words?
3. What is the price elasticity of supply? Can you explain it in your own words?

CRITICAL THINKING QUESTIONS

1. Transatlantic air travel in business class has an estimated elasticity of demand of 0.40 less than transatlantic air travel in economy class, with an estimated price elasticity of 0.62. Why do you think this is the case?
2. What is the relationship between price elasticity and position on the demand curve? For example, as you move up the demand curve to higher prices and lower quantities, what happens to the measured elasticity? How would you explain that?

PROBLEMS

1. The equation for a demand curve is $P = 48 - 3Q$. What is the elasticity in moving from a quantity of 5 to a quantity of 6?
2. The equation for a demand curve is $P = 2/Q$. What is the elasticity of demand as price falls from 5 to 4? What is the elasticity of demand as the price falls from 9 to 8? Would you expect these answers to be the same?
3. The equation for a supply curve is $4P = Q$. What is the elasticity of supply as price rises from 3 to 4? What is the elasticity of supply as the price rises from 7 to 8? Would you expect these answers to be the same?
4. The equation for a supply curve is $P = 3Q - 8$. What is the elasticity in moving from a price of 4 to a price of 7?

GLOSSARY

elastic demand when the elasticity of demand is greater than one, indicating a high responsiveness of quantity demanded or supplied to changes in price

elastic supply when the elasticity of either supply is greater than one, indicating a high responsiveness of quantity demanded or supplied to changes in price

elasticity an economics concept that measures responsiveness of one variable to changes in another variable

inelastic demand when the elasticity of demand is less than one, indicating that a 1 percent increase in price paid by the consumer leads to less than a 1 percent change in purchases (and vice versa); this indicates a low responsiveness by consumers to price changes

inelastic supply when the elasticity of supply is less than one, indicating that a 1 percent increase in price paid to the firm will result in a less than 1 percent increase in production by the firm; this indicates a low responsiveness of the firm to price increases (and vice versa if prices drop)

price elasticity the relationship between the percent change in price resulting in a corresponding percentage change in the quantity demanded or supplied

price elasticity of demand percentage change in the quantity *demanded* of a good or service divided the percentage change in price

price elasticity of supply percentage change in the quantity *supplied* divided by the percentage change in price

unitary elasticity when the calculated elasticity is equal to one indicating that a change in the price of the good or service results in a proportional change in the quantity demanded or supplied

SOLUTIONS

Answers to Self-Check Questions

1. From point B to point C, price rises from \$70 to \$80, and Qd decreases from 2,800 to 2,600. So:

$$\% \text{ change in quantity} \quad \frac{2,600 - 2,800}{(2,600 + 2,800)/2} \times 100$$

$$\frac{-200}{2,700} \times 100$$

$$-7.41$$

$$\% \text{ change in price} \quad \frac{80 - 70}{(80 + 70)/2} \times 100$$

$$\frac{10}{75} \times 100$$

$$13.33$$

$$\text{Elasticity of Demand} \quad \frac{-7.41\%}{13.33\%}$$

$$0.56$$

The demand curve is inelastic in this area; that is, its elasticity value is less than one.

Answer from Point D to point E:

$$\% \text{ change in quantity} \quad \frac{2,200 - 2,400}{(2,200 + 2,400)/2} \times 100$$

$$\frac{-200}{2,300} \times 100$$

$$-8.7$$

$$\% \text{ change in price} \quad \frac{100 - 90}{(100 + 90)/2} \times 100$$

$$\frac{10}{95} \times 100$$

$$10.53$$

$$\text{Elasticity of Demand} \quad \frac{-8.7\%}{10.53\%}$$

$$0.83$$

The demand curve is inelastic in this area; that is, its elasticity value is less than one.

Answer from Point G to point H:

$$\% \text{ change in quantity} \quad \frac{1,600-1,800}{(1,600+1,800)/2} \times 100$$

$$\frac{-200}{1,700} \times 100$$

$$-11.76$$

$$\% \text{ change in price} \quad \frac{130-120}{(130+120)/2} \times 100$$

$$\frac{10}{125} \times 100$$

$$7.81$$

$$\text{Elasticity of Demand} \quad \frac{-11.76\%}{7.81\%}$$

$$-1.51$$

The demand curve is elastic in this interval.

2. From point J to point K, price rises from \$8 to \$9, and quantity rises from 50 to 70. So:

$$\% \text{ change in quantity} \quad \frac{70-50}{(70+50)/2} \times 100$$

$$\frac{20}{60} \times 100$$

$$33.33$$

$$\% \text{ change in price} \quad \frac{\$9-\$8}{(\$9+\$8)/2} \times 100$$

$$\frac{1}{8.5} \times 100$$

$$11.76$$

$$\text{Elasticity of Supply} \quad \frac{33.33\%}{11.76\%}$$

$$2.83$$

The supply curve is elastic in this area; that is, its elasticity value is greater than one.

From point L to point M, the price rises from \$10 to \$11, while the Qs rises from 80 to 88:

$$\% \text{ change in quantity} \quad \frac{88-80}{(88+80)/2} \times 100$$

$$\frac{8}{84} \times 100$$

$$9.52$$

$$\% \text{ change in price} \quad \frac{\$11-\$10}{(\$11+\$10)/2} \times 100$$

$$\frac{1}{10.5} \times 100$$

$$9.52$$

$$\text{Elasticity of Demand} \quad \frac{9.52\%}{9.52\%}$$

$$1.0$$

The supply curve has unitary elasticity in this area.

From point N to point P, the price rises from \$12 to \$13, and Qs rises from 95 to 100:

$$\% \text{ change in quantity} \quad \frac{100-95}{(100+95)/2} \times 100$$

$$\frac{5}{97.5} \times 100$$

$$5.13$$

$$\% \text{ change in price} \quad \frac{\$13-\$12}{(\$13+\$12)/2} \times 100$$

$$\frac{1}{12.5} \times 100$$

$$8.0$$

$$\text{Elasticity of Supply} \quad \frac{5.13\%}{8.0\%}$$

$$0.64$$

The supply curve is inelastic in this region of the supply curve.

6.2 POLAR CASES OF ELASTICITY AND CONSTANT ELASTICITY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Differentiate between infinite and zero elasticity
- Analyze graphs in order to classify elasticity as constant unitary, infinite, or zero

There are two extreme cases of elasticity: when elasticity equals zero and when it is infinite. A third case is that of constant unitary elasticity. We will describe each case.

Infinite elasticity or **perfect elasticity** refers to the extreme case where either the quantity demanded (Qd) or supplied (Qs) changes by an infinite amount in response to any change in price at all. In both cases, the supply and the **demand curve** are horizontal as shown in Figure 1. While perfectly elastic supply curves are unrealistic, goods with readily available inputs and whose production can be easily expanded will feature highly elastic supply curves. Examples include pizza, bread, books and pencils. Similarly, perfectly elastic demand is an extreme example. But luxury goods, goods that take a large share of individuals' income, and goods with many substitutes are likely to have highly elastic demand curves. Examples of such goods are Caribbean cruises and sports vehicles.

Zero elasticity or **perfect inelasticity**, as depicted in Figure 2 refers to the extreme case in which a percentage change in price, no matter how large, results in zero change in quantity. While a perfectly inelastic supply is an extreme example, goods with limited supply of inputs are likely to feature highly inelastic supply curves. Examples include diamond rings or housing in prime locations such as apartments facing Central Park in New York City. Similarly, while perfectly inelastic demand is an extreme case, necessities with no close substitutes are likely to have highly inelastic demand curves. This is the case of life-saving drugs and gasoline.

Constant unitary elasticity, in either a supply or demand curve, occurs when a price change of one percent results in a quantity change of one percent. Figure 3 shows a demand curve with constant unit elasticity. As we move down the demand curve from A to B, the price falls by 33% and quantity demanded rises by 33%; as you move from B to C, the price falls by 25% and the quantity demanded rises by 25%; as you move from C to D, the price falls by 16% and the quantity rises by 16%. Notice that in absolute value, the declines in price, as you step down the demand curve, are not identical. Instead, the price falls by \$3 from A to B, by a smaller amount of \$1.50 from B to C, and by a still smaller amount of \$0.75 from C to D. As a result, a demand curve with constant unitary elasticity moves from a steeper slope on the left and a flatter slope on the right—and a curved shape overall.

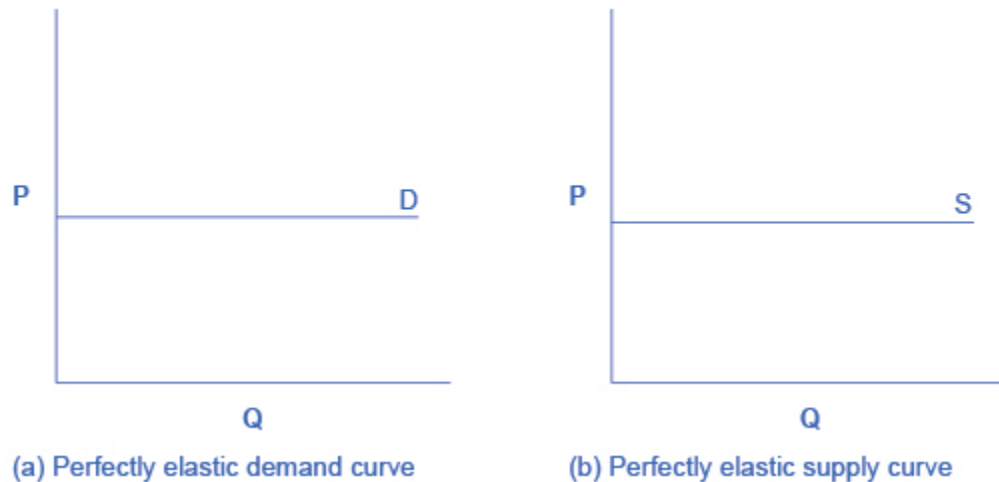


Figure 1. Infinite Elasticity. The horizontal lines show that an infinite quantity will be demanded or supplied at a specific price. This illustrates the cases of a perfectly (or infinitely) elastic demand curve and supply curve. The quantity supplied or demanded is extremely responsive to price changes, moving from zero for prices close to P to infinite when price reach P .

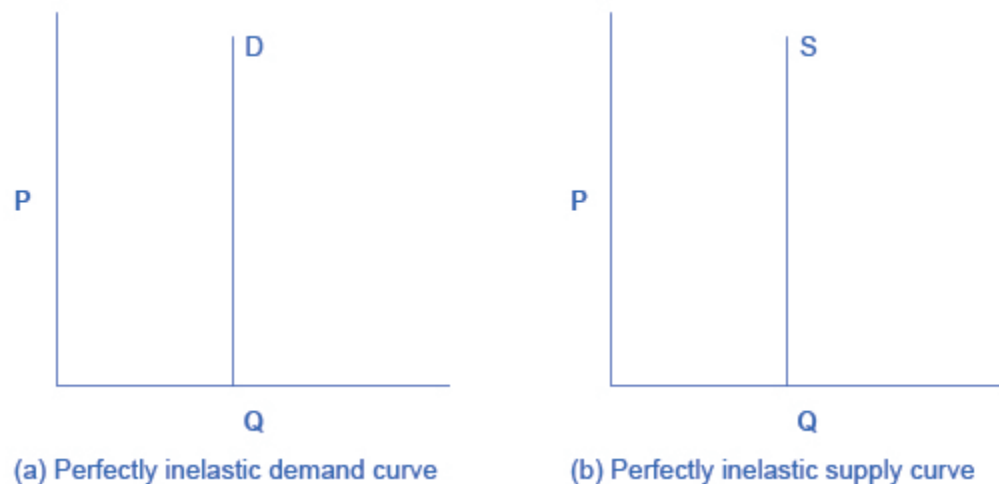


Figure 2. Zero Elasticity. The vertical supply curve and vertical demand curve show that there will be zero percentage change in quantity (a) demanded or (b) supplied, regardless of the price.

Unlike the demand curve with unitary elasticity, the supply curve with unitary elasticity is represented by a straight line. In moving up the supply curve from left to right, each increase in quantity of 30, from 90 to 120 to 150 to 180, is equal in absolute value. However, in percentage value, the steps are decreasing, from 33.3% to 25% to 16.7%, because the original quantity points in each percentage calculation are getting larger and larger, which expands the denominator in the elasticity calculation.

Consider the price changes moving up the supply curve in Figure 4. From points D to E to F and to G on the supply curve, each step of \$1.50 is the same in absolute value. However, if the price changes are measured in percentage change terms, they are also decreasing, from 33.3% to 25% to 16.7%, because the original price points in each percentage calculation are getting larger and larger in value. Along

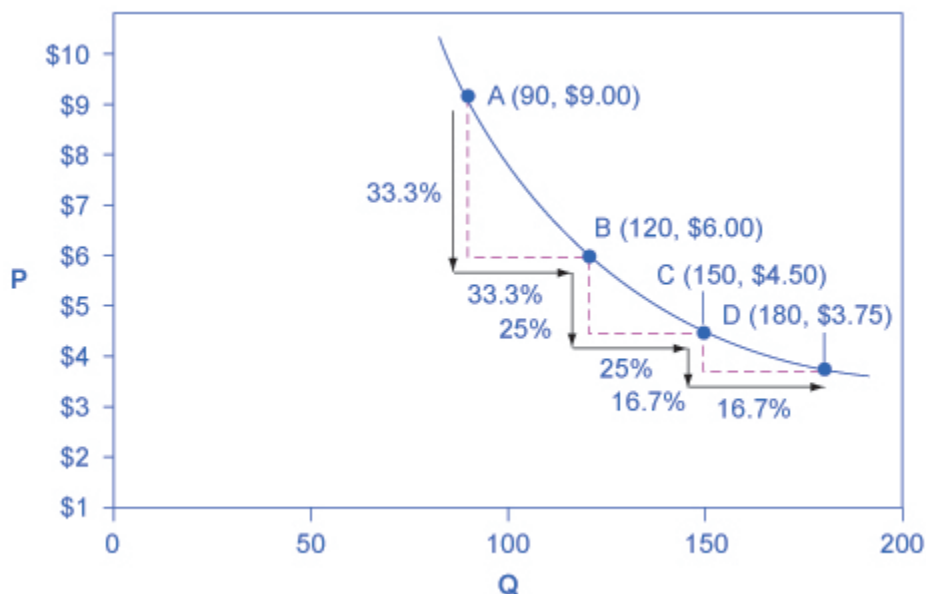


Figure 3. A Constant Unitary Elasticity Demand Curve. A demand curve with constant unitary elasticity will be a curved line. Notice how price and quantity demanded change by an identical amount in each step down the demand curve.

the constant unitary elasticity supply curve, the percentage quantity increases on the horizontal axis exactly match the percentage price increases on the vertical axis—so this supply curve has a constant unitary elasticity at all points.

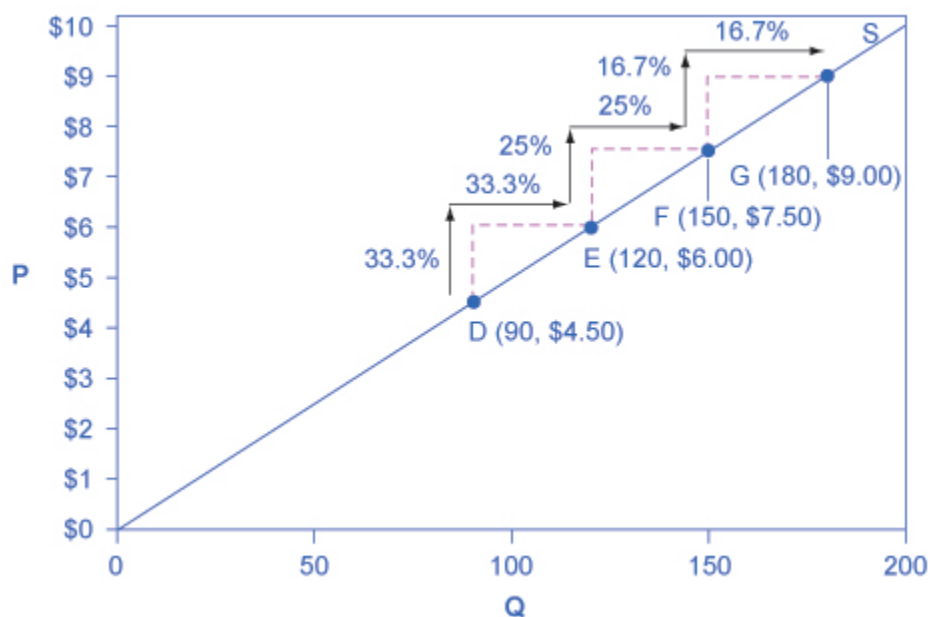


Figure 4. A Constant Unitary Elasticity Supply Curve. A constant unitary elasticity supply curve is a straight line reaching up from the origin. Between each point, the percentage increase in quantity supplied is the same as the percentage increase in price.

KEY CONCEPTS AND SUMMARY

Infinite or perfect elasticity refers to the extreme case where either the quantity demanded or supplied changes by an infinite amount in response to any change in price at all. Zero elasticity refers to the extreme case in which a percentage change in price, no matter how large, results in zero change in quantity. Constant unitary elasticity in either a supply or demand curve refers to a situation where a price change of one percent results in a quantity change of one percent.

SELF-CHECK QUESTIONS

1. Why is the demand curve with constant unitary elasticity concave?
2. Why is the supply curve with constant unitary elasticity a straight line?

REVIEW QUESTIONS

1. Describe the general appearance of a demand or a supply curve with zero elasticity.
2. Describe the general appearance of a demand or a supply curve with infinite elasticity.

CRITICAL THINKING QUESTIONS

Can you think of an industry (or product) with near infinite elasticity of supply in the short term? That is, what is an industry that could increase Q_s almost without limit in response to an increase in the price?

PROBLEMS

1. The supply of paintings by Leonardo Da Vinci, who painted the *Mona Lisa* and *The Last Supper* and died in 1519, is highly inelastic. Sketch a supply and demand diagram, paying attention to the appropriate elasticities, to illustrate that demand for these paintings will determine the price.
2. Say that a certain stadium for professional football has 70,000 seats. What is the shape of the supply curve for tickets to football games at that stadium? Explain.
3. When someone's kidneys fail, the person needs to have medical treatment with a dialysis machine (unless or until they receive a kidney transplant) or they will die. Sketch a supply and demand diagram, paying attention to the appropriate elasticities, to illustrate that the supply of such dialysis machines will primarily determine the price.

GLOSSARY

constant unitary elasticity when a given percent price change in price leads to an equal percentage change in quantity demanded or supplied

infinite elasticity the extremely elastic situation of demand or supply where quantity changes by an infinite amount in response to any change in price; horizontal in appearance

perfect elasticity see infinite elasticity

zero inelasticity the highly inelastic case of demand or supply in which a percentage change in price, no matter how large, results in zero change in the quantity; vertical in appearance

perfect inelasticity see zero elasticity

SOLUTIONS

Answers to Self-Check Questions

1. The demand curve with constant unitary elasticity is concave because at high prices, a one percent decrease in price results in more than a one percent increase in quantity. As we move down the demand curve, price drops and the one percent decrease in price causes less than a one percent increase in quantity.
2. The constant unitary elasticity is a straight line because the curve slopes upward and both price and quantity are increasing proportionally.

6.3 ELASTICITY AND PRICING

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Analyze how price elasticities impact revenue
- Evaluate how elasticity can cause shifts in demand and supply
- Predict how the long-run and short-run impacts of elasticity affect equilibrium
- Explain how the elasticity of demand and supply determine the incidence of a tax on buyers and sellers

Studying elasticities is useful for a number of reasons, pricing being most important. Let's explore how elasticity relates to revenue and pricing, both in the long run and short run. But first, let's look at the elasticities of some common goods and services.

Table 4 shows a selection of demand elasticities for different goods and services drawn from a variety of different studies by economists, listed in order of increasing elasticity.

Goods and Services	Elasticity of Price
Housing	0.12
Transatlantic air travel (economy class)	0.12
Rail transit (rush hour)	0.15
Electricity	0.20
Taxi cabs	0.22
Gasoline	0.35
Transatlantic air travel (first class)	0.40
Wine	0.55
Beef	0.59
Transatlantic air travel (business class)	0.62
Kitchen and household appliances	0.63
Cable TV (basic rural)	0.69
Chicken	0.64
Soft drinks	0.70
Beer	0.80
New vehicle	0.87
Rail transit (off-peak)	1.00
Computer	1.44
Cable TV (basic urban)	1.51
Cable TV (premium)	1.77
Restaurant meals	2.27

Table 4. Some Selected Elasticities of Demand

Note that necessities such as housing and electricity are inelastic, while items that are not necessities such as restaurant meals are more price-sensitive. If the price of the restaurant meal increases by 10%, the quantity demanded will decrease by 22.7%. A 10% increase in the price of housing will cause a slight decrease of 1.2% in the quantity of housing demanded.

Read this article for an example of price elasticity that may have affected you.



DOES RAISING PRICE BRING IN MORE REVENUE?

Imagine that a band on tour is playing in an indoor arena with 15,000 seats. To keep this example simple, assume that the band keeps all the money from ticket sales. Assume further that the band pays the costs for its appearance, but that these costs, like travel, setting up the stage, and so on, are the same regardless of how many people are in the audience. Finally, assume that all the tickets have the same price. (The same insights apply if ticket prices are more expensive for some seats than for others, but the calculations become more complicated.) The band knows that it faces a downward-sloping demand curve; that is, if the band raises the price of tickets, it will sell fewer tickets. How should the band set the price for tickets to bring in the most total revenue, which in this example, because costs are fixed, will also mean the highest profits for the band? Should the band sell more tickets at a lower price or fewer tickets at a higher price?

The key concept in thinking about collecting the most revenue is the price elasticity of demand. Total revenue is price times the quantity of tickets sold. Imagine that the band starts off thinking about a certain price, which will result in the sale of a certain quantity of tickets. The three possibilities are laid out in Table 5. If demand is elastic at that price level, then the band should cut the price, because the percentage drop in price will result in an even larger percentage increase in the quantity sold—thus raising total revenue. However, if demand is inelastic at that original quantity level, then the band should raise the price of tickets, because a certain percentage increase in price will result in a smaller percentage decrease in the quantity sold—and total revenue will rise. If demand has a unitary elasticity at that quantity, then a moderate percentage change in the price will be offset by an equal percentage change in quantity—so the band will earn the same revenue whether it (moderately) increases or decreases the price of tickets.

If Demand Is ...	Then ...	Therefore ...
Elastic	% change in Qd > % change in P	A given % rise in P will be more than offset by a larger % fall in Q so that total revenue ($P \times Q$) falls.
Unitary	% change in Qd = % change in P	A given % rise in P will be exactly offset by an equal % fall in Q so that total revenue ($P \times Q$) is unchanged.
Inelastic	% change in Qd < % change in P	A given % rise in P will cause a smaller % fall in Q so that total revenue ($P \times Q$) rises.

Table 5. Will the Band Earn More Revenue by Changing Ticket Prices?

What if the band keeps cutting price, because demand is elastic, until it reaches a level where all 15,000 seats in the available arena are sold? If demand remains elastic at that quantity, the band might try to move to a bigger arena, so that it could cut ticket prices further and see a larger percentage increase in the quantity of tickets sold. Of course, if the 15,000-seat arena is all that is available or if a larger arena would add substantially to costs, then this option may not work.

Conversely, a few bands are so famous, or have such fanatical followings, that demand for tickets may be inelastic right up to the point where the arena is full. These bands can, if they wish, keep raising the price of tickets. Ironically, some of the most popular bands could make more revenue by setting prices so high that the arena is not filled—but those who buy the tickets would have to pay very high prices. However, bands sometimes choose to sell tickets for less than the absolute maximum they might be able to charge, often in the hope that fans will feel happier and spend more on recordings, T-shirts, and other paraphernalia.

CAN COSTS BE PASSED ON TO CONSUMERS?

Most businesses face a day-to-day struggle to figure out ways to produce at a lower cost, as one pathway to their goal of earning higher profits. However, in some cases, the price of a key input over which the firm has no control may rise. For example, many chemical companies use petroleum as a **key input**, but they have no control over the world market price for crude oil. Coffee shops use coffee as a key input, but they have no control over the world market price of coffee. If the cost of a key input rises, can the firm pass those higher costs along to consumers in the form of higher prices? Conversely, if new and less expensive ways of producing are invented, can the firm keep the benefits in the form of higher profits, or will the market pressure them to pass the gains along to consumers in the form of lower prices? The price elasticity of demand plays a key role in answering these questions.

Imagine that as a consumer of legal pharmaceutical products, you read a newspaper story that a technological breakthrough in the production of aspirin has occurred, so that every aspirin factory can now make aspirin more cheaply than it did before. What does this discovery mean to you? Figure 1 illustrates two possibilities. In Figure 1 (a), the demand curve is drawn as highly inelastic. In this case, a technological breakthrough that shifts supply to the right, from S_0 to S_1 , so that the equilibrium shifts from E_0 to E_1 , creates a substantially lower price for the product with relatively little impact on the quantity sold. In Figure 1 (b), the demand curve is drawn as highly elastic. In this case, the technological breakthrough leads to a much greater quantity being sold in the market at very close to the original price. Consumers benefit more, in general, when the demand curve is more inelastic because the shift in the supply results in a much lower price for consumers.

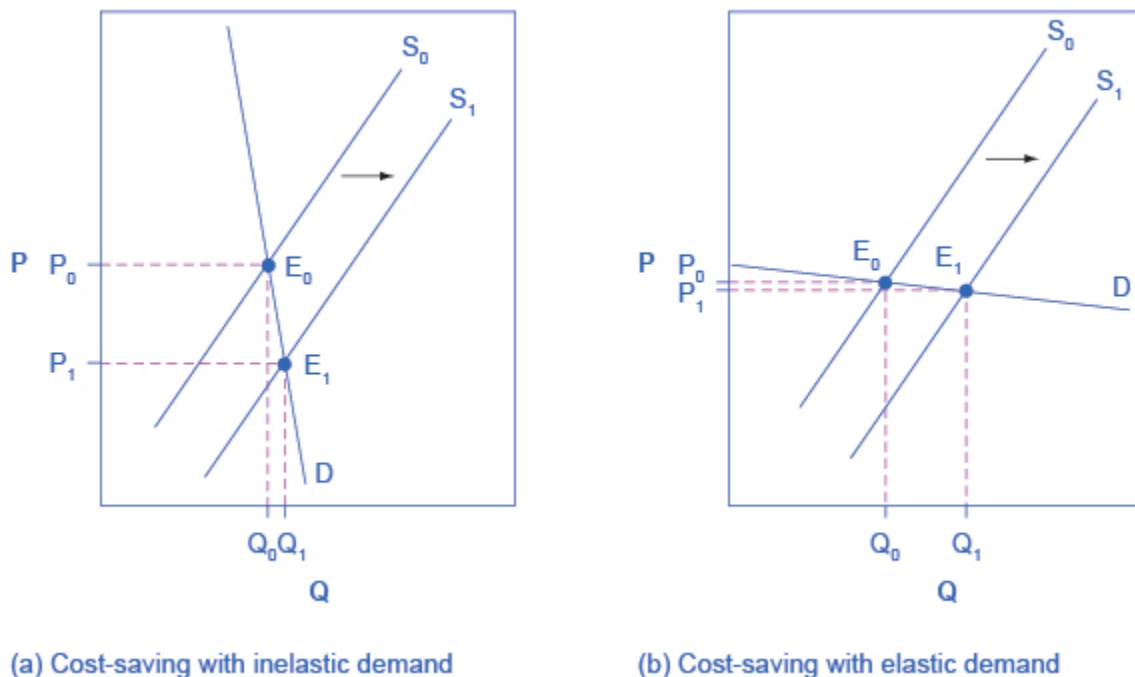


Figure 1. Passing along Cost Savings to Consumers. Cost-saving gains cause supply to shift out to the right from S_0 to S_1 ; that is, at any given price, firms will be willing to supply a greater quantity.

If demand is inelastic, as in (a), the result of this cost-saving technological improvement will be substantially lower prices. If demand is elastic, as in (b), the result will be only slightly lower prices. Consumers benefit in either case, from a greater quantity at a lower price, but the benefit is greater when demand is inelastic, as in (a).

Producers of aspirin may find themselves in a nasty bind here. The situation shown in Figure 1, with extremely inelastic demand, means that a new invention may cause the price to drop dramatically while quantity changes little. As a result, the new production technology can lead to a drop in the revenue that firms earn from sales of aspirin. However, if strong competition exists between producers of aspirin, each producer may have little choice but to search for and implement any breakthrough that allows it to reduce production costs. After all, if one firm decides not to implement such a cost-saving technology, it can be driven out of business by other firms that do.

Since demand for food is generally inelastic, farmers may often face the situation in Figure 1 (a). That is, a surge in production leads to a severe drop in price that can actually decrease the total revenue received by farmers. Conversely, poor weather or other conditions that cause a terrible year for farm production can sharply raise prices so that the total revenue received increases. The Clear It Up box discusses how these issues relate to coffee.

HOW DO COFFEE PRICES FLUCTUATE?

Coffee is an international crop. The top five coffee-exporting nations are Brazil, Vietnam, Colombia, Indonesia, and Ethiopia. In these nations and others, 20 million families depend on selling coffee beans as their main source of income. These families are exposed to enormous risk, because the world price of coffee bounces up and down. For example, in 1993, the world price of coffee was about 50 cents per pound; in 1995 it was four times as high, at \$2 per pound. By 1997 it had fallen by half to \$1.00 per pound. In 1998 it leaped back up to \$2 per pound. By 2001 it had fallen back to 46 cents a pound; by early 2011 it went back up to about \$2.31 per pound. By the end of 2012, the price had fallen back to about \$1.31 per pound.

The reason for these price bounces lies in a combination of inelastic demand and shifts in supply. The elasticity of coffee demand is only about 0.3; that is, a 10% rise in the price of coffee leads to a decline of about 3% in the quantity of coffee consumed. When a major frost hit the Brazilian coffee crop in 1994, coffee supply shifted to the left with an inelastic demand curve, leading to much higher prices. Conversely, when Vietnam entered the world coffee market as a major producer in the late 1990s, the supply curve shifted out to the right. With a highly inelastic demand curve, coffee prices fell dramatically. This situation is shown in Figure 1 (a).

Elasticity also reveals whether firms can pass higher costs that they incur on to consumers. Addictive substances tend to fall into this category. For example, the demand for cigarettes is relatively inelastic among regular smokers who are somewhat addicted; economic research suggests that increasing the price of cigarettes by 10% leads to about a 3% reduction in the quantity of cigarettes smoked by adults, so the elasticity of demand for cigarettes is 0.3. If society increases taxes on companies that make cigarettes, the result will be, as in Figure 2 (a), that the supply curve shifts from S_0 to S_1 . However, as the equilibrium moves from E_0 to E_1 , these taxes are mainly passed along to consumers in the form of higher prices. These higher taxes on cigarettes will raise tax revenue for the government, but they will not much affect the quantity of smoking.

If the goal is to reduce the quantity of cigarettes demanded, it must be achieved by shifting this inelastic demand back to the left, perhaps with public programs to discourage the use of cigarettes or to help people to quit. For example, anti-smoking advertising campaigns have shown some ability to reduce smoking. However, if demand for cigarettes was more elastic, as in Figure 2 (b), then an increase in taxes that shifts supply from S_0 to S_1 and equilibrium from E_0 to E_1 would reduce the quantity of cigarettes smoked substantially. Youth smoking seems to be more elastic than adult smoking—that is,

the quantity of youth smoking will fall by a greater percentage than the quantity of adult smoking in response to a given percentage increase in price.

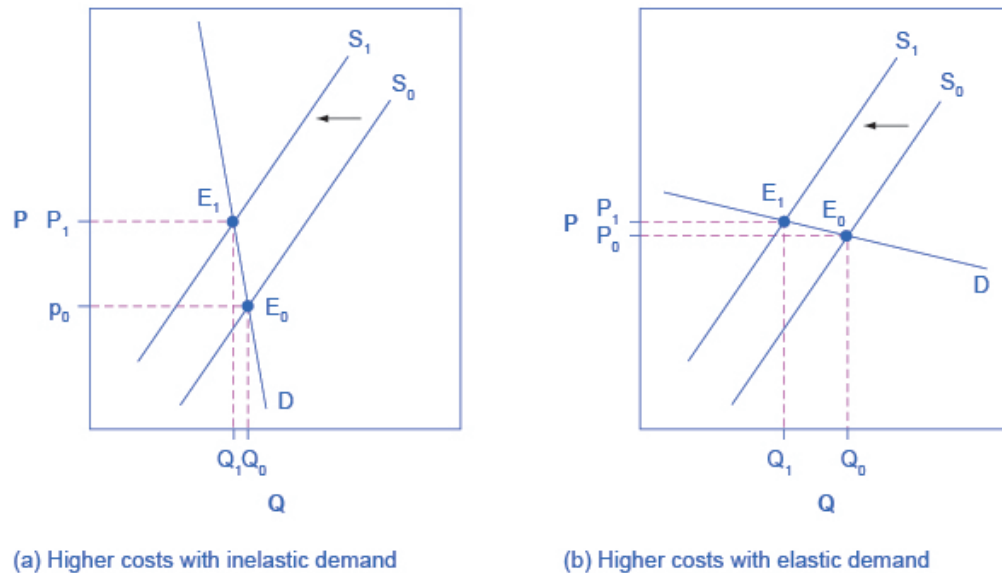


Figure 2. Passing along Higher Costs to Consumers. Higher costs, like a higher tax on cigarette companies for the example given in the text, lead supply to shift to the left. This shift is identical in (a) and (b). However, in (a), where demand is inelastic, the cost increase can largely be passed along to consumers in the form of higher prices, without much of a decline in equilibrium quantity. In (b), demand is elastic, so the shift in supply results primarily in a lower equilibrium quantity. Consumers suffer in either case, but in (a), they suffer from paying a higher price for the same quantity, while in (b), they suffer from buying a lower quantity (and presumably needing to shift their consumption elsewhere).

ELASTICITY AND TAX INCIDENCE

The example of cigarette taxes showed that because demand is inelastic, taxes are not effective at reducing the equilibrium quantity of smoking, and they are mainly passed along to consumers in the form of higher prices. The analysis, or manner, of how the burden of a tax is divided between consumers and producers is called **tax incidence**. Typically, the incidence, or burden, of a tax falls both on the consumers and producers of the taxed good. But if one wants to predict which group will bear most of the burden, all one needs to do is examine the elasticity of demand and supply. In the tobacco example, the tax burden falls on the most inelastic side of the market.

If demand is more inelastic than supply, consumers bear most of the tax burden, and if supply is more inelastic than demand, sellers bear most of the tax burden.

The intuition for this is simple. When the demand is inelastic, consumers are not very responsive to price changes, and the quantity demanded remains relatively constant when the tax is introduced. In the case of smoking, the demand is inelastic because consumers are addicted to the product. The government can then pass the tax burden along to consumers in the form of higher prices, without much of a decline in the equilibrium quantity.

Similarly, when a tax is introduced in a market with an inelastic supply, such as, for example, beach-front hotels, and sellers have no alternative than to accept lower prices for their business, taxes do not

greatly affect the equilibrium quantity. The tax burden is now passed on to the sellers. If the supply was elastic and sellers had the possibility of reorganizing their businesses to avoid supplying the taxed good, the tax burden on the sellers would be much smaller. The tax would result in a much lower quantity sold instead of lower prices received. Figure 3 illustrates this relationship between the tax incidence and elasticity of demand and supply.

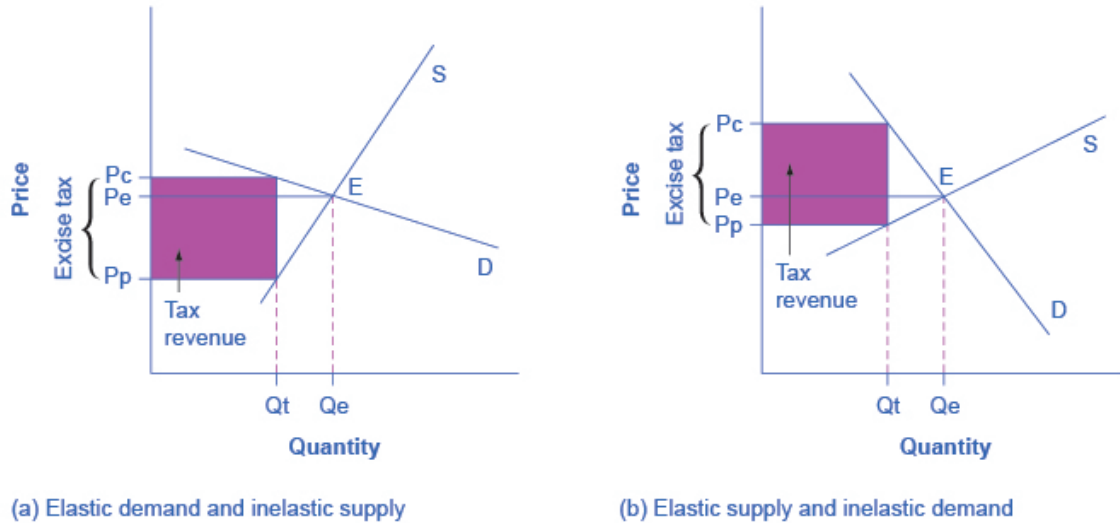


Figure 3. Elasticity and Tax Incidence. An excise tax introduces a wedge between the price paid by consumers (P_c) and the price received by producers (P_p). (a) When the demand is more elastic than supply, the tax incidence on consumers $P_c - P_e$ is lower than the tax incidence on producers $P_e - P_p$. (b) When the supply is more elastic than demand, the tax incidence on consumers $P_c - P_e$ is larger than the tax incidence on producers $P_e - P_p$. The more elastic the demand and supply curves are, the lower the tax revenue.

In Figure 3 (a), the supply is inelastic and the demand is elastic, such as in the example of beachfront hotels. While consumers may have other vacation choices, sellers can't easily move their businesses. By introducing a tax, the government essentially creates a wedge between the price paid by consumers P_c and the price received by producers P_p . In other words, of the total price paid by consumers, part is retained by the sellers and part is paid to the government in the form of a tax. The distance between P_c and P_p is the tax rate. The new market price is P_c , but sellers receive only P_p per unit sold, as they pay $P_c - P_p$ to the government. Since a tax can be viewed as raising the costs of production, this could also be represented by a leftward shift of the supply curve, where the new supply curve would intercept the demand at the new quantity Q_t . For simplicity, Figure 3 omits the shift in the supply curve.

The tax revenue is given by the shaded area, which is obtained by multiplying the tax per unit by the total quantity sold Q_t . The tax incidence on the consumers is given by the difference between the price paid P_c and the initial equilibrium price P_e . The tax incidence on the sellers is given by the difference between the initial equilibrium price P_e and the price they receive after the tax is introduced P_p . In Figure 3 (a), the tax burden falls disproportionately on the sellers, and a larger proportion of the tax revenue (the shaded area) is due to the resulting lower price received by the sellers than by the resulting higher prices paid by the buyers. The example of the tobacco excise tax could be described by Figure 3 (b) where the supply is more elastic than demand. The tax incidence now falls disproportionately on consumers, as shown by the large difference between the price they pay, P_c , and the initial equilibrium price, P_e . Sellers receive a lower price than before the tax, but this difference is

much smaller than the change in consumers' price. From this analysis one can also predict whether a tax is likely to create a large revenue or not. The more elastic the demand curve, the easier it is for consumers to reduce quantity instead of paying higher prices. The more elastic the supply curve, the easier it is for sellers to reduce the quantity sold, instead of taking lower prices. In a market where both the demand and supply are very elastic, the imposition of an excise tax generates low revenue.

Excise taxes tend to be thought to hurt mainly the specific industries they target. For example, the medical device excise tax, in effect since 2013, has been controversial for it can delay industry profitability and therefore hamper start-ups and medical innovation. But ultimately, whether the tax burden falls mostly on the medical device industry or on the patients depends simply on the elasticity of demand and supply.

LONG-RUN VS. SHORT-RUN IMPACT

Elasticities are often lower in the short run than in the long run. On the demand side of the market, it can sometimes be difficult to change Q_d in the short run, but easier in the long run. Consumption of energy is a clear example. In the short run, it is not easy for a person to make substantial changes in the energy consumption. Maybe you can carpool to work sometimes or adjust your home thermostat by a few degrees if the cost of energy rises, but that is about all. However, in the long-run you can purchase a car that gets more miles to the gallon, choose a job that is closer to where you live, buy more energy-efficient home appliances, or install more insulation in your home. As a result, the elasticity of demand for energy is somewhat inelastic in the short run, but much more elastic in the long run.

Figure 4 is an example, based roughly on historical experience, for the responsiveness of Q_d to price changes. In 1973, the price of crude oil was \$12 per barrel and total consumption in the U.S. economy was 17 million barrels per day. That year, the nations who were members of the Organization of Petroleum Exporting Countries (OPEC) cut off oil exports to the United States for six months because the Arab members of OPEC disagreed with the U.S. support for Israel. OPEC did not bring exports back to their earlier levels until 1975—a policy that can be interpreted as a shift of the supply curve to the left in the U.S. petroleum market. Figure 4 (a) and Figure 4 (b) show the same original equilibrium point and the same identical shift of a supply curve to the left from S_0 to S_1 .

Figure 4 (a) shows inelastic demand for oil in the short run similar to that which existed for the United States in 1973. In Figure 4 (a), the new equilibrium (E_1) occurs at a price of \$25 per barrel, roughly double the price before the OPEC shock, and an equilibrium quantity of 16 million barrels per day. Figure 4 (b) shows what the outcome would have been if the U.S. demand for oil had been more elastic, a result more likely over the long term. This alternative equilibrium (E_1) would have resulted in a smaller price increase to \$14 per barrel and larger reduction in equilibrium quantity to 13 million barrels per day. In 1983, for example, U.S. petroleum consumption was 15.3 million barrels a day, which was lower than in 1973 or 1975. U.S. petroleum consumption was down even though the U.S. economy was about one-fourth larger in 1983 than it had been in 1973. The primary reason for the lower quantity was that higher energy prices spurred conservation efforts, and after a decade of home insulation, more fuel-efficient cars, more efficient appliances and machinery, and other fuel-conserving choices, the demand curve for energy had become more elastic.

On the supply side of markets, producers of goods and services typically find it easier to expand production in the long term of several years rather than in the short run of a few months. After all, in

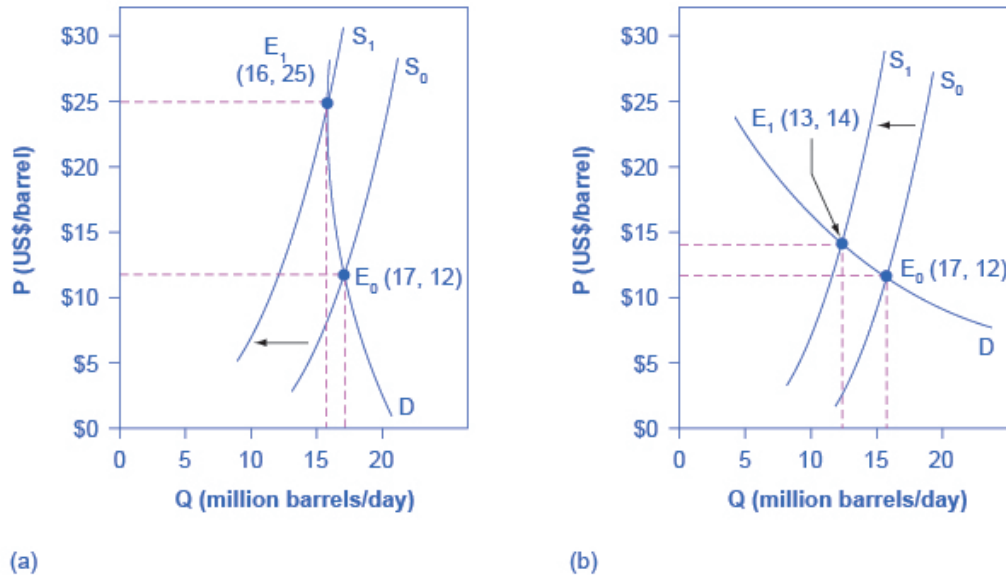


Figure 4. How a Shift in Supply Can Affect Price or Quantity. The intersection (E_0) between demand curve D and supply curve S_0 is the same in both (a) and (b). The shift of supply to the left from S_0 to S_1 is identical in both (a) and (b). The new equilibrium (E_1) has a higher price and a lower quantity than the original equilibrium (E_0) in both (a) and (b). However, the shape of the demand curve D is different in (a) and (b). As a result, the shift in supply can result either in a new equilibrium with a much higher price and an only slightly smaller quantity, as in (a), or in a new equilibrium with only a small increase in price and a relatively larger reduction in quantity, as in (b).

the short run it can be costly or difficult to build a new factory, hire many new workers, or open new stores. But over a few years, all of these are possible.

Indeed, in most markets for goods and services, prices bounce up and down more than quantities in the short run, but quantities often move more than prices in the long run. The underlying reason for this pattern is that supply and demand are often inelastic in the short run, so that shifts in either demand or supply can cause a relatively greater change in prices. But since supply and demand are more elastic in the long run, the long-run movements in prices are more muted, while quantity adjusts more easily in the long run.

KEY CONCEPTS AND SUMMARY

In the market for goods and services, quantity supplied and quantity demanded are often relatively slow to react to changes in price in the short run, but react more substantially in the long run. As a result, demand and supply often (but not always) tend to be relatively inelastic in the short run and relatively elastic in the long run. The tax incidence depends on the relative price elasticity of supply and demand. When supply is more elastic than demand, buyers bear most of the tax burden, and when demand is more elastic than supply, producers bear most of the cost of the tax. Tax revenue is larger the more inelastic the demand and supply are.

SELF-CHECK QUESTIONS

1. The federal government decides to require that automobile manufacturers install new anti-pollution equipment that costs \$2,000 per car. Under what conditions can carmakers pass almost all of this cost along to car buyers? Under what conditions can carmakers pass very little of this cost along to car buyers?
2. Suppose you are in charge of sales at a pharmaceutical company, and your firm has a new drug that causes bald men to grow hair. Assume that the company wants to earn as much revenue as possible from this drug. If the elasticity of demand for your company's product at the current price is 1.4, would you advise the company to raise the price, lower the price, or to keep the price the same? What if the elasticity were 0.6? What if it were 1? Explain your answer.

REVIEW QUESTIONS

1. If demand is elastic, will shifts in supply have a larger effect on equilibrium quantity or on price?
2. If demand is inelastic, will shifts in supply have a larger effect on equilibrium price or on quantity?
3. If supply is elastic, will shifts in demand have a larger effect on equilibrium quantity or on price?
4. If supply is inelastic, will shifts in demand have a larger effect on equilibrium price or on quantity?
5. Would you usually expect elasticity of demand or supply to be higher in the short run or in the long run? Why?
6. Under which circumstances does the tax burden fall entirely on consumers?

CRITICAL THINKING QUESTIONS

1. Would you expect supply to play a more significant role in determining the price of a basic necessity like food or a luxury like perfume? Explain. *Hint:* Think about how the price elasticity of demand will differ between necessities and luxuries.
2. A city has built a bridge over a river and it decides to charge a toll to everyone who crosses. For one year, the city charges a variety of different tolls and records information on how many drivers cross the bridge. The city thus gathers information about elasticity of demand. If the city wishes to raise as much revenue as possible from the tolls, where will the city decide to charge a toll: in the inelastic portion of the demand curve, the elastic portion of the demand curve, or the unit elastic portion? Explain.
3. In a market where the supply curve is perfectly inelastic, how does an excise tax affect the price paid by consumers and the quantity bought and sold?

PROBLEMS

Assume that the supply of low-skilled workers is fairly elastic, but the employers' demand for such workers is fairly inelastic. If the policy goal is to expand employment for low-skilled workers, is it better to focus on policy tools to shift

the supply of unskilled labor or on tools to shift the demand for unskilled labor? What if the policy goal is to raise wages for this group? Explain your answers with supply and demand diagrams.

GLOSSARY

tax incidence manner in which the tax burden is divided between buyers and sellers

SOLUTIONS

Answers to Self-Check Questions

1. Carmakers can pass this cost along to consumers if the demand for these cars is inelastic. If the demand for these cars is elastic, then the manufacturer must pay for the equipment.
2. If the elasticity is 1.4 at current prices, you would advise the company to lower its price on the product, since a decrease in price will be offset by the increase in the amount of the drug sold. If the elasticity were 0.6, then you would advise the company to increase its price. Increases in price will offset the decrease in number of units sold, but increase your total revenue. If elasticity is 1, the total revenue is already maximized, and you would advise that the company maintain its current price level.

6.4 ELASTICITY IN AREAS OTHER THAN PRICE

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Calculate the income elasticity of demand and the cross-price elasticity of demand
- Calculate the elasticity in labor and financial capital markets through an understanding of the elasticity of labor supply and the elasticity of savings
- Apply concepts of price elasticity to real-world situations

The basic idea of elasticity—how a percentage change in one variable causes a percentage change in another variable—does not just apply to the responsiveness of supply and demand to changes in the price of a product. Recall that quantity demanded (Q_d) depends on income, tastes and preferences, the prices of related goods, and so on, as well as price. Similarly, quantity supplied (Q_s) depends on the cost of production, and so on, as well as price. Elasticity can be measured for any determinant of supply and demand, not just the price.

INCOME ELASTICITY OF DEMAND

The **income elasticity of demand** is the percentage change in quantity demanded divided by the percentage change in income.

$$\text{Income elasticity of demand} = \frac{\% \text{ change in quantity demanded}}{\% \text{ change in income}}$$

For most products, most of the time, the income elasticity of demand is positive: that is, a rise in income will cause an increase in the quantity demanded. This pattern is common enough that these goods are referred to as **normal goods**. However, for a few goods, an increase in income means that one might purchase less of the good; for example, those with a higher income might buy fewer hamburgers, because they are buying more steak instead, or those with a higher income might buy less cheap wine and more imported beer. When the income elasticity of demand is negative, the good is called an **inferior good**.

The concepts of normal and inferior goods were introduced in Demand and Supply. A higher level of income for a normal good causes a demand curve to shift to the right for a normal good, which means that the income elasticity of demand is positive. How far the demand shifts depends on the income

elasticity of demand. A higher income elasticity means a larger shift. However, for an inferior good, that is, when the income elasticity of demand is negative, a higher level of income would cause the demand curve for that good to shift to the left. Again, how much it shifts depends on how large the (negative) income elasticity is.

CROSS-PRICE ELASTICITY OF DEMAND

A change in the price of one good can shift the quantity demanded for another good. If the two goods are complements, like bread and peanut butter, then a drop in the price of one good will lead to an increase in the quantity demanded of the other good. However, if the two goods are substitutes, like plane tickets and train tickets, then a drop in the price of one good will cause people to substitute toward that good, and to reduce consumption of the other good. Cheaper plane tickets lead to fewer train tickets, and vice versa.

The **cross-price elasticity of demand** puts some meat on the bones of these ideas. The term “cross-price” refers to the idea that the price of one good is affecting the quantity demanded of a different good. Specifically, the cross-price elasticity of demand is the percentage change in the quantity of good A that is demanded as a result of a percentage change in the price of good B.

$$\text{Cross} - \text{price elasticity of demand} = \frac{\% \text{ change in } Q_d \text{ of good } A}{\% \text{ change in price of good } B}$$

Substitute goods have positive cross-price elasticities of demand: if good A is a substitute for good B, like coffee and tea, then a higher price for B will mean a greater quantity consumed of A. Complement goods have negative cross-price elasticities: if good A is a complement for good B, like coffee and sugar, then a higher price for B will mean a lower quantity consumed of A.

ELASTICITY IN LABOR AND FINANCIAL CAPITAL MARKETS

The concept of elasticity applies to any market, not just markets for goods and services. In the labor market, for example, the **wage elasticity of labor supply**—that is, the percentage change in hours worked divided by the percentage change in wages—will determine the shape of the labor supply curve. Specifically:

$$\text{Elasticity of labor supply} = \frac{\% \text{ change in quantity of labor supplied}}{\% \text{ change in wage}}$$

The wage elasticity of labor supply for teenage workers is generally thought to be fairly elastic: that is, a certain percentage change in wages will lead to a larger percentage change in the quantity of hours worked. Conversely, the wage elasticity of labor supply for adult workers in their thirties and forties is thought to be fairly inelastic. When wages move up or down by a certain percentage amount, the quantity of hours that adults in their prime earning years are willing to supply changes but by a lesser percentage amount.

In markets for financial capital, the **elasticity of savings**—that is, the percentage change in the quantity of savings divided by the percentage change in interest rates—will describe the shape of the supply curve for financial capital. That is:

$$\text{Elasticity of savings} = \frac{\% \text{ change in quantity of financial savings}}{\% \text{ change in interest rate}}$$

Sometimes laws are proposed that seek to increase the quantity of savings by offering tax breaks so that the return on savings is higher. Such a policy will increase the quantity if the supply curve for financial capital is elastic, because then a given percentage increase in the return to savings will cause a higher percentage increase in the quantity of savings. However, if the supply curve for financial capital is highly inelastic, then a percentage increase in the return to savings will cause only a small increase in the quantity of savings. The evidence on the supply curve of financial capital is controversial but, at least in the short run, the elasticity of savings with respect to the interest rate appears fairly inelastic.

EXPANDING THE CONCEPT OF ELASTICITY

The elasticity concept does not even need to relate to a typical supply or demand curve at all. For example, imagine that you are studying whether the Internal Revenue Service should spend more money on auditing tax returns. The question can be framed in terms of the elasticity of tax collections with respect to spending on tax enforcement; that is, what is the percentage change in tax collections derived from a percentage change in spending on tax enforcement?

With all of the elasticity concepts that have just been described, some of which are listed in Table 6, the possibility of confusion arises. When you hear the phrases “elasticity of demand” or “elasticity of supply,” they refer to the elasticity with respect to price. Sometimes, either to be extremely clear or because a wide variety of elasticities are being discussed, the elasticity of demand or the demand elasticity will be called the price elasticity of demand or the “elasticity of demand with respect to price.” Similarly, elasticity of supply or the supply elasticity is sometimes called, to avoid any possibility of confusion, the price elasticity of supply or “the elasticity of supply with respect to price.” But in whatever context elasticity is invoked, the idea always refers to percentage change in one variable, almost always a price or money variable, and how it causes a percentage change in another variable, typically a quantity variable of some kind.

$$\text{Income elasticity of demand} = \frac{\% \text{ change in } Q_d}{\% \text{ change in income}}$$

$$\text{Cross – price elasticity of demand} = \frac{\% \text{ change in } Q_d \text{ of good } A}{\% \text{ change in price of good } B}$$

$$\text{Wage elasticity of labor supply} = \frac{\% \text{ change in quantity of labor supplied}}{\% \text{ change in wage}}$$

$$\text{Wage elasticity of labor demand} = \frac{\% \text{ change in quantity of labor demanded}}{\% \text{ change in wage}}$$

$$\text{Interest rate elasticity of savings} = \frac{\% \text{ change in quantity of savings}}{\% \text{ change in interest rate}}$$

$$\text{Interest rate elasticity of borrowing} = \frac{\% \text{ change in quantity of borrowing}}{\% \text{ change in interest rate}}$$

Table 6. Formulas for Calculating Elasticity

THAT WILL BE HOW MUCH?

How did the 60% price increase in 2011 end up for Netflix? It has been a very bumpy ride.

Before the price increase, there were about 24.6 million U.S. subscribers. After the price increase, 810,000 infuriated U.S. consumers canceled their Netflix subscriptions, dropping the total number of subscribers to 23.79 million. Fast forward to June 2013, when there were 36 million streaming Netflix subscribers in the United States. This was an increase of 11.4 million subscribers since the price increase—an average per quarter growth of about 1.6 million. This growth is less than the 2 million per quarter increases Netflix experienced in the fourth quarter of 2010 and the first quarter of 2011.

During the first year after the price increase, the firm's stock price (a measure of future expectations for the firm) fell from about \$300 per share to just under \$54. In 2015, however, the stock price is at \$448 per share. Today, Netflix has 57 million subscribers in fifty countries.

What happened? Obviously, Netflix company officials understood the law of demand. Company officials reported, when announcing the price increase, this could result in the loss of about 600,000 existing subscribers. Using the elasticity of demand formula, it is easy to see company officials expected an inelastic response:

$$\frac{-600,000/[(24 \text{ million} + 24.6 \text{ million})/2]}{\$6/[(\$10 + \$16)/2]}$$

$$\frac{-600,000/24.3 \text{ million}}{\$6/\$13}$$

$$\frac{-0.025}{0.46}$$

$$-0.05$$

In addition, Netflix officials had anticipated the price increase would have little impact on attracting new customers. Netflix anticipated adding up to 1.29 million new subscribers in the third quarter of 2011. It is true this was slower growth than the firm had experienced—about 2 million per quarter.

Why was the estimate of customers leaving so far off? In the 18 years since Netflix had been founded, there was an increase in the number of close, but not perfect, substitutes. Consumers now had choices ranging from Vudu, Amazon Prime, Hulu, and Redbox, to retail stores. Jaime Weinman reported in *Maclean's* that Redbox kiosks are “a five-minute drive for less from 68 percent of Americans, and it seems that many people still find a five-minute drive more convenient than loading up a movie online.” It seems that in 2012, many consumers still preferred a physical DVD disk over streaming video.

What missteps did the Netflix management make? In addition to misjudging the elasticity of demand, by failing to account for close substitutes, it seems they may have also misjudged customers' preferences and tastes. Yet, as the population increases, the preference for streaming video may overtake physical DVD disks. Netflix, the source of numerous late night talk show laughs and jabs in 2011, may yet have the last laugh.

KEY CONCEPTS AND SUMMARY

Elasticity is a general term, referring to percentage change of one variable divided by percentage change of a related variable that can be applied to many economic connections. For instance, the income elasticity of demand is the percentage change in quantity demanded divided by the percentage change in income. The cross-price elasticity of demand is the percentage change in the quantity demanded of a good divided by the percentage change in the price of another good. Elasticity applies in labor markets and financial capital markets just as it does in markets for goods and services. The wage elasticity of labor supply is the percentage change in the quantity of hours supplied divided by the percentage change in the wage. The elasticity of savings with respect to interest rates is the percentage change in the quantity of savings divided by the percentage change in interest rates.

SELF-CHECK QUESTIONS

1. What would the gasoline price elasticity of supply mean to UPS or FedEx?
2. The average annual income rises from \$25,000 to \$38,000, and the quantity of bread consumed in a year by the average person falls from 30 loaves to 22 loaves. What is the income elasticity of bread consumption? Is bread a normal or an inferior good?
3. Suppose the cross-price elasticity of apples with respect to the price of oranges is 0.4, and the price of oranges falls by 3%. What will happen to the demand for apples?

REVIEW QUESTIONS

1. What is the formula for the income elasticity of demand?
2. What is the formula for the cross-price elasticity of demand?
3. What is the formula for the wage elasticity of labor supply?
4. What is the formula for elasticity of savings with respect to interest rates?

CRITICAL THINKING QUESTIONS

1. Normal goods are defined as having a positive income elasticity. We can divide normal goods into two types: Those whose income elasticity is less than one and those whose income elasticity is greater than one. Think about products that would fall into each category. Can you come up with a name for each category?
2. Suppose you could buy shoes one at a time, rather than in pairs. What do you predict the cross-price elasticity for left shoes and right shoes would be?

REFERENCES

Abkowitz, A. "How Netflix got started: Netflix founder and CEO Reed Hastings tells Fortune how he got the idea for the DVD-by-mail service that now has more than eight million customers." *CNN Money*. Last Modified January 28, 2009. http://archive.fortune.com/2009/01/27/news/newsmakers/hastings_netflix.fortune/index.htm.

Associated Press (a). "Analyst: Coinstar gains from Netflix pricing moves." *Boston Globe Media Partners, LLC*. Accessed June 24, 2013. http://www.boston.com/business/articles/2011/10/12/analyst_coinstar_gains_from_netflix_pricing_moves/.

Associated Press (b). "Netflix loses 800,000 US subscribers in tough 3Q." *ABC Inc*. Accessed June 24, 2013. <http://abclocal.go.com/wpvi/story?section=news/business&id=8403368>

Baumgardner, James. 2014. "Presentation on Raising the Excise Tax on Cigarettes: Effects on Health and the Federal Budget." Congressional Budget Office. Accessed March 27, 2015. http://www.cbo.gov/sites/default/files/45214-ICA_Presentation.pdf.

Funding Universe. 2015. "Netflix, Inc. History." Accessed March 11, 2015. <http://www.fundinguniverse.com/company-histories/netflix-inc-history/>.

Laporte, Nicole. "A tale of two Netflix." *Fast Company* 177 (July 2013) 31-32. Accessed December 3 2013. <http://www.fastcompany-digital.com/fastcompany/20130708?pg=33#pg33>

Liedtke, Michael, The Associated Press. "Investors bash Netflix stock after slower growth forecast – fee hikes expected to take toll on subscribers most likely to shun costly bundled Net, DVD service." *The Seattle Times*. Accessed June 24, 2013 from NewsBank on-line database (Access World News).

Netflix, Inc. 2013. "A Quick Update On Our Streaming Plans And Prices." Netflix (blog). Accessed March 11, 2015. <http://blog.netflix.com/2014/05/a-quick-update-on-our-streaming-plans.html>.

Organization for Economic Co-Operation and Development (OECD). n.d. "Average annual hours actually worked per worker." Accessed March 11, 2015. <https://stats.oecd.org/Index.aspx?DataSetCode=ANHRS>.

Savitz, Eric. "Netflix Warns DVD Subs Eroding; Q4 View Weak; Losses Ahead; Shrs Plunge." *Forbes.com*, 2011. Accessed December 3, 2013. <http://www.forbes.com/sites/ericsavitz/2011/10/24/netflix-q3-top-ests-but-shares-hit-by-weak-q4-outlook/>.

Statistica.com. 2014. "Coffee Export Volumes Worldwide in November 2014, by Leading Countries (in 60-kilo sacks)." Accessed March 27, 2015. <http://www.statista.com/statistics/268135/ranking-of-coffee-exporting-countries/>.

Stone, Marcie. "Netflix responds to customers angry with price hike; Netflix stock falls 9%." *News & Politics Examiner*, 2011. Clarity Digital Group. Accessed June 24, 2013. <http://www.examiner.com/article/netflix-responds-to-customers-angry-with-price-hike-netflix-stock-falls-9>.

Weinman, J. (2012). Die hard, hardly dying. *Maclean's*, 125(18), 44.

The World Bank Group. 2015. "Gross Savings (% of GDP)." Accessed March 11, 2015. <http://data.worldbank.org/indicator/NY.GNS.ICTR.ZS>.

Yahoo Finance. Retrieved from <http://finance.yahoo.com/q?s=NFLX>

GLOSSARY

cross-price elasticity of demand the percentage change in the quantity of good A that is demanded as a result of a percentage change in good B

elasticity of savings the percentage change in the quantity of savings divided by the percentage change in interest rates

wage elasticity of labor supply the percentage change in hours worked divided by the percentage change in wages

SOLUTIONS

Answers to Self-Check Questions

1. The percentage change in quantity supplied as a result of a given percentage change in the price of gasoline.

Percentage change in quantity demanded $[(\text{change in quantity})/(\text{original quantity})] \times 100$

$$[22 - 30]/[(22 + 30)/2] \times 100$$

$$\frac{-8}{26} \times 100$$

$$-30.77$$

Percentage change in income $[(\text{change in income})/(\text{original income})] \times 100$

$$[38,000 - 25,000]/[(38,000 + 25,000)/2] \times 100$$

$$\frac{13}{31.5} \times 100$$

$$41.27$$

2.

In this example, bread is an inferior good because its consumption falls as income rises.

3. The formula for cross-price elasticity is % change in Qd for apples / % change in P of oranges. Multiplying both sides by % change in P of oranges yields:

$$\% \text{ change in } Qd \text{ for apples} = \text{cross-price elasticity} \times \% \text{ change in } P \text{ of oranges}$$

$$0.4 \times (-3\%)$$

$$-1.2\%$$

,

or a 1.2% decrease in demand for apples.

CHAPTER 7. CONSUMER CHOICES

INTRODUCTION TO CONSUMER CHOICES



Figure 1. Investment Choices. Higher education is generally viewed as a good investment, if one can afford it, regardless of the state of the economy. (Credit: modification of work by Jason Bache/Flickr Creative Commons)

“EENY, MEENY, MINEY, MOE”—MAKING CHOICES

The Great Recession of 2008–2009 touched families around the globe. In too many countries, workers found themselves out of a job. In developed countries, unemployment compensation provided a safety net, but families still saw a marked decrease in disposable income and had to make tough spending decisions. Of course, non-essential, discretionary spending was the first to go.

Even so, there was one particular category that saw a universal increase in spending world-wide during that time—an 18% uptick in the United States, specifically. You might guess that consumers began eating more meals at home, increasing spending at the grocery store. But the Bureau of Labor Statistics’ Consumer Expenditure Survey, which tracks U.S. food spending over time, showed “real total food spending by U.S. households declined five percent between 2006 and 2009.” So, it was not groceries. Just what product would people around the world demand more of during tough economic times, and more importantly, why? (Find out at chapter’s end.)

That question leads us to this chapter’s topic—analyzing how consumers make choices. For most consumers, using “eeny,

meeny, miney, moe” is not how they make decisions; their decision-making processes have been educated far beyond a children’s rhyme.

CHAPTER OBJECTIVES

Introduction to Consumer Choices

In this chapter, you will learn about:

- Consumption Choices
- How Changes in Income and Prices Affect Consumption Choices
- Labor-Leisure Choices
- Intertemporal Choices in Financial Capital Markets

Microeconomics seeks to understand the behavior of individual economic agents such as individuals and businesses. Economists believe that individuals’ decisions, such as what goods and services to buy, can be analyzed as choices made within certain budget constraints. Generally, consumers are trying to get the most for their limited budget. In economic terms they are trying to maximize total utility, or satisfaction, given their budget constraint.

Everyone has their own personal tastes and preferences. The French say: *Chacun à son goût*, or “Each to his own taste.” An old Latin saying states, *De gustibus non est disputandum* or “There’s no disputing about taste.” If people’s decisions are based on their own tastes and personal preferences, however, then how can economists hope to analyze the choices consumers make?

An economic explanation for why people make different choices begins with accepting the proverbial wisdom that tastes are a matter of personal preference. But economists also believe that the choices people make are influenced by their incomes, by the prices of goods and services they consume, and by factors like where they live. This chapter introduces the economic theory of how consumers make choices about what to buy, how much to work, and how much to save.

The analysis in this chapter will build on the three budget constraints introduced in the Choice in a World of Scarcity chapter. These were the consumption choice budget constraint, the labor-leisure budget constraint, and the intertemporal budget constraint. This chapter will also illustrate how economic theory provides a tool to systematically look at the full range of possible consumption choices to predict how consumption responds to changes in prices or incomes. After reading this chapter, consult the appendix Indifference Curves to learn more about representing utility and choice through indifference curves.

7.1 CONSUMPTION CHOICES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Calculate total utility
- Propose decisions that maximize utility
- Explain marginal utility and the significance of diminishing marginal utility

Information on the **consumption** choices of Americans is available from the Consumer Expenditure Survey carried out by the U.S. Bureau of Labor Statistics. Table 1 shows spending patterns for the average U.S. household. The first row shows income and, after taxes and personal savings are subtracted, it shows that, in 2015, the average U.S. household spent \$48,109 on consumption. The table then breaks down consumption into various categories. The average U.S. household spent roughly one-third of its consumption on shelter and other housing expenses, another one-third on food and vehicle expenses, and the rest on a variety of items, as shown. Of course, these patterns will vary for specific households by differing levels of family income, by geography, and by preferences.

Average Household Income before Taxes	\$62,481
Average Annual Expenditures	\$48,109
Food at home	\$3,264
Food away from home	\$2,505
Housing	\$16,557
Apparel and services	\$1,700
Transportation	\$7,677
Healthcare	\$3,157
Entertainment	\$2,504
Education	\$1,074
Personal insurance and pensions	\$5,357
All else: alcohol, tobacco, reading, personal care, cash contributions, miscellaneous	\$3,356

Table 1. U.S. Consumption Choices in 2015(Source: <http://www.bls.gov/cex/csxann13.pdf>)

TOTAL UTILITY AND DIMINISHING MARGINAL UTILITY

To understand how a household will make its choices, economists look at what consumers can afford, as shown in a **budget constraint line**, and the **total utility** or satisfaction derived from those choices. In a budget constraint line, the quantity of one good is measured on the horizontal axis and the quantity of the other good is measured on the vertical axis. The budget constraint line shows the various combinations of two goods that are affordable given consumer income. Consider the situation of José, shown in Figure 1. José likes to collect T-shirts and watch movies.

In Figure 1, the quantity of T-shirts is shown on the horizontal axis, while the quantity of movies is shown on the vertical axis. If José had unlimited income or goods were free, then he could consume without limit. But José, like all of us, faces a **budget constraint**. José has a total of \$56 to spend. The price of T-shirts is \$14 and the price of movies is \$7. Notice that the vertical intercept of the budget constraint line is at eight movies and zero T-shirts ($\$56/\$7=8$). The horizontal intercept of the budget constraint is four, where José spends all of his money on T-shirts and no movies ($\$56/\$14=4$). The slope of the budget constraint line is rise/run or $-8/4=-2$. The specific choices along the budget constraint line show the combinations of T-shirts and movies that are affordable.

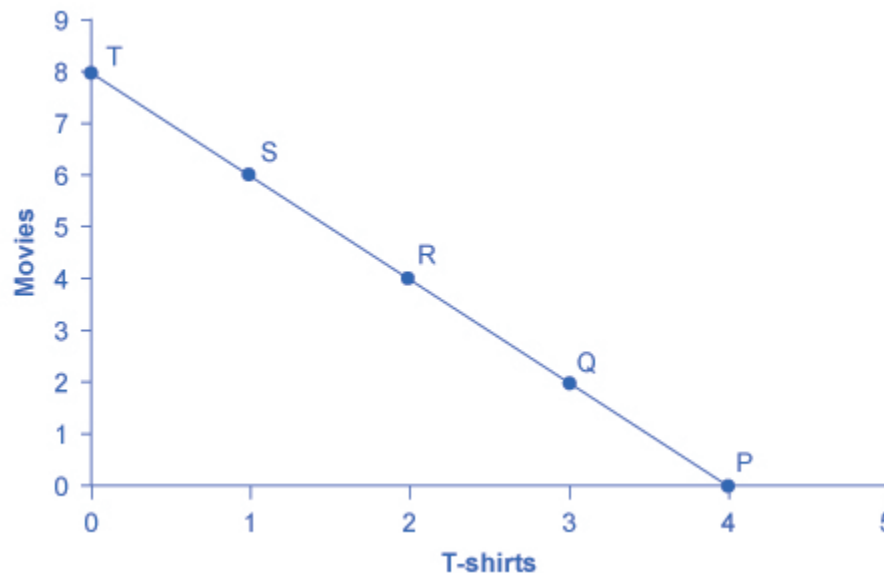


Figure 1. A Choice between Consumption Goods. José has income of \$56. Movies cost \$7 and T-shirts cost \$14. The points on the budget constraint line show the combinations of movies and T-shirts that are affordable.

José wishes to choose the combination that will provide him with the greatest **utility**, which is the term economists use to describe a person's level of satisfaction or happiness with his or her choices.

Let's begin with an assumption, which will be discussed in more detail later, that José can measure his own utility with something called *utils*. (It is important to note that you cannot make comparisons between the utils of individuals; if one person gets 20 utils from a cup of coffee and another gets 10 utils, this does not mean that the first person gets more enjoyment from the coffee than the other or that they enjoy the coffee twice as much.) Table 2 shows how José's utility is connected with his consumption of T-shirts or movies. The first column of the table shows the quantity of T-shirts consumed. The second column shows the total utility, or total amount of satisfaction, that José receives

from consuming that number of T-shirts. The most common pattern of total utility, as shown here, is that consuming additional goods leads to greater total utility, but at a decreasing rate. The third column shows **marginal utility**, which is the additional utility provided by one additional unit of consumption. This equation for marginal utility is:

$$MU = \frac{\text{change in total utility}}{\text{change in quantity}}$$

Notice that marginal utility diminishes as additional units are consumed, which means that each subsequent unit of a good consumed provides less *additional* utility. For example, the first T-shirt José picks is his favorite and it gives him an addition of 22 utils. The fourth T-shirt is just to something to wear when all his other clothes are in the wash and yields only 18 additional utils. This is an example of the law of **diminishing marginal utility**, which holds that the additional utility decreases with each unit added.

The rest of Table 2 shows the quantity of movies that José attends, and his total and marginal utility from seeing each movie. Total utility follows the expected pattern: it increases as the number of movies seen rises. Marginal utility also follows the expected pattern: each additional movie brings a smaller gain in utility than the previous one. The first movie José attends is the one he wanted to see the most, and thus provides him with the highest level of utility or satisfaction. The fifth movie he attends is just to kill time. Notice that total utility is also the sum of the marginal utilities. Read the next Work It Out feature for instructions on how to calculate total utility.

T-Shirts (Quantity)	Total Utility	Marginal Utility	Movies (Quantity)	Total Utility	Marginal Utility
1	22	22	1	16	16
2	43	21	2	31	15
3	63	20	3	45	14
4	81	18	4	58	13
5	97	16	5	70	12
6	111	14	6	81	11
7	123	12	7	91	10
8	133	10	8	100	9

Table 2. Total and Marginal Utility

Table 3 looks at each point on the budget constraint in Figure 1, and adds up José's total utility for five possible combinations of T-shirts and movies.

Point	T-Shirts	Movies	Total Utility
P	4	0	$81 + 0 = 81$
Q	3	2	$63 + 31 = 94$
R	2	4	$43 + 58 = 101$
S	1	6	$22 + 81 = 103$
T	0	8	$0 + 100 = 100$

Table 3. Finding the Choice with the Highest Utility

CALCULATING TOTAL UTILITY

Let's look at how José makes his decision in more detail.

Step 1. Observe that, at point Q (for example), José consumes three T-shirts and two movies.

Step 2. Look at Table 2. You can see from the fourth row/second column that three T-shirts are worth 63 utils. Similarly, the second row/fifth column shows that two movies are worth 31 utils.

Step 3. From this information, you can calculate that point Q has a total utility of 94 ($63 + 31$).

Step 4. You can repeat the same calculations for each point on Table 3, in which the total utility numbers are shown in the last column.

For José, the highest total utility for all possible combinations of goods occurs at point S, with a total utility of 103 from consuming one T-shirt and six movies.

CHOOSING WITH MARGINAL UTILITY

Most people approach their utility-maximizing combination of choices in a step-by-step way. This step-by-step approach is based on looking at the tradeoffs, measured in terms of marginal utility, of consuming less of one good and more of another.

For example, say that José starts off thinking about spending all his money on T-shirts and choosing point P, which corresponds to four T-shirts and no movies, as illustrated in Figure 1. José chooses this starting point randomly; he has to start somewhere. Then he considers giving up the last T-shirt, the one that provides him the least marginal utility, and using the money he saves to buy two movies instead. Table 4 tracks the step-by-step series of decisions José needs to make (*Key*: T-shirts are \$14, movies are \$7, and income is \$56). The following Work It Out feature explains how marginal utility can effect decision making.

Try	Which Has	Total Utility	Marginal Gain and Loss of Utility, Compared with Previous Choice	Conclusion
Choice 1: P	4 T-shirts and 0 movies	81 from 4 T-shirts + 0 from 0 movies = 81	–	–
Choice 2: Q	3 T-shirts and 2 movies	63 from 3 T-shirts + 31 from 0 movies = 94	Loss of 18 from 1 less T-shirt, but gain of 31 from 2 more movies, for a net utility gain of 13	Q is preferred over P
Choice 3: R	2 T-shirts and 4 movies	43 from 2 T-shirts + 58 from 4 movies = 101	Loss of 20 from 1 less T-shirt, but gain of 27 from two more movies for a net utility gain of 7	R is preferred over Q
Choice 4: S	1 T-shirt and 6 movies	22 from 1 T-shirt + 81 from 6 movies = 103	Loss of 21 from 1 less T-shirt, but gain of 23 from two more movies, for a net utility gain of 2	S is preferred over R
Choice 5: T	0 T-shirts and 8 movies	0 from 0 T-shirts + 100 from 8 movies = 100	Loss of 22 from 1 less T-shirt, but gain of 19 from two more movies, for a net utility loss of 3	S is preferred over T

Table 4. A Step-by-Step Approach to Maximizing Utility

DECISION MAKING BY COMPARING MARGINAL UTILITY

José could use the following thought process (if he thought in utils) to make his decision regarding how many T-shirts and movies to purchase:

Step 1. From Table 2, José can see that the marginal utility of the fourth T-shirt is 18. If José gives up the fourth T-shirt, then he loses 18 utils.

Step 2. Giving up the fourth T-shirt, however, frees up \$14 (the price of a T-shirt), allowing José to buy the first two movies (at \$7 each).

Step 3. José knows that the marginal utility of the first movie is 16 and the marginal utility of the second movie is 15. Thus, if José moves from point P to point Q, he gives up 18 utils (from the T-shirt), but gains 31 utils (from the movies).

Step 4. Gaining 31 utils and losing 18 utils is a net gain of 13. This is just another way of saying that the total utility at Q (94 according to the last column in Table 3) is 13 more than the total utility at P (81).

Step 5. So, for José, it makes sense to give up the fourth T-shirt in order to buy two movies.

José clearly prefers point Q to point P. Now repeat this step-by-step process of decision making with marginal utilities. José thinks about giving up the third T-shirt and surrendering a marginal utility of 20, in exchange for purchasing two more movies that promise a combined marginal utility of 27. José prefers point R to point Q. What if José thinks about going beyond R to point S? Giving up the second T-shirt means a marginal utility loss of 21, and the marginal utility gain from the fifth and sixth movies would combine to make a marginal utility gain of 23, so José prefers point S to R.

However, if José seeks to go beyond point S to point T, he finds that the loss of marginal utility from giving up the first T-shirt is 22, while the marginal utility gain from the last two movies is only a total of 19. If José were to choose point T, his utility would fall to 100. Through these stages of thinking about marginal tradeoffs, José again concludes that S, with one T-shirt and six movies, is the choice that will provide him with the highest level of total utility. This step-by-step approach will reach the same conclusion regardless of José's starting point.

Another way to look at this is by focusing on satisfaction per dollar. **Marginal utility per dollar** is the amount of additional utility José receives given the price of the product. For José's T-shirts and movies, the marginal utility per dollar is shown in Table 5.

$$\text{marginal utility per dollar} = \frac{\text{marginal utility}}{\text{price}}$$

José's first purchase will be a movie. Why? Because it gives him the highest marginal utility per dollar and it is affordable. José will continue to purchase the good which gives him the highest marginal utility per dollar until he exhausts the budget. José will keep purchasing movies because they give him a greater "bang or the buck" until the sixth movie is equivalent to a T-shirt purchase. José can afford to purchase that T-shirt. So José will choose to purchase six movies and one T-shirt.

Quantity of T-Shirts	Total Utility	Marginal Utility	Marginal Utility per Dollar	Quantity of Movies	Total Utility	Marginal Utility	Marginal Utility per Dollar
1	22	22	22/\$14=1.6	1	16	16	16/\$7=2.3
2	43	21	21/\$14=1.5	2	31	15	15/\$7=2.14
3	63	20	20/\$14=1.4	3	45	14	14/\$7=2
4	81	18	18/\$14=1.3	4	58	13	13/\$7=1.9
5	97	16	16/\$14=1.1	5	70	12	12/\$7=1.7
6	111	14	14/\$14=1	6	81	11	11/\$7=1.6
7	123	12	12/\$14=1.2	7	91	10	10/\$7=1.4

Table 5. Marginal Utility per Dollar

A RULE FOR MAXIMIZING UTILITY

This process of decision making suggests a rule to follow when **maximizing utility**. Since the price of T-shirts is twice as high as the price of movies, to maximize utility the last T-shirt chosen needs to provide exactly twice the marginal utility (MU) of the last movie. If the last T-shirt provides less than twice the marginal utility of the last movie, then the T-shirt is providing less “bang for the buck” (i.e., marginal utility per dollar spent) than if the same money were spent on movies. If this is so, José should trade the T-shirt for more movies to increase his total utility. Marginal utility per dollar measures the additional utility that José will enjoy given what he has to pay for the good.

If the last T-shirt provides more than twice the marginal utility of the last movie, then the T-shirt is providing more “bang for the buck” or marginal utility per dollar, than if the money were spent on movies. As a result, José should buy more T-shirts. Notice that at José’s optimal choice of point S, the marginal utility from the first T-shirt, of 22 is exactly twice the marginal utility of the sixth movie, which is 11. At this choice, the marginal utility per dollar is the same for both goods. This is a tell-tale signal that José has found the point with highest total utility.

This argument can be written as a general rule: the utility-maximizing choice between consumption goods occurs where the marginal utility per dollar is the same for both goods.

$$\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$$

A sensible economizer will pay twice as much for something only if, in the marginal comparison, the item confers twice as much utility. Notice that the formula for the table above is:

$$\frac{22}{\$14} = \frac{11}{\$7}$$

$$1.6 = 1.6$$

The following Work It Out feature provides step by step guidance for this concept of utility-maximizing choices.

MAXIMIZING UTILITY

$$\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$$

The general rule, $\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$, means that the last dollar spent on each good provides exactly the same marginal utility. So:

Step 1. If we traded a dollar more of movies for a dollar more of T-shirts, the marginal utility gained from T-shirts would exactly offset the marginal utility lost from fewer movies. In other words, the net gain would be zero.

Step 2. Products, however, usually cost more than a dollar, so we cannot trade a dollar's worth of movies. The best we can do is trade two movies for another T-shirt, since in this example T-shirts cost twice what a movie does.

Step 3. If we trade two movies for one T-shirt, we would end up at point R (two T-shirts and four movies).

Step 4. Choice 4 in Table 4 shows that if we move to point S, we would lose 21 utils from one less T-shirt, but gain 23 utils from two more movies, so we would end up with more total utility at point S.

In short, the general rule shows us the utility-maximizing choice.

There is another, equivalent way to think about this. The general rule can also be expressed as *the ratio of the prices of the two goods should be equal to the ratio of the marginal utilities*. When the price of good 1 is divided by the price of good 2, at the utility-maximizing point this will equal the marginal utility of good 1 divided by the marginal utility of good 2. This rule, known as the **consumer equilibrium**, can be written in algebraic form:

$$\frac{P_1}{P_2} = \frac{MU_1}{MU_2}$$

Along the budget constraint, the total price of the two goods remains the same, so the ratio of the prices does not change. However, the marginal utility of the two goods changes with the quantities consumed. At the optimal choice of one T-shirt and six movies, point S, the ratio of marginal utility to price for T-shirts (22:14) matches the ratio of marginal utility to price for movies (of 11:7).

MEASURING UTILITY WITH NUMBERS

This discussion of utility started off with an assumption that it is possible to place numerical values on utility, an assumption that may seem questionable. You can buy a thermometer for measuring temperature at the hardware store, but what store sells an “utilimometer” for measuring utility? However, while measuring utility with numbers is a convenient assumption to clarify the explanation, the key assumption is not that utility can be measured by an outside party, but only that individuals can decide which of two alternatives they prefer.

To understand this point, think back to the step-by-step process of finding the choice with highest total utility by comparing the marginal utility that is gained and lost from different choices along the budget constraint. As José compares each choice along his budget constraint to the previous choice, what matters is not the specific numbers that he places on his utility—or whether he uses any numbers at all—but only that he personally can identify which choices he prefers.

In this way, the step-by-step process of choosing the highest level of utility resembles rather closely how many people make consumption decisions. We think about what will make us the happiest; we think about what things cost; we think about buying a little more of one item and giving up a little

of something else; we choose what provides us with the greatest level of satisfaction. The vocabulary of comparing the points along a budget constraint and total and marginal utility is just a set of tools for discussing this everyday process in a clear and specific manner. It is welcome news that specific utility numbers are not central to the argument, since a good utilimometer is hard to find. Do not worry—while we cannot measure utils, by the end of the next module, we will have transformed our analysis into something we can measure—demand.

KEY CONCEPTS AND SUMMARY

Economic analysis of household behavior is based on the assumption that people seek the highest level of utility or satisfaction. Individuals are the only judge of their own utility. In general, greater consumption of a good brings higher total utility. However, the additional utility received from each unit of greater consumption tends to decline in a pattern of diminishing marginal utility.

The utility-maximizing choice on a consumption budget constraint can be found in several ways. You can add up total utility of each choice on the budget line and choose the highest total. You can choose a starting point at random and compare the marginal utility gains and losses of moving to neighboring points—and thus eventually seek out the preferred choice. Alternatively, you can compare the ratio of the marginal utility to price of good 1 with the marginal utility to price of good 2 and apply the rule that at the optimal choice, the two ratios should be equal:

$$\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$$

SELF-CHECK QUESTIONS

1. Jeremy is deeply in love with Jasmine. Jasmine lives where cell phone coverage is poor, so he can either call her on the land-line phone for five cents per minute or he can drive to see her, at a round-trip cost of \$2 in gasoline money. He has a total of \$10 per week to spend on staying in touch. To make his preferred choice, Jeremy uses a handy utilimometer that measures his total utility from personal visits and from phone minutes. Using the values given in Table 6, figure out the points on Jeremy's consumption choice budget constraint (it may be helpful to do a sketch) and identify his utility-maximizing point.

Round Trips	Total Utility	Phone Minutes	Total Utility
0	0	0	0
1	80	20	200
2	150	40	380
3	210	60	540
4	260	80	680
5	300	100	800
6	330	120	900
7	200	140	980
8	180	160	1040
9	160	180	1080
10	140	200	1100

Table 6.

- Take Jeremy's total utility information in Self-Check Question 1, and use the marginal utility approach to confirm the choice of phone minutes and round trips that maximize Jeremy's utility.

REVIEW QUESTIONS

- Who determines how much utility an individual will receive from consuming a good?
- Would you expect total utility to rise or fall with additional consumption of a good? Why?
- Would you expect marginal utility to rise or fall with additional consumption of a good? Why?
- Is it possible for total utility to increase while marginal utility diminishes? Explain.
- If people do not have a complete mental picture of total utility for every level of consumption, how can they find their utility-maximizing consumption choice?
- What is the rule relating the ratio of marginal utility to prices of two goods at the optimal choice? Explain why, if this rule does not hold, the choice cannot be utility-maximizing.

CRITICAL THINKING QUESTIONS

- Think back to a purchase that you made recently. How would you describe your thinking before you made that purchase?
- The rules of politics are not always the same as the rules of economics. In discussions of setting budgets for government agencies, there is a strategy called "closing the Washington monument." When an agency faces the unwelcome prospect of a budget cut, it may decide to close a high-visibility attraction enjoyed by many people (like the Washington monument). Explain in terms of diminishing marginal utility why the Washington monument strategy is so misleading. *Hint:* If you are really trying to make the best of a budget cut, should you cut the items in your budget with the highest marginal utility or the lowest marginal utility?

Does the Washington monument strategy cut the items with the highest marginal utility or the lowest marginal utility?

PROBLEMS

Praxilla, who lived in ancient Greece, derives utility from reading poems and from eating cucumbers. Praxilla gets 30 units of marginal utility from her first poem, 27 units of marginal utility from her second poem, 24 units of marginal utility from her third poem, and so on, with marginal utility declining by three units for each additional poem. Praxilla gets six units of marginal utility for each of her first three cucumbers consumed, five units of marginal utility for each of her next three cucumbers consumed, four units of marginal utility for each of the following three cucumbers consumed, and so on, with marginal utility declining by one for every three cucumbers consumed. A poem costs three bronze coins but a cucumber costs only one bronze coin. Praxilla has 18 bronze coins. Sketch Praxilla's budget set between poems and cucumbers, placing poems on the vertical axis and cucumbers on the horizontal axis. Start off with the choice of zero poems and 18 cucumbers, and calculate the changes in marginal utility of moving along the budget line to the next choice of one poem and 15 cucumbers. Using this step-by-step process based on marginal utility, create a table and identify Praxilla's utility-maximizing choice. Compare the marginal utility of the two goods and the relative prices at the optimal choice to see if the expected relationship holds. *Hint:* Label the table columns: 1) Choice, 2) Marginal Gain from More Poems, 3) Marginal Loss from Fewer Cucumbers, 4) Overall Gain or Loss, 5) Is the previous choice optimal? Label the table rows: 1) 0 Poems and 18 Cucumbers, 2) 1 Poem and 15 Cucumbers, 3) 2 Poems and 12 Cucumbers, 4) 3 Poems and 9 Cucumbers, 5) 4 Poems and 6 Cucumbers, 6) 5 Poems and 3 Cucumbers, 7) 6 Poems and 0 Cucumbers.

REFERENCES

- U.S. Bureau of Labor Statistics. 2015. "Consumer Expenditures in 2013." Accessed March 11, 2015. <http://www.bls.gov/cex/csxann13.pdf>.
- U.S. Bureau of Labor Statistics. 2015. "Employer Costs for Employee Compensation—December 2014." Accessed March 11, 2015. <http://www.bls.gov/news.release/pdf/ecec.pdf>.
- U.S. Bureau of Labor Statistics. 2015. "Labor Force Statistics from the Current Population Survey." Accessed March 11, 2015. <http://www.bls.gov/cps/cpsaat22.htm>.

GLOSSARY

budget constraint line shows the possible combinations of two goods that are affordable given a consumer's limited income

consumer equilibrium when the ratio of the prices of goods is equal to the ratio of the marginal utilities (point at which the consumer can get the most satisfaction)

diminishing marginal utility the common pattern that each marginal unit of a good consumed provides less of an addition to utility than the previous unit

marginal utility the additional utility provided by one additional unit of consumption

marginal utility per dollar the additional satisfaction gained from purchasing a good given the price of the product; $MU/Price$

total utility satisfaction derived from consumer choices

SOLUTIONS

Answers to Self-Check Questions

1. The rows of the table in the problem do not represent the actual choices available on the budget set; that is, the combinations of round trips and phone minutes that Jeremy can afford with his budget. One of the choices listed in the problem, the six round trips, is not even available on the budget set. If Jeremy has only \$10 to spend and a round trip costs \$2 and phone calls cost \$0.05 per minute, he could spend his entire budget on five round trips but no phone calls or 200 minutes of phone calls, but no round trips or any combination of the two in between. It is easy to see all of his budget options with a little algebra. The equation for a budget line is:

$$\text{Budget} = P_{RT} \times Q_{RT} + P_{PC} \times Q_{PC}$$

where P and Q are price and quantity of round trips (RT) and phone calls (PC) (per minute). In Jeremy's case the equation for the budget line is:

$$\$10 = \$2 \times Q_{RT} + \$0.05 \times Q_{PC}$$

$$\frac{\$10}{\$0.05} = \frac{\$2Q_{RT} + \$0.05Q_{PC}}{\$0.05}$$

$$200 = 40Q_{RT} + Q_{PC}$$

$$Q_{PC} = 200 - 40Q_{RT}$$

If we choose zero through five round trips (column 1), the table below shows how many phone minutes can be afforded with the budget (column 3). The total utility figures are given in the table below.

Round Trips	Total Utility for Trips	Phone Minutes	Total Utility for Minutes	Total Utility
0	0	200	1100	1100
1	80	160	1040	1120
2	150	120	900	1050
3	210	80	680	890
4	260	40	380	640
5	300	0	0	300

Table 7.

Adding up total utility for round trips and phone minutes at different points on the budget line gives total utility at each point on the budget line. The highest possible utility is at the combination of one trip and 160 minutes of phone time, with a total utility of 1120.

2. The first step is to use the total utility figures, shown in the table below, to calculate marginal utility, remembering that marginal utility is equal to the change in total utility divided by the change in trips or minutes.

Round Trips	Total Utility	Marginal Utility (per trip)	Phone Minutes	Total Utility	Marginal Utility (per minute)
0	0	–	200	1100	–
1	80	80	160	1040	$60/40 = 1.5$
2	150	70	120	900	$140/40 = 3.5$
3	210	60	80	680	$220/40 = 5.5$
4	260	50	40	380	$300/40 = 7.5$
5	300	40	0	0	$380/40 = 9.5$

Table 8.

Note that we cannot directly compare marginal utilities, since the units are trips versus phone minutes. We need a common denominator for comparison, which is price. Dividing MU by the price, yields columns 4 and 8 in the table below.

Round Trips	Total Utility	Marginal Utility (per trip)	MU/P	Phone Minutes	Total Utility	Marginal utility (per minute)	MU/P
0	0	–	–	200	1100	$60/40 = 1.5$	$1.5/\$0.05 = 30$
1	80	80	$80/\$2 = 40$	160	1040	$140/40 = 3.5$	$3.5/\$0.05 = 70$
2	150	70	$70/\$2 = 35$	120	900	$220/40 = 5.5$	$5.5/\$0.05 = 110$
3	210	60	$60/\$2 = 30$	80	680	$300/40 = 7.5$	$7.5/\$0.05 = 150$
4	260	50	$50/\$2 = 25$	40	380	$380/40 = 9.5$	$9.5/\$0.05 = 190$
5	300	40	$40/\$2 = 20$	0	0	–	–

Table 9.

Start at the bottom of the table where the combination of round trips and phone minutes is (5, 0). This starting point is arbitrary, but the numbers in this example work best starting from the bottom. Suppose we consider moving to the next point up. At (4, 40), the marginal utility per dollar spent on a round trip is 25. The marginal utility per dollar spent on phone minutes is 190.

Since $25 < 190$, we are getting much more utility per dollar spent on phone minutes, so let's choose more of those. At (3, 80), MU/PRT is $30 < 150$ (the MU/PM), but notice that the difference is narrowing. We keep trading round trips for phone minutes until we get to (1, 160), which is the best we can do. The MU/P comparison is as close as it is going to get (40 vs. 70). Often in the real world, it is not possible to get MU/P exactly equal for both products, so you get as close as you can.

7.2 HOW CHANGES IN INCOME AND PRICES AFFECT CONSUMPTION CHOICES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how income, prices, and preferences affect consumer choices
- Contrast the substitution effect and the income effect
- Utilize concepts of demand to analyze consumer choices
- Apply utility-maximizing choices to governments and businesses

Just as utility and marginal utility can be used to discuss making consumer choices along a budget constraint, these ideas can also be used to think about how consumer choices change when the budget constraint shifts in response to changes in income or price. Indeed, because the budget constraint framework can be used to analyze how quantities demanded change because of price movements, the budget constraint model can illustrate the underlying logic behind demand curves.

HOW CHANGES IN INCOME AFFECT CONSUMER CHOICES

Let's begin with a concrete example illustrating how changes in income level affect consumer choices. Figure 1 shows a **budget constraint** that represents Kimberly's choice between concert tickets at \$50 each and getting away overnight to a bed-and-breakfast for \$200 per night. Kimberly has \$1,000 per year to spend between these two choices. After thinking about her total utility and marginal utility and applying the decision rule that the ratio of the marginal utilities to the prices should be equal between the two products, Kimberly chooses point M, with eight concerts and three overnight get-aways as her utility-maximizing choice.

Now, assume that the income Kimberly has to spend on these two items rises to \$2,000 per year, causing her budget constraint to shift out to the right. How does this rise in income alter her utility-maximizing choice? Kimberly will again consider the utility and marginal utility that she receives from concert tickets and overnight getaways and seek her **utility-maximizing choice** on the new budget line. But how will her new choice relate to her original choice?

The possible choices along the new budget constraint can be divided into three groups, which are divided up by the dashed horizontal and vertical lines that pass through the original choice M in the figure. All choices on the upper left of the new budget constraint that are to the left of the vertical dashed line, like choice P with two overnight stays and 32 concert tickets, involve less of the good

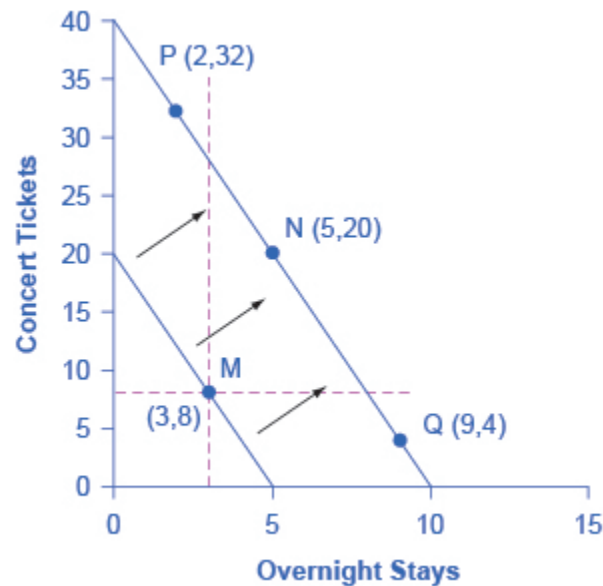


Figure 1. How a Change in Income Affects Consumption Choices. The utility-maximizing choice on the original budget constraint is M. The dashed horizontal and vertical lines extending through point M allow you to see at a glance whether the quantity consumed of goods on the new budget constraint is higher or lower than on the original budget constraint. On the new budget constraint, a choice like N will be made if both goods are normal goods. If overnight stays is an inferior good, a choice like P will be made. If concert tickets are an inferior good, a choice like Q will be made.

on the horizontal axis but much more of the good on the vertical axis. All choices to the right of the vertical dashed line and above the horizontal dashed line—like choice N with five overnight getaways and 20 concert tickets—have more consumption of both goods. Finally, all choices that are to the right of the vertical dashed line but below the horizontal dashed line, like choice Q with four concerts and nine overnight getaways, involve less of the good on the vertical axis but much more of the good on the horizontal axis.

All of these choices are theoretically possible, depending on Kimberly's personal preferences as expressed through the total and marginal utility she would receive from consuming these two goods. When income rises, the most common reaction is to purchase more of both goods, like choice N, which is to the upper right relative to Kimberly's original choice M, although exactly how much more of each good will vary according to personal taste. Conversely, when income falls, the most typical reaction is to purchase less of both goods. As defined in the chapter on Demand and Supply and again in the chapter on Elasticity, goods and services are called **normal goods** when a rise in income leads to a rise in the quantity consumed of that good and a fall in income leads to a fall in quantity consumed.

However, depending on Kimberly's preferences, a rise in income could cause consumption of one good to increase while consumption of the other good declines. A choice like P means that a rise in

income caused her quantity consumed of overnight stays to decline, while a choice like Q would mean that a rise in income caused her quantity of concerts to decline. Goods where demand declines as income rises (or conversely, where the demand rises as income falls) are called “inferior goods.” An **inferior good** occurs when people trim back on a good as income rises, because they can now afford the more expensive choices that they prefer. For example, a higher-income household might eat fewer hamburgers or be less likely to buy a used car, and instead eat more steak and buy a new car.

HOW PRICE CHANGES AFFECT CONSUMER CHOICES

For analyzing the possible effect of a change in price on consumption, let’s again use a concrete example. Figure 2 represents the consumer choice of Sergei, who chooses between purchasing baseball bats and cameras. A price increase for baseball bats would have no effect on the ability to purchase cameras, but it would reduce the number of bats Sergei could afford to buy. Thus a price increase for baseball bats, the good on the horizontal axis, causes the budget constraint to rotate inward, as if on a hinge, from the vertical axis. As in the previous section, the point labeled M represents the originally preferred point on the original budget constraint, which Sergei has chosen after contemplating his total utility and marginal utility and the tradeoffs involved along the budget constraint. In this example, the units along the horizontal and vertical axes are not numbered, so the discussion must focus on whether more or less of certain goods will be consumed, not on numerical amounts.

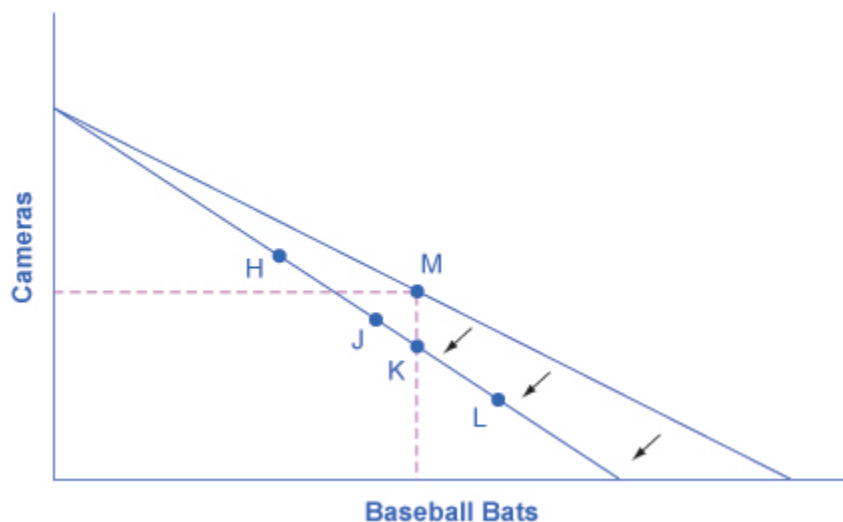


Figure 2. How a Change in Price Affects Consumption Choices. The original utility-maximizing choice is M. When the price rises, the budget constraint shifts in to the left. The dashed lines make it possible to see at a glance whether the new consumption choice involves less of both goods, or less of one good and more of the other. The new possible choices would be fewer baseball bats and more cameras, like point H, or less of both goods, as at point J. Choice K would mean that the higher price of bats led to exactly the same quantity of bats being consumed, but fewer cameras. Choices like L are ruled out as theoretically possible but highly unlikely in the real world, because they would mean that a higher price for baseball bats means a greater quantity consumed of baseball bats.

After the price increase, Sergei will make a choice along the new budget constraint. Again, his choices can be divided into three segments by the dashed vertical and horizontal lines. In the upper left por-

tion of the new budget constraint, at a choice like H, Sergei consumes more cameras and fewer bats. In the central portion of the new budget constraint, at a choice like J, he consumes less of both goods. At the right-hand end, at a choice like L, he consumes more bats but fewer cameras.

The typical response to higher prices is that a person chooses to consume less of the product with the higher price. This occurs for two reasons, and both effects can occur simultaneously. The **substitution effect** occurs when a price changes and consumers have an incentive to consume less of the good with a relatively higher price and more of the good with a relatively lower price. The **income effect** is that a higher price means, in effect, the buying power of income has been reduced (even though actual income has not changed), which leads to buying less of the good (when the good is normal). In this example, the higher price for baseball bats would cause Sergei to buy a fewer bats for both reasons. Exactly how much will a higher price for bats cause Sergei consumption of bats to fall? Figure 2 suggests a range of possibilities. Sergei might react to a higher price for baseball bats by purchasing the same quantity of bats, but cutting his consumption of cameras. This choice is the point K on the new budget constraint, straight below the original choice M. Alternatively, Sergei might react by dramatically reducing his purchases of bats and instead buy more cameras.

The key is that it would be imprudent to assume that a change in the price of baseball bats will only or primarily affect the good whose price is changed, while the quantity consumed of other goods remains the same. Since Sergei purchases all his products out of the same budget, a change in the price of one good can also have a range of effects, either positive or negative, on the quantity consumed of other goods.

In short, a higher price typically causes reduced consumption of the good in question, but it can affect the consumption of other goods as well.

Read this article about the potential of variable prices in vending machines.



THE FOUNDATIONS OF DEMAND CURVES

Changes in the price of a good lead the budget constraint to shift. A shift in the budget constraint means that when individuals are seeking their highest utility, the quantity that is demanded of that good will change. In this way, the logical foundations of demand curves—which show a connection between prices and quantity demanded—are based on the underlying idea of individuals seeking utility. Figure 3 (a) shows a budget constraint with a choice between housing and “everything else.” (Putting “everything else” on the vertical axis can be a useful approach in some cases, especially when the focus of the analysis is on one particular good.) The preferred choice on the original budget constraint that provides the highest possible utility is labeled M_0 . The other three budget constraints rep-

resent successively higher prices for housing of P_1 , P_2 , and P_3 . As the budget constraint rotates in, and in, and in again, the utility-maximizing choices are labeled M_1 , M_2 , and M_3 , and the quantity demanded of housing falls from Q_0 to Q_1 to Q_2 to Q_3 .

So, as the price of housing rises, the budget constraint shifts to the left, and the quantity consumed of housing falls, *ceteris paribus* (meaning, with all other things being the same). This relationship—the price of housing rising from P_0 to P_1 to P_2 to P_3 , while the quantity of housing demanded falls from Q_0 to Q_1 to Q_2 to Q_3 —is graphed on the demand curve in Figure 3 (b). Indeed, the vertical dashed lines stretching between the top and bottom of Figure 3 show that the quantity of housing demanded at each point is the same in both (a) and (b). The shape of a demand curve is ultimately determined by the underlying choices about maximizing utility subject to a budget constraint. And while economists may not be able to measure “utils,” they can certainly measure price and quantity demanded.

APPLICATIONS IN GOVERNMENT AND BUSINESS

The budget constraint framework for making utility-maximizing choices offers a reminder that people can react to a change in price or income in a range of different ways. For example, in the winter months of 2005, costs for heating homes increased significantly in many parts of the country as prices for natural gas and electricity soared, due in large part to the disruption caused by Hurricanes Katrina and Rita. Some people reacted by reducing the quantity demanded of energy; for example, by turning down the thermostats in their homes by a few degrees and wearing a heavier sweater inside. Even so, many home heating bills rose, so people adjusted their consumption in other ways, too. As you learned in the chapter on Elasticity, the short run demand for home heating is generally inelastic. Each household cut back on what it valued least on the margin; for some it might have been some dinners out, or a vacation, or postponing buying a new refrigerator or a new car. Indeed, sharply higher energy prices can have effects beyond the energy market, leading to a widespread reduction in purchasing throughout the rest of the economy.

A similar issue arises when the government imposes taxes on certain products, like it does on gasoline, cigarettes, and alcohol. Say that a tax on alcohol leads to a higher price at the liquor store, the higher price of alcohol causes the budget constraint to pivot left, and consumption of alcoholic beverages is likely to decrease. However, people may also react to the higher price of alcoholic beverages by cutting back on other purchases. For example, they might cut back on snacks at restaurants like chicken wings and nachos. It would be unwise to assume that the liquor industry is the only one affected by the tax on alcoholic beverages. Read the next Clear It Up to learn about how buying decisions are influenced by who controls the household income.

DOES WHO CONTROLS HOUSEHOLD INCOME MAKE A DIFFERENCE?

In the mid-1970s, the United Kingdom made an interesting policy change in its “child allowance” policy. This program provides a fixed amount of money per child to every family, regardless of family income. Traditionally, the child allowance had been distributed to families by withholding less in taxes from the paycheck of the family wage earner—typically the father in this time period. The new policy instead provided the child allowance as a cash payment to the mother. As a result of this change, households have the same level of income and face the same prices in the market, but the money is more likely to be in the purse of the mother than in the wallet of the father.

Should this change in policy alter household consumption patterns? Basic models of consumption decisions, of the sort

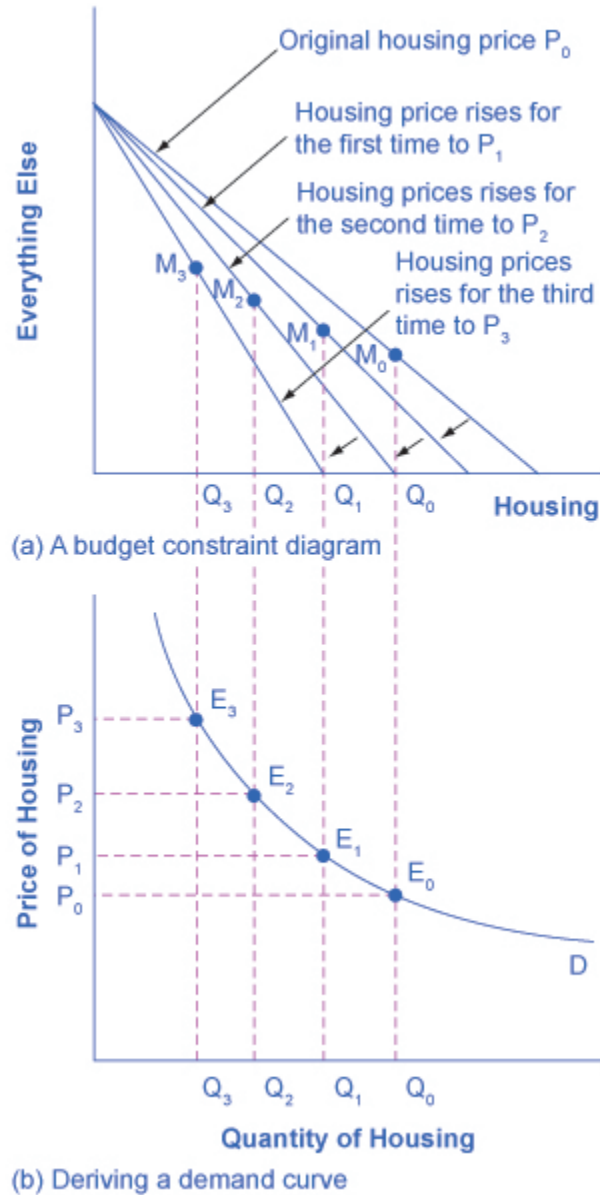


Figure 3. The Foundations of a Demand Curve: An Example of Housing. (a) As the price increases from P_0 to P_1 to P_2 to P_3 , the budget constraint on the upper part of the diagram shifts to the left. The utility-maximizing choice changes from M_0 to M_1 to M_2 to M_3 . As a result, the quantity demanded of housing shifts from Q_0 to Q_1 to Q_2 to Q_3 , *ceteris paribus*. (b) The demand curve graphs each combination of the price of housing and the quantity of housing demanded, *ceteris paribus*. Indeed, the quantities of housing are the same at the points on both (a) and (b). Thus, the original price of housing (P_0) and the original quantity of housing (Q_0) appear on the demand curve as point E_0 . The higher price of

housing (P_1) and the corresponding lower quantity demanded of housing (Q_1) appear on the demand curve as point E_1 .

examined in this chapter, assume that it does not matter whether the mother or the father receives the money, because both parents seek to maximize the utility of the family as a whole. In effect, this model assumes that everyone in the family has the same preferences.

In reality, the share of income controlled by the father or the mother does affect what the household consumes. When the mother controls a larger share of family income a number of studies, in the United Kingdom and in a wide variety of other countries, have found that the family tends to spend more on restaurant meals, child care, and women's clothing, and less on alcohol and tobacco. As the mother controls a larger share of household resources, children's health improves, too. These findings suggest that when providing assistance to poor families, in high-income countries and low-income countries alike, the monetary amount of assistance is not all that matters: it also matters which member of the family actually receives the money.

The budget constraint framework serves as a constant reminder to think about the full range of effects that can arise from changes in income or price, not just effects on the one product that might seem most immediately affected.

KEY CONCEPTS AND SUMMARY

The budget constraint framework suggest that when income or price changes, a range of responses are possible. When income rises, households will demand a higher quantity of normal goods, but a lower quantity of inferior goods. When the price of a good rises, households will typically demand less of that good—but whether they will demand a much lower quantity or only a slightly lower quantity will depend on personal preferences. Also, a higher price for one good can lead to more or less of the other good being demanded.

SELF-CHECK QUESTIONS

1. Explain all the reasons why a decrease in the price of a product would lead to an increase in purchases of the product.
2. As a college student you work at a part-time job, but your parents also send you a monthly “allowance.” Suppose one month your parents forgot to send the check. Show graphically how your budget constraint is affected. Assuming you only buy normal goods, what would happen to your purchases of goods?

REVIEW QUESTIONS

1. As a general rule, is it safe to assume that a change in the price of a good will always have its most significant impact on the quantity demanded of that good, rather than on the quantity demanded of other goods? Explain.
2. Why does a change in income cause a parallel shift in the budget constraint?

CRITICAL THINKING QUESTIONS

Income effects depend on the income elasticity of demand for each good that you buy. If one of the goods you buy has a negative income elasticity, that is, it is an inferior good, what must be true of the income elasticity of the other good you buy?

PROBLEMS

If a 10% decrease in the price of one product that you buy causes an 8% increase in quantity demanded of that product, will another 10% decrease in the price cause another 8% increase (no more and no less) in quantity demanded?

GLOSSARY

income effect a higher price means that, in effect, the buying power of income has been reduced, even though actual income has not changed; always happens simultaneously with a substitution effect

substitution effect when a price changes, consumers have an incentive to consume less of the good with a relatively higher price and more of the good with a relatively lower price; always happens simultaneously with an income effect

SOLUTIONS

Answers to Self-Check Questions

1. This is the opposite of the example explained in the text. A decrease in price has a substitution effect and an income effect. The substitution effect says that because the product is cheaper relative to other things the consumer purchases, he or she will tend to buy more of the product (and less of the other things). The income effect says that after the price decline, the consumer could purchase the same goods as before, and still have money left over to purchase more. For both reasons, a decrease in price causes an increase in quantity demanded.
2. This is a negative income effect. Because your parents' check failed to arrive, your monthly income is less than normal and your budget constraint shifts in toward the origin. If you only buy normal goods, the decrease in your income means you will buy less of every product.

7.3 LABOR-LEISURE CHOICES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Interpret labor-leisure budget constraint graphs
- Predict consumer choices based on wages and other compensation
- Explain the backward-bending supply curve of labor

People do not obtain utility just from products they purchase. They also obtain utility from leisure time. Leisure time is time not spent at work. The decision-making process of a utility-maximizing household applies to what quantity of hours to work in much the same way that it applies to purchases of goods and services. Choices made along the **labor-leisure budget constraint**, as wages shift, provide the logical underpinning for the labor supply curve. The discussion also offers some insights about the range of possible reactions when people receive higher wages, and specifically about the claim that if people are paid higher wages, they will work a greater quantity of hours—assuming that they have a say in the matter.

According to the Bureau of Labor Statistics, U.S. workers averaged 38.6 hours per week on the job in 2014. This average includes part-time workers; for full-time workers only, the average was 42.5 hours per week. Table 10 shows that more than half of all workers are on the job 35 to 48 hours per week, but significant proportions work more or less than this amount.

Table 11 breaks down the average hourly compensation received by private industry workers, including wages and benefits. Wages and salaries are about three-quarters of total compensation received by workers; the rest is in the form of health insurance, vacation pay, and other benefits. The compensation workers receive differs for many reasons, including experience, education, skill, talent, membership in a labor union, and the presence of discrimination against certain groups in the labor market. Issues surrounding the inequality of incomes in a market-oriented economy are explored in the chapters on Poverty and Economic Inequality and Issues in Labor Markets: Unions, Discrimination, Immigration.

Hours Worked per Week	Number of Workers	Percentage of Workforce
1–14 hours	6.9 million	5.0%
15–34 hours	27.6 million	20.1%
35–40 hours	68.5 million	49.9%
41–48 hours	11.9 million	8.6%
49–59 hours	13.3 million	9.6%
60 hours and over	9.3 million	6.8%

Table 10. Persons at Work, by Average Hours Worked per Week in 2013 (Total number of workers: 137.7 million). (Source: <http://www.bls.gov/news.release/empst.t18.htm>)

Compensation, Wage, Salary, and Benefits	\$30.92 per hour
Wages and Salaries	\$20.92
Benefits	
Vacation	\$2.09
Supplemental Pay	\$0.84
Insurance	\$2.15
Health Benefits	\$2.36
Retirement and Savings	\$1.24
Defined Benefit	\$0.57
Defined Contribution	\$0.064
Legally Required	\$2.46

Table 11. Hourly Compensation: Wages, Benefits, and Taxes in 2014. (Source: <http://www.bls.gov/news.release/pdf/ecec.pdf>)

THE LABOR-LEISURE BUDGET CONSTRAINT

How do workers make decisions about the number of hours to work? Again, let's proceed with a concrete example. The economic logic is precisely the same as in the case of a **consumption choice budget constraint**, but the labels are different on a labor-leisure budget constraint.

Vivian has 70 hours per week that she could devote either to work or to leisure, and her wage is \$10/hour. The lower budget constraint in Figure 1 shows Vivian's possible choices. The horizontal axis of this diagram measures both leisure and labor, by showing how Vivian's time is divided between leisure and labor. Hours of leisure are measured from left to right on the horizontal axis, while hours of labor are measured from right to left. Vivian will compare choices along this budget constraint, ranging from 70 hours of leisure and no income at point S to zero hours of leisure and \$700 of income at point L. She will choose the point that provides her with the highest total utility. For this example, let's assume that Vivian's utility-maximizing choice occurs at O, with 30 hours of leisure, 40 hours of work, and \$400 in weekly income.

For Vivian to discover the labor-leisure choice that will maximize her utility, she does not have to place numerical values on the total and marginal utility that she would receive from every level of income and leisure. All that really matters is that Vivian can compare, in her own mind, whether she would prefer more leisure or more income, given the tradeoffs she faces. If Vivian can say to herself:

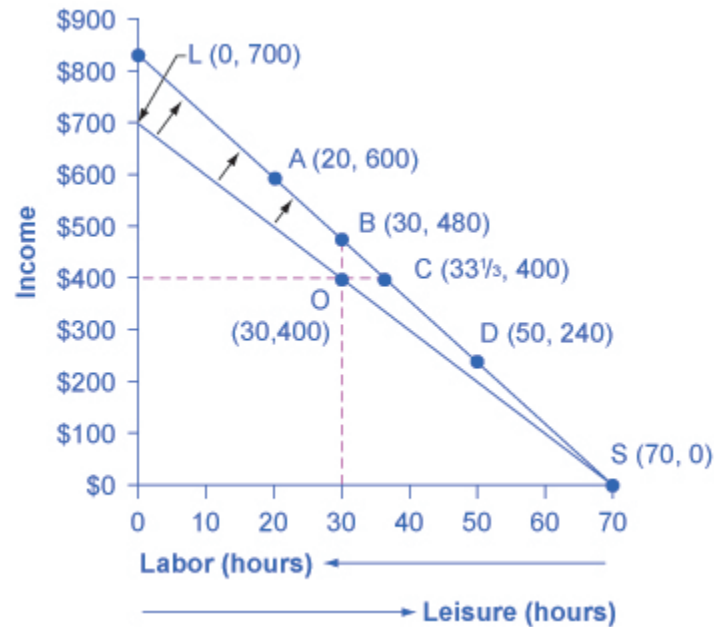


Figure 1. How a Rise in Wages Alters the Utility-Maximizing Choice. Vivian's original choice is point O on the lower opportunity set. A rise in her wage causes her opportunity set to swing upward. In response to the increase in wages, Vivian can make a range of different choices available to her: a choice like D, which involves less work; and a choice like B, which involves the same amount of work but more income; or a choice like A, which involves more work and considerably more income. Vivian's personal preferences will determine which choice she makes.

"I'd really rather work a little less and have more leisure, even if it means less income," or "I'd be willing to work more hours to make some extra income," then as she gradually moves in the direction of her preferences, she will seek out the utility-maximizing choice on her labor-leisure budget constraint.

Now imagine that Vivian's wage level increases to \$12/hour. A higher wage will mean a new budget constraint that tilts up more steeply; conversely, a lower wage would have led to a new budget constraint that was flatter. How will a change in the wage and the corresponding shift in the budget constraint affect Vivian's decisions about how many hours to work?

Vivian's choices of quantity of hours to work and income along her new budget constraint can be divided into several categories, using the dashed horizontal and vertical lines in Figure 1 that go through her original choice (O). One set of choices in the upper-left portion of the new budget constraint involves more hours of work (that is, less leisure) and more income, at a point like A with 20 hours of leisure, 50 hours of work, and \$600 of income (that is, 50 hours of work multiplied by the new wage of \$12 per hour). A second choice would be to work exactly the same 40 hours, and to take the benefits of the higher wage in the form of income that would now be \$480, at choice B. A third choice would involve more leisure and the same income at point C (that is, 33-1/3 hours of work multiplied by the new wage of \$12 per hour equals \$400 of total income). A fourth choice would involve

less income and much more leisure at a point like D, with a choice like 50 hours of leisure, 20 hours of work, and \$240 in income.

In effect, Vivian can choose whether to receive the benefits of her wage increase in the form of more income, or more leisure, or some mixture of these two. With this range of possibilities, it would be unwise to assume that Vivian (or anyone else) will necessarily react to a wage increase by working substantially more hours. Maybe they will; maybe they will not.

APPLICATIONS OF UTILITY MAXIMIZING WITH THE LABOR-LEISURE BUDGET CONSTRAINT

The theoretical insight that higher wages will sometimes cause an increase in hours worked, sometimes cause hours worked not to change by much, and sometimes cause hours worked to decline, has led to labor supply curves that look like the one in Figure 2. The bottom-left portion of the labor supply curve slopes upward, which reflects the situation of a person who reacts to a higher wage by supplying a greater quantity of labor. The middle, close-to-vertical portion of the labor supply curve reflects the situation of a person who reacts to a higher wage by supplying about the same quantity of labor. The very top portion of the labor supply curve is called a **backward-bending supply curve for labor**, which is the situation of high-wage people who can earn so much that they respond to a still-higher wage by working fewer hours. Read the following Clear It Up feature for more on the number of hours the average person works each year.

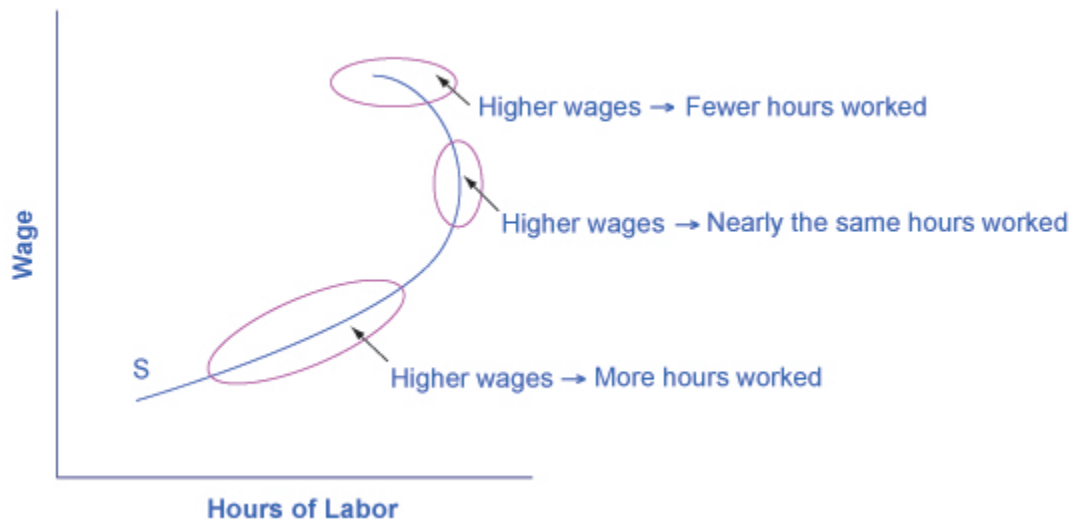


Figure 2. A Backward-Bending Supply Curve of Labor. The bottom upward-sloping portion of the labor supply curve shows that as wages increase over this range, the quantity of hours worked also increases. The middle, nearly vertical portion of the labor supply curve shows that as wages increase over this range, the quantity of hours worked changes very little. The backward-bending portion of the labor supply curve at the top shows that as wages increase over this range, the quantity of hours worked actually decreases. All three of these possibilities can be derived from how a change in wages causes movement in the labor-leisure budget constraint, and thus different choices by individuals.

IS AMERICA A NATION OF WORKAHOLICS?

Americans work a lot. Table 12 shows average hours worked per year in the United States, Canada, Japan, and several European countries, with data from 2013. To get a perspective on these numbers, someone who works 40 hours per week for 50 weeks per year, with two weeks off, would work 2,000 hours per year. The gap in hours worked is a little astonishing; the 250 to 300 hour gap between how much Americans work and how much Germans or the French work amounts to roughly six to seven weeks less of work per year. Economists who study these international patterns debate the extent to which average Americans and Japanese have a preference for working more than, say, Germans, or whether German workers and employers face particular kinds of taxes and regulations that lead to fewer hours worked. Many countries have laws that regulate the work week and dictate holidays and the standards of “normal” vacation time vary from country to country. It is also interesting to take the amount of time spent working in context; it is estimated that in the late nineteenth century in the United States, the average work week was over 60 hours per week—leaving little to no time for leisure.

Country	Average Annual Hours Actually Worked per Employed Person
United States	1,824
Spain	1,799
Japan	1,759
Canada	1,751
United Kingdom	1,669
Sweden	1,585
Germany	1,443
France	1,441

Table 12. Average Hours Worked Per Year in Select Countries. (Source: <http://stats.oecd.org/Index.aspx?DataSetCode=ANHRS>)

The different responses to a rise in wages—more hours worked, the same hours worked, or fewer hours worked—are patterns exhibited by different groups of workers in the U.S. economy. Many full-time workers have jobs where the number of hours is held relatively fixed, partly by their own choice and partly by their employer’s practices. These workers do not much change their hours worked as wages rise or fall, so their supply curve of labor is inelastic. However, part-time workers and younger workers tend to be more flexible in their hours, and more ready to increase hours worked when wages are high or cut back when wages fall.

The backward-bending supply curve for labor, when workers react to higher wages by working fewer hours and having more income, is not observed often in the short run. However, some well-paid professionals, like dentists or accountants, may react to higher wages by choosing to limit the number of hours, perhaps by taking especially long vacations, or taking every other Friday off. Over a long-term perspective, the backward-bending supply curve for labor is common. Over the last century, Americans have reacted to gradually rising wages by working fewer hours; for example, the length of the average work-week has fallen from about 60 hours per week in 1900 to the present average of less than 40 hours per week.

Recognizing that workers have a range of possible reactions to a change in wages casts some fresh insight on a perennial political debate: the claim that a reduction in income taxes—which would, in effect, allow people to earn more per hour—will encourage people to work more. The leisure-income budget set points out that this connection will not hold true for all workers. Some people, especially part-timers, may react to higher wages by working more. Many will work the same number of hours.

Some people, especially those whose incomes are already high, may react to the tax cut by working *fewer* hours. Of course, cutting taxes may be a good or a bad idea for a variety of reasons, not just because of its impact on work incentives, but the specific claim that tax cuts will lead people to work more hours is only likely to hold for specific groups of workers and will depend on how and for whom taxes are cut.

KEY CONCEPTS AND SUMMARY

When making a choice along the labor-leisure budget constraint, a household will choose the combination of labor, leisure, and income that provides the most utility. The result of a change in wage levels can be higher work hours, the same work hours, or lower work hours.

SELF-CHECK QUESTIONS

1. Siddhartha has 50 hours per week to devote to work or leisure. He has been working for \$8 per hour. Based on the information in Table 13, calculate his utility-maximizing choice of labor and leisure time.

Leisure Hours	Total Utility from Leisure	Work Hours	Income	Total Utility from Income
0	0	0	0	0
10	200	10	80	500
20	350	20	160	800
30	450	30	240	1,040
40	500	40	320	1,240
50	530	50	400	1,400

Table 13.

2. In Siddhartha's problem, calculate marginal utility for income and for leisure. Now, start off at the choice with 50 hours of leisure and zero income, and a wage of \$8 per hour, and explain, in terms of marginal utility how Siddhartha could reason his way to the optimal choice, using marginal thinking only.

REVIEW QUESTIONS

1. How will a utility-maximizer find the choice of leisure and income that provides the greatest utility?
2. As a general rule, is it safe to assume that a higher wage will encourage significantly more hours worked for all individuals? Explain.

CRITICAL THINKING QUESTIONS

1. In the labor-leisure choice model, what is the price of leisure?
2. Think about the backward-bending part of the labor supply curve. Why would someone work less as a result of a higher wage rate?

3. What would be the substitution effect and the income effect of a wage increase?
4. Visit the BLS website and determine if education level, race/ethnicity, or gender appear to impact labor versus leisure choices.

GLOSSARY

backward-bending supply curve for labor the situation when high-wage people can earn so much that they respond to a still-higher wage by working fewer hours

SOLUTIONS

Answers to Self-Check Questions

1. This problem is straightforward if you remember leisure hours plus work hours are limited to 50 hours total. If you reverse the order of the last three columns so that more leisure corresponds to less work and income, you can add up columns two and five to find utility is maximized at 10 leisure hours and 40 work hours:

Leisure Hours	Total Utility from Leisure	Work Hours	Income	Total Utility from Income	Total Utility from Both
0	0	50	400	1,400	1,400
10	200	40	320	1,240	1,440
20	350	30	240	1,040	1,390
30	450	20	160	800	1,250
40	500	10	80	500	1,000
50	530	0	0	0	530

Table 14.

2. Begin from the last table and compute marginal utility from leisure and work:

Leisure Hours	Total Utility from Leisure	MU from Leisure	Work Hours	Income	Total Utility from Income	MU from Income
0	0	–	50	400	1,400	160
10	200	200	40	320	1,240	200
20	350	150	30	240	1,040	240
30	450	100	20	160	800	300
40	500	50	10	80	500	500
50	530	30	0	0	0	–

Table 15.

Suppose Sid starts with 50 hours of leisure and 0 hours of work. As Sid moves up the table, he trades 10 hours of leisure for 10 hours of work at each step. At (40, 10), his $MU_{\text{Leisure}} = 50$, which is substantially less than his MU_{Income} of 500. This shortfall signals Sid to keep trading leisure for work/income until at (10, 40) the marginal utility of both is equal at 200. This is the sign that he should stop here, confirming the answer in question 1.

7.4 INTERTEMPORAL CHOICES IN FINANCIAL CAPITAL MARKETS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Evaluate the reasons for making intertemporal choices
- Interpret an intertemporal budget constraint
- Analyze why people in America tend to save such a small percentage of their income

Rates of saving in America have never been especially high, but they seem to have dipped even lower in recent years, as the data from the Bureau of Economic Analysis in Figure 1 show. A decision about how much to save can be represented using an **intertemporal budget constraint**. Household decisions about the quantity of financial savings show the same underlying pattern of logic as the consumption choice decision and the labor-leisure decision.



Figure 1. Personal Savings as a Percentage of Personal Income. Personal savings were about 7 to 11% of personal income for most of the years from the late 1950s up to the early 1990s. Since then, the rate of personal savings has fallen substantially, although it seems to have bounced back a bit since 2008. (Source: <http://www.bea.gov/newsreleases/national/pi/pinewsrelease.htm>)

The discussion of financial saving here will not focus on the specific financial investment choices, like bank accounts, stocks, bonds, mutual funds, or owning a house or gold coins. The characteristics of

these specific financial investments, along with the risks and tradeoffs they pose, are detailed in the Labor and Financial Markets chapter. Here, the focus is saving in total—that is, on how a household determines how much to consume in the present and how much to save, given the expected rate of return (or interest rate), and how the quantity of saving alters when the rate of return changes.

USING MARGINAL UTILITY TO MAKE INTERTEMPORAL CHOICES

Savings behavior varies considerably across households. One factor is that households with higher incomes tend to save a larger percentage of their income. This pattern makes intuitive sense; a well-to-do family has the flexibility in its budget to save 20–25% of income, while a poor family struggling to keep food on the table will find it harder to put money aside.

Another factor that causes personal saving to vary is personal preferences. Some people may prefer to consume more now, and let the future look after itself. Others may wish to enjoy a lavish retirement, complete with expensive vacations, or to pile up money that they can pass along to their grandchildren. There are savers and spendthrifts among the young, middle-aged, and old, and among those with high, middle, and low income levels.

Consider this example: Yelberton is a young man starting off at his first job. He thinks of the “present” as his working life and the “future” as after retirement. Yelberton’s plan is to save money from ages 30 to 60, retire at age 60, and then live off his retirement money from ages 60 to 85. On average, therefore, he will be saving for 30 years. If the rate of return that he can receive is 6% per year, then \$1 saved in the present would build up to \$5.74 after 30 years (using the formula for compound interest, $\$1(1 + 0.06)^{30} = \5.74). Say that Yelberton will earn \$1,000,000 over the 30 years from age 30 to age 60 (this amount is approximately an annual salary of \$33,333 multiplied by 30 years). The question for Yelberton is how much of those lifetime earnings to consume during his working life, and how much to put aside until after retirement. This example is obviously built on simplifying assumptions, but it does convey the basic life-cycle choice of saving during working life for future consumption after retirement.

Figure 2 and Table 16 show Yelberton’s **intertemporal budget constraint**. Yelberton’s choice involves comparing the utility of present consumption during his working life and future consumption after retirement. The rate of return that determines the slope of the intertemporal budget line between present consumption and future consumption in this example is the annual interest rate that he would earn on his savings, compounded over the 30 years of his working life. (For simplicity, we are assuming that any savings from current income will compound for 30 years.) Thus, in the lower budget constraint line on the figure, future consumption grows by increments of \$574,000, because each time \$100,000 is saved in the present, it compounds to \$574,000 after 30 years at a 6% interest rate. If some of the numbers on the future consumption axis look bizarrely large, remember that this occurs because of the power of compound interest over substantial periods of time, and because the figure is grouping together all of Yelberton’s saving for retirement over his lifetime.

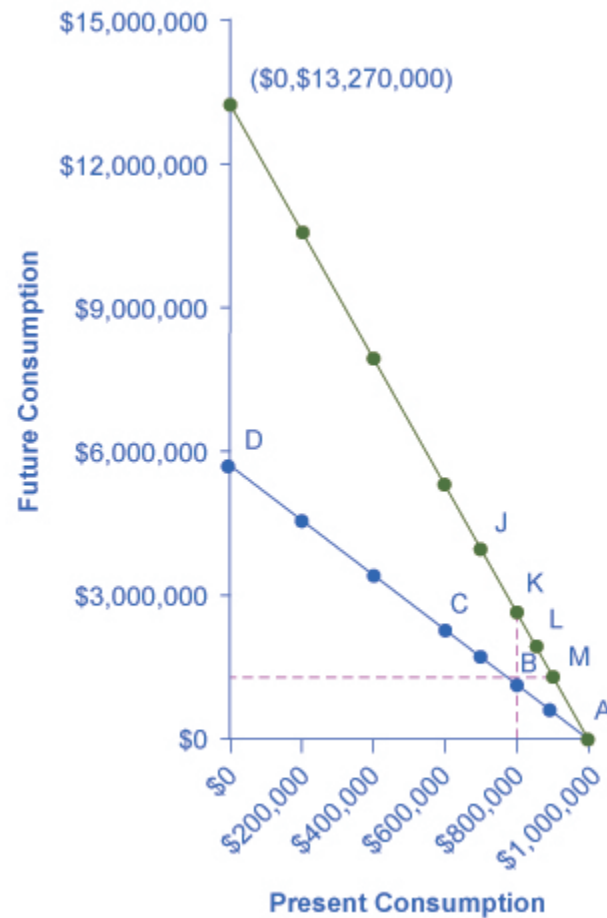


Figure 2. Yelberton's Choice: The Intertemporal Budget Set.

Yelberton will make a choice between present and future consumption. With an annual rate of return of 6%, he decides that his utility will be highest at point B, which represents a choice of \$800,000 in present consumption and \$1,148,000 in future consumption. When the annual rate of return rises to 9%, the intertemporal budget constraint pivots up. Yelberton could choose to take the gains from this higher rate of return in several forms: more present saving and much higher future consumption (J), the same present saving and higher future consumption (K), more present consumption and more future consumption (L), or more present consumption and the same future consumption (M).

Present Consumption	Present Savings	Future Consumption (6% annual return)	Future Consumption (9% annual return)
\$1,000,000	0	0	0
\$900,000	\$100,000	\$574,000	\$1,327,000
\$800,000	\$200,000	\$1,148,000	\$2,654,000
\$700,000	\$300,000	\$1,722,000	\$3,981,000
\$600,000	\$400,000	\$2,296,000	\$5,308,000
\$400,000	\$600,000	\$3,444,000	\$7,962,000
\$200,000	\$800,000	\$4,592,000	\$10,616,000
0	\$1,000,000	\$5,740,000	\$13,270,000

Table 16. Yelburton's Intertemporal Budget Constraint

Yelberton will compare the different choices along the **budget constraint** and choose the one that provides him with the highest utility. For example, he will compare the utility he would receive from a choice like point A, with consumption of \$1 million in the present, zero savings, and zero future consumption; point B, with present consumption of \$800,000, savings of \$200,000, and future consumption of \$1,148,000; point C, with present consumption of \$600,000, savings of \$400,000, and future consumption of \$2,296,000; or even choice D, with present consumption of zero, savings of \$1,000,000, and future consumption of \$5,740,000. Yelberton will also ask himself questions like these: "Would I prefer to consume a little less in the present, save more, and have more future consumption?" or "Would I prefer to consume a little more in the present, save less, and have less future consumption?" By considering marginal changes toward more or less consumption, he can seek out the choice that will provide him with the highest level of utility.

Let us say that Yelberton's preferred choice is B. Imagine that Yelberton's annual rate of return raises from 6% to 9%. In this case, each time he saves \$100,000 in the present, it will be worth \$1,327,000 in 30 years from now (using the formula for compound interest that $\$100,000 (1 + 0.09)^{30} = \$1,327,000$). A change in rate of return alters the slope of the intertemporal budget constraint: a higher rate of return or interest rate will cause the budget line to pivot upward, while a lower rate of return will cause it to pivot downward. If Yelberton were to consume nothing in the present and save all \$1,000,000, with a 9% rate of return, his future consumption would be \$13,270,000, as shown on Figure 2.

As the rate of return rises, Yelberton considers a range of choices on the new intertemporal budget constraint. The dashed vertical and horizontal lines running through the original choice B help to illustrate his range of options. One choice is to reduce present consumption (that is, to save more) and to have considerably higher future consumption at a point like J above and to the left of his original choice B. A second choice would be to keep the level of present consumption and savings the same, and to receive the benefits of the higher rate of return entirely in the form of higher future consumption, which would be choice K.

As a third choice Yelberton could have both more present consumption—that is, less savings—but still have higher future consumption because of the higher interest rate, which would be choice like L, above and to the right of his original choice B. Thus, the higher rate of return might cause Yelberton to save more, or less, or the same amount, depending on his own preferences. A fourth choice would be that Yelberton could react to the higher rate of return by increasing his current consumption and

leaving his future consumption unchanged, as at point M directly to the right of his original choice B. The actual choice of what quantity to save and how saving will respond to changes in the rate of return will vary from person to person, according to the choice that will maximize each person's utility.

APPLICATIONS OF THE MODEL OF INTERTEMPORAL CHOICE

The theoretical model of the intertemporal budget constraint suggests that when the rate of return rises, the quantity of saving may rise, fall, or remain the same, depending on the preferences of individuals. For the U.S. economy as a whole, the most common pattern seems to be that the quantity of savings does not adjust much to changes in the rate of return. As a practical matter, many households either save at a fairly steady pace, by putting regular contributions into a retirement account or by making regular payments as they buy a house, or they do not save much at all. Of course, some people will have preferences that cause them to react to a higher rate of return by increasing their quantity of saving; others will react to a higher rate of return by noticing that with a higher rate of return, they can save less in the present and still have higher future consumption.

One prominent example in which a higher rate of return leads to a lower savings rate occurs when firms save money because they have promised to pay workers a certain fixed level of pension benefits after retirement. When rates of return rise, those companies can save less money in the present in their pension fund and still have enough to pay the promised retirement benefits in the future.

This insight suggests some skepticism about political proposals to encourage higher savings by providing savers with a higher rate of return. For example, **Individual Retirement Accounts (IRAs)** and **401(k)** accounts are special savings accounts where the money going into the account is not taxed until it is taken out many years later, after retirement. The main difference between these accounts is that an IRA is usually set up by an individual, while a 401(k) needs to be set up through an employer. By not taxing savings in the present, the effect of an IRA or a 401(k) is to increase the return to saving in these accounts.

IRA and 401(k) accounts have attracted a large quantity of savings since they became common in the late 1980s and early 1990s. In fact, the amount of IRAs rose from \$239 million in 1992 to \$3.7 billion in 2005 to over \$5 billion in 2012, as per the Investment Company Institute, a national association of U.S. investment companies. However, overall U.S. personal savings, as discussed earlier, actually dropped from low to lower in the late 1990s and into the 2000s. Evidently, the larger amounts in these retirement accounts are being offset, in the economy as a whole, either by less savings in other kinds of accounts, or by a larger amount of borrowing (that is, negative savings). The following Clear It Up further explores America's saving rates.

A rise in interest rates makes it easier for people to enjoy higher future consumption. But it also allows them to enjoy higher present consumption, if that is what these individuals desire. Again, a change in prices—in this case, in interest rates—leads to a range of possible outcomes.

HOW DOES AMERICA'S SAVING RATES COMPARE TO OTHER COUNTRIES?

By international standards, Americans do not save a high proportion of their income, as Table 17 shows. The rate of gross national saving includes saving by individuals, businesses, and government. By this measure, U.S. national savings amount

to 17% of the size of the U.S. GDP, which measures the size of the U.S. economy. The comparable world average rate of savings is 22%.

Country	Gross Domestic Savings as a Percentage of GDP
China	51%
India	30%
Russia	28%
Mexico	22%
Germany	26%
Japan	22%
Canada	21%
France	21%
Brazil	15%
United States	17%
United Kingdom	13%

Table 17. National Savings in Select Countries. (Source: <http://data.worldbank.org/indicator/NY.GNS.ICTR.ZS>)

THE UNIFYING POWER OF THE UTILITY-MAXIMIZING BUDGET SET FRAMEWORK

The choices of households are determined by an interaction between prices, budget constraints, and personal preferences. The flexible and powerful terminology of utility-maximizing gives economists a vocabulary for bringing these elements together.

Not even economists believe that people walk around mumbling about their marginal utilities before they walk into a shopping mall, accept a job, or make a deposit in a savings account. However, economists do believe that individuals seek their own satisfaction or utility and that people often decide to try a little less of one thing and a little more of another. If these assumptions are accepted, then the idea of utility-maximizing households facing budget constraints becomes highly plausible.

BEHAVIORAL ECONOMICS: AN ALTERNATIVE VIEWPOINT

As we know, people sometimes make decisions that seem “irrational” and not in their own best interest. People’s decisions can seem inconsistent from one day to the next and they even deliberately ignore ways to save money or time. The traditional economic models assume rationality, which means that people take all available information and make consistent and informed decisions that are in their best interest. (In fact, economics professors often delight in pointing out so-called “irrational behavior” each semester to their new students, and present economics as a way to become more rational.)

But a new group of economists, known as behavioral economists, argue that the traditional method leaves out something important: people’s state of mind. For example, one can think differently about money if one is feeling revenge, optimism, or loss. These are not necessarily irrational states of mind, but part of a range of emotions that can affect anyone on a given day. And what’s more, actions under these conditions are indeed predictable, if the underlying environment is better understood. So, **behavioral economics** seeks to enrich the understanding of decision-making by integrating the insights of psychology into economics. It does this by investigating how given dollar amounts can

mean different things to individuals depending on the situation. This can lead to decisions that appear outwardly inconsistent, or irrational, to the outside observer.

The way the mind works, according to this view, may seem inconsistent to traditional economists but is actually far more complex than an unemotional cost-benefit adding machine.

For example, a traditional economist would say that if you lost a \$10 bill today, and also got an extra \$10 in your paycheck, you should feel perfectly neutral. After all, $-\$10 + \$10 = \$0$. You are the same financially as you were before. However, behavioral economists have done research that shows many people will feel some negative emotion—anger, frustration, and so forth—after those two things happen. We tend to focus more on the loss than the gain. This is known as **loss aversion**, where a \$1 loss pains us 2.25 times more than a \$1 gain helps us, according to the economists Daniel Kahneman and Amos Tversky in a famous 1979 article in the journal *Econometrica*. This insight has implications for investing, as people tend to “overplay” the stock market by reacting more to losses than to gains. Indeed, this behavior looks irrational to traditional economists, but is consistent once we understand better how the mind works, these economists argue.

Traditional economists also assume human beings have complete self-control. But, for instance, people will buy cigarettes by the pack instead of the carton even though the carton saves them money, to keep usage down. They purchase locks for their refrigerators and overpay on taxes to force themselves to save. In other words, we protect ourselves from our worst temptations but pay a price to do so. One way behavioral economists are responding to this is by setting up ways for people to keep themselves free of these temptations. This includes what are called “nudges” toward more rational behavior rather than mandatory regulations from government. For example, up to 20 percent of new employees do not enroll in retirement savings plans immediately, because of procrastination or feeling overwhelmed by the different choices. Some companies are now moving to a new system, where employees are automatically enrolled unless they “opt out.” Almost no-one opts out in this program and employees begin saving at the early years, which are most critical for retirement.

Another area that seems illogical is the idea of mental accounting, or putting dollars in different mental categories where they take different values. Economists typically consider dollars to be **fungible**, or having equal value to the individual, regardless of the situation.

You might, for instance, think of the \$25 you found in the street differently from the \$25 you earned from three hours working in a fast food restaurant. The street money might well be treated as “mad money” with little rational regard to getting the best value. This is in one sense strange, since it is still equivalent to three hours of hard work in the restaurant. Yet the “easy come-easy go” mentality replaces the rational economizer because of the situation, or context, in which the money was attained.

In another example of mental accounting that seems inconsistent to a traditional economist, a person could carry a credit card debt of \$1,000 that has a 15% yearly interest cost, and simultaneously have a \$2,000 savings account that pays only 2% per year. That means she pays \$150 a year to the credit card company, while collecting only \$40 annually in bank interest, so she loses \$130 a year. That doesn't seem wise.

The “rational” decision would be to pay off the debt, since a \$1,000 savings account with \$0 in debt is the equivalent net worth, and she would now net \$20 per year. But curiously, it is not uncommon for people to ignore this advice, since they will treat a loss to their savings account as higher than the

benefit of paying off their credit card. The dollars are not being treated as fungible so it looks irrational to traditional economists.

Which view is right, the behavioral economists' or the traditional view? Both have their advantages, but behavioral economists have at least shed a light on trying to describe and explain behavior that has historically been dismissed as irrational. If most of us are engaged in some "irrational behavior," perhaps there are deeper underlying reasons for this behavior in the first place.

"EENY, MEENY, MINEY, MOE"—MAKING CHOICES

In what category did consumers worldwide increase their spending during the recession? Higher education. According to the United Nations Educational, Scientific, and Cultural Organization (UNESCO), enrollment in colleges and universities rose one-third in China and almost two-thirds in Saudi Arabia, nearly doubled in Pakistan, tripled in Uganda, and surged by three million—18 percent—in the United States. Why were consumers willing to spend on education during lean times? Both individuals and countries view higher education as the way to prosperity. Many feel that increased earnings are a significant benefit of attending college.

Bureau of Labor Statistics data from May 2012 supports this view, as shown in Figure 3. They show a positive correlation between earnings and education. The data also indicate that unemployment rates fall with higher levels of education and training.

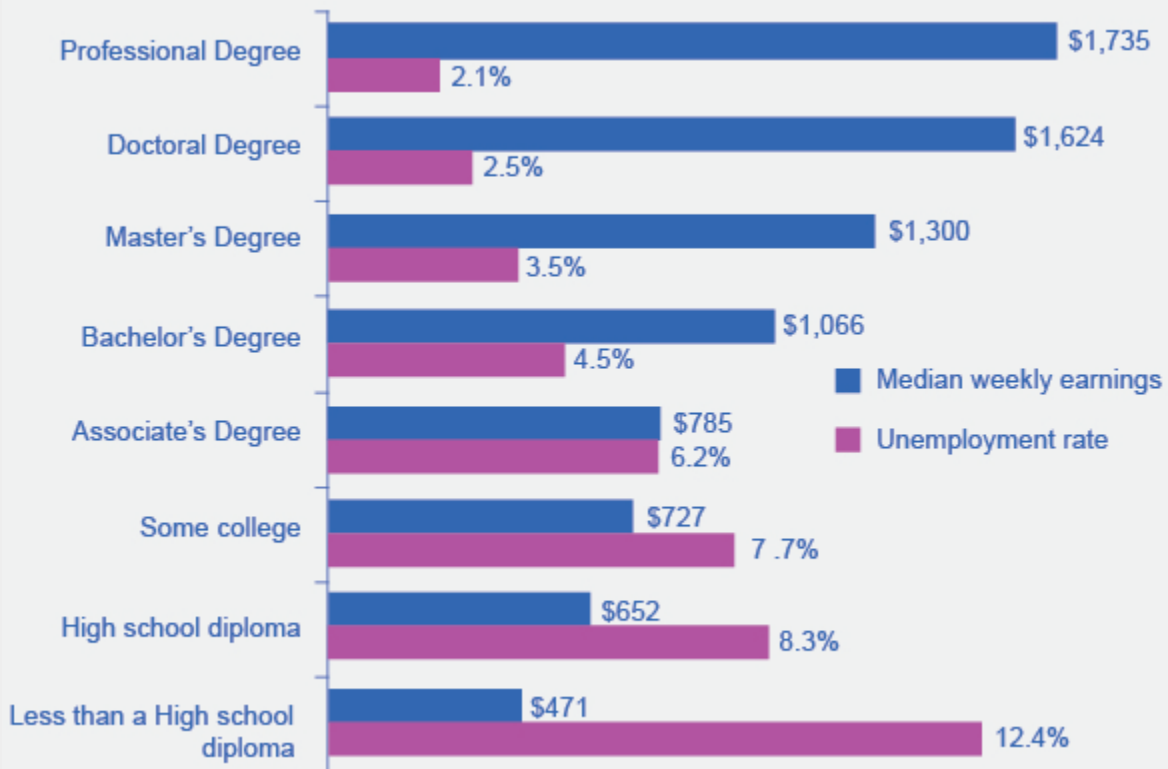


Figure 3. The Impact of Education on Earnings and Unemployment Rates, 2012. Those with the highest degrees in 2012 had substantially lower unemployment rates whereas those with the least formal education suffered from the highest unemployment rates. The national median average weekly income was \$815, and the nation unemployment average in 2012 was 6.8%. (Source: Bureau of Labor Statistics, May 22, 2013)

KEY CONCEPTS AND SUMMARY

When making a choice along the intertemporal budget constraint, a household will choose the combination of present consumption, savings, and future consumption that provides the most utility. The result of a higher rate of return (or higher interest rates) can be a higher quantity of saving, the same quantity of saving, or a lower quantity of saving, depending on preferences about present and future consumption. Behavioral economics is a branch of economics that seeks to understand and explain the “human” factors that drive what traditional economists see as people’s irrational spending decisions.

SELF-CHECK QUESTIONS

1. How would an increase in expected income over one’s lifetime affect one’s intertemporal budget constraint? How would it affect one’s consumption/saving decision?
2. How would a decrease in expected interest rates over one’s working life affect one’s intertemporal budget constraint? How would it affect one’s consumption/saving decision?

REVIEW QUESTIONS

1. According to the model of intertemporal choice, what are the major factors which determine how much saving an individual will do? What factors might a behavioral economist use to explain savings decisions?
2. As a general rule, is it safe to assume that a lower interest rate will encourage significantly lower financial savings for all individuals? Explain.

CRITICAL THINKING QUESTIONS

1. What do you think accounts for the wide range of savings rates in different countries?
2. What assumptions does the model of intertemporal choice make that are not likely true in the real world and would make the model harder to use in practice?

REFERENCES

Holden, Sarah, and Daniel Schrass. 2012. “The rise of IRAs in U.S Households’ Saving for Retirement, 2012.” *ICI Research Perspective* 18.8 (2012). <http://www.ici.org/pdf/per18-08.pdf>.

Kahneman, Daniel and Amos Tversky. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47.2 (March 1979) 263-291.

Thaler, Richard H. “Shifting Our Retirement Savings into Automatic.” *The New York Times*, April 6, 2013.

<http://www.nytimes.com/2013/04/07/business/an-automatic-solution-for-the-retirement-savings-problem.html?pagewanted=all>.

UNESCO Institute for Statistics. “Statistics in Brief / Profiles” Accessed August 2013. [http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=121](http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=121&IF_Language=en&BR_Country=5580)

&IF_Lan-

guage=en&BR_Country=5580.

GLOSSARY

behavioral economics a branch of economics that seeks to enrich the understanding of decision-making by integrating the insights of psychology and by investigating how given dollar amounts can mean different things to individuals depending on the situation.

fungible the idea that units of a good, such as dollars, ounces of gold, or barrels of oil are capable of mutual substitution with each other and carry equal value to the individual.

SOLUTIONS

Answers to Self-Check Questions

1. An increase in expected income would cause an outward shift in the intertemporal budget constraint. This would likely increase both current consumption and saving, but the answer would depend on one's time preference, that is, how much one is willing to wait to forgo current consumption. Children are notoriously bad at this, which is to say they might simply consume more, and not save any. Adults, because they think about the future, are generally better at time preference—that is, they are more willing to wait to receive a reward.
2. Lower interest rates would make lending cheaper and saving less rewarding. This would be reflected in a flatter intertemporal budget line, a rotation around the amount of current income. This would likely cause a decrease in saving and an increase in current consumption, though the results for any individual would depend on time preference.

CHAPTER 8. CHALLENGING THE ROLE OF UTILITARIANISM

THE ROLE OF VALUE(S) IN THE ECONOMICS DISCIPLINE

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Value Theory in Economics
- Value as it is Defined by Orthodox Economics
- The Relationship between Utilitarian Philosophy and the Economics discipline.
- Critique of the Utilitarian Approach.
- The Application in Economics of Philosophies other than Utilitarianism.

8.1 ECONOMICS AND VALUE

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define the term value.
- Identify different interpretations of the term value.

The term **value** and the concept of **value theory** have taken on two distinct definitions in relation to economic theory. The first definition of the term value pertains to the role of normative analysis within economic theory. As discussed in chapter 1, for orthodox economic thinkers there is little room for normative thinking in the rigorous world of scientific economics. Subjective or moral judgments are discouraged in favor of utilizing a positive, value-free approach that is not a statement of anyone's value judgment or subjective feelings, but rather of what "is." By the standard orthodox textbook definition, the term "value" relates to personal biases and subjective beliefs. Also, recall from Chapter 2, the positive versus normative conflict presented by orthodox economics represents a false dichotomy. As has been discussed by non-orthodox economic theorists, subjective values tend to be present in all economic paradigms and are unavoidable.

The second definition of value that is common in economics is the one most people think of when they think of the term "value." For most people when they think of the value of something, such as a product, they are simultaneously thinking of the price of that product. Therefore, the term value and the term price tend to be used interchangeably. As a result, for many economists it is important to develop a theory of value because the theory of value provides the theorist with a sense of the origin of prices. After all, if an economist can identify the root source of value, then the same economist will also understand the root origin of prices. Since prices are incredibly important informational signals, being able to explain from where prices originate adds an important depth to economic theory.

On the surface both usages of the term "value," as described above, appear to be mutually exclusive. However, closer inspection reveals two interrelated issues. Scientists, economists included, already tend to, whether knowingly or unknowingly, assert subjective values within their theories. Values not only generate a necessary entanglement of facts and values, but they then become present throughout the very act of theorizing. In the case of economics, this means that the economic ideas that theorists present will include elements of their basic worldview or, as some would say, their philosophy.

The remainder of this chapter is laid out in such a manner as to explore the role of philosophy in economics. First, the philosophical foundation of orthodox economics, and its history, is explored. As will be shown, orthodox economics relies upon utilitarian philosophy for its value-based under-

standing of human action and motivation. As such, orthodox economics believes that the concept of utility holds the key to understanding the origin of prices. Additionally, utilitarianism forms the basis for the ethical perspective presented by orthodox economics. Upon closely examining the orthodox story of utility, the chapter moves on to identify important limitations that develop as a result of the use of utility. The chapter then moves on to provide alternative examples from heterodox economics, pointing out philosophical worldviews outside of utilitarianism that can form the basis of economic thought. Importantly, the alternative approaches presented do not suffer the pitfalls that are present within the orthodox, utilitarian approach.

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify the history of utility and its relationship to economic ideas.
- Define use-value
- Contrast Adam Smith and Jeremy Bentham with regard to their view of utility.
- Contrast cardinal and ordinal utility
- Define revealed preference theory

For those interested in the study of economics the concept of utility has a long history. Going back to the time of antiquity, great thinkers such as Aristotle made reference to the importance of the creation of use-value. From an Aristotelian standpoint use-value is the idea that human material production ought to be directed toward the creation of things that benefit human beings. Aristotle prioritized usefulness and discouraged wastefulness.

The role of the utility theory of value as a central component of the preferred ideological vision of orthodox economists of today, in some ways dates back to the time of Adam Smith. Smith, continuing the Aristotelian line of thinking, also saw significance in the concept of use-value.

Use-Value – The idea that the value of an object is based on how useful the product is to the consumer. Products have an intrinsic value to their consumers.

For Smith though, use-value represented the perceived benefit that people would derive from consumption. For Smith, consumers would be more apt to buy products that they perceived as useful as opposed to products that they believed are not useful. Still, Smith's analysis remains a long way from how utility is utilized by orthodox economists.

Although Smith does not go into much further development beyond his acknowledgement that consumers attach some kind of use-value to useful items, it is clear that Smith opens the door to the potential development of the idea that utility plays a role in determining the prices of products. Consider, if utility, or use-value, is important to consumers, then, presumably, utility must play a role in determining consumers' interest in purchasing a product. Because consumer demand is a reflection of consumer purchasing interests, consumer demand will be essential to the determination of the prices of products. It is with the evolution of economic ideas in the 19th century that the relationship between utility and prices is cemented within orthodox economic thought.

In order to understand the role of utility in contemporary orthodox economics, the place to begin is

with the early 19th century philosopher and political economist Jeremy Bentham. Bentham possibly had the most prominent influence on the philosophical development of utility. Bentham's perspective is known as utilitarianism. Bentham describes utility as the cornerstone of the "Greatest Happiness Principle." In Bentham's usage, utility evolves beyond a description of usefulness into the embodiment of happiness. Utilitarianism espouses the belief that people are pleasure maximizers and pain minimizers. Therefore, any action that a person engages in must be driven by the desire either to gain happiness or to reduce pain. Taken to the field of economics, in Bentham's view, people desire products because those products bring the person happiness. Clearly, Bentham's perspective had a profound effect on orthodox economics, given the many references that this textbook has made in Chapters 2 and 6 to consumers being driven to maximize their utility.

Bentham's story, for all of its applicability to contemporary orthodox economics, is still only a partial analysis. One area in which Bentham's argument is incomplete is in his treatment of the measurability of utility. When describing how a consumer may assign utility to a product, Bentham freely employed both **cardinal** and **ordinal** measures of utility.

Cardinal utility – Is the idea that the utility a consumer assigns to a product is objective and, therefore, quantifiably measurable.

Ordinal utility – Is the idea that the utility that a consumer assigns to a product is subjective, in the mind of the consumer, and is, therefore, not directly quantifiably measurable.

As a rule Bentham seemed to generally acknowledge ordinal utility as being conceptually sound in comparison to cardinal utility. Bentham understood that cardinal utility cannot be objectively measured, a person cannot be attached to a device that measures units of happy (utils). There are, however, indirect measures that can allude to cardinal utility, such as the order in which consumers purchase products. Given two products with an identical price, if a consumer purchases one of the products prior to the other, it stands to reason that the first product purchased retain greater utility than the second product. Regardless, in absolute terms, how much utility a consumer actually assigns to a product cannot be empirically measured, rendering it subjective to the consumer. Limitations aside, Bentham frequently developed examples that required cardinal measures of utility.

Another area in which Bentham's analysis was incomplete pertained to the relationship between utility and the origins of prices of products. Recall the principle of diminishing marginal utility described in Chapter 7. Bentham anticipates the idea of diminishing marginal utility by describing instances in which a person can acquire more of something but feel less satisfaction from the additional unit than they received from earlier units. Bentham's depiction, however, remains abstract rather than appearing "scientifically" concrete.

Resolving some of the incomplete features of Bentham's story are the marginalists of the 1870s. Independent of one another, a group of thinkers triggered what is now referred to as the marginalist revolution. The marginalists famously apply differential calculus to utility analysis. Differential calculus focuses on rates of change between two variables. In the case of utility, differential calculus can be used to tell a story about how, as a consumer consumes more of a product, the happiness derived from that product will change. The marginalists attempted to demonstrate, in a more concrete way, the idea that as more of something is consumed, the satisfaction or utility per unit would decline. By applying calculus, and drawing pictures to reflect the conclusions drawn from the mathematical presentation, the marginalists create the appearance of a scientific advancement in the study of utility in

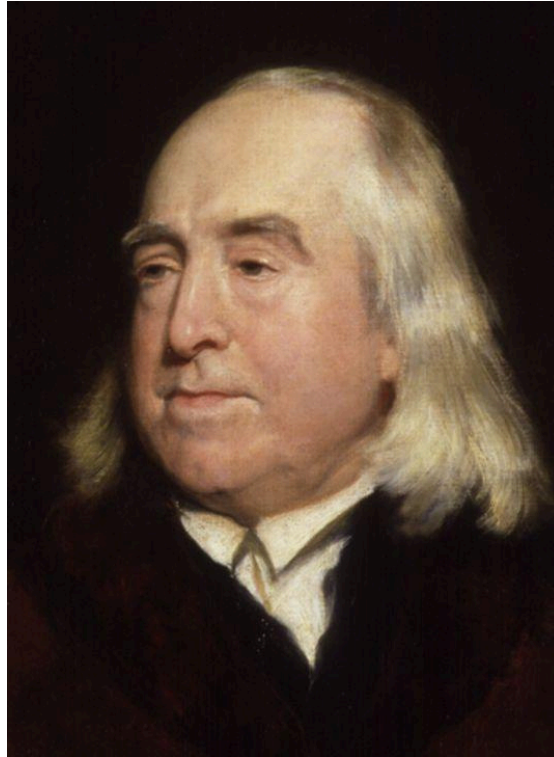


Figure 1. Portrait of Jeremy Bentham (Source: Wellcome Images, Wikimedia, CC-BY 4.0)

economics. In terms of legacy, the marginalists use of calculus is a hallmark of modern orthodox economic theory.

As much as the marginalists influence the apparatus of modern orthodox economics, their analysis still remained incomplete in that the marginalists do not resolve the issue of ordinal versus cardinal utility. It is not until the 20th century that orthodox economic theorists settle on the use of ordinal utility as the foundation of their measurement of utility. The thinker most responsible for solidifying ordinal utility's place within economics is Paul Samuelson. Paul Samuelson develops a theory known as Revealed Preference. On the surface, revealed preference theory articulates the idea that the amount of utility that any one person derives from the consumption of a product is purely subjective to the consumer in question. However, once a consumer identifies that which brings them the greater happiness, the consumer then reveals their preferences to everyone by consuming the product that they desire. In this respect, revealed preference theory seeks to quantify utility by identifying utility as a consequence of a consumer's actions.

THE APPLICATION OF UTILITARIANS WITHIN ORTHODOX ECONOMICS AND ITS IMPLICATIONS

Whether an unwitting or a conscious choice, utility continues to be the preferred philosophical starting point for contemporary orthodox economists. Utility also carries with it a specific interpretation of human actions and interactions. Recall, for orthodox economics, the human being is a hedonist continually in search of greater pleasure or diminished pain. Any socially determined characteristics are given (not analyzed, merely accepted), yielding an isolated utility maximizer. Also, recall, that this notion of a hedonistic person is deeply controversial with respect to its application in economics. There simply does not exist irrefutable evidence that individuals only pursue products because

those products provide the individual with pleasure. Human motive can be defined in many ways with hedonism being just one of many interpretations.

Regardless of its accuracy or inaccuracy, orthodox economists perceive that, for individuals, utility is maximized through the act of exchange. At any particular point in time, an individual, possessing an initial endowment (income), begins engaging in the act of exchange with some other individual economic agent. As all exchanges are assumed to be voluntary, any exchange that two parties are willing to engage in must be an exchange that will make both parties better off, otherwise neither party would engage in the exchange. Because being made better off means, according to orthodox economic theory, acquiring more utility, the act of exchanging is the act increasing one's holdings of utility. Any and all exchanges then must be increasing individual, and subsequently, society's total utility. Presumably, once some kind of round of exchanges is complete, any and all parties that engaged in exchanges are now better off because now they have more utility as a result of their exchanges.

Of course, no one really knows whether or not the result of these exchanges really has made the participants better off or worse off because there is no device that measures utility. Utility is purely subjective and non-empirical, which means that the revealed preference argument yields an outcome in which it must be assumed that all exchange participants are made better, otherwise they would not have participated in the exchange. Assuming people are better off does not mean that people are actually better off.

GLOSSARY

use-value

the idea that the value of an object is based on how useful the product is to the consumer. Products have an intrinsic value to their consumers

cardinal utility

the idea that the utility a consumer assigns to a product is objective and, therefore, quantifiably measurable

ordinal utility

the idea that the utility that a consumer assigns to a product is subjective, in the mind of the consumer, and is, therefore, not directly quantifiably measurable

8.3 UTILITY AND PARETO OPTIMALITY: THE ORTHODOX ECONOMIC VIEW OF SOCIAL WELFARE

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define Pareto optimality
- Explain how orthodox determines the conditions for societal well-being to be maximized.
- Analyze the ethical issues surrounding Pareto optimality.

For orthodox economists the ideal outcome for an economy is an outcome in which Pareto optimality is achieved. The concept of Pareto optimality owes its origins to a 19th century Italian mathematician Vilfredo Pareto. Stated simply, the Pareto criterion for determining whether an economy has produced the “best” or “ideal” outcome is fulfilled when economic outcomes are such that there is no way to make any one or many people better off without making any one person or many worse off.

On its own the Pareto criterion for social well-being is an attractive proposition. After all if someone is harmed in order to improve the plight of someone else or many others, then it appears obvious that someone is being granted preferential treatment at the expense of someone else or many others. The granting of preferential treatment hardly seems fair or equitable. In this context, the utilization of the Pareto criterion eliminates the need to make those choices.

Additionally, the granting of preferential treatment opens the door to a long series of ethical questions that can be avoided by applying the Pareto criterion. On what basis is the decision to harm or benefit being made? How or when are interventions that inflict harm or bestow benefits decided? If an intervention does take place, who is making the decision to inflict harm or bestow benefits? What is the degree of harm or benefit triggered by an intervention? What is the ethical basis for intervening? In many ways, whenever a government must make budgetary decisions, it is asking and answering these questions. For example, perhaps government policymakers would like to expand the size of the military. Expanding the size of the military requires that the government finance the expansion of the military. Financing military expansion may require that other government spending programs be reduced. Alternatively, perhaps taxes will be raised to pay for the military expansion. Either way, whomever is responsible for financing the military expansion is directly paying for someone else to benefit. In defense of expanding the military, policymakers may have to justify to the public why the public will, presumably, benefit from the expansion of the military. The Pareto criterion appears to clearly answer these questions. If a society knows when it is in a position of maximum benefit, then,

on the basis of the Pareto criterion, no justification can be made to either harm or benefit anyone, causing all of the above questions become moot.

Something important has now been revealed. The Pareto criterion only becomes applicable when there is a measure of what it means for someone, or many, to benefit and for someone, or many, to be harmed. If some kind of metric exists, then, once measured, a society will know when no one person or many can be made better off without someone or many being harmed. Therefore, what is essential to the Pareto criterion is the mechanism a theorist uses to measure and determine a society's overall well being. For orthodox economists, the measure of individual and social well-being is the amount of utility that individuals and society has derived.

UTILITY MAXIMIZATION AND PARETO OPTIMALITY: ETHICAL IMPLICATIONS

Given the nature of utility maximization, human social interaction is reduced to exchange between individuals. Clearly the utility assumption generates a distinct depiction of human beings, human needs, and human society. The utility maximizing assumption also generates a specific conception of what is an economically "just" society. In other words, orthodox economics has a specific view of what it believes is the best way for society to be organized economically.

The orthodox worldview can be neatly summarized. If people desire utility, then people can acquire utility by exchanging with others. Once people exchange with others, the act of exchanging makes them better off. Once all exchanges are complete, all people are as well off as they can become economically. If everyone is as well off as they can become, then no one, or many, can be made better off without someone, or many, being made worse off. As far as orthodox economics is concerned, society has reached its ideal, Pareto optimal, outcome because people were free to make choices within an economy that facilitates exchanges, namely a market based economy.

At this point it should be clear that utilitarianism provides the philosophical foundation supporting orthodox economics while Pareto optimality represents the final outcome of the orthodox social welfare theory. If utility is the initial proposition on which orthodox theory is built, and Pareto optimality is the culmination of orthodox economic theory, then the utility and Pareto optimality must be intrinsically linked.

The dilemma for orthodox economics is that by accepting the utility theory of value and adopting the Pareto criterion as a social welfare measure, orthodox economics produces a very specific, and ethically limited, framework. For example, by focusing exclusively on the point of exchange, orthodox economics tends to ignore situations in which people are in disagreement with one another, circumstances of conflict. The reason why conflict is generally ignored is because the conditions of exchange are assumed in advance. For orthodox economics, the institutions of exchange, such as markets, contract law, and contract enforcement, are assumed to be in place prior to economic agents exchanging. Additionally, by assuming that initial endowments are given, orthodox economics assumes the existence of income on the part of trading parties, but generally fails to recognize where income originates.

The trouble for orthodox economics is that it is not hard to imagine a world in which one's income is affected by discriminatory factors such as racism, sexism, sexual orientation or any other divisive perception. In fact, imagination is not necessary, as the real world is filled with examples of discrimination and exploitation. Regardless of form, the outcome of discrimination is the same, some are able to exploit others by virtue of their power to discriminate. Or, in another example, it is clear that

conditions of inequality can empower some relative to others. People that disproportionately benefit from economic outcomes are endowed with more income and have the power to acquire resources. In instances in which the acquisition of resources becomes concentrated in the hands of a small number of people, those with great wealth and/or income may have the power to exploit their positions of power at the expense of others. Of course, with exploitation, conditions of inequality may be worsened allowing unequal situations to become further entrenched.

Conscientious observers of a capitalist market system, including utilitarians, are aware that situations of exploitation are not only present, but virtually unavoidable. Given the existence of conflict and inherent disparities of power, it is easy to imagine that someone might want to intervene in economic affairs in order to level the playing field. For example, someone might advocate to pass laws to penalize discriminatory behavior so as to reduce its prevalence, or others might seek to impose taxes on some so as to redistribute income or wealth in the hopes of reducing inequality. The interventions in questions would certainly be designed to make someone, or many, better off at the expense of someone else, or many others. As a result, no matter how justified the interventionist actions may be, and they are certainly justifiable in the examples provided above, within the utilitarian and Pareto criterion context the interventionist actions cannot be supported!

If the significance of utilitarianism and the Pareto criterion remains obscure, then perhaps less abstract and much more obvious examples may be useful. The following situations, using subjective measures of utility, are Pareto sub-optimal and cannot be justified:

1. The defeat of Nazi Germany because Adolf Hitler was made worse off.
2. The end of slavery in the American South because slave owners were made worse off.

Clearly, to any reasonable observer the defeat of Nazi Germany and the end of slavery in the American South are justifiable as they ended horrible chapters in human history. However, by following the utilitarian belief and coupling it to the Pareto criterion, orthodox economics cannot really make any argument to defend the destruction of Nazi Germany and the conclusion of slavery in the American South. Subjectively speaking, Hitler's loss of utility may have more than offset the world's gain in utility and thus made the world worse off. Or, subjectively speaking, the loss of utility to slave owners in the South may have more than offset the gain in utility experienced by the former slaves. After all, if no quantifiable measure is possible it is actually impossible to know whether or not total utility has become larger or smaller so there is no way for a utilitarian to justify an intervention in the above, albeit extreme, cases. Additionally, any intervention in the above cases required someone or many being made worse off in order to make some or many better off. The utilitarian tradition has no adequate way of applying anything more than some kind of hedonistic, happiness seeking, value judgment to human action. Obviously, there are serious moral and ethical weaknesses associated with the acceptance of utilitarianism.

In summary, the significance and value-laden implications of accepting utility, appears to elude most orthodox economists. In neoclassical theory price and utility are equalized, specifically price and marginal utility. Therefore, price equals value and this concludes the story. The neoclassical oversimplification of value theory, which treats value and price as synonyms, fails to recognize that by believing in utility, the orthodox economist is believing in a set of pre-ordained value judgments. The importance of this point cannot be overstated. Observe the significance of prices in a capitalist economy as described in this textbook. Marveling at how fluidly consumers and producers respond to price changes in markets the textbook states:

“...adjustments in response to price changes happen all of the time in a market economy, often so smoothly and rapidly that we barely notice them...For both the U.S. economy and the world economy as a whole, markets – that is, demand and supply – are the primary social mechanism for answering the basic questions about what is produced, how it is produced, and for whom it is produced.” (pg. 71)

Because prices send the signals that determine the allocation of resources in a market economy, and the allocation of resources has tremendous social ramifications, then the question of whether or not the allocation of resources is “socially just” must be addressed. Because value theory provides the foundation for which the quantitative origin of prices, it is value theory then that also provides the basis for the normative criterion from which prices and the allocation of resources may be interpreted. Clearly, the choice of value theory is extremely important.

Unfortunately, for orthodox economics, the normative implications of the utilitarian assumption generally lay dormant.

As non-orthodox economists are well aware, value theory acts as a cohesive, it is the glue that connects economic theory to a society’s moral and ethical norms. Any complete theoretical depiction of economic interaction necessarily requires a theory of value. As such, it is important to self-reflect on what values have been adopted by a theory and how those values impact how economic outcomes are interpreted. Given the importance of value theory, the one dominant proposition presented throughout this paper is that it is absolutely necessary for any economic theorist to anchor their theoretical ideas to a theory of value. Within this context, however, the utility theory of value presents orthodox economists with a dual horned dilemma. To accept utility as the basis for their analysis means that orthodox economists are implicitly valuing any and all market driven economic outcomes, regardless of the negative impacts that those outcomes may have on large segments of the society. To the contrary, to argue against any and all market outcomes as just, an economist, orthodox or otherwise, necessarily must begin their theorizing applying a different set of values outside of the value of endless hedonism that is central to utilitarian beliefs.

GLOSSARY

Pareto optimality (the Pareto criterion)

when economic outcomes are such that there is no way to make any one or many people better off without making any one person or many worse off

8.4 ABANDONING THE NORMATIVE CONSTRAINTS OF UTILITARIANISM

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the basic elements of Amartya Sen's Capabilities approach
- Define capabilities and functionings
- Explain the Marxian Labor Theory of Value and its implications

Given the limitations and challenges associated with utilitarianism, many students of economics seek out alternative approaches for measuring economic well-being and interpreting the societal implications of economic outcomes. The following describes, two alternative, heterodox, approaches for interpreting economic conditions.

TWO NORMATIVE EXAMPLES FROM HETERODOX ECONOMICS

ABANDONING UTILITY: THE CAPABILITIES APPROACH AND THE ETHICAL CONSIDERATIONS OF AMARTYA SEN

Where do the seeds of discontent sprout? In 1998 Amartya Sen was awarded the Nobel Prize in economics. Sen is a fascinating case study as he provides an example of an economist who is not a utilitarian. Early in his career Sen had abandoned utilitarianism. What prompted Sen to abandon utility as a first principle? The answer to this question can be found at the beginning of Sen's career when Sen encountered a utilitarian tradition in economics that had shed nearly all the conventional characteristics of ethical analysis. While the early 19th century utilitarian thinkers such as Bentham had identified and described some of the ethical issues surrounding the idea of diminishing marginal utility, by the late 20th century utilitarianism in economics had been reduced to revealed preference theory. For Sen, utilitarianism provides entirely too-narrow an assessment of what it means to be a person. The result is that there are far too many limits associated with utilitarianism for the utilitarian to be able to make effective, and informed, moral evaluations.

Taken further, Sen also recognizes the link between the moral vacuum produced by utilitarian philosophy and the neoclassical application of the Pareto criterion. For Sen, utilitarianism and the Pareto criterion are like a form of normative handcuffs. Neither the distribution of income nor initial utilities factor in to the choice making, exchange story, promoted by orthodox economics. By ignoring the initial endowments of income or utility, Sen argues that a condition of Pareto optimal may also be

one of deep inequality and hardship. Those without property and with little income often times find themselves without property and little income even though they have engaged in market exchange.

Admittedly, a person's actual freedom to exchange may provide a person with a degree of additional utility. Clearly being free to make choices is empowering and valuable to individuals. However, one's freedom to act does not mean that a person is in a remotely decent position in the market. The fact is, a market outcome can be Pareto optimal and still be less than socially optimal. For example, instead of demanding change people in poor situations often simply adjust to their negative circumstances. The deprived find some area of existence that gives them pleasure and make do. The result is a disadvantaged yet acquiescent populace that not only continues to function, but may be falsely perceived as happy or not too sad. In this example, utilitarianism generates a persistently unfair outcome within a population of people ill-equipped to effectively do anything to alter that outcome. This utilitarian picture highlights a tendency to under-estimate the plight of people in negative circumstances and causes Sen to conclude that neoclassical welfare economics is incapable of addressing the ethical issues that accompany the existence of poverty and oppression.

In terms of the problems associated with the distribution of utilities, Sen criticizes the utilitarian notion that every action in life is based on pleasure or happiness. There is no equality to be found in adding up everyone's utility to find a sum of total utility. It is possible for a maximum of total social utility to be generated under a circumstance in which one person receives the largest share of the available utility while many others receive very little. Further, it is not feasible to correct for an unequal distribution of utilities generated by market outcomes. The only way to rectify the impasse requires that the theorist be able to engage in interpersonal comparisons of well-being. In other words, people need to be able to compare how well off they are relative to other people. Relative comparisons are not something to which utility is adept at explaining. Given all of the problems Sen identifies, he departs from utilitarian precepts and develops an alternative measure of economic well-being.

THE CAPABILITIES APPROACH

The capabilities approach represents Sen's alternative to the neoclassical welfare model. The capabilities approach is a theoretical device that allows for interpersonal welfare comparisons, while also maintaining the ability and freedom of people to make choices. For Sen, the interpersonal comparisons of well-being should be based on a group of what Sen calls "functionings."

Functionings – Represents a state of being for an individual as well as the collection of all things an individual can be doing. Functionings can be as basic as being well-nourished and properly sheltered to as challenging as flying an airplane.

In turn, the measure of social welfare is the "capability" of people to access and develop their "functionings."

Capability – Is the availability of functionings coupled with the ability of the individual to achieve a combination of functionings.

The combination of functionings that are available and the ability of an individual to achieve a combination of functionings is the measure of an individual's well-being. Consider the following table.

	Limited Capabilities	Greater Capabilities
Limited Functionings	Low Welfare	Moderate Welfare
Greater Functionings	Moderate Welfare	High Welfare

Based on the capabilities approach, having many options available, such as states of being or activities to pursue, and then having the freedom to pursue those options is the ideal. The highest social welfare is achieved when a society produces an environment that meets basic needs while also provide people with ample opportunities to pursue challenging and life enriching activities. To the contrary, having limited access to functionings, and limited freedom in the pursuit of those functionings is the least advantageous position. In the case of moderate welfare outcomes, either individuals have many possible functionings available, but a limited freedom to pursue those functionings, or a limited set of available functionings but substantial freedom to pursue those goals.

CAPABILITIES AND FUNCTIONINGS: EXAMPLES OF WELFARE OUTCOMES

Limited Functionings/Limited Capabilities

Represents a case in which a person or a group of people are living with clearly definable limits to their **functionings** as defined by abject poverty, malnourishment and/or starvation and with little to no medical care. In this scenario the person or people are simply seeking to survive from one day to the next, there is little to no long term planning or goals. Possibly the result of social and political unrest such as a civil war (the current situation in Syria for example), or a weather calamity such as a drought, or a socioeconomic order such as the slave economy of the pre-Civil War American South.

Given the horrendous living conditions and limited functioning, any person living under these conditions would likely attempt make changes in their lives in order to try and reduce their immediate suffering. The ability to make changes, their **capabilities**, may not be possible as they may be living under conditions of political, military, and economic repression. Absent the freedom to maneuver, the person or people in this circumstance are rendered virtually helpless to the situation in which they have befallen. For example, unless the slave is freed, as property, the slave simply cannot pursue other options

Limited Functionings/Greater Capabilities

Represents a case in which a person or people are again deeply limited in terms of having basic needs and few wants met. In this instance, however, the person or people in this circumstance have some options to escape their plight. For example, consider the scenario of political unrest, if greater capabilities are present, the person or people can possibly vacate the region in which the unrest is unfolding. Evacuation of refugees to other more stable regions of the world create capabilities for the now displaced residents from a place devoid of functionings to places with both functionings and resources to expand capabilities. Another example could be the freeing of slaves in the American South during the post-Civil War period.

Greater Functionings/Limited Capabilities

Represents a case in which a person or people have their basic needs and wants met, while enjoying the luxury of having many other possible functionings available within the society such as educational, health care, and employment opportunities. In this instance, however, the person or people in this circumstance do not have options available to access the abundance of functionings. For example, consider the scenario of repressive political forces that are the result of situations of patriarchal social structures, or general totalitarian principles. A society characterized by repressive political regimes will restrict the freedom of its populace to enjoy the varied functionings that exist. A historic example would be the Jewish population that was subjected to the ruling principles of Nazi Germany.

Greater Functionings/Greater Capabilities

Represents a case in which a person or people have their basic needs and wants met, while enjoying the luxury of having many other possible functionings available within the society such as educational, health care, and employment opportuni-

ties. In this instance, however, the person or people in this circumstance also have a great many options available to access the abundance of functionings. For example, consider the example of the majority of people living in economies characterized as advanced developed economies with largely democratic political institutions. Certainly there remain conditions of repression such as discrimination, sexism, racism, and any number of limiting conditions. However, given democratic political structures and civil society outlets, people are able to continue to press for ever-expanding capabilities.

The capabilities approach then is not just a story of the freedom to pursue well-being, such as is the case in the story provided by orthodox economics. Instead, the capabilities approach is asking questions about initial endowments, available functionings, as well as the distributional outcomes associated with the pursuit of those functionings. Given the capabilities approach, the social welfare analysis of economics is no longer reduced to the action of choice making, but rather to both the conditions preceding choice as well as the outcomes associated with choices. While functionings can vary for different people across social circumstances, as well as time and place, so long as people have the ability to obtain the outlined functionings then they have an opportunity to elevate their well-being. Clearly, the capabilities approach represents for Sen, and others, a framework capable of generating more ethically pleasing results than that of the utility-tainted Pareto criterion.

In conclusion, Sen has managed to escape the ethical handcuffs of utilitarianism. The capabilities approach offers Sen, and other economists interested in exploring the capabilities approach, a new ethical theory. Sen has clearly avoided the fate of many neoclassical economists, i.e. being able to study economics absent ethics.

THE MARXIAN APPROACH AND THE LABOR THEORY OF VALUE

Another approach to applying normative reasoning toward economic phenomena is the Marxian labor theory of value. In stark contrast to the utility theory of value, the preanalytic vision associated with the labor theory of value produces a view of humanity and human need that is virtually antithetical to the utilitarian depiction. Rather than experiencing social interaction through exchange, within the labor theory of value human beings are presented as engaged in the act of production.

The basic proposition on which the labor theory of value is constructed is the idea that before something can be exchanged it must be produced. As a result, the value of products is generated at the point of production, where the products are created. At the point of production, labor is the essential component of production processes. While machines can enhance the productivity of human labor, machines only exist as a result of the collective accumulation of human labor over time. Importantly, machines themselves do not create labor. After all, a nail cannot hammer itself. The implication is that the process of production has always required human labor power as even the most sophisticated machines require human operators.

Additionally, production has always been a social endeavor as human beings work with one another to produce those things that satisfy the material wants of people. Consider, for example, a loaf of bread that you see on a grocery store shelf. The production and distribution process necessary to bring the loaf of bread to the grocery store shelf is extensive, requiring the contribution of a great many people. For starters, the agricultural process that provides the wheat, and/or oats, and/or grains that are the primary ingredients for making bread have to be grown by farmers. The number of farmers involved may number in the hundreds or thousands. Additionally, the farmers will use tools and other inputs provided to them by other producers. Those producers, the producers of farm equipment, will also bring together laborers and equipment in a manufacturing process that will build farm equipment. By continuation, the tools and equipment used in the manufacturing process that builds farm equip-

ment will also have to be produced by laborers using other tools and equipment in other production processes, and so on and so forth. On the other end of the production process, after harvesting, the basic agricultural inputs used in making bread will be transported to processing facilities. The act of transportation requires that laborers operate transportation equipment like trains and trucks. Just as was the case with farm equipment, transportation equipment will be made by many laborers that will be utilizing many machines that will have been produced by many other laborers using many other machines, and so on and so forth. Of course, transportation equipment generally requires fuel. If the fuel is a petroleum byproduct, then the extraction of petroleum will once again require many laborers using significant amounts of equipment through many stages of production until a fuel is available that will be used to operate the transportation equipment. Over and over again, each element of production requires many other laborers in many other production processes in a seemingly endless and nearly incomprehensible scale of human interaction. Clearly, production is an inherently social process that only takes place when human beings apply themselves to a multitude of production processes

Because the sole universally necessary element in the production process is labor, labor becomes the only true and relevant source of value. Human history demonstrates that human survival necessitates that human beings engage first and foremost in the process of providing for material needs. If human beings exchange, it is only because they first produced an excess that they are willing to exchange. Exchange is not a necessary condition for human survival, whereas production is required for human existence.

When juxtaposed, the differences between the vision generated by the utility theory of value and the labor theory of value could not be more visible. Let's examine four primary differences.

First, orthodox economics downplays any necessary human social relationships, as individuals are viewed as isolated utility maximizers within orthodox economic thought. To the contrary, adherents to the labor theory of value see human behavior itself, including exchange relationships, as being the result of social relationships. By continuation, a strong argument can be made that there are also important historical understandings and divisions between the two value theories. For utility theorists, because human behavior is perceived as a universal constant across time, understanding economic history is not a necessary component of their theory. Labor theorists, however, must study history closely in order to attempt to understand and identify the important changes in social arrangements that have historically arisen in different socioeconomic systems.

Second, due to the fact that exchange is a voluntary form of social interaction, harmonious outcomes appear normal within utilitarian thinking. To the contrary, since labor is the source of value and production, if there are class differences in a society, those class differences will inevitably end in conflict. The result is that conflict, not harmony, is front and center for labor theorists.

Third, while neoclassical economists take as given individual preferences, and by implication assert that the act of choosing is always efficient, labor theorists argue that preferences arise from social relationships and social interactions. As such, adherents to the labor theory of value do not accept the choice-driven revealed preference argument on social welfare.

Fourth, whereas orthodox economists identify individual incomes as originating through a voluntary process of exchange between individuals whose preferences are given, the labor theorist argues that the income that flows to non-laboring classes is the byproduct of a process of **exploitation**.

Exploitation – The idea that a producer or producers of a product receive compensation that is less than the value of what the producer or producers contributed to the production process.

Clearly the utility theory of value and the labor theory of value yield two different and competing conceptions of human existence, producing two different normative criteria for evaluating human interaction.

It would appear that the choice of value theory and recognition of its normative impact should be of the utmost importance to a theorist. This is not always the case. Perhaps due to the obscure relationship between price and value in neoclassical theory, there are many neoclassical theorists who directly violate the philosophical foundations of their accepted theory of value and do so by attempting to temper the less-than-savory ethical outcomes present in a capitalist market system. Alternatively, while most heterodox thinkers openly challenge the ethical outcomes of the market system and the conclusions brought about by the use of utility, many very prominent heterodox thinkers also either reject the need for a theory of value or do not explicitly endorse a formal theory of value. Sen falls under the latter category. In both instances value theory's importance is being inadequately recognized. While continued criticism of neoclassical theory is undoubtedly warranted, heterodox economists, regardless of their lofty accomplishments, prestige, and respectability, should not be absolved of deserved criticism. Within this premise, an evaluation of the relationship between value theory and Sen's important contribution to the critique of neoclassical welfare economics will be the focus of the remainder of this paper.

GLOSSARY

functionings

a state of being for an individual as well as the collection of all things an individual can be doing

capability

the availability of functionings coupled with the ability of the individual to achieve a combination of functionings

exploitation

the idea that a producer or producers of a product receive compensation that is less than the value of what the producer or producers contributed to the production process

8.5 CONCLUSION

Many students of economics are drawn to the discipline out of an interest in understanding the social and moral meaning and significance of economic outcomes. Interpreting the social and moral meaning of economic outcomes, however, requires a set of philosophical beliefs that can be effectively integrated and tied to economic activity. In this chapter we have seen how orthodox economics entwines utilitarian philosophy to economic activity. We have also seen the limitations to the utilitarian approach in that it gives the adherents to this tradition little to no room to challenge the outcomes associated with market economic activity. In this chapter we have also seen how heterodox economists, by not accepting the utilitarian tradition have utilized other philosophical approaches, such as the capabilities approach or the labor theory of value, to interpret economic events. In the case of these alternative philosophical approaches, market outcomes are not assumed to be socially just. For heterodox economists, as a result of alternative philosophical visions, questions of conflict, such as exploitation and discrimination, are directly confronted and possible solutions are provided.

CHAPTER 9. AN INSTITUTIONAL ANALYSIS OF MODERN CONSUMPTION

INTRODUCTION TO AN INSTITUTIONAL ANALYSIS OF MODERN CONSUMPTION



“The damned” by Lorenzo Paolini, depicting consumerism and the barbarization of modern culture, metaphorically. “Jailed where they think to freely run, they can’t stop because a delay means death.” (Caption quote from Paolini, Wikimedia, CC0 1.0)

WHO ARE WE AND WHAT DO WE WANT?

One of the more notorious elements of the neoclassical perspective in economics is its treatment of people as isolated, self-

interested individuals. These ‘economic agents’ are endowed with their own personal ‘utility functions’ and *de gustibus non est disputandum*: tastes are simply a matter of personal preference. The choices we make, then, are nothing more than the manifestation of our tastes as we calculate the most satisfaction we can get from our scarce resources—most importantly, our money, which we begrudgingly acquire through work.

Pause for a moment to think about the following questions: what does it mean to be human? What is our atomistic approach to utility maximizing agents saying about human nature? And finally, does this offer a complete explanation of why we consume the things we consume (and often times aspire to consume more in the future)?

CHAPTER OBJECTIVES

In this chapter, you will learn about

- Institutional Analysis
- Conspicuous Consumption
- The Complex World of Modern Consumption

The ‘hedonistic man’ (or *homo economicus*) of neoclassical economics is, as institutional economist Thorstein Veblen observed,

a lightning calculator of pleasures and pains, who oscillates like a homogeneous globule of desire of happiness under the impulse of stimuli that shift him about the area, but leave him intact. He has neither antecedent nor consequent. ... The hedonistic man is not a prime mover. He is not the seat of a process of living...

In point of fact, however, it seems obvious that

1. People do behave deliberately and consciously, in ways that affect the world around them – that is, they *are* prime movers, not simply reacting to external changes.
2. People change, they develop, in ways that cannot simply be construed as reallocations of scarce resources – that is, people *have* antecedents (where they came from) and consequents (where they’re going).

And, perhaps most importantly, people are social. *Homo sapiens* is a species that, unlike say *lumbricus terrestris* (earthworms), must live in groups. This fact has been long recognized. Take, for instance, Aristotle:

Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human. Society is something that precedes the individual. Anyone who either cannot lead the common life or is so self-sufficient as not to need to, and therefore does not partake of society, is either a beast or a god.

While human societies are made up of individuals, any given individual must come out of a society

which existed before she did. This would suggest that our approach to understanding individual behavior should start with an understanding of society, not the other way around. In this chapter you will be introduced to an alternative, heterodox analysis along these lines: the institutional approach; and you will explore the new insights that this analysis brings to the study of consumption in modern economies.

9.1 INSTITUTIONAL ANALYSIS

An economics that treats institutions as primary is necessary to understand the economic behavior of group-dependent individuals—that is, people making decisions within the context of their families, communities, industries, and other forms of organization, formal and informal. It is these social contexts and organizations to which heterodox economists are referring when they use the term **institutions**. A broad definition of the term might be stated as follows:

Institutions are collectively shared habits of thought—of knowing, doing, and valuing—that control, expand, and liberate individual action.

As Aristotle recognized some 24 centuries ago, these institutions exist prior to the individuals born into them. They are the rules we learn that tell us what to do and what not to do. But as constricting as that may sound, institutions also constitute a powerful tool for the human race: they allow individuals to forego solving problems that others have already solved (so that we don't have to 'reinvent the wheel', as it were); and they create reasonable expectations of others' behavior, allowing us to act collectively. Take, for instance, the simple collectively shared habit of thought concerning which side of the road to drive on. Knowing ahead of time which side we're to drive on saves us from the dangerous requirement of learning through trial and error. Likewise, this simple institution doesn't just restrict our decisions; it also produces expectations of what others will be doing, allowing us to travel more safely than we could without those expectations. In this manner, our actions are expanded beyond what they could otherwise be—we would, in fact, be more limited without this institution.

Of course, individuals are not—and could not be—made to conform precisely to the institutions into which they are born—society is not a rigid mold into which each person is poured at birth. We are 'prime movers' capable of making our own decisions and, to some extent, of changing the institutions of our society. But, it is the collective power that institutions give human communities as a whole which ultimately liberates us, that allows us to shape our own destinies and to participate in the collective shaping of our community's destiny as well.

Finally, as habits, institutions clearly must stay the same over some period of time. Yet, the technologies, laws, culture, and so on—that is, the institutions—of any given society evolve over time, too. This occurs through the accidental and intentional adaptation of how we think—our 'habits of thought'—to an ever-changing world around us. But, this creates an inherent tension within the concept of institutions: as the essential means by which humans collectively survive, institutions must constantly evolve to meet the proliferating exigencies we face; yet, as habits, they must have a resilience to change, a constancy, to be of any use at all. Recognizing this tension offers an important insight into the essential nature of present institutions: our collectively shared habits of thought are always a combination of useful ways of solving problems and the useless remnants of old ways of thinking.

With the basic tools and insights of institutional analysis in hand, we turn now to analyzing consump-

tion in modern society. As you'll see, the decisions we make when it comes to what we consume and why we make those decisions can look very different once we take this fundamentally distinct perspective as our starting point. Our understanding of how we act as consumers depends on our understanding of who we are as a species.

9.2 CONSPICUOUS CONSUMPTION

An institutional analysis of modern consumption patterns would, of course, look first and foremost to a society's institutions to explain these behaviors. Our tastes and preferences, in this view, are primarily the result, not of some abstract lightning calculation of marginal utilities per dollar, but rather of the common habits of thought shared among our peers, our family, and our society more generally. They do not emerge miraculously from each of our individual constitutions. Instead, they are the result of an institutional evolution, occurring in part by unplanned drift, and in part by intentional acts of problem solving, ceremonial observance, and persuasion.

In this view, we're consuming the things we consume today largely because we learned that those things are appropriate to consume. And, importantly, the institutions that define what is and is not appropriate do not entirely reflect what is most conducive to nurturing individual well-being, social cooperation, and sustainable uses of technology. Rather, they will reflect in part the imbecile and anachronistic habits of thought inherited from our past.

Take for instance one of Veblen's best-known concepts, conspicuous consumption: consuming goods not for their capacity to produce personal satisfaction (utility), but rather because they allow the consumer to demonstrate or enhance her prestige. Surely, a host of examples come to mind readily: designer clothing that serves no better as clothing than equivalents without the brand recognition; upscale dining establishments serving food that is no more nutritional than their more modest counterparts; sports cars that, while perhaps hypothetically capable of winning high-speed races, are no better at what they'll actually be used for (driving around town and sitting in traffic) than an ordinary automobile. What do all of these have in common?

While one could argue that each of these examples is a good or service of a higher quality than their cheaper alternatives, most people would easily recognize that there is an element in each of allowing one to conspicuously consume—to 'show off' the fact that one has the money to make purchases that others simply could not afford. The question becomes, why would someone go out of their way to spend money on things that, in fact, *don't* provide additional utility or value in use? And, as you might have guessed, answering this question requires an analysis of the institutions that guide our consumption decisions in modern society. The following breakout box points the way toward Veblen's ultimate conclusion.

Sometimes referred to as the first feminist economist, Thorstein Veblen published his "Economic Theory of Woman's Dress" in 1894, breaking down the characteristics of the apparel of his time that offered the consumer something more than simply protection from the elements. His three cardinal principles of women's dress, each of which was "advertising the fiction that they live without any gainful occupation," were:

1. Expensiveness: clothing “must be uneconomical”—that is, demonstrating that the wearer, or her family, has the means to waste money on items of little functional value. High-dollar brand names make this easier to recognize for the uninitiated; but ultimately the expense of the work and material that went into producing these items is a more subtle, sophisticated way of indicating the same.
2. Novelty: to the extent that clothing is useful, the wearer must not get the full use value from it. She must discard articles well before even insignificant wear starts to show. Observing seasonal changes in fashion helps to coordinate and enforce this sort of consumption observance.
3. Ineptitude: clothing must show that the wearer doesn’t have to participate in useful work. For instance, clothing might be severely binding, or footwear might restrict movement.

Can you think of any examples of consumption goods outside of women’s apparel that fit this description? My personal favorite comes from men’s professional clothing: the necktie. As you’re probably aware, neckties don’t serve much of a function—they’re certainly not providing warmth, and in fact they’re not even covering up an otherwise naked part of the body. So why do some men find themselves compelled not only to buy and wear these products, but to periodically discard the old ones and purchase new ones?

A common response from students is: they make you look professional. But, that just begs the question, *why* do they make you look professional—that is, why do we associate neckties with professionalism? Surely men in most professions don’t actually make use of the tie in their work (though I’ve found they make a convenient wipe for smudged eyeglasses). Consider Veblen’s cardinal principles above as applied to the necktie.

1. Expensiveness: while presumably whatever function neckties serve could be equally well-served were they made from a relatively inexpensive material like cotton (or, for that matter, paper), they tend instead to be made of more expensive materials like silk.
2. Novelty: those who follow men’s fashion will know that neckties are subject to similar style changes as other types of clothing—in terms, especially, of patterns and widths. Most people are probably able to discern a tie that was purchased in the 1980s versus the 2010s, much as most people can give a rough date to when a car was produced just by looking at it.
3. Ineptitude: paradoxically, to demonstrate professionalism, men are often expected to wear apparel that does more to hinder than to aid their capacity for work. The delicate and expensive fabric, for instance, indicates that the ‘professional man’ is not likely doing work that would involve ‘getting his hands dirty’ in the literal sense. Likewise, it would not typically be wise to wear a long necktie in a manufacturing context in which machines are being used, lest the tie get caught in the machine.

These are strange principles by which to guide your consumption goods, indeed; and certainly not ones that are limited to women’s dress—or even clothing in general. A little reflection on modern culture and the reader could probably see these principles in much of what we consume today. But, again, why are these apparently useless and even wasteful elements so desirable to the modern consumer? Why would someone want to waste money, or to show off they they don’t contribute to society? The remainder of this section will give a summary explanation.

It goes almost without saying that, as social creatures, we humans feel a sense of self-esteem based in part on how we think others see us. We strive to fit in, to avoid feeling inferior, and at times to

show our superiority to others—which, of course, must make others feel inferior. But, how we do so is dictated by the accepted social norms—the institutions—that we were born into; and those institutions have mutated through generations to produce a ‘strange’ set of criteria for who is superior and who is inferior.

In much earlier times one’s prowess, the respect afforded by one’s peers, would reflect chiefly the capacity to ably accomplish something that was useful to the group. As humans’ technologies improved, however, it became increasingly the case that:

1. The greatest contribution to current productivity came from the countless generations of workers, inventors, engineers, and so on from whom society had inherited its industrial arts. And,
2. As technologies became increasingly sophisticated, more and more people, each having increasingly specialized skills, was necessary to efficiently produce things with these more sophisticated technologies.

The problem with these developments is that it becomes very difficult to differentiate one person’s contribution from another’s, or, more accurately, from the general productivity of the group acting collectively.

Consequently, over a very long time the presumption developed that how much you contributed relative to others would simply be reflected in how much money you made (whether this is true or not is another matter altogether). But this presented a problem for establishing the amount of prestige others should afford you: how much you have in your bank account isn’t typically public information, and politeness dictates that you shouldn’t go around waving your most recent ATM account balance slip in people’s faces. The solution? Demonstrate your prowess, as evidenced by your vast riches, by wasting money wherever possible.

Part of this anthropological history, which can only be given in the most rudimentary of forms here, also demonstrates a division of employments in terms of prestige. In particular, early humans garnered prestige from exercising their power over the natural world—hunting and killing large game, or controlling the floods through irrigation channels, for instance. However, once technologies improved, it became increasingly the case that demonstrating power meant controlling or extracting from the productive community itself—that is, from people—more so than from nature. As a result, the higher classes came to be associated with occupations of predation, control, and leisure—for instance, politics and war, management and finance, celebrity and fashion, . With consumption as the chief way to demonstrate, and to reinforce in the minds of one’s peers, that one belongs to the ‘better’ class of people, our institutions began to prescribe an element of ineptitude, showing exemption from useful labor.

9.3 THE COMPLEX WORLD OF MODERN CONSUMPTION

The institutional analysis of conspicuous consumption demonstrates the importance of moving past notions of utility maximization, and beginning to look at the historical and cultural reasons behind individual consumers' decisions. Yet, consider the following

- Americans are exposed to hundreds, perhaps thousands, of advertisements every day
- Some consumers are willing to camp outside of stores for days just to be the first to purchase a new product
- The traditional holiday of Thanksgiving has now been paired with (if not eclipsed by) the so-called 'Black Friday' it precedes

These and many other examples might be taken to suggest that it is not only the individual decisions of consumers concerning what to buy that needs to be explained. Rather, we should also be looking to understand *consumerism*—the social preoccupation with acquiring and, most importantly, identifying with consumer goods—as well.

To understand the importance of advertising and consumerism today, the best place to start is with the work of institutional economist John Kenneth Galbraith. In *The Affluent Society* (1958) and later in *The New Industrial State* (1967), Galbraith argued forcefully that the historical conditions that classical and neoclassical economists were chiefly concerned with—namely, production of necessities amidst widespread poverty and the allocation of scarce resources—had been superseded by an age of relative affluence. Mass production, as you'll see in a later chapter of this textbook, had largely solved the problem of *producing* enough to adequately meet the needs of the people (though ensuring everyone has *access* to basic necessities remains a problem to this day). Of course, solving one problem presents new ones, and for Galbraith one of the most important problems society had come to face was how to ensure that there would be demand for what could be produced once our basic needs were met.

Enter the adman. As Galbraith argued, advertising and salesmanship became important once our basic needs were, for the most part, easily met, and our remaining desires were no long immediately evident to us. Advertising and salesmanship functioned, then, to 'synthesize, elaborate, and nurture' our desires for goods and services beyond those serving simply to keep us fed, clothed, sheltered, and so on. Another way to look at it: where businesses had come increasingly to serve customers in ways that went beyond providing the goods and services to meet basic needs, these businesses required a way to maintain demand—the willingness to buy—where hunger, cold, and so on would not. In this manner, advertising and salesmanship became a form of 'demand management', designed to keep sales up, to keep people buying even after they would have otherwise felt satisfied.

With the ascendancy of marketing, the products we buy come to take on special characteristics as their design increasingly reflects the needs of businesses to sell as much as our own personal wants and



The 1959 Cadillac Eldorado, sporting the decorative tailfins that reflected the increasing influence of marketing on automobile design in the mid-20th century. (Wikimedia user Veesixteen. Public Domain)

needs. More or less superficial options for colors and frequent style changes, for instance, demonstrate that consumption is, increasingly, a conspicuous act.

But more than this, these new characteristics indicate that people are associating their very identities with the consumption choices they make. Consider, for instance, what kind of car you would prefer to drive versus what kind of car you wouldn't want to 'be caught dead in'. Or, ask yourself why it is common to see groups of people who are all dressed similarly. It's likely not difficult to see that consumption and identity have become significantly intertwined in modern culture.

But what of the rebels—the people who refuse to conform? As you might have guessed, today there are plenty of products available for her as well. Consider the New York advertising firm Doyle Dane Bernbach's campaign to sell the Volkswagen Beetle in the 1960s. The ads famously used wit and humor to target drivers who favored reliability and affordability over the increasingly common conspicuous consumption elements of American automobiles. A 1963 magazine spread, for instance, depicted 15 identical images of the Beetle with model years from 1945 to 1963 under each image, indicating the utter lack of changes in the body design over this period. With advertising like this Volkswagen was able to successfully market its products to the consumers who didn't accommodate themselves to the mainstream preferences of the time.

The Volkswagen ad campaign of the 1960s might be taken as marketing naturally restraining the potential excesses of conspicuous consumption: needless model changes and functionless decorations were explicitly rejected, even mocked, in these ads. However, the important thing to recognize is that it also allowed consumers to display their independence from the cultural trends while still maintaining their relationship with the businesses that provide these goods.

Over the decades this form of marketing has evolved into what institutional economist Philip Mirowski calls “murketing”: a form of marketing in which businesses promote the belief that consumers are ‘too smart’ to be fooled or persuaded by advertising, brands, and so on. With that belief firmly held, consumers are free to buy what they want, safe in the knowledge that their decisions are their own. Self-reflection becomes unnecessary, and (potentially all) differences of opinion, whether political, cultural, or otherwise, become expressible through the conspicuous consumption of goods and services. In this manner, the very concept of Self is reduced to selections from a menu of shopping opportunities; the boundaries between a person’s identity and businesses’ marketing become murky—hence, “murketing.”

The curious result of all of this is that our consumption choices, our preferences, our very identities seem more and more to be adjusted to fit the goods and services available for purchase, rather the reverse. To be sure, this is not a criticism of individuals who prefer one branded identity over another. It is, rather, an observation of the extent to which the complexity of the practices of conspicuous consumption and corporate demand management have coevolved.

9.4 CONCLUSION

The neoclassical approach to understanding consumption decisions appears to make a lot of sense, given the specific (perhaps peculiar) way the question is posed. Why do people buy what they buy? Because they prefer these things over others. Of course! But how well does this approach answer the truly interesting questions of modern consumption decisions? Why, for instance, do some people buy sports cars only to park them with everyone else in rush hour traffic? Why is it that, when we go to buy clothes, we spend so little time comparing the warmth and durability of the materials, and so much time on the finer details of ‘how it looks’? *De gustibus non est disputandum* leads the neoclassical analysis to conclude simply that these are the elements which provide the most utility, leaving the critical thinker to ask: well, why are these the major elements of utility? (Naturally, to argue that we prefer the things that give us utility, and the things that give us utility are the things we prefer, is little more than an exercise in circular reasoning after all.)

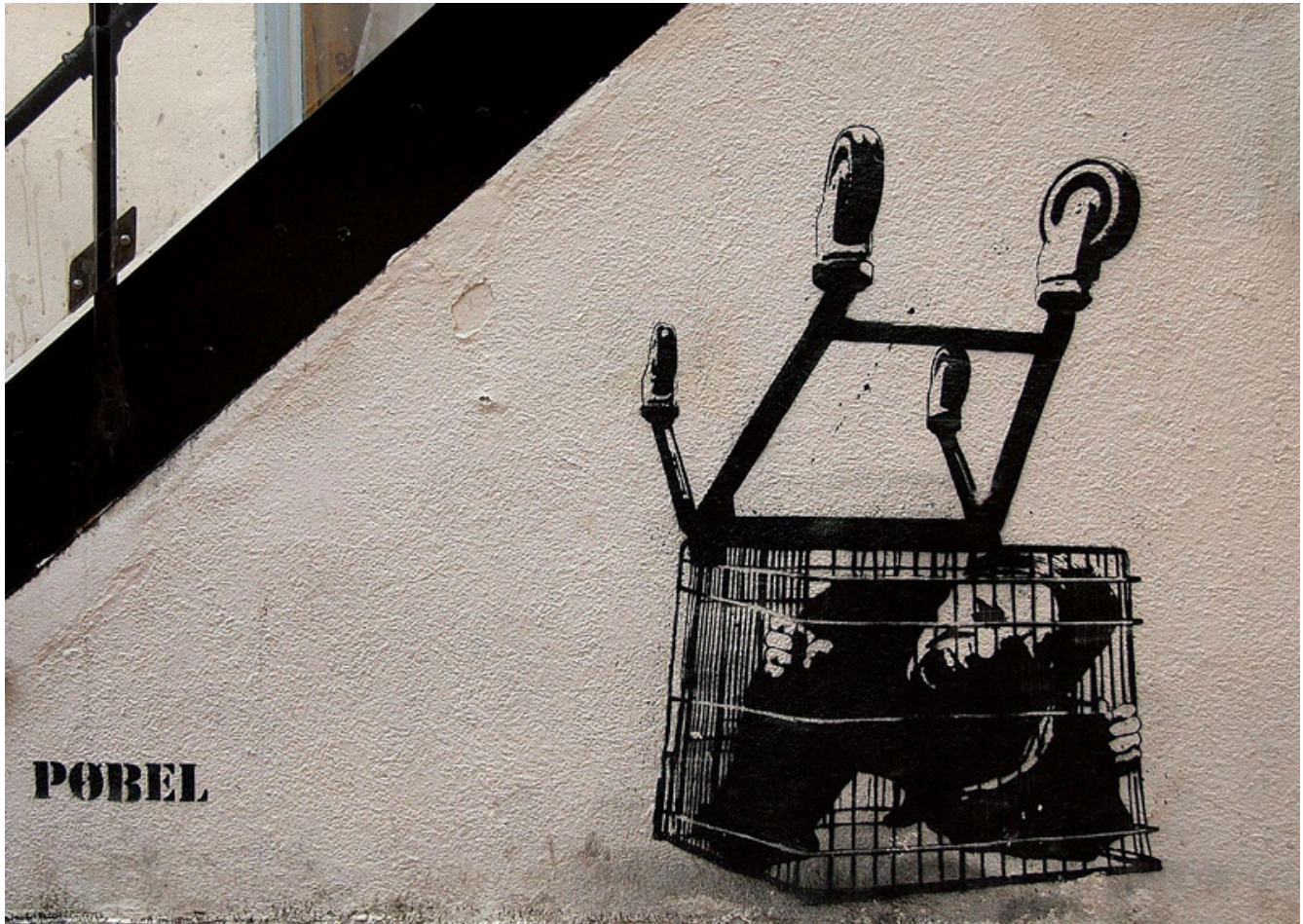
Jettisoning the overly constricting assumption of *de gustibus non est disputandum* allows us to move on to an institutional analysis; and with that analysis we can begin to answer the really interesting questions. Once we accept that humans are naturally, inherently social creatures, and that a society’s institutions evolve over time, we can begin to understand how consumption patterns reflect partly the needs and desire of individuals, but also in part the received ‘habits of thought’ of the societies these individuals were born into. We can develop plausible explanations, like conspicuous consumption, for why consumers appear to be purchasing things for their *disutility* or their wastefulness.

These theories, in turn, bring to light new questions. For instance, what is the relationship between the producers of consumer goods and the desires of consumers if not simply for the former to cater to the latter? What is the nature of the Self, how is that constrained by a culture of consumerism, and how might it break free of those constraints? These may be difficult questions to answer within the standard framework of neoclassical economics. Because of this, heterodox economics often turn to an institutional analysis instead.

WHO ARE WE AND WHAT DO WE WANT?

Certainly, the basic purpose of this textbook is to introduce the student to how economists explain the way our economy works. But it’s worth pausing for a moment to reflect on the following question: are you, the student, being trained in something more than theories that explain the consumption, production, and distribution of goods and services? Does it go deeper than that?

In a recent interview, linguist Noam Chomsky was asked if he agreed with the view that “bedrock human nature” indicated that we, as people, are “selfish and really seeking material comfort...that we can never get away from that—that’s what we are.” Chomsky’s response was that such a view was a product of the last century of advertising, rather than any scientific study of actual human nature. “The natural thing for humans,” Chomsky argued, “is to want to be independent, creative...



Anti-consumerism graffiti in Oslo, Norway. (Wikimedia user Anne-Sofie Ofrim. CC BY-SA 3.0)

People, I think, want dignity and a sense of self-worth, and a sense of creating and doing something that's important; that's what we are. I think it's taken huge efforts, enormous efforts—a huge part of the economy is devoted to trying to drive these things out of people's heads, to make you think that all you want is more commodities...”

It is worth considering where an introductory economics class (and textbook) fits into the broader culture in which we live. While economists surely strive to be objective social scientists, it is impossible for us to completely divorce ourselves from the societies in which we were brought up. Is it possible, therefore, that the stories we tell about, for instance, consumer choices have embedded in them certain assumptions about what it is to be human? And, furthermore, is it possible that those assumptions are in part a reflection of the very societies we're trying to understand? As this chapter has demonstrated, regardless of your personal beliefs about human nature, asking these questions is important if we're to properly understand our economy as it evolves through time.

CHAPTER 10. COST AND INDUSTRY STRUCTURE

INTRODUCTION TO COST AND INDUSTRY STRUCTURE



Figure 1. Amazon is an American international electronic commerce company that sells books, among many other things, shipping them directly to the consumer. There is no brick-and-mortar Amazon store. (Credit: modification of work by William Christiansen/Flickr Creative Commons)

AMAZON

In less than two decades, Amazon.com has transformed the way books are sold, bought, and even read. Prior to Amazon, books were primarily sold through independent bookstores with limited inventories in small retail locations. There were exceptions, of course; Borders and Barnes & Noble offered larger stores in urban areas. In the last decade, however, independent bookstores have become few and far between, Borders has gone out of business, and Barnes & Noble is struggling. Online delivery and purchase of books has indeed overtaken the more traditional business models. How has Amazon changed the book selling industry? How has it managed to crush its competition?

A major reason for the giant retailer's success is its production model and cost structure, which has enabled Amazon to undercut the prices of its competitors even when factoring in the cost of shipping. Read on to see how firms great (like Amazon) and small (like your corner deli) determine what to sell, at what output and price.

CHAPTER OBJECTIVES

Introduction to Cost and Industry Structure

In this chapter, you will learn about:

- Explicit and Implicit Costs, and Accounting and Economic Profit
- The Structure of Costs in the Short Run
- The Structure of Costs in the Long Run

This chapter is the first of four chapters that explore the *theory of the firm*. This theory explains that firms behave in much the same way as consumers behave. What does that mean? Let's define what is meant by the firm. A **firm** (or business) combines inputs of labor, capital, land, and raw or finished component materials to produce outputs. If the firm is successful, the outputs are more valuable than the inputs. This activity of **production** goes beyond manufacturing (i.e., making things). It includes any process or service that creates value, including transportation, distribution, wholesale and retail sales. Production involves a number of important decisions that define the behavior of firms. These decisions include, but are not limited to:

- What product or products should the firm produce?
- How should the products be produced (i.e., what production process should be used)?
- How much output should the firm produce?
- What price should the firm charge for its products?
- How much labor should the firm employ?

The answers to these questions depend on the production and cost conditions facing each firm. The answers also depend on the structure of the market for the product(s) in question. Market structure is a multidimensional concept that involves how competitive the industry is. It is defined by questions such as these:

- How much market power does each firm in the industry possess?
- How similar is each firm's product to the products of other firms in the industry?
- How difficult is it for new firms to enter the industry?
- Do firms compete on the basis of price, advertising, or other product differences?

Figure 2 illustrates the range of different market structures, which we will explore in Perfect Competition, Monopoly, and Monopolistic Competition and Oligopoly.

First let's take a look at how firms determine their costs and desired profit levels. Then we will discuss costs in the short run and long run and the factors that can influence each.



Figure 2. The Spectrum of Competition. Firms face different competitive situations. At one extreme—perfect competition—many firms are all trying to sell identical products. At the other extreme—monopoly—only one firm is selling the product, and this firm faces no competition. Monopolistic competition and oligopoly fall between the extremes of perfect competition and monopoly. Monopolistic competition is a situation with many firms selling similar, but not identical, products. Oligopoly is a situation with few firms that sell identical or similar products.

10.1 EXPLICIT AND IMPLICIT COSTS, AND ACCOUNTING AND ECONOMIC PROFIT

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the difference between explicit costs and implicit costs
- Understand the relationship between cost and revenue

Private enterprise, the ownership of businesses by private individuals, is a hallmark of the U.S. economy. When people think of businesses, often giants like Wal-Mart, Microsoft, or General Motors come to mind. But firms come in all sizes, as shown in Table 1. The vast majority of American firms have fewer than 20 employees. As of 2010, the U.S. Census Bureau counted 5.7 million firms with employees in the U.S. economy. Slightly less than half of all the workers in private firms are at the 17,000 large firms, meaning they employ more than 500 workers. Another 35% of workers in the U.S. economy are at firms with fewer than 100 workers. These small-scale businesses include everything from dentists and lawyers to businesses that mow lawns or clean houses. Indeed, Table 1 does not include a separate category for the millions of small “non-employer” businesses where a single owner or a few partners are not officially paid wages or a salary, but simply receive whatever they can earn.

Number of Employees	Firms (% of total firms)	Number of Paid Employees (% of total employment)
Total	5,734,538	112.0 million
0–9	4,543,315 (79.2%)	12.3 million (11.0%)
10–19	617,089 (10.8%)	8.3 million (7.4%)
20–99	475,125 (8.3%)	18.6 million (16.6%)
100–499	81,773 (1.4%)	15.9 million (14.2%)
500 or more	17,236 (0.30%)	50.9 million (49.8%)

Table 1. Range in Size of U.S. Firms. (Source: U.S. Census, 2010 www.census.gov)

Each of these businesses, regardless of size or complexity, tries to earn a profit:

$$\text{Profit} = \text{Total Revenue} - \text{Total Cost}$$

Total **revenue** is the income brought into the firm from selling its products. It is calculated by multiplying the price of the product times the quantity of output sold:

$$\text{Total Revenue} = \text{Price} \times \text{Quantity}$$

We will see in the following chapters that revenue is a function of the demand for the firm's products.

We can distinguish between two types of cost: explicit and implicit. **Explicit costs** are out-of-pocket costs, that is, payments that are actually made. Wages that a firm pays its employees or rent that a firm pays for its office are explicit costs. **Implicit costs** are more subtle, but just as important. They represent the opportunity cost of using resources already owned by the firm. Often for small businesses, they are resources contributed by the owners; for example, working in the business while not getting a formal salary, or using the ground floor of a home as a retail store. Implicit costs also allow for depreciation of goods, materials, and equipment that are necessary for a company to operate. (See the Work it Out feature for an extended example.)

These two definitions of cost are important for distinguishing between two conceptions of profit, accounting profit and economic profit. **Accounting profit** is a cash concept. It means total revenue minus explicit costs—the difference between dollars brought in and dollars paid out. **Economic profit** is total revenue minus total cost, including both explicit and implicit costs. The difference is important because even though a business pays income taxes based on its accounting profit, whether or not it is economically successful depends on its economic profit.

CALCULATING IMPLICIT COSTS

Consider the following example. Fred currently works for a corporate law firm. He is considering opening his own legal practice, where he expects to earn \$200,000 per year once he gets established. To run his own firm, he would need an office and a law clerk. He has found the perfect office, which rents for \$50,000 per year. A law clerk could be hired for \$35,000 per year. If these figures are accurate, would Fred's legal practice be profitable?

Step 1. First you have to calculate the costs. You can take what you know about explicit costs and total them:

<i>Office rental :</i>	\$50,000
<i>Law clerk's salary :</i>	+\$35,000
<i>Total explicit costs :</i>	\$85,000

Step 2. Subtracting the explicit costs from the revenue gives you the accounting profit.

<i>Revenues :</i>	\$200,000
<i>Explicit costs :</i>	−\$85,000
<i>Accounting profit :</i>	\$115,000

But these calculations consider only the explicit costs. To open his own practice, Fred would have to quit his current job, where he is earning an annual salary of \$125,000. This would be an implicit cost of opening his own firm.

Step 3. You need to subtract both the explicit and implicit costs to determine the true economic profit:

$$\text{Economic profit} = \text{total revenues} - \text{explicit costs} - \text{implicit costs}$$

$$\$200,000 - \$85,000 - \$125,000$$

$$-\$10,000 \text{ per year}$$

Fred would be losing \$10,000 per year. That does not mean he would not want to open his own business, but it does mean he would be earning \$10,000 less than if he worked for the corporate firm.

Implicit costs can include other things as well. Maybe Fred values his leisure time, and starting his own firm would require him to put in more hours than at the corporate firm. In this case, the lost leisure would also be an implicit cost that would subtract from economic profits.

Now that we have an idea about the different types of costs, let's look at cost structures. A firm's cost structure in the long run may be different from that in the short run. We turn to that distinction in the next section.

KEY CONCEPTS AND SUMMARY

Privately owned firms are motivated to earn profits. Profit is the difference between revenues and costs. While accounting profit considers only explicit costs, economic profit considers both explicit and implicit costs.

SELF-CHECK QUESTIONS

1. A firm had sales revenue of \$1 million last year. It spent \$600,000 on labor, \$150,000 on capital and \$200,000 on materials. What was the firm's accounting profit?
2. Continuing from Self-Check Question 1, the firm's factory sits on land owned by the firm that could be rented out for \$30,000 per year. What was the firm's economic profit last year?

REVIEW QUESTIONS

1. What are explicit and implicit costs?
2. Would an interest payment on a loan to a firm be considered an explicit or implicit cost?
3. What is the difference between accounting and economic profit?

CRITICAL THINKING QUESTIONS

Small "Mom and Pop firms," like inner city grocery stores, sometimes exist even though they do not earn economic profits. How can you explain this?

PROBLEMS

A firm is considering an investment that will earn a 6% rate of return. If it were to borrow the money, it would have to pay 8% interest on the loan, but it currently has the cash, so it will not need to borrow. Should the firm make the investment? Show your work.

REFERENCES

2010 U.S. Census. www.census.gov.

GLOSSARY

accounting profit total revenues minus explicit costs, including depreciation

economic profit total revenues minus total costs (explicit plus implicit costs)

explicit costs out-of-pocket costs for a firm, for example, payments for wages and salaries, rent, or materials

firm an organization that combines inputs of labor, capital, land, and raw or finished component materials to produce outputs.

implicit costs opportunity cost of resources already owned by the firm and used in business, for example, expanding a factory onto land already owned

private enterprise the ownership of businesses by private individuals

production the process of combining inputs to produce outputs, ideally of a value greater than the value of the inputs

revenue income from selling a firm's product; defined as price times quantity sold

SOLUTIONS

Answers to Self-Check Questions

Accounting profit $\text{total revenues} - \text{explicit costs}$

$$\$1,000,000 - (\$600,000 + \$150,000 + \$200,000)$$

1. $\$50,000$

Economic profit $\text{accounting profit} - \text{implicit cost}$

$$\$50,000 - \$30,000$$

2. $\$20,000$

10.2 THE STRUCTURE OF COSTS IN THE SHORT RUN

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Analyze short-run costs as influenced by total cost, fixed cost, variable cost, marginal cost, and average cost.
- Calculate average profit
- Evaluate patterns of costs to determine potential profit

The cost of producing a firm's output depends on how much **labor** and **physical capital** the firm uses. A list of the costs involved in producing cars will look very different from the costs involved in producing computer software or haircuts or fast-food meals. However, the cost structure of all firms can be broken down into some common underlying patterns. When a firm looks at its **total costs** of production in the short run, a useful starting point is to divide total costs into two categories: fixed costs that cannot be changed in the short run and variable costs that can be changed.

FIXED AND VARIABLE COSTS

Fixed costs are expenditures that do not change regardless of the level of production, at least not in the short term. Whether you produce a lot or a little, the fixed costs are the same. One example is the rent on a factory or a retail space. Once you sign the lease, the rent is the same regardless of how much you produce, at least until the lease runs out. Fixed costs can take many other forms: for example, the cost of machinery or equipment to produce the product, research and development costs to develop new products, even an expense like advertising to popularize a brand name. The level of fixed costs varies according to the specific line of business: for instance, manufacturing computer chips requires an expensive factory, but a local moving and hauling business can get by with almost no fixed costs at all if it rents trucks by the day when needed.

Variable costs, on the other hand, are incurred in the act of producing—the more you produce, the greater the variable cost. Labor is treated as a variable cost, since producing a greater quantity of a good or service typically requires more workers or more work hours. Variable costs would also include raw materials.

As a concrete example of fixed and variable costs, consider the barber shop called “The Clip Joint” shown in Figure 1. The data for output and costs are shown in Table 2. The fixed costs of operating the barber shop, including the space and equipment, are \$160 per day. The variable costs are the costs of hiring barbers, which in our example is \$80 per barber each day. The first two columns of the table

show the quantity of haircuts the barbershop can produce as it hires additional barbers. The third column shows the fixed costs, which do not change regardless of the level of production. The fourth column shows the variable costs at each level of output. These are calculated by taking the amount of labor hired and multiplying by the wage. For example, two barbers cost: $2 \times \$80 = \160 . Adding together the fixed costs in the third column and the variable costs in the fourth column produces the total costs in the fifth column. So, for example, with two barbers the total cost is: $\$160 + \$160 = \$320$.

Labor	Quantity	Fixed Cost	Variable Cost	Total Cost
1	16	\$160	\$80	\$240
2	40	\$160	\$160	\$320
3	60	\$160	\$240	\$400
4	72	\$160	\$320	\$480
5	80	\$160	\$400	\$560
6	84	\$160	\$480	\$640
7	82	\$160	\$560	\$720

Table 2. Output and Total Costs

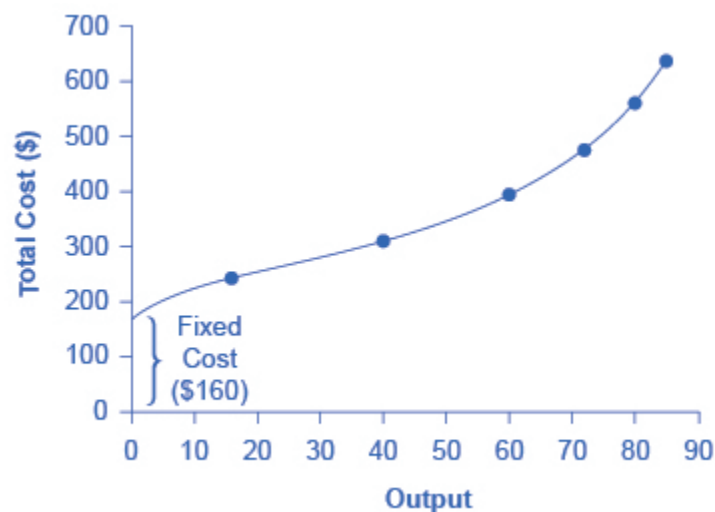


Figure 1. How Output Affects Total Costs. At zero production, the fixed costs of \$160 are still present. As production increases, variable costs are added to fixed costs, and the total cost is the sum of the two.

The relationship between the quantity of output being produced and the cost of producing that output is shown graphically in the figure. The fixed costs are always shown as the vertical intercept of the total cost curve; that is, they are the costs incurred when output is zero so there are no variable costs.

You can see from the graph that once production starts, total costs and variable costs rise. While variable costs may initially increase at a decreasing rate, at some point they begin increasing at an increasing rate. This is caused by diminishing marginal returns, discussed in the chapter on Choice in a World of Scarcity, which is easiest to see with an example. As the number of barbers increases from zero to one in the table, output increases from 0 to 16 for a marginal gain of 16; as the number rises from one to two barbers, output increases from 16 to 40, a marginal gain of 24. From that point on,

though, the marginal gain in output diminishes as each additional barber is added. For example, as the number of barbers rises from two to three, the marginal output gain is only 20; and as the number rises from three to four, the marginal gain is only 12.

To understand the reason behind this pattern, consider that a one-man barber shop is a very busy operation. The single barber needs to do everything: say hello to people entering, answer the phone, cut hair, sweep up, and run the cash register. A second barber reduces the level of disruption from jumping back and forth between these tasks, and allows a greater division of labor and specialization. The result can be greater increasing marginal returns. However, as other barbers are added, the advantage of each additional barber is less, since the specialization of labor can only go so far. The addition of a sixth or seventh or eighth barber just to greet people at the door will have less impact than the second one did. This is the pattern of diminishing marginal returns. As a result, the total costs of production will begin to rise more rapidly as output increases. At some point, you may even see negative returns as the additional barbers begin bumping elbows and getting in each other's way. In this case, the addition of still more barbers would actually cause output to decrease, as shown in the last row of Table 2.

This pattern of **diminishing marginal returns** is common in production. As another example, consider the problem of irrigating a crop on a farmer's field. The plot of land is the fixed factor of production, while the water that can be added to the land is the key variable cost. As the farmer adds water to the land, output increases. But adding more and more water brings smaller and smaller increases in output, until at some point the water floods the field and actually reduces output. Diminishing marginal returns occur because, at a given level of fixed costs, each additional input contributes less and less to overall production.

AVERAGE TOTAL COST, AVERAGE VARIABLE COST, MARGINAL COST

The breakdown of total costs into fixed and variable costs can provide a basis for other insights as well. The first five columns of Table 3 duplicate the previous table, but the last three columns show average total costs, average variable costs, and marginal costs. These new measures analyze costs on a per-unit (rather than a total) basis and are reflected in the curves shown in Figure 2.

Labor	Quantity	Fixed Cost	Variable Cost	Total Cost	Marginal Cost	Average Total Cost	Average Variable Cost
1	16	\$160	\$80	\$240	\$5.00	\$15.00	\$5.00
2	40	\$160	\$160	\$320	\$3.30	\$8.00	\$4.00
3	60	\$160	\$240	\$400	\$4.00	\$6.60	\$4.00
4	72	\$160	\$320	\$480	\$6.60	\$6.60	\$4.40
5	80	\$160	\$400	\$560	\$10.00	\$7.00	\$5.00
6	84	\$160	\$480	\$640	\$20.00	\$7.60	\$5.70

Table 3. Different Types of Costs

Average total cost (sometimes referred to simply as average cost) is total cost divided by the quantity of output. Since the total cost of producing 40 haircuts is \$320, the average total cost for producing each of 40 haircuts is $\$320/40$, or \$8 per haircut. Average cost curves are typically U-shaped, as Figure 2 shows. Average total cost starts off relatively high, because at low levels of output total costs are dominated by the fixed cost; mathematically, the denominator is so small that average total cost is

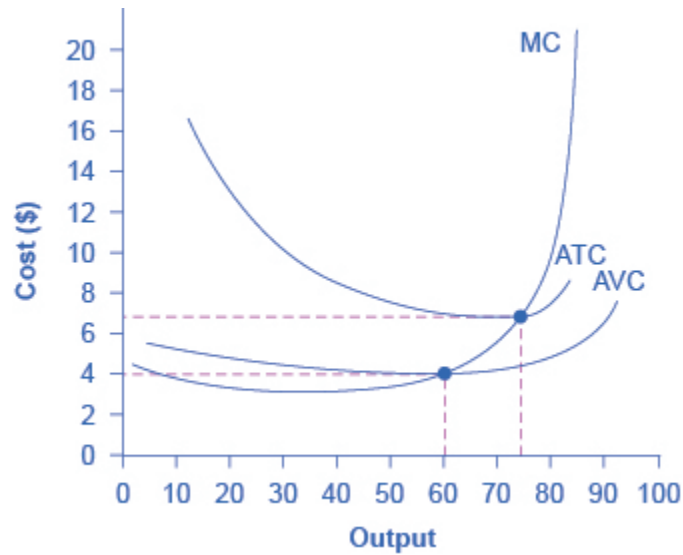


Figure 2. Cost Curves at the Clip Joint. The information on total costs, fixed cost, and variable cost can also be presented on a per-unit basis. Average total cost (ATC) is calculated by dividing total cost by the total quantity produced. The average total cost curve is typically U-shaped. Average variable cost (AVC) is calculated by dividing variable cost by the quantity produced. The average variable cost curve lies below the average total cost curve and is typically U-shaped or upward-sloping. Marginal cost (MC) is calculated by taking the change in total cost between two levels of output and dividing by the change in output. The marginal cost curve is upward-sloping.

large. Average total cost then declines, as the fixed costs are spread over an increasing quantity of output. In the average cost calculation, the rise in the numerator of total costs is relatively small compared to the rise in the denominator of quantity produced. But as output expands still further, the average cost begins to rise. At the right side of the average cost curve, total costs begin rising more rapidly as diminishing returns kick in.

Average variable cost obtained when variable cost is divided by quantity of output. For example, the variable cost of producing 80 haircuts is \$400, so the average variable cost is $\$400/80$, or \$5 per haircut. Note that at any level of output, the average variable cost curve will always lie below the curve for average total cost, as shown in Figure 2. The reason is that average total cost includes average variable cost and average fixed cost. Thus, for $Q = 80$ haircuts, the average total cost is \$8 per haircut, while the average variable cost is \$5 per haircut. However, as output grows, fixed costs become relatively less important (since they do not rise with output), so average variable cost sneaks closer to average cost.

Average total and variable costs measure the average costs of producing some quantity of output. Marginal cost is somewhat different. **Marginal cost** is the additional cost of producing one more unit of output. So it is not the cost per unit of *all* units being produced, but only the next one (or next few).

Marginal cost can be calculated by taking the change in total cost and dividing it by the change in quantity. For example, as quantity produced increases from 40 to 60 haircuts, total costs rise by 400 – 320, or 80. Thus, the marginal cost for each of those marginal 20 units will be $80/20$, or \$4 per haircut. The marginal cost curve is generally upward-sloping, because diminishing marginal returns implies that additional units are more costly to produce. A small range of increasing marginal returns can be seen in the figure as a dip in the marginal cost curve before it starts rising. There is a point at which marginal and average costs meet, as the following Clear it Up feature discusses.

WHERE DO MARGINAL AND AVERAGE COSTS MEET?

The marginal cost line intersects the average cost line exactly at the bottom of the average cost curve—which occurs at a quantity of 72 and cost of \$6.60 in Figure 2. The reason why the intersection occurs at this point is built into the economic meaning of marginal and average costs. If the marginal cost of production is below the average cost for producing previous units, as it is for the points to the left of where MC crosses ATC, then producing one more additional unit will reduce average costs overall—and the ATC curve will be downward-sloping in this zone. Conversely, if the marginal cost of production for producing an additional unit is above the average cost for producing the earlier units, as it is for points to the right of where MC crosses ATC, then producing a marginal unit will increase average costs overall—and the ATC curve must be upward-sloping in this zone. The point of transition, between where MC is pulling ATC down and where it is pulling it up, must occur at the minimum point of the ATC curve.

This idea of the marginal cost “pulling down” the average cost or “pulling up” the average cost may sound abstract, but think about it in terms of your own grades. If the score on the most recent quiz you take is lower than your average score on previous quizzes, then the marginal quiz pulls down your average. If your score on the most recent quiz is higher than the average on previous quizzes, the marginal quiz pulls up your average. In this same way, low marginal costs of production first pull down average costs and then higher marginal costs pull them up.

The numerical calculations behind average cost, average variable cost, and marginal cost will change from firm to firm. However, the general patterns of these curves, and the relationships and economic intuition behind them, will not change.

LESSONS FROM ALTERNATIVE MEASURES OF COSTS

Breaking down total costs into fixed cost, marginal cost, average total cost, and average variable cost is useful because each statistic offers its own insights for the firm.

Whatever the firm’s quantity of production, total revenue must exceed total costs if it is to earn a profit. As explored in the chapter *Choice in a World of Scarcity*, fixed costs are often **sunk costs** that cannot be recouped. In thinking about what to do next, sunk costs should typically be ignored, since this spending has already been made and cannot be changed. However, variable costs can be changed, so they convey information about the firm’s ability to cut costs in the present and the extent to which costs will increase if production rises.

WHY ARE TOTAL COST AND AVERAGE COST NOT ON THE SAME GRAPH?

Total cost, fixed cost, and variable cost each reflect different aspects of the cost of production over the entire quantity of output being produced. These costs are measured in dollars. In contrast, marginal cost, average cost, and average variable

cost are costs per unit. In the previous example, they are measured as cost per haircut. Thus, it would not make sense to put all of these numbers on the same graph, since they are measured in different units (\$ versus \$ per unit of output). It would be as if the vertical axis measured two different things. In addition, as a practical matter, if they were on the same graph, the lines for marginal cost, average cost, and average variable cost would appear almost flat against the horizontal axis, compared to the values for total cost, fixed cost, and variable cost. Using the figures from the previous example, the total cost of producing 40 haircuts is \$320. But the average cost is \$320/40, or \$8. If you graphed both total and average cost on the same axes, the average cost would hardly show.

Average cost tells a firm whether it can earn profits given the current price in the market. If we divide profit by the quantity of output produced we get **average profit**, also known as the firm's *profit margin*. Expanding the equation for profit gives:

$$\begin{aligned}
 \text{average profit} &= \frac{\text{profit}}{\text{quantity produced}} \\
 &= \frac{\text{total revenue} - \text{total cost}}{\text{quantity produced}} \\
 &= \frac{\text{total revenue}}{\text{quantity produced}} - \frac{\text{total cost}}{\text{quantity produced}} \\
 &= \text{average revenue} - \text{average cost}
 \end{aligned}$$

But note that:

$$\begin{aligned}
 \text{average revenue} &= \frac{\text{price} \times \text{quantity produced}}{\text{quantity produced}} \\
 &= \text{price}
 \end{aligned}$$

Thus:

$$\text{average profit} = \text{price} - \text{average cost}$$

This is the firm's **profit margin**. This definition implies that if the market price is above average cost, average profit, and thus total profit, will be positive; if price is below average cost, then profits will be negative.

The marginal cost of producing an additional unit can be compared with the marginal revenue gained by selling that additional unit to reveal whether the additional unit is adding to total profit—or not. Thus, marginal cost helps producers understand how profits would be affected by increasing or decreasing production.

A VARIETY OF COST PATTERNS

The pattern of costs varies among industries and even among firms in the same industry. Some businesses have high fixed costs, but low marginal costs. Consider, for example, an Internet company that provides medical advice to customers. Such a company might be paid by consumers directly, or perhaps hospitals or healthcare practices might subscribe on behalf of their patients. Setting up the

website, collecting the information, writing the content, and buying or leasing the computer space to handle the web traffic are all fixed costs that must be undertaken before the site can work. However, when the website is up and running, it can provide a high quantity of service with relatively low variable costs, like the cost of monitoring the system and updating the information. In this case, the total cost curve might start at a high level, because of the high fixed costs, but then might appear close to flat, up to a large quantity of output, reflecting the low variable costs of operation. If the website is popular, however, a large rise in the number of visitors will overwhelm the website, and increasing output further could require a purchase of additional computer space.

For other firms, fixed costs may be relatively low. For example, consider firms that rake leaves in the fall or shovel snow off sidewalks and driveways in the winter. For fixed costs, such firms may need little more than a car to transport workers to homes of customers and some rakes and shovels. Still other firms may find that diminishing marginal returns set in quite sharply. If a manufacturing plant tried to run 24 hours a day, seven days a week, little time remains for routine maintenance of the equipment, and marginal costs can increase dramatically as the firm struggles to repair and replace overworked equipment.

Every firm can gain insight into its task of earning profits by dividing its total costs into fixed and variable costs, and then using these calculations as a basis for average total cost, average variable cost, and marginal cost. However, making a final decision about the profit-maximizing quantity to produce and the price to charge will require combining these perspectives on cost with an analysis of sales and revenue, which in turn requires looking at the market structure in which the firm finds itself. Before we turn to the analysis of market structure in other chapters, we will analyze the firm's cost structure from a long-run perspective.

KEY CONCEPTS AND SUMMARY

In a short-run perspective, a firm's total costs can be divided into fixed costs, which a firm must incur before producing any output, and variable costs, which the firm incurs in the act of producing. Fixed costs are sunk costs; that is, because they are in the past and cannot be altered, they should play no role in economic decisions about future production or pricing. Variable costs typically show diminishing marginal returns, so that the marginal cost of producing higher levels of output rises.

Marginal cost is calculated by taking the change in total cost (or the change in variable cost, which will be the same thing) and dividing it by the change in output, for each possible change in output. Marginal costs are typically rising. A firm can compare marginal cost to the additional revenue it gains from selling another unit to find out whether its marginal unit is adding to profit.

Average total cost is calculated by taking total cost and dividing by total output at each different level of output. Average costs are typically U-shaped on a graph. If a firm's average cost of production is lower than the market price, a firm will be earning profits.

Average variable cost is calculated by taking variable cost and dividing by the total output at each level of output. Average variable costs are typically U-shaped. If a firm's average variable cost of production is lower than the market price, then the firm would be earning profits if fixed costs are left out of the picture.

SELF-CHECK QUESTIONS

1. The WipeOut Ski Company manufactures skis for beginners. Fixed costs are \$30. Fill in Table 4 for total cost, average variable cost, average total cost, and marginal cost.

Quantity	Variable Cost	Fixed Cost	Total Cost	Average Total Cost	Average Variable Cost	Marginal Cost
0	0	\$30				
1	\$10	\$30				
2	\$25	\$30				
3	\$45	\$30				
4	\$70	\$30				
5	\$100	\$30				
6	\$135	\$30				

Table 4.

2. Based on your answers to the WipeOut Ski Company in Self-Check Question 1, now imagine a situation where the firm produces a quantity of 5 units that it sells for a price of \$25 each.
- What will be the company's profits or losses?
 - How can you tell at a glance whether the company is making or losing money at this price by looking at average cost?
 - At the given quantity and price, is the marginal unit produced adding to profits?

REVIEW QUESTIONS

- What is the difference between fixed costs and variable costs?
- Are there fixed costs in the long-run? Explain briefly.
- Are fixed costs also sunk costs? Explain.
- What are diminishing marginal returns as they relate to costs?
- Which costs are measured on per-unit basis: fixed costs, average cost, average variable cost, variable costs, and marginal cost?
- How is each of the following calculated: marginal cost, average total cost, average variable cost?

CRITICAL THINKING QUESTIONS

- A common name for fixed cost is "overhead." If you divide fixed cost by the quantity of output produced, you get average fixed cost. Supposed fixed cost is \$1,000. What does the average fixed cost curve look like? Use your response to explain what "spreading the overhead" means.
- How does fixed cost affect marginal cost? Why is this relationship important?
- Average cost curves (except for average fixed cost) tend to be U-shaped, decreasing and then increasing.

Marginal cost curves have the same shape, though this may be harder to see since most of the marginal cost curve is increasing. Why do you think that average and marginal cost curves have the same general shape?

PROBLEMS

- Return to Figure 1. What is the marginal gain in output from increasing the number of barbers from 4 to 5 and from 5 to 6? Does it continue the pattern of diminishing marginal returns?
- Compute the average total cost, average variable cost, and marginal cost of producing 60 and 72 haircuts. Draw the graph of the three curves between 60 and 72 haircuts.

GLOSSARY

average profit profit divided by the quantity of output produced; profit margin

average total cost total cost divided by the quantity of output

average variable cost variable cost divided by the quantity of output

fixed cost expenditure that must be made before production starts and that does not change regardless of the level of production

marginal cost the additional cost of producing one more unit

total cost the sum of fixed and variable costs of production

variable cost cost of production that increases with the quantity produced

EXERCISES

Answers to Self-Check Questions

1.	Quantity	Variable Cost	Fixed Cost	Total Cost	Average Total Cost	Average Variable Cost	Marginal Cost
	0	0	\$30	\$30	–	–	
	1	\$10	\$30	\$40	\$10.00	\$40.00	\$10
	2	\$25	\$30	\$55	\$12.50	\$27.50	\$15
	3	\$45	\$30	\$75	\$15.00	\$25.00	\$20
	4	\$70	\$30	\$100	\$17.50	\$25.00	\$25
	5	\$100	\$30	\$130	\$20.00	\$26.00	\$30
	6	\$135	\$30	\$165	\$22.50	\$27.50	\$35

Table 5.

- Total revenues in this example will be a quantity of five units multiplied by the price of \$25/unit, which equals \$125. Total costs when producing five units are \$130. Thus, at this level of quantity and output the firm experiences losses (or negative profits) of \$5.
 - If price is less than average cost, the firm is not making a profit. At an output of five units, the average cost is \$26/unit. Thus, at a glance you can see the firm is making losses. At a second glance, you can see that it must be losing \$1 for each unit produced (that is, average cost of \$26/

unit minus the price of \$25/unit). With five units produced, this observation implies total losses of \$5.

- c. When producing five units, marginal costs are \$30/unit. Price is \$25/unit. Thus, the marginal unit is not adding to profits, but is actually subtracting from profits, which suggests that the firm should reduce its quantity produced.

10.3 THE STRUCTURE OF COSTS IN THE LONG RUN

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Calculate total cost
- Identify economies of scale, diseconomies of scale, and constant returns to scale
- Interpret graphs of long-run average cost curves and short-run average cost curves
- Analyze cost and production in the long run and short run

The long run is the period of time when all costs are variable. The long run depends on the specifics of the firm in question—it is not a precise period of time. If you have a one-year lease on your factory, then the long run is any period longer than a year, since after a year you are no longer bound by the lease. No costs are fixed in the long run. A firm can build new factories and purchase new machinery, or it can close existing facilities. In planning for the long run, the firm will compare alternative **production technologies** (or processes).

In this context, technology refers to all alternative methods of combining inputs to produce outputs. It does not refer to a specific new invention like the tablet computer. The firm will search for the production technology that allows it to produce the desired level of output at the lowest cost. After all, lower costs lead to higher profits—at least if total revenues remain unchanged. Moreover, each firm must fear that if it does not seek out the lowest-cost methods of production, then it may lose sales to competitor firms that find a way to produce and sell for less.

CHOICE OF PRODUCTION TECHNOLOGY

Many tasks can be performed with a range of combinations of labor and physical capital. For example, a firm can have human beings answering phones and taking messages, or it can invest in an automated voicemail system. A firm can hire file clerks and secretaries to manage a system of paper folders and file cabinets, or it can invest in a computerized recordkeeping system that will require fewer employees. A firm can hire workers to push supplies around a factory on rolling carts, it can invest in motorized vehicles, or it can invest in robots that carry materials without a driver. Firms often face a choice between buying a many small machines, which need a worker to run each one, or buying one larger and more expensive machine, which requires only one or two workers to operate it. In short, physical capital and labor can often substitute for each other.

Consider the example of a private firm that is hired by local governments to clean up public parks.

Three different combinations of labor and physical capital for cleaning up a single average-sized park appear in Table 6. The first production technology is heavy on workers and light on machines, while the next two technologies substitute machines for workers. Since all three of these production methods produce the same thing—one cleaned-up park—a profit-seeking firm will choose the production technology that is least expensive, given the prices of labor and machines.

Production technology 1	10 workers	2 machines
Production technology 2	7 workers	4 machines
Production technology 3	3 workers	7 machines

Table 6. Three Ways to Clean a Park

Production technology 1 uses the most labor and least machinery, while production technology 3 uses the least labor and the most machinery. Table 7 outlines three examples of how the total cost will change with each production technology as the cost of labor changes. As the cost of labor rises from example A to B to C, the firm will choose to substitute away from labor and use more machinery.

Example A: Workers cost \$40, machines cost \$80

	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	$10 \times \$40 = \400	$2 \times \$80 = \160	\$560
Cost of technology 2	$7 \times \$40 = \280	$4 \times \$80 = \320	\$600
Cost of technology 3	$3 \times \$40 = \120	$7 \times \$80 = \560	\$680

Example B: Workers cost \$55, machines cost \$80

	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	$10 \times \$55 = \550	$2 \times \$80 = \160	\$710
Cost of technology 2	$7 \times \$55 = \385	$4 \times \$80 = \320	\$705
Cost of technology 3	$3 \times \$55 = \165	$7 \times \$80 = \560	\$725

Example C: Workers cost \$90, machines cost \$80

	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	$10 \times \$90 = \900	$2 \times \$80 = \160	\$1,060
Cost of technology 2	$7 \times \$90 = \630	$4 \times \$80 = \320	\$950
Cost of technology 3	$3 \times \$90 = \270	$7 \times \$80 = \560	\$830

Table 7. Total Cost with Rising Labor Costs

Example A shows the firm's cost calculation when wages are \$40 and machines costs are \$80. In this case, technology 1 is the low-cost production technology. In example B, wages rise to \$55, while the cost of machines does not change, in which case technology 2 is the low-cost production technology. If wages keep rising up to \$90, while the cost of machines remains unchanged, then technology 3 clearly becomes the low-cost form of production, as shown in example C.

This example shows that as an input becomes more expensive (in this case, the labor input), firms will attempt to conserve on using that input and will instead shift to other inputs that are relatively less expensive. This pattern helps to explain why the demand curve for labor (or any input) slopes down; that is, as labor becomes relatively more expensive, profit-seeking firms will seek to substitute the use of other inputs. When a multinational employer like Coca-Cola or McDonald's sets up a bottling plant

or a restaurant in a high-wage economy like the United States, Canada, Japan, or Western Europe, it is likely to use production technologies that conserve on the number of workers and focuses more on machines. However, that same employer is likely to use production technologies with more workers and less machinery when producing in a lower-wage country like Mexico, China, or South Africa.

ECONOMIES OF SCALE

Once a firm has determined the least costly production technology, it can consider the optimal scale of production, or quantity of output to produce. Many industries experience economies of scale. **Economies of scale** refers to the situation where, as the quantity of output goes up, the cost per unit goes down. This is the idea behind “warehouse stores” like Costco or Walmart. In everyday language: a larger factory can produce at a lower average cost than a smaller factory.

Figure 1 illustrates the idea of economies of scale, showing the average cost of producing an alarm clock falling as the quantity of output rises. For a small-sized factory like S, with an output level of 1,000, the average cost of production is \$12 per alarm clock. For a medium-sized factory like M, with an output level of 2,000, the average cost of production falls to \$8 per alarm clock. For a large factory like L, with an output of 5,000, the average cost of production declines still further to \$4 per alarm clock.

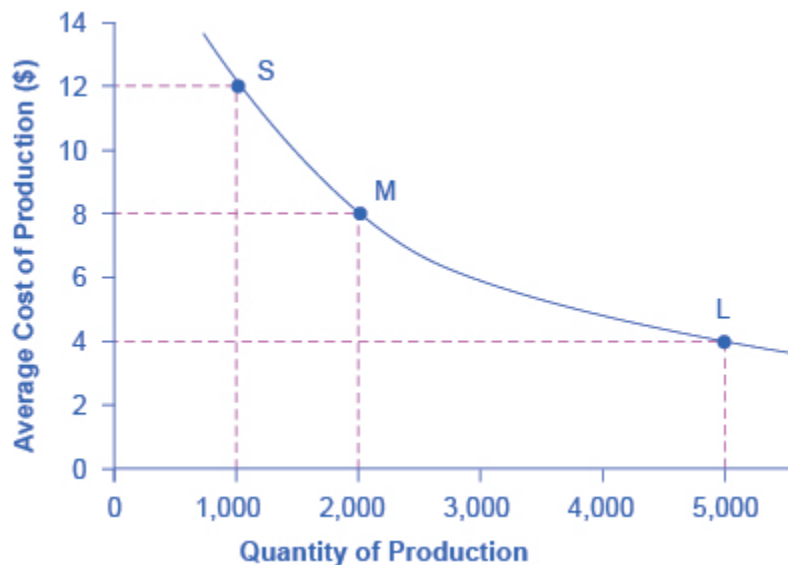


Figure 1. Economies of Scale. A small factory like S produces 1,000 alarm clocks at an average cost of \$12 per clock. A medium factory like M produces 2,000 alarm clocks at a cost of \$8 per clock. A large factory like L produces 5,000 alarm clocks at a cost of \$4 per clock. Economies of scale exist because the larger scale of production leads to lower average costs.

The average cost curve in Figure 1 may appear similar to the average cost curves presented earlier in this chapter, although it is downward-sloping rather than U-shaped. But there is one major difference. The economies of scale curve is a long-run average cost curve, because it allows all factors of production to change. The short-run average cost curves presented earlier in this chapter assumed the existence of fixed costs, and only variable costs were allowed to change.

One prominent example of economies of scale occurs in the chemical industry. Chemical plants have

a lot of pipes. The cost of the materials for producing a pipe is related to the circumference of the pipe and its length. However, the volume of chemicals that can flow through a pipe is determined by the cross-section area of the pipe. The calculations in Table 8 show that a pipe which uses twice as much material to make (as shown by the circumference of the pipe doubling) can actually carry four times the volume of chemicals because the cross-section area of the pipe rises by a factor of four (as shown in the Area column).

	Circumference ($2\pi r$)	Area (πr^2)
4-inch pipe	12.5 inches	12.5 square inches
8-inch pipe	25.1 inches	50.2 square inches
16-inch pipe	50.2 inches	201.1 square inches

Table 8. Comparing Pipes: Economies of Scale in the Chemical Industry

A doubling of the cost of producing the pipe allows the chemical firm to process four times as much material. This pattern is a major reason for economies of scale in chemical production, which uses a large quantity of pipes. Of course, economies of scale in a chemical plant are more complex than this simple calculation suggests. But the chemical engineers who design these plants have long used what they call the “six-tenths rule,” a rule of thumb which holds that increasing the quantity produced in a chemical plant by a certain percentage will increase total cost by only six-tenths as much.

SHAPES OF LONG-RUN AVERAGE COST CURVES

While in the short run firms are limited to operating on a single average cost curve (corresponding to the level of fixed costs they have chosen), in the long run when all costs are variable, they can choose to operate on any average cost curve. Thus, the **long-run average cost (LRAC) curve** is actually based on a group of **short-run average cost (SRAC) curves**, each of which represents one specific level of fixed costs. More precisely, the long-run average cost curve will be the least expensive average cost curve for any level of output. Figure 2 shows how the long-run average cost curve is built from a group of short-run average cost curves. Five short-run-average cost curves appear on the diagram. Each SRAC curve represents a different level of fixed costs. For example, you can imagine SRAC₁ as a small factory, SRAC₂ as a medium factory, SRAC₃ as a large factory, and SRAC₄ and SRAC₅ as very large and ultra-large. Although this diagram shows only five SRAC curves, presumably there are an infinite number of other SRAC curves between the ones that are shown. This family of short-run average cost curves can be thought of as representing different choices for a firm that is planning its level of investment in fixed cost physical capital—knowing that different choices about capital investment in the present will cause it to end up with different short-run average cost curves in the future.

The long-run average cost curve shows the cost of producing each quantity in the long run, when the firm can choose its level of fixed costs and thus choose which short-run average costs it desires. If the firm plans to produce in the long run at an output of Q_3 , it should make the set of investments that will lead it to locate on SRAC₃, which allows producing q_3 at the lowest cost. A firm that intends to produce Q_3 would be foolish to choose the level of fixed costs at SRAC₂ or SRAC₄. At SRAC₂ the level of fixed costs is too low for producing Q_3 at lowest possible cost, and producing q_3 would require adding a very high level of variable costs and make the average cost very high. At SRAC₄, the level of fixed costs is too high for producing q_3 at lowest possible cost, and again average costs would be very high as a result.

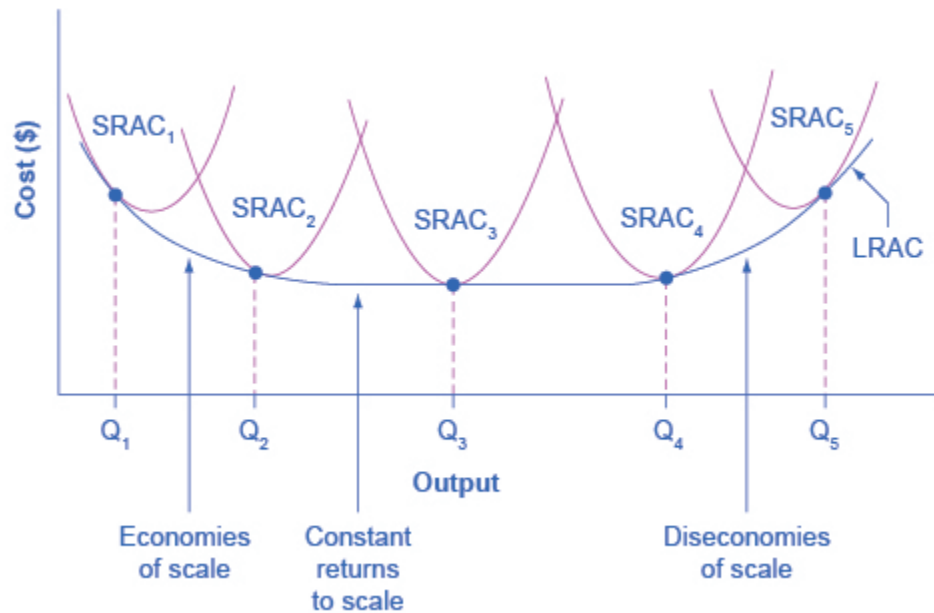


Figure 2. From Short-Run Average Cost Curves to Long-Run Average Cost Curves. The five different short-run average cost (SRAC) curves each represents a different level of fixed costs, from the low level of fixed costs at SRAC₁ to the high level of fixed costs at SRAC₅. Other SRAC curves, not shown in the diagram, lie between the ones that are shown here. The long-run average cost (LRAC) curve shows the lowest cost for producing each quantity of output when fixed costs can vary, and so it is formed by the bottom edge of the family of SRAC curves. If a firm wished to produce quantity Q₃, it would choose the fixed costs associated with SRAC₃.

The shape of the long-run cost curve, as drawn in Figure 2, is fairly common for many industries. The left-hand portion of the long-run average cost curve, where it is downward-sloping from output levels Q₁ to Q₂ to Q₃, illustrates the case of economies of scale. In this portion of the long-run average cost curve, larger scale leads to lower average costs. This pattern was illustrated earlier in Figure 1.

In the middle portion of the long-run average cost curve, the flat portion of the curve around Q₃, economies of scale have been exhausted. In this situation, allowing all inputs to expand does not much change the average cost of production, and it is called **constant returns to scale**. In this range of the LRAC curve, the average cost of production does not change much as scale rises or falls. The following Clear it Up feature explains where diminishing marginal returns fit into this analysis.

HOW DO ECONOMIES OF SCALE COMPARE TO DIMINISHING MARGINAL RETURNS?

The concept of economies of scale, where average costs decline as production expands, might seem to conflict with the idea of diminishing marginal returns, where marginal costs rise as production expands. But diminishing marginal returns refers only to the short-run average cost curve, where one variable input (like labor) is increasing, but other inputs (like capital) are fixed. Economies of scale refers to the long-run average cost curve where all inputs are being allowed to increase together. Thus, it is quite possible and common to have an industry that has both diminishing marginal returns when only one input is allowed to change, and at the same time has increasing or constant economies of scale when all inputs change together to produce a larger-scale operation.

Finally, the right-hand portion of the long-run average cost curve, running from output level Q_4 to Q_5 , shows a situation where, as the level of output and the scale rises, average costs rise as well. This situation is called **diseconomies of scale**. A firm or a factory can grow so large that it becomes very difficult to manage, resulting in unnecessarily high costs as many layers of management try to communicate with workers and with each other, and as failures to communicate lead to disruptions in the flow of work and materials. Not many overly large factories exist in the real world, because with their very high production costs, they are unable to compete for long against plants with lower average costs of production. However, in some planned economies, like the economy of the old Soviet Union, plants that were so large as to be grossly inefficient were able to continue operating for a long time because government economic planners protected them from competition and ensured that they would not make losses.

Diseconomies of scale can also be present across an entire firm, not just a large factory. The **leviathan effect** can hit firms that become too large to run efficiently, across the entirety of the enterprise. Firms that shrink their operations are often responding to finding itself in the diseconomies region, thus moving back to a lower average cost at a lower output level.

Visit this website to read an article about the complexity of the belief that banks can be “too-big-to-fail.”



THE SIZE AND NUMBER OF FIRMS IN AN INDUSTRY

The shape of the long-run average cost curve has implications for how many firms will compete in an industry, and whether the firms in an industry have many different sizes, or tend to be the same size. For example, say that one million dishwashers are sold every year at a price of \$500 each and the long-run average cost curve for dishwashers is shown in Figure 3 (a). In Figure 3 (a), the lowest point of the LRAC curve occurs at a quantity of 10,000 produced. Thus, the market for dishwashers will consist of 100 different manufacturing plants of this same size. If some firms built a plant that produced 5,000 dishwashers per year or 25,000 dishwashers per year, the average costs of production at such plants would be well above \$500, and the firms would not be able to compete.

HOW CAN CITIES BE VIEWED AS EXAMPLES OF ECONOMIES OF SCALE?

Why are people and economic activity concentrated in cities, rather than distributed evenly across a country? The fundamental reason must be related to the idea of economies of scale—that grouping economic activity is more productive in many cases than spreading it out. For example, cities provide a large group of nearby customers, so that businesses can produce at an efficient economy of scale. They also provide a large group of workers and suppliers, so that business can hire easily and purchase whatever specialized inputs they need. Many of the attractions of cities, like sports stadiums and

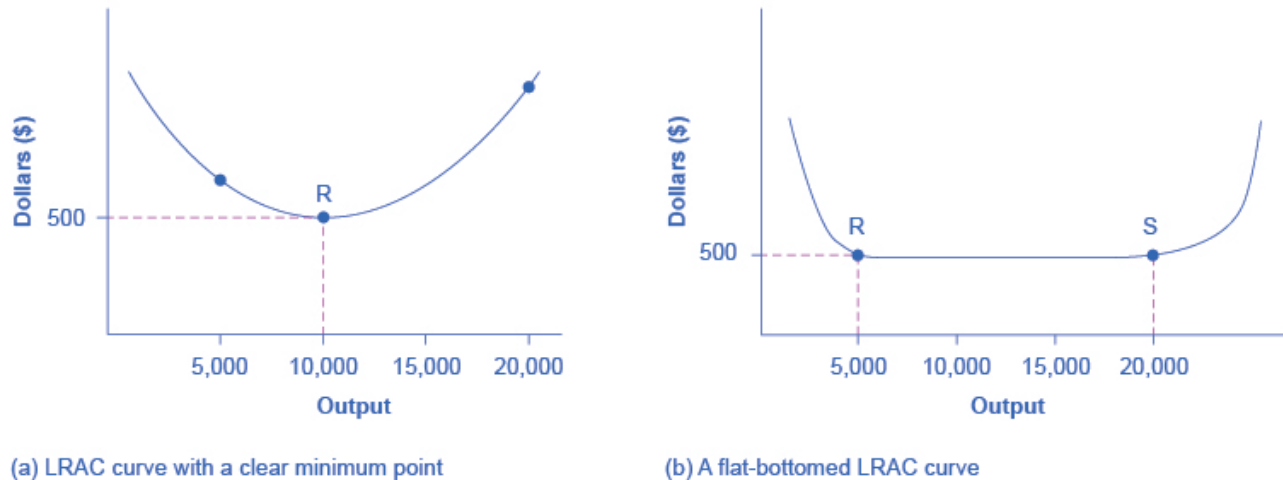


Figure 3. The LRAC Curve and the Size and Number of Firms. (a) Low-cost firms will produce at output level R. When the LRAC curve has a clear minimum point, then any firm producing a different quantity will have higher costs. In this case, a firm producing at a quantity of 10,000 will produce at a lower average cost than a firm producing, say, 5,000 or 20,000 units. (b) Low-cost firms will produce between output levels R and S. When the LRAC curve has a flat bottom, then firms producing at any quantity along this flat bottom can compete. In this case, any firm producing a quantity between 5,000 and 20,000 can compete effectively, although firms producing less than 5,000 or more than 20,000 would face higher average costs and be unable to compete.

museums, can operate only if they can draw on a large nearby population base. Cities are big enough to offer a wide variety of products, which is what many shoppers are looking for.

These factors are not exactly economies of scale in the narrow sense of the production function of a single firm, but they are related to growth in the overall size of population and market in an area. Cities are sometimes called “agglomeration economies.”

These agglomeration factors help to explain why every economy, as it develops, has an increasing proportion of its population living in urban areas. In the United States, about 80% of the population now lives in metropolitan areas (which include the suburbs around cities), compared to just 40% in 1900. However, in poorer nations of the world, including much of Africa, the proportion of the population in urban areas is only about 30%. One of the great challenges for these countries as their economies grow will be to manage the growth of the great cities that will arise.

If cities offer economic advantages that are a form of economies of scale, then why don't all or most people live in one giant city? At some point, agglomeration economies must turn into diseconomies. For example, traffic congestion may reach a point where the gains from being geographically nearby are counterbalanced by how long it takes to travel. High densities of people, cars, and factories can mean more garbage and air and water pollution. Facilities like parks or museums may become overcrowded. There may be economies of scale for negative activities like crime, because high densities of people and businesses, combined with the greater impersonality of cities, make it easier for illegal activities as well as legal ones. The future of cities, both in the United States and in other countries around the world, will be determined by their ability to benefit from the economies of agglomeration and to minimize or counterbalance the corresponding diseconomies.

A more common case is illustrated in Figure 3 (b), where the LRAC curve has a flat-bottomed area of constant returns to scale. In this situation, any firm with a level of output between 5,000 and 20,000 will be able to produce at about the same level of average cost. Given that the market will demand one million dishwashers per year at a price of \$500, this market might have as many as 200 producers (that is, one million dishwashers divided by firms making 5,000 each) or as few as 50 producers (one

million dishwashers divided by firms making 20,000 each). The producers in this market will range in size from firms that make 5,000 units to firms that make 20,000 units. But firms that produce below 5,000 units or more than 20,000 will be unable to compete, because their average costs will be too high. Thus, if we see an industry where almost all plants are the same size, it is likely that the long-run average cost curve has a unique bottom point as in Figure 3 (a). However, if the long-run average cost curve has a wide flat bottom like Figure 3 (b), then firms of a variety of different sizes will be able to compete with each other.

The flat section of the long-run average cost curve in Figure 3 (b) can be interpreted in two different ways. One interpretation is that a single manufacturing plant producing a quantity of 5,000 has the same average costs as a single manufacturing plant with four times as much capacity that produces a quantity of 20,000. The other interpretation is that one firm owns a single manufacturing plant that produces a quantity of 5,000, while another firm owns four separate manufacturing plants, which each produce a quantity of 5,000. This second explanation, based on the insight that a single firm may own a number of different manufacturing plants, is especially useful in explaining why the long-run average cost curve often has a large flat segment—and thus why a seemingly smaller firm may be able to compete quite well with a larger firm. At some point, however, the task of coordinating and managing many different plants raises the cost of production sharply, and the long-run average cost curve slopes up as a result.

In the examples to this point, the quantity demanded in the market is quite large (one million) compared with the quantity produced at the bottom of the long-run average cost curve (5,000, 10,000 or 20,000). In such a situation, the market is set for competition between many firms. But what if the bottom of the long-run average cost curve is at a quantity of 10,000 and the total market demand at that price is only slightly higher than that quantity—or even somewhat lower?

Return to Figure 3 (a), where the bottom of the long-run average cost curve is at 10,000, but now imagine that the total quantity of dishwashers demanded in the market at that price of \$500 is only 30,000. In this situation, the total number of firms in the market would be three. A handful of firms in a market is called an “oligopoly,” and the chapter on Monopolistic Competition and Oligopoly will discuss the range of competitive strategies that can occur when oligopolies compete.

Alternatively, consider a situation, again in the setting of Figure 3 (a), where the bottom of the long-run average cost curve is 10,000, but total demand for the product is only 5,000. (For simplicity, imagine that this demand is highly inelastic, so that it does not vary according to price.) In this situation, the market may well end up with a single firm—a monopoly—producing all 5,000 units. If any firm tried to challenge this monopoly while producing a quantity lower than 5,000 units, the prospective competitor firm would have a higher average cost, and so it would not be able to compete in the longer term without losing money. The chapter on Monopoly discusses the situation of a monopoly firm.

Thus, the shape of the long-run average cost curve reveals whether competitors in the market will be different sizes. If the LRAC curve has a single point at the bottom, then the firms in the market will be about the same size, but if the LRAC curve has a flat-bottomed segment of constant returns to scale, then firms in the market may be a variety of different sizes.

The relationship between the quantity at the minimum of the long-run average cost curve and the quantity demanded in the market at that price will predict how much competition is likely to exist in the market. If the quantity demanded in the market far exceeds the quantity at the minimum of the

LRAC, then many firms will compete. If the quantity demanded in the market is only slightly higher than the quantity at the minimum of the LRAC, a few firms will compete. If the quantity demanded in the market is less than the quantity at the minimum of the LRAC, a single-producer monopoly is a likely outcome.

SHIFTING PATTERNS OF LONG-RUN AVERAGE COST

New developments in production technology can shift the long-run average cost curve in ways that can alter the size distribution of firms in an industry.

For much of the twentieth century, the most common change has been to see alterations in technology, like the assembly line or the large department store, where large-scale producers seemed to gain an advantage over smaller ones. In the long-run average cost curve, the downward-sloping economies of scale portion of the curve stretched over a larger quantity of output.

However, new production technologies do not inevitably lead to a greater average size for firms. For example, in recent years some new technologies for generating electricity on a smaller scale have appeared. The traditional coal-burning electricity plants needed to produce 300 to 600 megawatts of power to exploit economies of scale fully. However, high-efficiency turbines to produce electricity from burning natural gas can produce electricity at a competitive price while producing a smaller quantity of 100 megawatts or less. These new technologies create the possibility for smaller companies or plants to generate electricity as efficiently as large ones. Another example of a technology-driven shift to smaller plants may be taking place in the tire industry. A traditional mid-size tire plant produces about six million tires per year. However, in 2000, the Italian company Pirelli introduced a new tire factory that uses many robots. The Pirelli tire plant produced only about one million tires per year, but did so at a lower average cost than a traditional mid-sized tire plant.

Controversy has simmered in recent years over whether the new information and communications technologies will lead to a larger or smaller size for firms. On one side, the new technology may make it easier for small firms to reach out beyond their local geographic area and find customers across a state, or the nation, or even across international boundaries. This factor might seem to predict a future with a larger number of small competitors. On the other side, perhaps the new information and communications technology will create “winner-take-all” markets where one large company will tend to command a large share of total sales, as Microsoft has done in the production of software for personal computers or Amazon has done in online bookselling. Moreover, improved information and communication technologies might make it easier to manage many different plants and operations across the country or around the world, and thus encourage larger firms. This ongoing battle between the forces of smallness and largeness will be of great interest to economists, businesspeople, and policymakers.

AMAZON

Traditionally, bookstores have operated in retail locations with inventories held either on the shelves or in the back of the store. These retail locations were very pricey in terms of rent. Amazon has no retail locations; it sells online and delivers by mail. Amazon offers almost any book in print, convenient purchasing, and prompt delivery by mail. Amazon holds its inventories in huge warehouses in low-rent locations around the world. The warehouses are highly computerized using robots and relatively low-skilled workers, making for low average costs per sale. Amazon demonstrates the significant advantages economies of scale can offer to a firm that exploits those economies.

KEY CONCEPTS AND SUMMARY

A production technology refers to a specific combination of labor, physical capital, and technology that makes up a particular method of production.

In the long run, firms can choose their production technology, and so all costs become variable costs. In making this choice, firms will try to substitute relatively inexpensive inputs for relatively expensive inputs where possible, so as to produce at the lowest possible long-run average cost.

Economies of scale refers to a situation where as the level of output increases, the average cost decreases. Constant returns to scale refers to a situation where average cost does not change as output increases. Diseconomies of scale refers to a situation where as output increases, average costs increase also.

The long-run average cost curve shows the lowest possible average cost of production, allowing all the inputs to production to vary so that the firm is choosing its production technology. A downward-sloping LRAC shows economies of scale; a flat LRAC shows constant returns to scale; an upward-sloping LRAC shows diseconomies of scale. If the long-run average cost curve has only one quantity produced that results in the lowest possible average cost, then all of the firms competing in an industry should be the same size. However, if the LRAC has a flat segment at the bottom, so that a range of different quantities can be produced at the lowest average cost, the firms competing in the industry will display a range of sizes. The market demand in conjunction with the long-run average cost curve determines how many firms will exist in a given industry.

If the quantity demanded in the market of a certain product is much greater than the quantity found at the bottom of the long-run average cost curve, where the cost of production is lowest, the market will have many firms competing. If the quantity demanded in the market is less than the quantity at the bottom of the LRAC, there will likely be only one firm.

SELF-CHECK QUESTIONS

1. Return to the problem explained in Table 6 and Table 7. If the cost of labor remains at \$40, but the cost of a machine decreases to \$50, what would be the total cost of each method of production? Which method should the firm use, and why?
2. Suppose the cost of machines increases to \$55, while the cost of labor stays at \$40. How would that affect the total cost of the three methods? Which method should the firm choose now?
3. Automobile manufacturing is an industry subject to significant economies of scale. Suppose there are four domestic auto manufacturers, but the demand for domestic autos is no more than 2.5 times the quantity produced at the bottom of the long-run average cost curve. What do you expect will happen to the domestic auto industry in the long run?

REVIEW QUESTIONS

1. What shapes would you generally expect each of the following cost curves to have: fixed costs, variable costs, marginal costs, average total costs, and average variable costs?
2. What is a production technology?
3. In choosing a production technology, how will firms react if one input becomes relatively more expensive?
4. What is a long-run average cost curve?
5. What is the difference between economies of scale, constant returns to scale, and diseconomies of scale?
6. What shape of a long-run average cost curve illustrates economies of scale, constant returns to scale, and diseconomies of scale?
7. Why will firms in most markets be located at or close to the bottom of the long-run average cost curve?

CRITICAL THINKING QUESTIONS

1. It is clear that businesses operate in the short run, but do they ever operate in the long run? Discuss.
2. How would an improvement in technology, like the high-efficiency gas turbines or Pirelli tire plant, affect the long-run average cost curve of a firm? Can you draw the old curve and the new one on the same axes? How might such an improvement affect other firms in the industry?
3. Do you think that the taxicab industry in large cities would be subject to significant economies of scale? Why or why not?

PROBLEMS

PROBLEMS

A small company that shovels sidewalks and driveways has 100 homes signed up for its services this winter. It can use various combinations of capital and labor: lots of labor with hand shovels, less labor with snow blowers, and still less labor with a pickup truck that has a snowplow on front. To summarize, the method choices are:

Method 1: 50 units of labor, 10 units of capital

Method 2: 20 units of labor, 40 units of capital

Method 3: 10 units of labor, 70 units of capital

If hiring labor for the winter costs \$100/unit and a unit of capital costs \$400, what production method should be chosen? What method should be chosen if the cost of labor rises to \$200/unit?

GLOSSARY

constant returns to scale expanding all inputs proportionately does not change the average cost of production

diseconomies of scale the long-run average cost of producing each individual unit increases as total output increases

long-run average cost (LRAC) curve shows the lowest possible average cost of production, allowing all the inputs to production to vary so that the firm is choosing its production technology

production technologies alternative methods of combining inputs to produce output

short-run average cost (SRAC) curve the average total cost curve in the short term; shows the total of the average fixed costs and the average variable costs

SOLUTIONS

Answers to Self-Check Questions

1. The new table should look like this:

	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	$10 \times \$40 = \400	$2 \times \$50 = \100	\$500
Cost of technology 2	$7 \times \$40 = \280	$4 \times \$50 = \200	\$480
Cost of technology 3	$3 \times \$40 = \120	$7 \times \$50 = \350	\$470

Table 9.

The firm should choose production technology 3 since it has the lowest total cost. This makes sense since, with cheaper machine hours, one would expect a shift in the direction of more machines and less labor.

- 2.

	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	$10 \times \$40 = \400	$2 \times \$55 = \110	\$510
Cost of technology 2	$7 \times \$40 = \280	$4 \times \$55 = \220	\$500
Cost of technology 3	$3 \times \$40 = \120	$7 \times \$55 = \385	\$505

Table 10.

The firm should choose production technology 2 since it has the lowest total cost. Because the cost of machines increased (relative to the previous question), you would expect a shift toward less capital and more labor.

3. This is the situation that existed in the United States in the 1970s. Since there is only demand enough for 2.5 firms to reach the bottom of the average cost curve, you would expect one firm will not be around in the long run, and at least one firm will be struggling.

CHAPTER 11. PERFECT COMPETITION

INTRODUCTION TO PERFECT COMPETITION



Figure 1. Depending upon the competition and prices offered, a wheat farmer may choose to grow a different crop. (Credit: modification of work by Daniel X. O’Neil/Flickr Creative Commons)

A DIME A DOZEN

When you were younger did you babysit, deliver papers, or mow the lawn for money? If so, you faced stiff competition from a lot of other competitors who offered identical services. There was nothing to stop others from offering their services too.

All of you charged the “going rate.” If you tried to charge more, your customers would simply buy from someone else. These conditions are very similar to the conditions agricultural growers face.

Growing a crop may be more difficult to start than a babysitting or lawn mowing service, but growers face the same fierce competition. In the grand scale of world agriculture, farmers face competition from thousands of others because they sell an identical product. After all, winter wheat is winter wheat. But it is relatively easy for farmers to leave the marketplace for another crop. In this case, they do not sell the family farm, they switch crops.

Take the case of the upper Midwest region of the United States—for many generations the area was called “King Wheat.”

According to the United States Department of Agriculture National Agricultural Statistics Service, statistics by state, in 1997, 11.6 million acres of wheat and 780,000 acres of corn were planted in North Dakota. In the intervening 15 or so years has the mix of crops changed? Since it is relatively easy to switch crops, did farmers change what was planted as the relative crop prices changed? We will find out at chapter's end.

In the meantime, let's consider the topic of this chapter—the perfectly competitive market. This is a market in which entry and exit are relatively easy and competitors are “a dime a dozen.”

CHAPTER OBJECTIVES

Introduction to Perfect Competition

In this chapter, you will learn about:

- Perfect Competition and Why It Matters
- How Perfectly Competitive Firms Make Output Decisions
- Entry and Exit Decisions in the Long Run
- Efficiency in Perfectly Competitive Markets

All businesses face two realities: no one is required to buy their products, and even customers who might want those products may buy from other businesses instead. Firms that operate in perfectly competitive markets face this reality. In this chapter, you will learn how such firms make decisions about how much to produce, how much profit they make, whether to stay in business or not, and many others. Industries differ from one another in terms of how many sellers there are in a specific market, how easy or difficult it is for a new firm to enter, and the type of products that are sold. This is referred to as the **market structure** of the industry. In this chapter, we focus on perfect competition. However, in other chapters we will examine other industry types: Monopoly and Monopolistic Competition and Oligopoly.

11.1 PERFECT COMPETITION AND WHY IT MATTERS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the characteristics of a perfectly competitive market
- Discuss how perfectly competitive firms react in the short run and in the long run

Firms are said to be in **perfect competition** when the following conditions occur: (1) many firms produce identical products; (2) many buyers are available to buy the product, and many sellers are available to sell the product; (3) sellers and buyers have all relevant information to make rational decisions about the product being bought and sold; and (4) firms can enter and leave the market without any restrictions—in other words, there is free entry and exit into and out of the market.

A **perfectly competitive firm** is known as a **price taker**, because the pressure of competing firms forces them to accept the prevailing equilibrium price in the market. If a firm in a perfectly competitive market raises the price of its product by so much as a penny, it will lose all of its sales to competitors. When a wheat grower, as discussed in the Bring it Home feature, wants to know what the going price of wheat is, he or she has to go to the computer or listen to the radio to check. The market price is determined solely by supply and demand in the entire market and not the individual farmer. Also, a perfectly competitive firm must be a very small player in the overall market, so that it can increase or decrease output without noticeably affecting the overall quantity supplied and price in the market.

A perfectly competitive market is a hypothetical extreme; however, producers in a number of industries do face many competitor firms selling highly similar goods, in which case they must often act as price takers. Agricultural markets are often used as an example. The same crops grown by different farmers are largely interchangeable. According to the United States Department of Agriculture monthly reports, in 2015, U.S. corn farmers received an average price of \$6.00 per bushel and wheat farmers received an average price of \$6.00 per bushel. A corn farmer who attempted to sell at \$7.00 per bushel, or a wheat grower who attempted to sell for \$8.00 per bushel, would not have found any buyers. A perfectly competitive firm will not sell below the equilibrium price either. Why should they when they can sell all they want at the higher price? Other examples of agricultural markets that operate in close to perfectly competitive markets are small roadside produce markets and small organic farmers.

Visit this website that reveals the current value of various commodities.



This chapter examines how profit-seeking firms decide how much to produce in perfectly competitive markets. Such firms will analyze their costs as discussed in the chapter on Cost and Industry Structure. In the short run, the perfectly competitive firm will seek the quantity of output where profits are highest or, if profits are not possible, where losses are lowest. In this example, the “short run” refers to a situation in which firms are producing with one fixed input and incur fixed costs of production. (In the real world, firms can have many fixed inputs.)

In the long run, perfectly competitive firms will react to profits by increasing production. They will respond to losses by reducing production or exiting the market. Ultimately, a long-run *equilibrium* will be attained when no new firms want to enter the market and existing firms do not want to leave the market, as economic profits have been driven down to zero.

KEY CONCEPTS AND SUMMARY

A perfectly competitive firm is a price taker, which means that it must accept the equilibrium price at which it sells goods. If a perfectly competitive firm attempts to charge even a tiny amount more than the market price, it will be unable to make any sales. In a perfectly competitive market there are thousands of sellers, easy entry, and identical products. A short-run production period is when firms are producing with some fixed inputs. Long-run equilibrium in a perfectly competitive industry occurs after all firms have entered and exited the industry and seller profits are driven to zero.

Perfect competition means that there are many sellers, there is easy entry and exiting of firms, products are identical from one seller to another, and sellers are price takers.

SELF-CHECK QUESTIONS

1. Firms in a perfectly competitive market are said to be “price takers”—that is, once the market determines an equilibrium price for the product, firms must accept this price. If you sell a product in a perfectly competitive market, but you are not happy with its price, would you raise the price, even by a cent?
2. Would independent trucking fit the characteristics of a perfectly competitive industry?

REVIEW QUESTIONS

1. A single firm in a perfectly competitive market is relatively small compared to the rest of the market. What does this mean? How “small” is “small”?

2. What are the four basic assumptions of perfect competition? Explain in words what they imply for a perfectly competitive firm.
3. What is a “price taker” firm?

CRITICAL THINKING QUESTIONS

1. Finding a life partner is a complicated process that may take many years. It is hard to think of this process as being part of a very complex market, with a demand and a supply for partners. Think about how this market works and some of its characteristics, such as search costs. Would you consider it a perfectly competitive market?
2. Can you name five examples of perfectly competitive markets? Why or why not?

GLOSSARY

market structure the conditions in an industry, such as number of sellers, how easy or difficult it is for a new firm to enter, and the type of products that are sold

perfect competition each firm faces many competitors that sell identical products

price taker a firm in a perfectly competitive market that must take the prevailing market price as given

SOLUTIONS

Answers to Self-Check Questions

1. No, you would not raise the price. Your product is exactly the same as the product of the many other firms in the market. If your price is greater than that of your competitors, then your customers would switch to them and stop buying from you. You would lose all your sales.
2. Possibly. Independent truckers are by definition small and numerous. All that is required to get into the business is a truck (not an inexpensive asset, though) and a commercial driver’s license. To exit, one need only sell the truck. All trucks are essentially the same, providing transportation from point A to point B. (We’re assuming we not talking about specialized trucks.) Independent truckers must take the going rate for their service, so independent trucking does seem to have most of the characteristics of perfect competition.

11.2 HOW PERFECTLY COMPETITIVE FIRMS MAKE OUTPUT DECISIONS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Calculate profits by comparing total revenue and total cost
- Identify profits and losses with the average cost curve
- Explain the shutdown point
- Determine the price at which a firm should continue producing in the short run

A **perfectly competitive firm** has only one major decision to make—namely, what quantity to produce. To understand why this is so, consider a different way of writing out the basic definition of **profit**:

$$\text{Profit} = \text{Total revenue} - \text{Total cost}$$

$$(\text{Price})(\text{Quantity produced}) - (\text{Average cost})(\text{Quantity produced})$$

Since a perfectly competitive firm must accept the price for its output as determined by the product's market demand and supply, it cannot choose the price it charges. This is already determined in the profit equation, and so the perfectly competitive firm can sell any number of units at exactly the same price. It implies that the firm faces a perfectly elastic demand curve for its product: buyers are willing to buy any number of units of output from the firm at the market price. When the perfectly competitive firm chooses what quantity to produce, then this quantity—along with the prices prevailing in the market for output and inputs—will determine the firm's total revenue, total costs, and ultimately, level of profits.

DETERMINING THE HIGHEST PROFIT BY COMPARING TOTAL REVENUE AND TOTAL COST

A perfectly competitive firm can sell as large a quantity as it wishes, as long as it accepts the prevailing market price. Total revenue is going to increase as the firm sells more, depending on the price of the product and the number of units sold. If you increase the number of units sold at a given price, then total revenue will increase. If the price of the product increases for every unit sold, then total revenue also increases. As an example of how a perfectly competitive firm decides what quantity to produce, consider the case of a small farmer who produces raspberries and sells them frozen for \$4 per pack. Sales of one pack of raspberries will bring in \$4, two packs will be \$8, three packs will be \$12, and so

on. If, for example, the price of frozen raspberries doubles to \$8 per pack, then sales of one pack of raspberries will be \$8, two packs will be \$16, three packs will be \$24, and so on.

Total revenue and **total costs** for the raspberry farm, broken down into fixed and variable costs, are shown in Table 1 and also appear in Figure 1. The horizontal axis shows the quantity of frozen raspberries produced in packs; the vertical axis shows both total revenue and total costs, measured in dollars. The total cost curve intersects with the vertical axis at a value that shows the level of fixed costs, and then slopes upward. All these cost curves follow the same characteristics as the curves covered in the Cost and Industry Structure chapter.

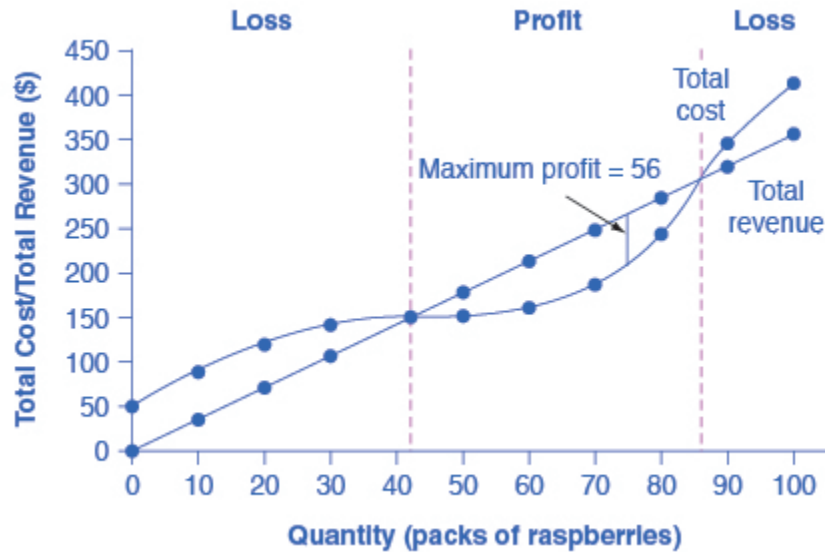


Figure 1. Total Cost and Total Revenue at the Raspberry Farm. Total revenue for a perfectly competitive firm is a straight line sloping up. The slope is equal to the price of the good. Total cost also slopes up, but with some curvature. At higher levels of output, total cost begins to slope upward more steeply because of diminishing marginal returns. The maximum profit will occur at the quantity where the gap of total revenue over total cost is largest.

Quantity	Total Cost	Fixed Cost	Variable Cost	Total Revenue	Profit
(Q)	(TC)	(FC)	(VC)	(TR)	
0	\$62	\$62	–	\$0	–\$62
10	\$90	\$62	\$28	\$40	–\$50
20	\$110	\$62	\$48	\$80	–\$30
30	\$126	\$62	\$64	\$120	–\$6
40	\$144	\$62	\$82	\$160	\$16
50	\$166	\$62	\$104	\$200	\$34
60	\$192	\$62	\$130	\$240	\$48
70	\$224	\$62	\$162	\$280	\$56
80	\$264	\$62	\$202	\$320	\$56
90	\$324	\$62	\$262	\$360	\$36
100	\$404	\$62	\$342	\$400	–\$4

Table 1. Total Cost and Total Revenue at the Raspberry Farm

Based on its total revenue and total cost curves, a perfectly competitive firm like the raspberry farm can calculate the quantity of output that will provide the highest level of profit. At any given quantity, total revenue minus total cost will equal profit. One way to determine the most profitable quantity to produce is to see at what quantity total revenue exceeds total cost by the largest amount. On Figure 1, the vertical gap between total revenue and total cost represents either profit (if total revenues are greater than total costs at a certain quantity) or losses (if total costs are greater than total revenues at a certain quantity). In this example, total costs will exceed total revenues at output levels from 0 to 40, and so over this range of output, the firm will be making losses. At output levels from 50 to 80, total revenues exceed total costs, so the firm is earning profits. But then at an output of 90 or 100, total costs again exceed total revenues and the firm is making losses. Total profits appear in the final column of Table 1. The highest total profits in the table, as in the figure that is based on the table values, occur at an output of 70–80, when profits will be \$56.

A higher price would mean that total revenue would be higher for every quantity sold. A lower price would mean that total revenue would be lower for every quantity sold. What happens if the price drops low enough so that the total revenue line is completely below the total cost curve; that is, at every level of output, total costs are higher than total revenues? In this instance, the best the firm can do is to suffer losses. But a profit-maximizing firm will prefer the quantity of output where total revenues come closest to total costs and thus where the losses are smallest.

(Later we will see that sometimes it will make sense for the firm to shutdown, rather than stay in operation producing output.)

COMPARING MARGINAL REVENUE AND MARGINAL COSTS

Firms often do not have the necessary data they need to draw a complete total cost curve for all levels of production. They cannot be sure of what total costs would look like if they, say, doubled production or cut production in half, because they have not tried it. Instead, firms experiment. They produce a slightly greater or lower quantity and observe how profits are affected. In economic terms, this prac-

tical approach to maximizing profits means looking at how changes in production affect marginal revenue and marginal cost.

Figure 2 presents the marginal revenue and marginal cost curves based on the total revenue and total cost in Table 1. The **marginal revenue** curve shows the additional revenue gained from selling one more unit. As mentioned before, a firm in perfect competition faces a perfectly elastic demand curve for its product—that is, the firm’s demand curve is a horizontal line drawn at the market price level. This also means that the firm’s marginal revenue curve is the same as the firm’s demand curve: Every time a consumer demands one more unit, the firm sells one more unit and revenue goes up by exactly the same amount equal to the market price. In this example, every time a pack of frozen raspberries is sold, the firm’s revenue increases by \$4. Table 2 shows an example of this. This condition only holds for price taking firms in perfect competition where:

$$\text{marginal revenue} = \text{price}$$

The formula for marginal revenue is:

$$\text{marginal revenue} = \frac{\text{change in total revenue}}{\text{change in quantity}}$$

Price	Quantity	Total Revenue	Marginal Revenue
\$4	1	\$4	–
\$4	2	\$8	\$4
\$4	3	\$12	\$4
\$4	4	\$16	\$4

Table 2. Marginal Revenue

Notice that marginal revenue does not change as the firm produces more output. That is because the price is determined by supply and demand and does not change as the farmer produces more (keeping in mind that, due to the relative small size of each firm, increasing their supply has no impact on the total market supply where price is determined).

Since a perfectly competitive firm is a price taker, it can sell whatever quantity it wishes at the market-determined price. Marginal cost, the cost per additional unit sold, is calculated by dividing the change in total cost by the change in quantity. The formula for marginal cost is:

$$\text{marginal cost} = \frac{\text{change in total cost}}{\text{change in quantity}}$$

Ordinarily, marginal cost changes as the firm produces a greater quantity.

In the raspberry farm example, shown in Figure 2, Figure 3 and Table 3, marginal cost at first declines as production increases from 10 to 20 to 30 packs of raspberries—which represents the area of increasing marginal returns that is not uncommon at low levels of production. But then marginal costs start to increase, displaying the typical pattern of diminishing marginal returns. If the firm is producing at a quantity where $MR > MC$, like 40 or 50 packs of raspberries, then it can increase profit by increasing output because the marginal revenue is exceeding the marginal cost. If the firm is producing at a quantity where $MC > MR$, like 90 or 100 packs, then it can increase profit by reducing

output because the reductions in marginal cost will exceed the reductions in marginal revenue. The firm's profit-maximizing choice of output will occur where $MR = MC$ (or at a choice close to that point). You will notice that what occurs on the production side is exemplified on the cost side. This is referred to as duality.

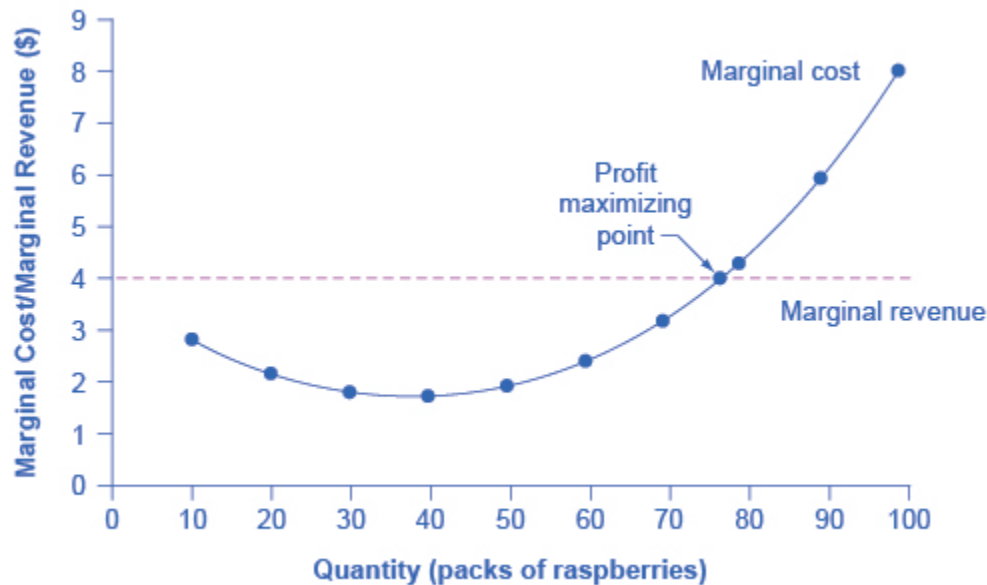


Figure 2. Marginal Revenues and Marginal Costs at the Raspberry Farm: Individual Farmer. For a perfectly competitive firm, the marginal revenue (MR) curve is a horizontal straight line because it is equal to the price of the good, which is determined by the market, shown in Figure 3. The marginal cost (MC) curve is sometimes first downward-sloping, if there is a region of increasing marginal returns at low levels of output, but is eventually upward-sloping at higher levels of output as diminishing marginal returns kick in.

Quantity	Total Cost	Fixed Cost	Variable Cost	Marginal Cost	Total Revenue	Marginal Revenue
0	\$62	\$62	–	–	–	–
10	\$90	\$62	\$28	\$2.80	\$40	\$4.00
20	\$110	\$62	\$48	\$2.00	\$80	\$4.00
30	\$126	\$62	\$64	\$1.60	\$120	\$4.00
40	\$144	\$62	\$82	\$1.80	\$160	\$4.00
50	\$166	\$62	\$104	\$2.20	\$200	\$4.00
60	\$192	\$62	\$130	\$2.60	\$240	\$4.00
70	\$224	\$62	\$162	\$3.20	\$280	\$4.00
80	\$264	\$62	\$202	\$4.00	\$320	\$4.00
90	\$324	\$62	\$262	\$6.00	\$360	\$4.00
100	\$404	\$62	\$342	\$8.00	\$400	\$4.00

Table 3. Marginal Revenues and Marginal Costs at the Raspberry Farm

In this example, the marginal revenue and **marginal cost** curves cross at a price of \$4 and a quantity of 80 produced. If the farmer started out producing at a level of 60, and then experimented with

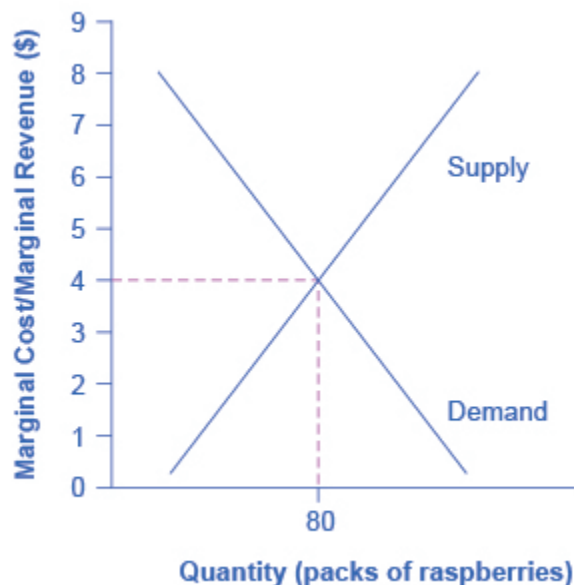


Figure 3. Marginal Revenues and Marginal Costs at the Raspberry Farm: Raspberry Market. The equilibrium price of raspberries is determined through the interaction of market supply and market demand at \$4.00.

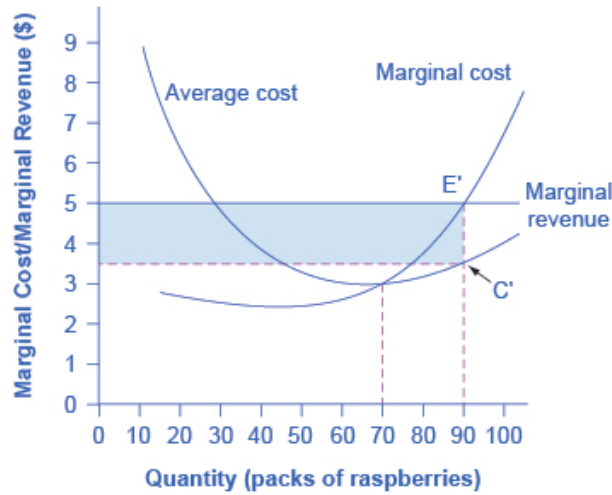
increasing production to 70, marginal revenues from the increase in production would exceed marginal costs—and so profits would rise. The farmer has an incentive to keep producing. From a level of 70 to 80, marginal cost and marginal revenue are equal so profit doesn't change. If the farmer then experimented further with increasing production from 80 to 90, he would find that marginal costs from the increase in production are greater than marginal revenues, and so profits would decline.

The profit-maximizing choice for a perfectly competitive firm will occur where marginal revenue is equal to marginal cost—that is, where $MR = MC$. A profit-seeking firm should keep expanding production as long as $MR > MC$. But at the level of output where $MR = MC$, the firm should recognize that it has achieved the highest possible level of economic profits. (In the example above, the profit maximizing output level is between 70 and 80 units of output, but the firm will not know they've maximized profit until they reach 80, where $MR = MC$.) Expanding production into the zone where $MR < MC$ will only reduce economic profits. Because the marginal revenue received by a perfectly competitive firm is equal to the price P , so that $P = MR$, the profit-maximizing rule for a perfectly competitive firm can also be written as a recommendation to produce at the quantity where $P = MC$.

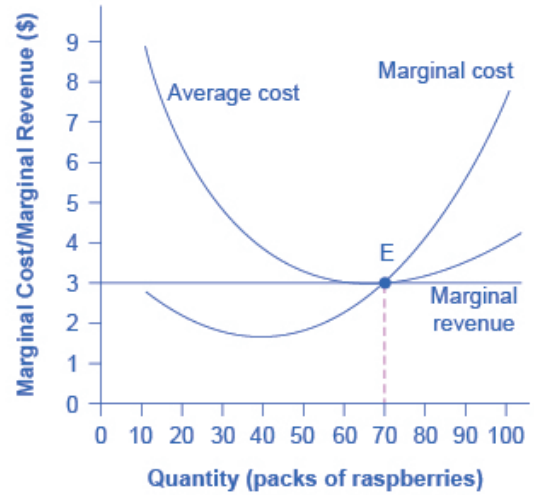
PROFITS AND LOSSES WITH THE AVERAGE COST CURVE

Does maximizing profit (producing where $MR = MC$) imply an actual economic profit? The answer depends on the relationship between price and average total cost. If the price that a firm charges is higher than its average cost of production for that quantity produced, then the firm will earn profits. Conversely, if the price that a firm charges is lower than its average cost of production, the firm will suffer losses. You might think that, in this situation, the farmer may want to shut down immediately. Remember, however, that the firm has already paid for fixed costs, such as equipment, so it may continue to produce and incur a loss. Figure 4 illustrates three situations: (a) where price intersects marginal cost at a level above the average cost curve, (b) where price intersects marginal cost at a level

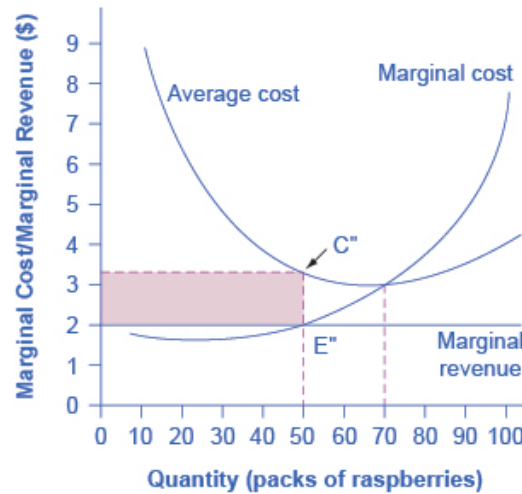
equal to the average cost curve, and (c) where price intersects marginal cost at a level below the average cost curve.



(a) Price is above average cost



(b) Price equals cost



(c) Price is below average cost

Figure 4. Price and Average Cost at the Raspberry Farm. In (a), price intersects marginal cost above the average cost curve. Since price is greater than average cost, the firm is making a profit. In (b), price intersects marginal cost at the minimum point of the average cost curve. Since price is equal to average cost, the firm is breaking even. In (c), price intersects marginal cost below the average cost curve. Since price is less than average cost, the firm is making a loss.

First consider a situation where the price is equal to \$5 for a pack of frozen raspberries. The rule for a profit-maximizing perfectly competitive firm is to produce the level of output where $\text{Price} = \text{MR} = \text{MC}$, so the raspberry farmer will produce a quantity of 90, which is labeled as e in Figure 4 (a). Remember that the area of a rectangle is equal to its base multiplied by its height. The farm's total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 90 packs (the base) up to point E' (the height), over to the price of \$5, and back to the origin. The

average cost of producing 80 packs is shown by point C or about \$3.50. Total costs will be the quantity of 80 times the average cost of \$3.50, which is shown by the area of the rectangle from the origin to a quantity of 90, up to point C, over to the vertical axis and down to the origin. It should be clear from examining the two rectangles that total revenue is greater than total cost. Thus, profits will be the blue shaded rectangle on top.

It can be calculated as:

$$\begin{aligned} \text{profit} &= \text{total revenue} - \text{total cost} \\ &= (90)(\$5.00) - (90)(\$3.50) \\ &= \$135 \end{aligned}$$

Or, it can be calculated as:

$$\begin{aligned} \text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\ &= (\$5.00 - \$3.50) \times 90 \\ &= \$135 \end{aligned}$$

Now consider Figure 4 (b), where the price has fallen to \$3.00 for a pack of frozen raspberries. Again, the perfectly competitive firm will choose the level of output where Price = MR = MC, but in this case, the quantity produced will be 70. At this price and output level, where the marginal cost curve is crossing the average cost curve, the price received by the firm is exactly equal to its average cost of production.

The farm's total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 70 packs (the base) up to point E (the height), over to the price of \$3, and back to the origin. The average cost of producing 70 packs is shown by point C'. Total costs will be the quantity of 70 times the average cost of \$3.00, which is shown by the area of the rectangle from the origin to a quantity of 70, up to point E, over to the vertical axis and down to the origin. It should be clear from that the rectangles for total revenue and total cost are the same. Thus, the firm is making zero profit. The calculations are as follows:

$$\begin{aligned} \text{profit} &= \text{total revenue} - \text{total cost} \\ &= (70)(\$3.00) - (70)(\$3.00) \\ &= \$0 \end{aligned}$$

Or, it can be calculated as:

$$\begin{aligned}
 \text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\
 &= (\$3.00 - \$3.00) \times 70 \\
 &= \$0
 \end{aligned}$$

In Figure 4 (c), the market price has fallen still further to \$2.00 for a pack of frozen raspberries. At this price, marginal revenue intersects marginal cost at a quantity of 50. The farm's total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 50 packs (the base) up to point E" (the height), over to the price of \$2, and back to the origin. The average cost of producing 50 packs is shown by point C" or about \$3.30. Total costs will be the quantity of 50 times the average cost of \$3.30, which is shown by the area of the rectangle from the origin to a quantity of 50, up to point C", over to the vertical axis and down to the origin. It should be clear from examining the two rectangles that total revenue is less than total cost. Thus, the firm is losing money and the loss (or negative profit) will be the rose-shaded rectangle.

The calculations are:

$$\begin{aligned}
 \text{profit} &= \text{total revenue} - \text{total cost} \\
 &= (50)(\$2.00) - (50)(\$3.30) \\
 &= -\$77.50
 \end{aligned}$$

Or:

$$\begin{aligned}
 \text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\
 &= (\$1.75 - \$3.30) \times 50 \\
 &= -\$77.50
 \end{aligned}$$

If the market price received by a perfectly competitive firm leads it to produce at a quantity where the price is greater than average cost, the firm will earn profits. If the price received by the firm causes it to produce at a quantity where price equals average cost, which occurs at the minimum point of the AC curve, then the firm earns zero profits. Finally, if the price received by the firm leads it to produce at a quantity where the price is less than average cost, the firm will earn losses. This is summarized in Table 4.

If...	Then...
Price > ATC	Firm earns an economic profit
Price = ATC	Firm earns zero economic profit
Price < ATC	Firm earns a loss

Table 4.

THE SHUTDOWN POINT

The possibility that a firm may earn losses raises a question: Why can the firm not avoid losses by shutting down and not producing at all? The answer is that shutting down can reduce variable costs to zero, but in the short run, the firm has already paid for fixed costs. As a result, if the firm produces a quantity of zero, it would still make losses because it would still need to pay for its fixed costs. So, when a firm is experiencing losses, it must face a question: should it continue producing or should it shut down?

As an example, consider the situation of the Yoga Center, which has signed a contract to rent space that costs \$10,000 per month. If the firm decides to operate, its marginal costs for hiring yoga teachers is \$15,000 for the month. If the firm shuts down, it must still pay the rent, but it would not need to hire labor. Table 5 shows three possible scenarios. In the first scenario, the Yoga Center does not have any clients, and therefore does not make any revenues, in which case it faces losses of \$10,000 equal to the fixed costs. In the second scenario, the Yoga Center has clients that earn the center revenues of \$10,000 for the month, but ultimately experiences losses of \$15,000 due to having to hire yoga instructors to cover the classes. In the third scenario, the Yoga Center earns revenues of \$20,000 for the month, but experiences losses of \$5,000.

In all three cases, the Yoga Center loses money. In all three cases, when the rental contract expires in the long run, assuming revenues do not improve, the firm should exit this business. In the short run, though, the decision varies depending on the level of losses and whether the firm can cover its variable costs. In scenario 1, the center does not have any revenues, so hiring yoga teachers would increase variable costs and losses, so it should shut down and only incur its fixed costs. In scenario 2, the center's losses are greater because it does not make enough revenue to offset the increased variable costs plus fixed costs, so it should shut down immediately. If price is below the minimum average variable cost, the firm must shut down. In contrast, in scenario 3 the revenue that the center can earn is high enough that the losses diminish when it remains open, so the center should remain open in the short run.

Scenario 1

If the center shuts down now, revenues are zero but it will not incur any variable costs and would only need to pay fixed costs of \$10,000.

$$\begin{aligned} \text{profit} &= \text{total revenue} - (\text{fixed costs} + \text{variable cost}) \\ &= 0 - \$10,000 \\ &= -\$10,000 \end{aligned}$$

Scenario 2

The center earns revenues of \$10,000, and variable costs are \$15,000. The center should shut down now.

$$\begin{aligned} \text{profit} &= \text{total revenue} - (\text{fixed costs} + \text{variable cost}) \\ &= \$10,000 - (\$10,000 + \$15,000) \\ &= -\$15,000 \end{aligned}$$

Scenario 3

The center earns revenues of \$20,000, and variable costs are \$15,000. The center should continue in business.

$$\begin{aligned} \text{profit} &= \text{total revenue} - (\text{fixed costs} + \text{variable cost}) \\ &= \$20,000 - (\$10,000 + \$15,000) \\ &= -\$5,000 \end{aligned}$$

Table 5. Should the Yoga Center Shut Down Now or Later?

This example suggests that the key factor is whether a firm can earn enough revenues to cover at least its variable costs by remaining open. Let's return now to our raspberry farm. Figure 5 illustrates this lesson by adding the average variable cost curve to the marginal cost and average cost curves. At a price of \$2.20 per pack, as shown in Figure 5 (a), the farm produces at a level of 50. It is making losses of \$56 (as explained earlier), but price is above average variable cost and so the firm continues to operate. However, if the price declined to \$1.80 per pack, as shown in Figure 5 (b), and if the firm applied its rule of producing where $P = MR = MC$, it would produce a quantity of 40. This price is below average variable cost for this level of output. If the farmer cannot pay workers (the variable costs), then it has to shut down. At this price and output, total revenues would be \$72 (quantity of 40 times price of \$1.80) and total cost would be \$144, for overall losses of \$72. If the farm shuts down, it must pay only its fixed costs of \$62, so shutting down is preferable to selling at a price of \$1.80 per pack.

Looking at Table 6, if the price falls below \$2.05, the minimum average variable cost, the firm must shut down.

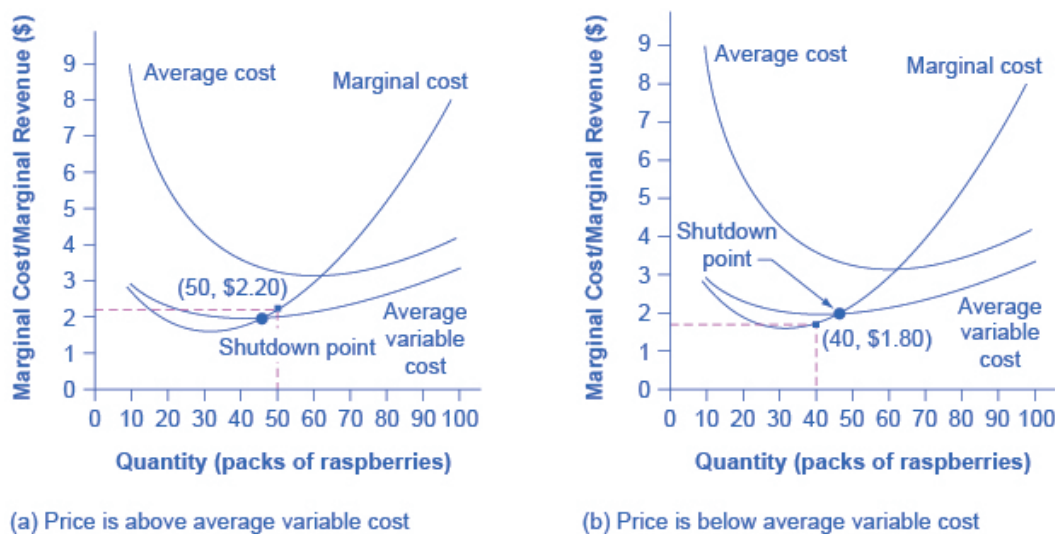


Figure 5. The Shutdown Point for the Raspberry Farm. In (a), the farm produces at a level of 50. It is making losses of \$56, but price is above average variable cost, so it continues to operate. In (b), total revenues are \$72 and total cost is \$144, for overall losses of \$72. If the farm shuts down, it must pay only its fixed costs of \$62. Shutting down is preferable to selling at a price of \$1.80 per pack.

Quantity	Total Cost	Fixed Cost	Variable Cost	Marginal Cost	Average Cost	Average Variable Cost
0	\$62	\$62	–	–	–	–
10	\$90	\$62	\$28	\$2.80	\$9.00	\$2.80
20	\$110	\$62	\$48	\$2.00	\$5.50	\$2.40
30	\$126	\$62	\$64	\$1.60	\$4.20	\$2.13
40	\$144	\$62	\$82	\$1.80	\$3.60	\$2.05
50	\$166	\$62	\$104	\$2.20	\$3.32	\$2.08
60	\$192	\$62	\$130	\$2.60	\$3.20	\$2.16
70	\$224	\$62	\$162	\$3.20	\$3.20	\$2.31
80	\$264	\$62	\$202	\$4.00	\$3.30	\$2.52
90	\$324	\$62	\$262	\$6.00	\$3.60	\$2.91
100	\$404	\$62	\$342	\$8.00	\$4.04	\$3.42

Table 6. Cost of Production for the Raspberry Farm

The intersection of the average variable cost curve and the marginal cost curve, which shows the price where the firm would lack enough revenue to cover its variable costs, is called the **shutdown point**. If the perfectly competitive firm can charge a price above the shutdown point, then the firm is at least covering its average variable costs. It is also making enough revenue to cover at least a portion of fixed costs, so it should limp ahead even if it is making losses in the short run, since at least those losses will be smaller than if the firm shuts down immediately and incurs a loss equal to total fixed costs. However, if the firm is receiving a price below the price at the shutdown point, then the firm is not even covering its variable costs. In this case, staying open is making the firm's losses larger, and it should shut down immediately. To summarize, if:

- price < minimum average variable cost, then firm shuts down
- price = minimum average variable cost, then firm stays in business

SHORT-RUN OUTCOMES FOR PERFECTLY COMPETITIVE FIRMS

The average cost and average variable cost curves divide the marginal cost curve into three segments, as shown in Figure 6. At the market price, which the perfectly competitive firm accepts as given, the profit-maximizing firm chooses the output level where price or marginal revenue, which are the same thing for a perfectly competitive firm, is equal to marginal cost: $P = MR = MC$.

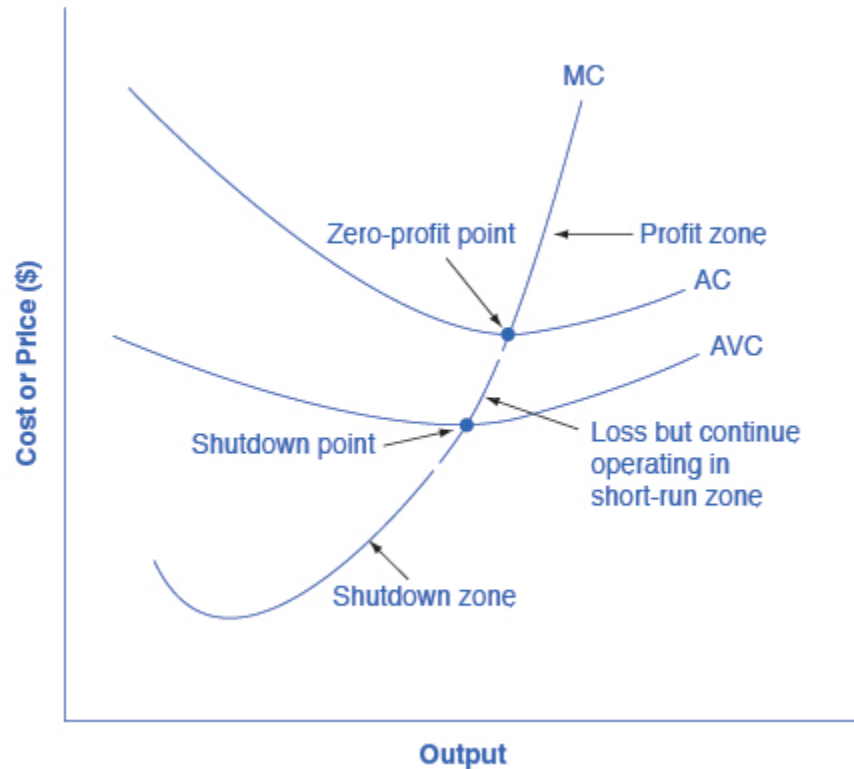


Figure 6. Profit, Loss, Shutdown. The marginal cost curve can be divided into three zones, based on where it is crossed by the average cost and average variable cost curves. The point where MC crosses AC is called the zero-profit point. If the firm is operating at a level of output where the market price is at a level higher than the zero-profit point, then price will be greater than average cost and the firm is earning profits. If the price is exactly at the zero-profit point, then the firm is making zero profits. If price falls in the zone between the shutdown point and the zero-profit point, then the firm is making losses but will continue to operate in the short run, since it is covering its variable costs. However, if price falls below the price at the shutdown point, then the firm will shut down immediately, since it is not even covering its variable costs.

First consider the upper zone, where prices are above the level where marginal cost (MC) crosses average cost (AC) at the zero profit point. At any price above that level, the firm will earn profits in the short run. If the price falls exactly on the zero profit point where the MC and AC curves cross, then the firm earns zero profits. If a price falls into the zone between the zero profit point, where MC crosses AC, and the shutdown point, where MC crosses AVC, the firm will be making losses in the

short run—but since the firm is more than covering its variable costs, the losses are smaller than if the firm shut down immediately. Finally, consider a price at or below the shutdown point where MC crosses AVC. At any price like this one, the firm will shut down immediately, because it cannot even cover its variable costs.

MARGINAL COST AND THE FIRM'S SUPPLY CURVE

For a perfectly competitive firm, the marginal cost curve is identical to the firm's supply curve starting from the minimum point on the average variable cost curve. To understand why this perhaps surprising insight holds true, first think about what the supply curve means. A firm checks the market price and then looks at its supply curve to decide what quantity to produce. Now, think about what it means to say that a firm will maximize its profits by producing at the quantity where $P = MC$. This rule means that the firm checks the market price, and then looks at its marginal cost to determine the quantity to produce—and makes sure that the price is greater than the minimum average variable cost. In other words, the marginal cost curve above the minimum point on the average variable cost curve becomes the firm's supply curve.

Watch this video that addresses how drought in the United States can impact food prices across the world. (Note that the story on the drought is the second one in the news report; you need to let the video play through the first story in order to watch the story on the drought.)



As discussed in the chapter on Demand and Supply, many of the reasons that supply curves shift relate to underlying changes in costs. For example, a lower price of key inputs or new technologies that reduce production costs cause supply to shift to the right; in contrast, bad weather or added government regulations can add to costs of certain goods in a way that causes supply to shift to the left. These shifts in the firm's supply curve can also be interpreted as shifts of the marginal cost curve. A shift in costs of production that increases marginal costs at all levels of output—and shifts MC to the left—will cause a perfectly competitive firm to produce less at any given market price. Conversely, a shift in costs of production that decreases marginal costs at all levels of output will shift MC to the right and as a result, a competitive firm will choose to expand its level of output at any given price. The following Work It Out feature will walk you through an example.

AT WHAT PRICE SHOULD THE FIRM CONTINUE PRODUCING IN THE SHORT RUN?

To determine the short-run economic condition of a firm in perfect competition, follow the steps outlined below. Use the data shown in Table 7.

Q	P	TFC	TVC	TC	AVC	ATC	MC	TR	Profits
0	\$28	\$20	\$0	-	-	-	-	-	-
1	\$28	\$20	\$20	-	-	-	-	-	-
2	\$28	\$20	\$25	-	-	-	-	-	-
3	\$28	\$20	\$35	-	-	-	-	-	-
4	\$28	\$20	\$52	-	-	-	-	-	-
5	\$28	\$20	\$80	-	-	-	-	-	-

Table 7.

Step 1. Determine the cost structure for the firm. For a given total fixed costs and variable costs, calculate total cost, average variable cost, average total cost, and marginal cost. Follow the formulas given in the Cost and Industry Structure chapter. These calculations are shown in Table 8.

Q	P	TFC	TVC	TC (TFC+TVC)	AVC (TVC/Q)	ATC (TC/Q)	MC (TC ₂ -TC ₁)/ (Q ₂ -Q ₁)
0	\$28	\$20	\$0	\$20+\$0=\$20	-	-	-
1	\$28	\$20	\$20	\$20+\$20=\$40	\$20/1=\$20.00	\$40/1=\$40.00	(\$40-\$20)/ (1-0)= \$20
2	\$28	\$20	\$25	\$20+\$25=\$45	\$25/2=\$12.50	\$45/2=\$22.50	(\$45-\$40)/ (2-1)= \$5
3	\$28	\$20	\$35	\$20+\$35=\$55	\$35/3=\$11.67	\$55/3=\$18.33	(\$55-\$45)/ (3-2)= \$10
4	\$28	\$20	\$52	\$20+\$52=\$72	\$52/4=\$13.00	\$72/4=\$18.00	(\$72-\$55)/ (4-3)= \$17
5	\$28	\$20	\$80	\$20+\$80=\$100	\$80/5=\$16.00	\$100/5=\$20.00	(\$100-\$72)/ (5-4)= \$28

Table 8.

Step 2. Determine the market price that the firm receives for its product. This should be given information, as the firm in perfect competition is a price taker. With the given price, calculate total revenue as equal to price multiplied by quantity for all output levels produced. In this example, the given price is \$30. You can see that in the second column of Table 9.

Quantity	Price	Total Revenue ($P \times Q$)
0	\$28	$\$28 \times 0 = \0
1	\$28	$\$28 \times 1 = \28
2	\$28	$\$28 \times 2 = \56
3	\$28	$\$28 \times 3 = \84
4	\$28	$\$28 \times 4 = \112
5	\$28	$\$28 \times 5 = \140

Table 9.

Step 3. Calculate profits as total cost subtracted from total revenue, as shown in Table 10.

Quantity	Total Revenue	Total Cost	Profits ($TR - TC$)
0	\$0	\$20	$\$0 - \$20 = -\$20$
1	\$28	\$40	$\$28 - \$40 = -\$12$
2	\$56	\$45	$\$56 - \$45 = \$11$
3	\$84	\$55	$\$84 - \$55 = \$29$
4	\$112	\$72	$\$112 - \$72 = \$40$
5	\$140	\$100	$\$140 - \$100 = \$40$

Table 10.

Step 4. To find the profit-maximizing output level, look at the Marginal Cost column (at every output level produced), as shown in Table 11, and determine where it is equal to the market price. The output level where price equals the marginal cost is the output level that maximizes profits.

Q	P	TFC	TVC	TC	AVC	ATC	MC	TR	Profits
0	\$28	\$20	\$0	\$20	–	–	–	\$0	–\$20
1	\$28	\$20	\$20	\$40	\$20.00	\$40.00	\$20	\$28	–\$12
2	\$28	\$20	\$25	\$45	\$12.50	\$22.50	\$5	\$56	\$11
3	\$28	\$20	\$35	\$55	\$11.67	\$18.33	\$10	\$84	\$29
4	\$28	\$20	\$52	\$72	\$13.00	\$18.00	\$17	\$112	\$40
5	\$28	\$20	\$80	\$100	\$16.40	\$20.40	\$30	\$140	\$40

Table 11.

Step 5. Once you have determined the profit-maximizing output level (in this case, output quantity 5), you can look at the amount of profits made (in this case, \$40).

Step 6. If the firm is making economic losses, the firm needs to determine whether it produces the output level where price equals marginal revenue and equals marginal cost or it shuts down and only incurs its fixed costs.

Step 7. For the output level where marginal revenue is equal to marginal cost, check if the market price is greater than the average variable cost of producing that output level.

- If $P > AVC$ but $P < ATC$, then the firm continues to produce in the short-run, making economic losses.
- If $P < AVC$, then the firm stops producing and only incurs its fixed costs.

In this example, the price of \$28 is greater than the AVC (\$16.40) of producing 5 units of output, so the firm continues producing.

KEY CONCEPTS AND SUMMARY

As a perfectly competitive firm produces a greater quantity of output, its total revenue steadily increases at a constant rate determined by the given market price. Profits will be highest (or losses will be smallest) at the quantity of output where total revenues exceed total costs by the greatest amount (or where total revenues fall short of total costs by the smallest amount). Alternatively, profits will be highest where marginal revenue, which is price for a perfectly competitive firm, is equal to marginal cost. If the market price faced by a perfectly competitive firm is above average cost at the profit-maximizing quantity of output, then the firm is making profits. If the market price is below average cost at the profit-maximizing quantity of output, then the firm is making losses.

If the market price is equal to average cost at the profit-maximizing level of output, then the firm is making zero profits. The point where the marginal cost curve crosses the average cost curve, at the minimum of the average cost curve, is called the “zero profit point.” If the market price faced by a perfectly competitive firm is below average variable cost at the profit-maximizing quantity of output, then the firm should shut down operations immediately. If the market price faced by a perfectly competitive firm is above average variable cost, but below average cost, then the firm should continue producing in the short run, but exit in the long run. The point where the marginal cost curve crosses the average variable cost curve is called the shutdown point.

SELF-CHECK QUESTIONS

1. Look at Table 12. What would happen to the firm's profits if the market price increases to \$6 per pack of raspberries?

Quantity	Total Cost	Fixed Cost	Variable Cost	Total Revenue	Profit
0	\$62	\$62	–	\$0	–\$62
10	\$90	\$62	\$28	\$60	–\$30
20	\$110	\$62	\$48	\$120	\$10
30	\$126	\$62	\$64	\$180	\$54
40	\$144	\$62	\$82	\$240	\$96
50	\$166	\$62	\$104	\$300	\$134
60	\$192	\$62	\$130	\$360	\$168
70	\$224	\$62	\$162	\$420	\$196
80	\$264	\$62	\$202	\$480	\$216
90	\$324	\$62	\$262	\$540	\$216
100	\$404	\$62	\$342	\$600	\$196

Table 12.

2. Suppose that the market price increases to \$6, as shown in Table 13. What would happen to the profit-maximizing output level?

Quantity	Total Cost	Fixed Cost	Variable Cost	Marginal Cost	Total Revenue	Marginal Revenue
0	\$62	\$62	–	–	\$0	–
10	\$90	\$62	\$28	\$2.80	\$60	\$6.00
20	\$110	\$62	\$48	\$2.00	\$120	\$6.00
30	\$126	\$62	\$64	\$1.60	\$180	\$6.00
40	\$144	\$62	\$82	\$1.80	\$240	\$6.00
50	\$166	\$62	\$104	\$2.20	\$300	\$6.00
60	\$192	\$62	\$130	\$2.60	\$360	\$6.00
70	\$224	\$62	\$162	\$3.20	\$420	\$6.00
80	\$264	\$62	\$202	\$4.00	\$480	\$6.00
90	\$324	\$62	\$262	\$6.00	\$540	\$6.00
100	\$404	\$62	\$342	\$8.00	\$600	\$6.00

Table 13.

3. Explain in words why a profit-maximizing firm will not choose to produce at a quantity where marginal cost exceeds marginal revenue.
4. A firm's marginal cost curve above the average variable cost curve is equal to the firm's individual supply curve. This means that every time a firm receives a price from the market it will be willing to supply the amount of output where the price equals marginal cost. What happens to the firm's individual supply curve if marginal costs increase?

REVIEW QUESTIONS

1. How does a perfectly competitive firm decide what price to charge?
2. What prevents a perfectly competitive firm from seeking higher profits by increasing the price that it charges?
3. How does a perfectly competitive firm calculate total revenue?
4. Briefly explain the reason for the shape of a marginal revenue curve for a perfectly competitive firm.
5. What two rules does a perfectly competitive firm apply to determine its profit-maximizing quantity of output?
6. How does the average cost curve help to show whether a firm is making profits or losses?
7. What two lines on a cost curve diagram intersect at the zero-profit point?
8. Should a firm shut down immediately if it is making losses?
9. How does the average variable cost curve help a firm know whether it should shut down immediately?
10. What two lines on a cost curve diagram intersect at the shutdown point?

CRITICAL THINKING QUESTIONS

1. Your company operates in a perfectly competitive market. You have been told that advertising can help you increase your sales in the short run. Would you create an aggressive advertising campaign for your product?
2. Since a perfectly competitive firm can sell as much as it wishes at the market price, why can the firm not simply increase its profits by selling an extremely high quantity?

PROBLEMS

1. The AAA Aquarium Co. sells aquariums for \$20 each. Fixed costs of production are \$20. The total variable costs are \$20 for one aquarium, \$25 for two units, \$35 for the three units, \$50 for four units, and \$80 for five units. In the form of a table, calculate total revenue, marginal revenue, total cost, and marginal cost for each output level (one to five units). What is the profit-maximizing quantity of output? On one diagram, sketch the total revenue and total cost curves. On another diagram, sketch the marginal revenue and marginal cost curves.
2. Perfectly competitive firm Doggies Paradise Inc. sells winter coats for dogs. Dog coats sell for \$72 each. The fixed costs of production are \$100. The total variable costs are \$64 for one unit, \$84 for two units, \$114 for three units, \$184 for four units, and \$270 for five units. In the form of a table, calculate total revenue, marginal revenue, total cost and marginal cost for each output level (one to five units). On one diagram, sketch the total revenue and total cost curves. On another diagram, sketch the marginal revenue and marginal cost curves. What is the profit maximizing quantity?
3. A computer company produces affordable, easy-to-use home computer systems and has fixed costs of \$250. The marginal cost of producing computers is \$700 for the first computer, \$250 for the second, \$300 for the third, \$350 for the fourth, \$400 for the fifth, \$450 for the sixth, and \$500 for the seventh.
 - a. Create a table that shows the company's output, total cost, marginal cost, average cost, variable cost, and average variable cost.
 - b. At what price is the zero-profit point? At what price is the shutdown point?
 - c. If the company sells the computers for \$500, is it making a profit or a loss? How big is the profit or loss? Sketch a graph with AC, MC, and AVC curves to illustrate your answer and show the profit or loss.
 - d. If the firm sells the computers for \$300, is it making a profit or a loss? How big is the profit or loss? Sketch a graph with AC, MC, and AVC curves to illustrate your answer and show the profit or loss.

GLOSSARY

marginal revenue the additional revenue gained from selling one more unit

shutdown point level of output where the marginal cost curve intersects the average variable cost curve at the minimum point of AVC; if the price is below this point, the firm should shut down immediately

SOLUTIONS

Answers to Self-Check Questions

1. Holding total cost constant, profits at every output level would increase.
2. When the market price increases, marginal revenue increases. The firm would then increase production up to the point where the new price equals marginal cost, at a quantity of 90.
3. If marginal costs exceeds marginal revenue, then the firm will reduce its profits for every additional unit of output it produces. Profit would be greatest if it reduces output to where $MR = MC$.
4. The firm will be willing to supply fewer units at every price level. In other words, the firm's individual supply curve decreases and shifts to the left.

11.3 ENTRY AND EXIT DECISIONS IN THE LONG RUN

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how entry and exit lead to zero profits in the long run
- Discuss the long-run adjustment process

The line between the short run and the long run cannot be defined precisely with a stopwatch, or even with a calendar. It varies according to the specific business. The distinction between the short run and the long run is therefore more technical: in the short run, firms cannot change the usage of fixed inputs, while in the long run, the firm can adjust all factors of production.

In a competitive market, profits are a red cape that incites businesses to charge. If a business is making a profit in the short run, it has an incentive to expand existing factories or to build new ones. New firms may start production, as well. When new firms enter the industry in response to increased industry profits it is called **entry**.

Losses are the black thundercloud that causes businesses to flee. If a business is making losses in the short run, it will either keep limping along or just shut down, depending on whether its revenues are covering its variable costs. But in the long run, firms that are facing losses will shut down at least some of their output, and some firms will cease production altogether. The long-run process of reducing production in response to a sustained pattern of losses is called **exit**. The following Clear It Up feature discusses where some of these losses might come from, and the reasons why some firms go out of business.

WHY DO FIRMS CEASE TO EXIST?

Can we say anything about what causes a firm to exit an industry? Profits are the measurement that determines whether a business stays operating or not. Individuals start businesses with the purpose of making profits. They invest their money, time, effort, and many other resources to produce and sell something that they hope will give them something in return. Unfortunately, not all businesses are successful, and many new startups soon realize that their “business adventure” must eventually end.

In the model of perfectly competitive firms, those that consistently cannot make money will “exit,” which is a nice, bloodless word for a more painful process. When a business fails, after all, workers lose their jobs, investors lose their money, and owners and managers can lose their dreams. Many businesses fail. The U.S. Small Business Administration indicates that in 2011, 409,040 new firms “entered,” and 470,376 firms failed.

Sometimes a business fails because of poor management or workers who are not very productive, or because of tough domestic or foreign competition. Businesses also fail from a variety of causes that might best be summarized as bad luck. For example, conditions of demand and supply in the market shift in an unexpected way, so that the prices that can be charged for outputs fall or the prices that need to be paid for inputs rise. With millions of businesses in the U.S. economy, even a small fraction of them failing will affect many people—and business failures can be very hard on the workers and managers directly involved. But from the standpoint of the overall economic system, business exits are sometimes a necessary evil if a market-oriented system is going to offer a flexible mechanism for satisfying customers, keeping costs low, and inventing new products.

HOW ENTRY AND EXIT LEAD TO ZERO PROFITS IN THE LONG RUN

No **perfectly competitive firm** acting alone can affect the market price. However, the combination of many firms entering or exiting the market will affect overall supply in the market. In turn, a shift in supply for the market as a whole will affect the market price. Entry and exit to and from the market are the driving forces behind a process that, in the long run, pushes the price down to minimum average total costs so that all firms are earning a zero profit.

To understand how short-run profits for a perfectly competitive firm will evaporate in the long run, imagine the following situation. The market is in **long-run equilibrium**, where all firms earn zero economic profits producing the output level where $P = MR = MC$ and $P = AC$. No firm has the incentive to enter or leave the market. Let's say that the product's demand increases, and with that, the market price goes up. The existing firms in the industry are now facing a higher price than before, so they will increase production to the new output level where $P = MR = MC$.

This will temporarily make the market price rise above the average cost curve, and therefore, the existing firms in the market will now be earning economic profits. However, these economic profits attract other firms to enter the market. Entry of many new firms causes the market supply curve to shift to the right. As the supply curve shifts to the right, the market price starts decreasing, and with that, economic profits fall for new and existing firms. As long as there are still profits in the market, entry will continue to shift supply to the right. This will stop whenever the market price is driven down to the zero-profit level, where no firm is earning economic profits.

Short-run losses will fade away by reversing this process. Say that the market is in long-run equilibrium. This time, instead, demand decreases, and with that, the market price starts falling. The existing firms in the industry are now facing a lower price than before, and as it will be below the average cost curve, they will now be making economic losses. Some firms will continue producing where the new $P = MR = MC$, as long as they are able to cover their average variable costs. Some firms will have to shut down immediately as they will not be able to cover their average variable costs, and will then only incur their fixed costs, minimizing their losses. Exit of many firms causes the market supply curve to shift to the left. As the supply curve shifts to the left, the market price starts rising, and economic losses start to be lower. This process ends whenever the market price rises to the zero-profit level, where the existing firms are no longer losing money and are at zero profits again. Thus, while a perfectly competitive firm can earn profits in the short run, in the long run the process of entry will push down prices until they reach the zero-profit level. Conversely, while a perfectly competitive firm may earn losses in the short run, firms will not continually lose money. In the long run, firms making losses are able to escape from their fixed costs, and their exit from the market will push the price back up to the zero-profit level. In the long run, this process of entry and exit will drive the price in

perfectly competitive markets to the zero-profit point at the bottom of the AC curve, where marginal cost crosses average cost.

THE LONG-RUN ADJUSTMENT AND INDUSTRY TYPES

Whenever there are expansions in an industry, costs of production for the existing and new firms could either stay the same, increase, or even decrease. Therefore, we can categorize an industry as being (1) a constant cost industry (as demand increases, the cost of production for firms stays the same), (2) an increasing cost industry (as demand increases, the cost of production for firms increases), or (3) a decreasing cost industry (as demand increases the costs of production for the firms decreases).

For a **constant cost industry**, whenever there is an increase in market demand and price, then the supply curve shifts to the right with new firms' entry and stops at the point where the new long-run equilibrium intersects at the same market price as before. But why will costs remain the same? In this type of industry, the supply curve is very elastic. Firms can easily supply any quantity that consumers demand. In addition, there is a perfectly elastic supply of inputs—firms can easily increase their demand for employees, for example, with no increase to wages. Tying in to our Bring it Home discussion, an increased demand for ethanol in recent years has caused the demand for corn to increase. Consequently, many farmers switched from growing wheat to growing corn. Agricultural markets are generally good examples of constant cost industries.

For an **increasing cost industry**, as the market expands, the old and new firms experience increases in their costs of production, which makes the new zero-profit level intersect at a higher price than before. Here companies may have to deal with limited inputs, such as skilled labor. As the demand for these workers rise, wages rise and this increases the cost of production for all firms. The industry supply curve in this type of industry is more inelastic.

For a **decreasing cost industry**, as the market expands, the old and new firms experience lower costs of production, which makes the new zero-profit level intersect at a lower price than before. In this case, the industry and all the firms in it are experiencing falling average total costs. This can be due to an improvement in technology in the entire industry or an increase in the education of employees. High tech industries may be a good example of a decreasing cost market.

Figure 1 (a) presents the case of an adjustment process in a constant cost industry. Whenever there are output expansions in this type of industry, the long-run outcome implies more output produced at exactly the same original price. Note that supply was able to increase to meet the increased demand. When we join the before and after long-run equilibriums, the resulting line is the long run supply (LRS) curve in perfectly competitive markets. In this case, it is a flat curve. Figure 1 (b) and Figure 1 (c) present the cases for an increasing cost and decreasing cost industry, respectively. For an increasing cost industry, the LRS is upward sloping, while for a decreasing cost industry, the LRS is downward sloping.

KEY CONCEPTS AND SUMMARY

In the long run, firms will respond to profits through a process of entry, where existing firms expand output and new firms enter the market. Conversely, firms will react to losses in the long run through a process of exit, in which existing firms reduce output or cease production altogether. Through the process of entry in response to profits and exit in response to losses, the price level in a perfectly com-

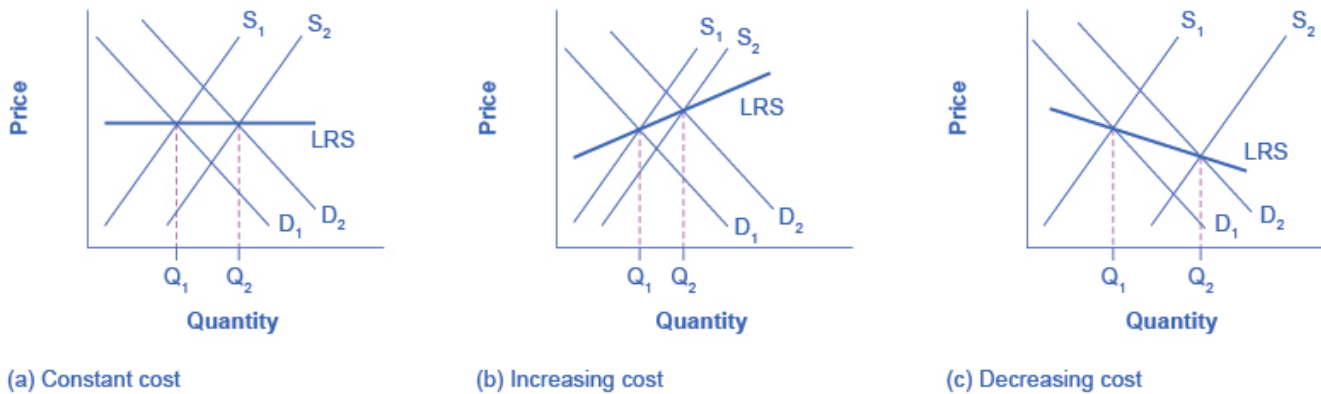


Figure 1. Adjustment Process in a Constant-Cost Industry. In (a), demand increased and supply met it. Notice that the supply increase is equal to the demand increase. The result is that the equilibrium price stays the same as quantity sold increases. In (b), notice that sellers were not able to increase supply as much as demand. Some inputs were scarce, or wages were rising. The equilibrium price rises. In (c), sellers easily increased supply in response to the demand increase. Here, new technology or economies of scale caused the large increase in supply, resulting in declining equilibrium price.

petitive market will move toward the zero-profit point, where the marginal cost curve crosses the AC curve, at the minimum of the average cost curve.

The long-run supply curve shows the long-run output supplied by firms in three different types of industries: constant cost, increasing cost, and decreasing cost.

SELF-CHECK QUESTIONS

1. If new technology in a perfectly competitive market brings about a substantial reduction in costs of production, how will this affect the market?
2. A market in perfect competition is in long-run equilibrium. What happens to the market if labor unions are able to increase wages for workers?

REVIEW QUESTIONS

1. Why does entry occur?
2. Why does exit occur?
3. Do entry and exit occur in the short run, the long run, both, or neither?
4. What price will a perfectly competitive firm end up charging in the long run? Why?

CRITICAL THINKING QUESTIONS

1. Many firms in the United States file for bankruptcy every year, yet they still continue operating. Why would they do this instead of completely shutting down?
2. Why will profits for firms in a perfectly competitive industry tend to vanish in the long run?
3. Why will losses for firms in a perfectly competitive industry tend to vanish in the long run?

GLOSSARY

entry the long-run process of firms entering an industry in response to industry profits

exit the long-run process of firms reducing production and shutting down in response to industry losses

long-run equilibrium where all firms earn zero economic profits producing the output level where $P = MR = MC$ and $P = AC$

SOLUTIONS

Answers to Self-Check Questions

1. With a technological improvement that brings about a reduction in costs of production, an adjustment process will take place in the market. The technological improvement will result in an increase in supply curves, by individual firms and at the market level. The existing firms will experience higher profits for a while, which will attract other firms into the market. This entry process will stop whenever the market supply increases enough (both by existing and new firms) so profits are driven back to zero.
2. When wages increase, costs of production increase. Some firms would now be making economic losses and would shut down. The supply curve then starts shifting to the left, pushing the market price up. This process ends when all firms remaining in the market earn zero economic profits. The result is a contraction in the output produced in the market.

11.4 EFFICIENCY IN PERFECTLY COMPETITIVE MARKETS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Apply concepts of productive efficiency and allocative efficiency to perfectly competitive markets
- Compare the model of perfect competition to real-world markets

When profit-maximizing firms in perfectly competitive markets combine with utility-maximizing consumers, something remarkable happens: the resulting quantities of outputs of goods and services demonstrate both productive and allocative efficiency (terms that were first introduced in *Choice in a World of Scarcity*).

Productive efficiency means producing without waste, so that the choice is on the production possibility frontier. In the long run in a perfectly competitive market, because of the process of entry and exit, the price in the market is equal to the minimum of the long-run average cost curve. In other words, goods are being produced and sold at the lowest possible average cost.

Allocative efficiency means that among the points on the production possibility frontier, the point that is chosen is socially preferred—at least in a particular and specific sense. In a perfectly competitive market, price will be equal to the marginal cost of production. Think about the price that is paid for a good as a measure of the social benefit received for that good; after all, willingness to pay conveys what the good is worth to a buyer. Then think about the marginal cost of producing the good as representing not just the cost for the firm, but more broadly as the social cost of producing that good. When perfectly competitive firms follow the rule that profits are maximized by producing at the quantity where price is equal to marginal cost, they are thus ensuring that the social benefits received from producing a good are in line with the social costs of production.

To explore what is meant by **allocative efficiency**, it is useful to walk through an example. Begin by assuming that the market for wholesale flowers is perfectly competitive, and so $P = MC$. Now, consider what it would mean if firms in that market produced a lesser quantity of flowers. At a lesser quantity, marginal costs will not yet have increased as much, so that price will exceed marginal cost; that is, $P > MC$. In that situation, the benefit to society as a whole of producing additional goods, as measured by the willingness of consumers to pay for marginal units of a good, would be higher than the cost of the inputs of labor and physical capital needed to produce the marginal good. In other words, the gains to society as a whole from producing additional marginal units will be greater than the costs.

Conversely, consider what it would mean if, compared to the level of output at the allocatively efficient choice when $P = MC$, firms produced a greater quantity of flowers. At a greater quantity, marginal costs of production will have increased so that $P < MC$. In that case, the marginal costs of producing additional flowers is greater than the benefit to society as measured by what people are willing to pay. For society as a whole, since the costs are outstripping the benefits, it will make sense to produce a lower quantity of such goods.

When perfectly competitive firms maximize their profits by producing the quantity where $P = MC$, they also assure that the benefits to consumers of what they are buying, as measured by the price they are willing to pay, is equal to the costs to society of producing the marginal units, as measured by the marginal costs the firm must pay—and thus that allocative efficiency holds.

The statements that a perfectly competitive market in the long run will feature both productive and allocative efficiency do need to be taken with a few grains of salt. Remember, economists are using the concept of “efficiency” in a particular and specific sense, not as a synonym for “desirable in every way.” For one thing, consumers’ ability to pay reflects the income distribution in a particular society. Thus, a homeless person may have no ability to pay for housing because they have insufficient income.

Perfect competition, in the long run, is a hypothetical benchmark. For market structures such as monopoly, monopolistic competition, and oligopoly, which are more frequently observed in the real world than perfect competition, firms will not always produce at the minimum of average cost, nor will they always set price equal to marginal cost. Thus, these other competitive situations will not produce productive and allocative efficiency.

Moreover, real-world markets include many issues that are assumed away in the model of perfect competition, including pollution, inventions of new technology, poverty which may make some people unable to pay for basic necessities of life, government programs like national defense or education, discrimination in labor markets, and buyers and sellers who must deal with imperfect and unclear information. These issues are explored in other chapters. However, the theoretical efficiency of perfect competition does provide a useful benchmark for comparing the issues that arise from these real-world problems.

A DIME A DOZEN

A quick glance at Table 14 reveals the dramatic increase in North Dakota corn production—more than double. Taking into consideration that corn typically yields two to three times as many bushels per acre as wheat, it is obvious there has been a significant increase in bushels of corn. Why the increase in corn acreage? Converging prices.

Year	Corn (millions of acres)	Wheat (millions of acres)
2014	91.6	56.82

Table 14. (Source: USDA National Agricultural Statistics Service)

Historically, wheat prices have been higher than corn prices, offsetting wheat’s lower yield per acre. However, in recent years wheat and corn prices have been converging. In April 2013, *Agweek* reported the gap was just 71 cents per bushel. As the difference in price narrowed, switching to the production of higher yield per acre of corn simply made good business sense. Erik Younggren, president of the National Association of Wheat Growers said in the *Agweek* article, “I don’t think we’re going to see mile after mile of waving amber fields [of wheat] anymore.” (Until wheat prices rise, we will probably be seeing field after field of tasseled corn.)

KEY CONCEPTS AND SUMMARY

Long-run equilibrium in perfectly competitive markets meets two important conditions: allocative efficiency and productive efficiency. These two conditions have important implications. First, resources are allocated to their best alternative use. Second, they provide the maximum satisfaction attainable by society.

SELF-CHECK QUESTIONS

1. Productive efficiency and allocative efficiency are two concepts achieved in the long run in a perfectly competitive market. These are the two reasons why we call them “perfect.” How would you use these two concepts to analyze other market structures and label them “imperfect?”
2. Explain how the profit-maximizing rule of setting $P = MC$ leads a perfectly competitive market to be allocatively efficient.

REVIEW QUESTIONS

1. Will a perfectly competitive market display productive efficiency? Why or why not?
2. Will a perfectly competitive market display allocative efficiency? Why or why not?

CRITICAL THINKING QUESTIONS

1. Assuming that the market for cigarettes is in perfect competition, what does allocative and productive efficiency imply in this case? What does it not imply?
2. In the argument for why perfect competition is allocatively efficient, the price that people are willing to pay represents the gains to society and the marginal cost to the firm represents the costs to society. Can you think of some social costs or issues that are not included in the marginal cost to the firm? Or some social gains that are not included in what people pay for a good?

REFERENCES

Index Mundi. n.d. “Wheat Monthly Price—U.S. Dollars per Metric Ton.” Accessed March 11, 2015. <http://www.indexmundi.com/commodities/?commodity=wheat>.

Knutson, J. “Wheat on the Defensive in the Northern Plains.” *Agweek, Associated Press State Wire: North Dakota (ND)*. April 14, 2013.

SBA Office of Advocacy. 2014. “Frequently Asked Questions: Advocacy: the voice of small business in government.” Accessed March 11, 2015. https://www.sba.gov/sites/default/files/advocacy/FAQ_March_2014_0.pdf.

SOLUTIONS

Answers to Self-Check Questions

1. Perfect competition is considered to be “perfect” because both allocative and productive efficiency are met at the same time in a long-run equilibrium. If a market structure results in long-run equilibrium that does not minimize average total costs and/or does not charge a price equal to marginal cost, then either allocative or productive (or both) efficiencies are not met, and therefore the market cannot be labeled “perfect.”
2. Think of the market price as representing the gain to society from a purchase, since it represents what someone is willing to pay. Think of the marginal cost as representing the cost to society from making the last unit of a good. If $P > MC$, then the benefits from producing more of a good exceed the costs, and society would gain from producing more of the good. If $P < MC$, then the social costs of producing the marginal good exceed the social benefits, and society should produce less of the good. Only if $P = MC$, the rule applied by a profit-maximizing perfectly competitive firm, will society’s costs and benefits be in balance. This choice will be the option that brings the greatest overall benefit to society.

CHAPTER 12. MONOPOLY

INTRODUCTION TO A MONOPOLY



Figure 1. Political Power from a Cotton Monopoly. In the mid-nineteenth century, the United States, specifically the Southern states, had a near monopoly in the cotton supplied to Great Britain. These states attempted to leverage this economic power into political power—trying to sway Great Britain to formally recognize the Confederate States of America. (Credit: modification of work by “ashleylovespizza”/Flickr Creative Commons)

THE REST IS HISTORY

Many of the opening case studies have focused on current events. This one steps into the past to observe how monopoly, or near monopolies, have helped shape history. In the spring of 1773, the East India Company, a firm that, in its time, was designated ‘too big to fail,’ was continuing to experience financial difficulties. To help shore up the failing firm, the British Parliament authorized the Tea Act. The act continued the tax on teas and made the East India Company the sole legal supplier of tea to the American colonies. By November, the citizens of Boston had had enough. They refused to permit the tea to be unloaded, citing their main complaint: “No taxation without representation.” Arriving tea-bearing ships were warned

via several newspapers, including *The Massachusetts Gazette*, “We are prepared, and shall not fail to pay them an unwelcome visit; by The Mohawks.”

Step forward in time to 1860—the eve of the American Civil War—to another near monopoly supplier of historical significance: the U.S. cotton industry. At that time, the Southern states provided the majority of the cotton Britain imported. The South, wanting to secede from the Union, hoped to leverage Britain’s high dependency on its cotton into formal diplomatic recognition of the Confederate States of America.

This leads us to the topic of this chapter: a firm that controls all (or nearly all) of the supply of a good or service—a monopoly. How do monopoly firms behave in the marketplace? Do they have “power?” Does this power potentially have unintended consequences? We’ll return to this case at the end of the chapter to see how the tea and cotton monopolies influenced U.S. history.

CHAPTER OBJECTIVES

Introduction to a Monopoly

In this chapter, you will learn about:

- How Monopolies form: Barriers to Entry
- How a Profit-Maximizing Monopoly Chooses Output and Price

There is a widespread belief that top executives at firms are the strongest supporters of market competition, but this belief is far from the truth. Think about it this way: If you very much wanted to win an Olympic gold medal, would you rather be far better than everyone else, or locked in competition with many athletes just as good as you are? Similarly, if you would like to attain a very high level of profits, would you rather manage a business with little or no competition, or struggle against many tough competitors who are trying to sell to your customers? By now, you might have read the chapter on Perfect Competition. In this chapter, we explore the opposite extreme: monopoly.

If perfect competition is a market where firms have no market power and they simply respond to the market price, monopoly is a market with no competition at all, and firms have complete market power. In the case of **monopoly**, one firm produces all of the output in a market. Since a monopoly faces no significant competition, it can charge any price it wishes. While a monopoly, by definition, refers to a single firm, in practice the term is often used to describe a market in which one firm merely has a very high market share. This tends to be the definition that the U.S. Department of Justice uses.

Even though there are very few true monopolies in existence, we do deal with some of those few every day, often without realizing it: The U.S. Postal Service, your electric and garbage collection companies are a few examples. Some new drugs are produced by only one pharmaceutical firm—and no close substitutes for that drug may exist.

From the mid-1990s until 2004, the U.S. Department of Justice prosecuted the Microsoft Corporation for including Internet Explorer as the default web browser with its operating system. The Justice Department’s argument was that, since Microsoft possessed an extremely high market share in the industry for operating systems, the inclusion of a free web browser constituted unfair competition to other browsers, such as Netscape Navigator. Since nearly everyone was using Windows, including Internet Explorer eliminated the incentive for consumers to explore other browsers and made it

impossible for competitors to gain a foothold in the market. In 2013, the Windows system ran on more than 90% of the most commonly sold personal computers. In 2015, a U.S. federal court tossed out antitrust charges that Google had an agreement with mobile device makers to set Google as the default search engine.

This chapter begins by describing how monopolies are protected from competition, including laws that prohibit competition, technological advantages, and certain configurations of demand and supply. It then discusses how a monopoly will choose its profit-maximizing quantity to produce and what price to charge. While a monopoly must be concerned about whether consumers will purchase its products or spend their money on something altogether different, the monopolist need not worry about the actions of other competing firms producing its products. As a result, a monopoly is not a price taker like a perfectly competitive firm, but instead exercises some power to choose its market price.

12.1 HOW MONOPOLIES FORM: BARRIERS TO ENTRY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Distinguish between a natural monopoly and a legal monopoly.
- Explain how economies of scale and the control of natural resources led to the necessary formation of legal monopolies
- Analyze the importance of trademarks and patents in promoting innovation
- Identify examples of predatory pricing

Because of the lack of competition, monopolies tend to earn significant economic profits. These profits should attract vigorous competition as described in Perfect Competition, and yet, because of one particular characteristic of monopoly, they do not. **Barriers to entry** are the legal, technological, or market forces that discourage or prevent potential competitors from entering a market. Barriers to entry can range from the simple and easily surmountable, such as the cost of renting retail space, to the extremely restrictive. For example, there are a finite number of radio frequencies available for broadcasting. Once the rights to all of them have been purchased, no new competitors can enter the market.

In some cases, barriers to entry may lead to monopoly. In other cases, they may limit competition to a few firms. Barriers may block entry even if the firm or firms currently in the market are earning profits. Thus, in markets with significant barriers to entry, it is *not* true that abnormally high profits will attract new firms, and that this entry of new firms will eventually cause the price to decline so that surviving firms earn only a normal level of profit in the long run.

There are two types of monopoly, based on the types of barriers to entry they exploit. One is **natural monopoly**, where the barriers to entry are something other than legal prohibition. The other is **legal monopoly**, where laws prohibit (or severely limit) competition.

NATURAL MONOPOLY

Economies of scale can combine with the size of the market to limit competition. (This theme was introduced in Cost and Industry Structure). Figure 1 presents a long-run average cost curve for the airplane manufacturing industry. It shows economies of scale up to an output of 8,000 planes per year and a price of P_0 , then constant returns to scale from 8,000 to 20,000 planes per year, and diseconomies of scale at a quantity of production greater than 20,000 planes per year.

Now consider the market demand curve in the diagram, which intersects the long-run average cost (LRAC) curve at an output level of 6,000 planes per year and at a price P_1 , which is higher than P_0 . In this situation, the market has room for only one producer. If a second firm attempts to enter the market at a smaller size, say by producing a quantity of 4,000 planes, then its average costs will be higher than the existing firm, and it will be unable to compete. If the second firm attempts to enter the market at a larger size, like 8,000 planes per year, then it could produce at a lower average cost—but it could not sell all 8,000 planes that it produced because of insufficient demand in the market.

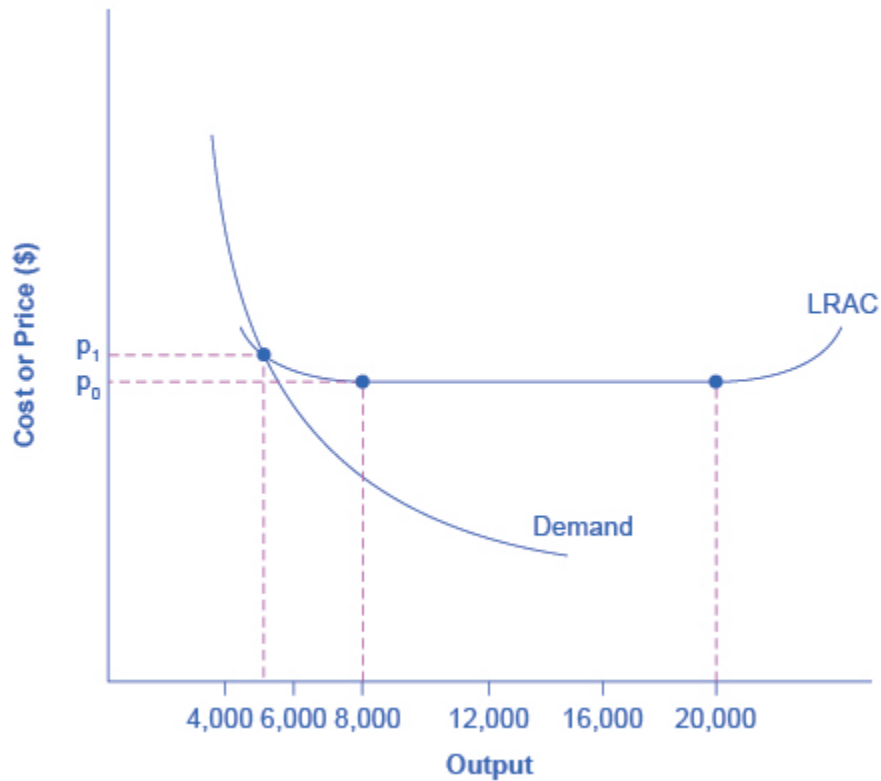


Figure 1. Economies of Scale and Natural Monopoly. In this market, the demand curve intersects the long-run average cost (LRAC) curve at its downward-sloping part. A natural monopoly occurs when the quantity demanded is less than the minimum quantity it takes to be at the bottom of the long-run average cost curve.

This situation, when economies of scale are large relative to the quantity demanded in the market, is called a natural monopoly. Natural monopolies often arise in industries where the marginal cost of adding an additional customer is very low, once the fixed costs of the overall system are in place. Once the main water pipes are laid through a neighborhood, the marginal cost of providing water service to another home is fairly low. Once electricity lines are installed through a neighborhood, the marginal cost of providing additional electrical service to one more home is very low. It would be costly and duplicative for a second water company to enter the market and invest in a whole second set of main water pipes, or for a second electricity company to enter the market and invest in a whole new set of electrical wires. These industries offer an example where, because of economies of scale, one producer can serve the entire market more efficiently than a number of smaller producers that would need to make duplicate physical capital investments.

A natural monopoly can also arise in smaller local markets for products that are difficult to transport.

For example, cement production exhibits economies of scale, and the quantity of cement demanded in a local area may not be much larger than what a single plant can produce. Moreover, the costs of transporting cement over land are high, and so a cement plant in an area without access to water transportation may be a natural monopoly.

CONTROL OF A PHYSICAL RESOURCE

Another type of natural monopoly occurs when a company has control of a scarce physical resource. In the U.S. economy, one historical example of this pattern occurred when ALCOA—the Aluminum Company of America—controlled most of the supply of bauxite, a key mineral used in making aluminum. Back in the 1930s, when ALCOA controlled most of the bauxite, other firms were simply unable to produce enough aluminum to compete.

As another example, the majority of global diamond production is controlled by DeBeers, a multinational company that has mining and production operations in South Africa, Botswana, Namibia, and Canada. It also has exploration activities on four continents, while directing a worldwide distribution network of rough cut diamonds. Though in recent years they have experienced growing competition, their impact on the rough diamond market is still considerable.

LEGAL MONOPOLY

For some products, the government erects barriers to entry by prohibiting or limiting competition. Under U.S. law, no organization but the U.S. Postal Service is legally allowed to deliver first-class mail. Many states or cities have laws or regulations that allow households a choice of only one electric company, one water company, and one company to pick up the garbage. Most legal monopolies are considered utilities—products necessary for everyday life—that are socially beneficial to have. As a consequence, the government allows producers to become regulated monopolies, to insure that an appropriate amount of these products is provided to consumers. Additionally, legal monopolies are often subject to economies of scale, so it makes sense to allow only one provider.

PROMOTING INNOVATION

Innovation takes time and resources to achieve. Suppose a company invests in research and development and finds the cure for the common cold. In this world of near ubiquitous information, other companies could take the formula, produce the drug, and because they did not incur the costs of research and development (R&D), undercut the price of the company that discovered the drug. Given this possibility, many firms would choose not to invest in research and development, and as a result, the world would have less innovation. To prevent this from happening, the Constitution of the United States specifies in Article I, Section 8: “The Congress shall have Power . . . To Promote the Progress of Science and Useful Arts, by securing for limited Times to Authors and Inventors the Exclusive Right to their Writings and Discoveries.” Congress used this power to create the U.S. Patent and Trademark Office, as well as the U.S. Copyright Office. A **patent** gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time; in the United States, exclusive patent rights last for 20 years. The idea is to provide limited monopoly power so that innovative firms can recoup their investment in R&D, but then to allow other firms to produce the product more cheaply once the patent expires.

A **trademark** is an identifying symbol or name for a particular good, like Chiquita bananas, Chevrolet

cars, or the Nike “swoosh” that appears on shoes and athletic gear. Roughly 1.9 million trademarks are registered with the U.S. government. A firm can renew a trademark over and over again, as long as it remains in active use.

A **copyright**, according to the U.S. Copyright Office, “is a form of protection provided by the laws of the United States for ‘original works of authorship’ including literary, dramatic, musical, architectural, cartographic, choreographic, pantomimic, pictorial, graphic, sculptural, and audiovisual creations.” No one can reproduce, display, or perform a copyrighted work without permission of the author. Copyright protection ordinarily lasts for the life of the author plus 70 years.

Roughly speaking, patent law covers inventions and copyright protects books, songs, and art. But in certain areas, like the invention of new software, it has been unclear whether patent or copyright protection should apply. There is also a body of law known as **trade secrets**. Even if a company does not have a patent on an invention, competing firms are not allowed to steal their secrets. One famous trade secret is the formula for Coca-Cola, which is not protected under copyright or patent law, but is simply kept secret by the company.

Taken together, this combination of patents, trademarks, copyrights, and trade secret law is called **intellectual property**, because it implies ownership over an idea, concept, or image, not a physical piece of property like a house or a car. Countries around the world have enacted laws to protect intellectual property, although the time periods and exact provisions of such laws vary across countries. There are ongoing negotiations, both through the World Intellectual Property Organization (WIPO) and through international treaties, to bring greater harmony to the intellectual property laws of different countries to determine the extent to which patents and copyrights in one country will be respected in other countries.

Government limitations on competition used to be even more common in the United States. For most of the twentieth century, only one phone company—AT&T—was legally allowed to provide local and long distance service. From the 1930s to the 1970s, one set of federal regulations limited which destinations airlines could choose to fly to and what fares they could charge; another set of regulations limited the interest rates that banks could pay to depositors; yet another specified what trucking firms could charge customers.

What products are considered utilities depends, in part, on the available technology. Fifty years ago, local and long distance telephone service was provided over wires. It did not make much sense to have multiple companies building multiple systems of wiring across towns and across the country. AT&T lost its monopoly on long distance service when the technology for providing phone service changed from wires to microwave and satellite transmission, so that multiple firms could use the same transmission mechanism. The same thing happened to local service, especially in recent years, with the growth in cellular phone systems.

The combination of improvements in production technologies and a general sense that the markets could provide services adequately led to a wave of **deregulation**, starting in the late 1970s and continuing into the 1990s. This wave eliminated or reduced government restrictions on the firms that could enter, the prices that could be charged, and the quantities that could be produced in many industries, including telecommunications, airlines, trucking, banking, and electricity.

Around the world, from Europe to Latin America to Africa and Asia, many governments continue to

control and limit competition in what those governments perceive to be key industries, including airlines, banks, steel companies, oil companies, and telephone companies.

Visit this website for examples of some pretty bizarre patents.



INTIMIDATING POTENTIAL COMPETITION

Businesses have developed a number of schemes for creating barriers to entry by deterring potential competitors from entering the market. One method is known as **predatory pricing**, in which a firm uses the threat of sharp price cuts to discourage competition. Predatory pricing is a violation of U.S. antitrust law, but it is difficult to prove.

Consider a large airline that provides most of the flights between two particular cities. A new, small start-up airline decides to offer service between these two cities. The large airline immediately slashes prices on this route to the bone, so that the new entrant cannot make any money. After the new entrant has gone out of business, the incumbent firm can raise prices again.

After this pattern is repeated once or twice, potential new entrants may decide that it is not wise to try to compete. Small airlines often accuse larger airlines of predatory pricing: in the early 2000s, for example, ValuJet accused Delta of predatory pricing, Frontier accused United, and Reno Air accused Northwest. In 2015, the Justice Department ruled against American Express and Mastercard for imposing restrictions on retailers who encouraged customers to use lower swipe fees on credit transactions.

In some cases, large advertising budgets can also act as a way of discouraging the competition. If the only way to launch a successful new national cola drink is to spend more than the promotional budgets of Coca-Cola and Pepsi Cola, not too many companies will try. A firmly established brand name can be difficult to dislodge.

SUMMING UP BARRIERS TO ENTRY

Table 1 lists the barriers to entry that have been discussed here. This list is not exhaustive, since firms have proved to be highly creative in inventing business practices that discourage competition. When barriers to entry exist, perfect competition is no longer a reasonable description of how an industry works. When barriers to entry are high enough, monopoly can result.

Barrier to Entry	Government Role?	Example
Natural monopoly	Government often responds with regulation (or ownership)	Water and electric companies
Control of a physical resource	No	DeBeers for diamonds
Legal monopoly	Yes	Post office, past regulation of airlines and trucking
Patent, trademark, and copyright	Yes, through protection of intellectual property	New drugs or software
Intimidating potential competitors	Somewhat	Predatory pricing; well-known brand names

Table 1. Barriers to Entry

KEY CONCEPTS AND SUMMARY

Barriers to entry prevent or discourage competitors from entering the market. These barriers include: economies of scale that lead to natural monopoly; control of a physical resource; legal restrictions on competition; patent, trademark and copyright protection; and practices to intimidate the competition like predatory pricing. Intellectual property refers to legally guaranteed ownership of an idea, rather than a physical item. The laws that protect intellectual property include patents, copyrights, trademarks, and trade secrets. A natural monopoly arises when economies of scale persist over a large enough range of output that if one firm supplies the entire market, no other firm can enter without facing a cost disadvantage.

SELF-CHECK QUESTIONS

- Classify the following as a government-enforced barrier to entry, a barrier to entry that is not government-enforced, or a situation that does not involve a barrier to entry.
 - A patented invention
 - A popular but easily copied restaurant recipe
 - An industry where economies of scale are very small compared to the size of demand in the market
 - A well-established reputation for slashing prices in response to new entry
 - A well-respected brand name that has been carefully built up over many years
- Classify the following as a government-enforced barrier to entry, a barrier to entry that is not government-enforced, or a situation that does not involve a barrier to entry.
 - A city passes a law on how many licenses it will issue for taxicabs
 - A city passes a law that all taxicab drivers must pass a driving safety test and have insurance
 - A well-known trademark
 - Owning a spring that offers very pure water
 - An industry where economies of scale are very large compared to the size of demand in the market
- Suppose the local electrical utility, a legal monopoly based on economies of scale, was split into four firms of

equal size, with the idea that eliminating the monopoly would promote competitive pricing of electricity.

What do you anticipate would happen to prices?

4. If Congress reduced the period of patent protection from 20 years to 10 years, what would likely happen to the amount of private research and development?

REVIEW QUESTIONS

1. How is monopoly different from perfect competition?
2. What is a barrier to entry? Give some examples.
3. What is a natural monopoly?
4. What is a legal monopoly?
5. What is predatory pricing?
6. How is intellectual property different from other property?
7. By what legal mechanisms is intellectual property protected?
8. In what sense is a natural monopoly “natural”?

CRITICAL THINKING QUESTIONS

1. ALCOA does not have the monopoly power it once had. How do you suppose their barriers to entry were weakened?
2. Why are generic pharmaceuticals significantly cheaper than name brand ones?
3. For many years, the Justice Department has tried to break up large firms like IBM, Microsoft, and most recently Google, on the grounds that their large market share made them essentially monopolies. In a global market, where U.S. firms compete with firms from other countries, would this policy make the same sense as it might in a purely domestic context?
4. Intellectual property laws are intended to promote innovation, but some economists, such as Milton Friedman, have argued that such laws are not desirable. In the United States, there is no intellectual property protection for food recipes or for fashion designs. Considering the state of these two industries, and bearing in mind the discussion of the inefficiency of monopolies, can you think of any reasons why intellectual property laws might hinder innovation in some cases?

PROBLEMS

Return to Figure 1. Suppose P_0 is \$10 and P_1 is \$11. Suppose a new firm with the same LRAC curve as the incumbent tries to break into the market by selling 4,000 units of output. Estimate from the graph what the new firm's average cost of producing output would be. If the incumbent continues to produce 6,000 units, how much output would be supplied to the market by the two firms? Estimate what would happen to the market price as a result of the supply of both the incumbent firm and the new entrant. Approximately how much profit would each firm earn?

GLOSSARY

barriers to entry the legal, technological, or market forces that may discourage or prevent potential competitors from entering a market

copyright a form of legal protection to prevent copying, for commercial purposes, original works of authorship, including books and music

deregulation removing government controls over setting prices and quantities in certain industries

intellectual property the body of law including patents, trademarks, copyrights, and trade secret law that protect the right of inventors to produce and sell their inventions

legal monopoly legal prohibitions against competition, such as regulated monopolies and intellectual property protection

monopoly a situation in which one firm produces all of the output in a market

natural monopoly economic conditions in the industry, for example, economies of scale or control of a critical resource, that limit effective competition

patent a government rule that gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time

predatory pricing when an existing firm uses sharp but temporary price cuts to discourage new competition

trade secrets methods of production kept secret by the producing firm

trademark an identifying symbol or name for a particular good and can only be used by the firm that registered that trademark

SOLUTIONS

Answers to Self-Check Questions

1.
 - a. A patent is a government-enforced barrier to entry.
 - b. This is not a barrier to entry.
 - c. This is not a barrier to entry.
 - d. This is a barrier to entry, but it is not government-enforced.
 - e. This is a barrier to entry, but it is not directly government enforced.
2.
 - a. This is a government-enforced barrier to entry.
 - b. This is an example of a government law, but perhaps it is not much of a barrier to entry if most people can pass the safety test and get insurance.
 - c. Trademarks are enforced by government, and therefore are a barrier to entry.
 - d. This is probably not a barrier to entry, since there are a number of different ways of getting pure water.
 - e. This is a barrier to entry, but it is not government-enforced.
3. Because of economies of scale, each firm would produce at a higher average cost than before. (They would each have to build their own power lines.) As a result, they would each have to raise prices to cover their higher costs. The policy would fail.
4. Shorter patent protection would make innovation less lucrative, so the amount of research and development would likely decline.

12.2 HOW A PROFIT-MAXIMIZING MONOPOLY CHOOSES OUTPUT AND PRICE

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the perceived demand curve for a perfect competitor and a monopoly
- Analyze a demand curve for a monopoly and determine the output that maximizes profit and revenue
- Calculate marginal revenue and marginal cost
- Explain allocative efficiency as it pertains to the efficiency of a monopoly

Consider a monopoly firm, comfortably surrounded by barriers to entry so that it need not fear competition from other producers. How will this monopoly choose its profit-maximizing quantity of output, and what price will it charge? Profits for the monopolist, like any firm, will be equal to total revenues minus total costs. The pattern of costs for the monopoly can be analyzed within the same framework as the costs of a **perfectly competitive firm**—that is, by using total cost, fixed cost, variable cost, marginal cost, average cost, and average variable cost. However, because a monopoly faces no competition, its situation and its decision process will differ from that of a perfectly competitive firm. (The Clear it Up feature discusses how hard it is sometimes to define “market” in a monopoly situation.)

DEMAND CURVES PERCEIVED BY A PERFECTLY COMPETITIVE FIRM AND BY A MONOPOLY

A perfectly competitive firm acts as a price taker, so its calculation of total revenue is made by taking the given market price and multiplying it by the quantity of output that the firm chooses. The demand curve *as it is perceived by a perfectly competitive firm* appears in Figure 1 (a). The flat perceived demand curve means that, from the viewpoint of the perfectly competitive firm, it could sell either a relatively low quantity like Q_l or a relatively high quantity like Q_h at the market price P .

WHAT DEFINES THE MARKET?

A monopoly is a firm that sells all or nearly all of the goods and services in a given market. But what defines the “market”? In a famous 1947 case, the federal government accused the DuPont company of having a monopoly in the cellophane market, pointing out that DuPont produced 75% of the cellophane in the United States. DuPont countered that even though it had a 75% market share in cellophane, it had less than a 20% share of the “flexible packaging materials,” which includes

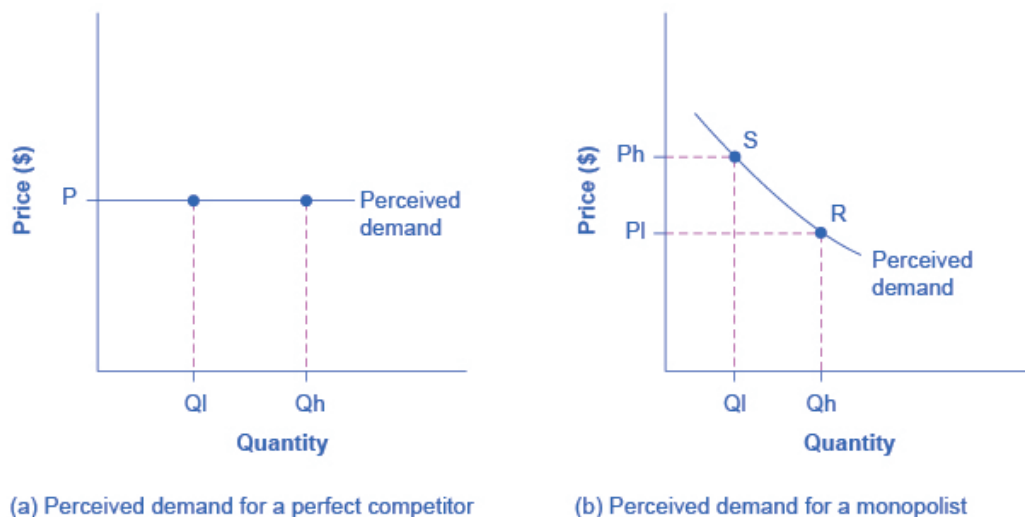


Figure 1. The Perceived Demand Curve for a Perfect Competitor and a Monopolist. (a) A perfectly competitive firm perceives the demand curve that it faces to be flat. The flat shape means that the firm can sell either a low quantity (Q_l) or a high quantity (Q_h) at exactly the same price (P). (b) A monopolist perceives the demand curve that it faces to be the same as the market demand curve, which for most goods is downward-sloping. Thus, if the monopolist chooses a high level of output (Q_h), it can charge only a relatively low price (P_l); conversely, if the monopolist chooses a low level of output (Q_l), it can then charge a higher price (P_h). The challenge for the monopolist is to choose the combination of price and quantity that maximizes profits.

all other moisture-proof papers, films, and foils. In 1956, after years of legal appeals, the U.S. Supreme Court held that the broader market definition was more appropriate, and the case against DuPont was dismissed.

Questions over how to define the market continue today. True, Microsoft in the 1990s had a dominant share of the software for computer operating systems, but in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 14% in 2014. The Greyhound bus company may have a near-monopoly on the market for intercity bus transportation, but it is only a small share of the market for intercity transportation if that market includes private cars, airplanes, and railroad service. DeBeers has a monopoly in diamonds, but it is a much smaller share of the total market for precious gemstones and an even smaller share of the total market for jewelry. A small town in the country may have only one gas station: is this gas station a “monopoly,” or does it compete with gas stations that might be five, 10, or 50 miles away?

In general, if a firm produces a product without close substitutes, then the firm can be considered a monopoly producer in a single market. But if buyers have a range of similar—even if not identical—options available from other firms, then the firm is not a monopoly. Still, arguments over whether substitutes are close or not close can be controversial.

While a monopolist can charge *any* price for its product, that price is nonetheless constrained by demand for the firm’s product. No monopolist, even one that is thoroughly protected by high barriers to entry, can require consumers to purchase its product. Because the monopolist is the only firm in the market, its demand curve is the same as the market demand curve, which is, unlike that for a perfectly competitive firm, downward-sloping.

Figure 1 illustrates this situation. The monopolist can either choose a point like R with a low price (P_l) and high quantity (Q_h), or a point like S with a high price (P_h) and a low quantity (Q_l), or some intermediate point. Setting the price too high will result in a low quantity sold, and will not bring in

much revenue. Conversely, setting the price too low may result in a high quantity sold, but because of the low price, it will not bring in much revenue either. The challenge for the monopolist is to strike a profit-maximizing balance between the price it charges and the quantity that it sells. But why isn't the perfectly competitive firm's demand curve also the market demand curve? See the following Clear it Up feature for the answer to this question.

WHAT IS THE DIFFERENCE BETWEEN PERCEIVED DEMAND AND MARKET DEMAND?

The demand curve as perceived by a perfectly competitive firm is not the overall market demand curve for that product. However, the firm's demand curve as perceived by a monopoly is the same as the market demand curve. The reason for the difference is that each perfectly competitive firm perceives the demand for its products in a market that includes many other firms; in effect, the demand curve perceived by a perfectly competitive firm is a tiny slice of the entire market demand curve. In contrast, a monopoly perceives demand for its product in a market where the monopoly is the only producer.

TOTAL COST AND TOTAL REVENUE FOR A MONOPOLIST

Profits for a monopolist can be illustrated with a graph of total revenues and total costs, as shown with the example of the hypothetical HealthPill firm in Figure 2. The total cost curve has its typical shape; that is, total costs rise and the curve grows steeper as output increases.

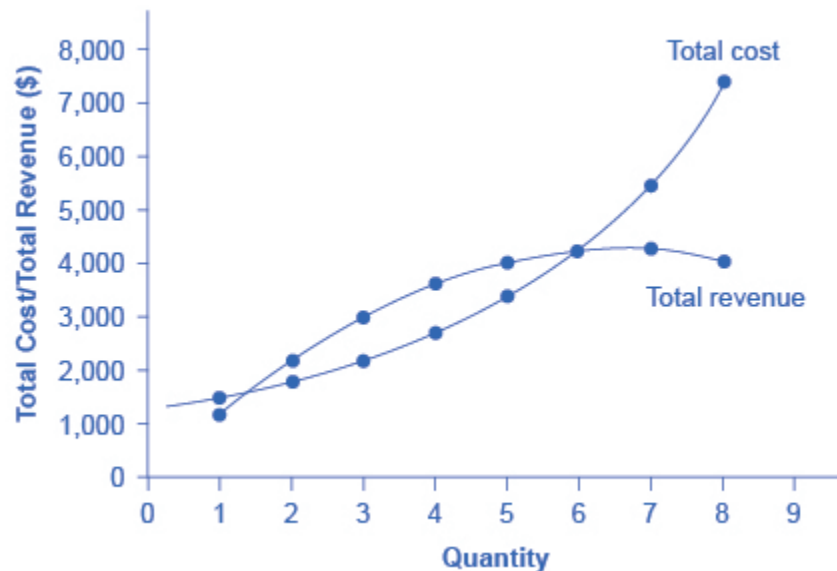


Figure 2. Total Revenue and Total Cost for the HealthPill Monopoly. Total revenue for the monopoly firm called HealthPill first rises, then falls. Low levels of output bring in relatively little total revenue, because the quantity is low. High levels of output bring in relatively less revenue, because the high quantity pushes down the market price. The total cost curve is upward-sloping. Profits will be highest at the quantity of output where total revenue is most above total cost. Of the choices in Table 2, the highest profits happen at an output of 4. The profit-maximizing level of output is not the same as the revenue-maximizing level of output, which should make sense, because profits take costs into account and revenues do not.

Quantity	Total Cost	Quantity	Price	Total Revenue	Profit = Total Revenue – Total Cost
1	1,500	1	1,200	1,200	–300
2	1,800	2	1,100	2,200	400
3	2,200	3	1,000	3,000	800
4	2,800	4	900	3,600	800
5	3,500	5	800	4,000	500
6	4,200	6	700	4,200	0
7	5,600	7	600	4,200	–1,400
8	7,400	8	500	4,000	–3,400

Table 2. Total Costs and Total Revenues of HealthPill

To calculate total revenue for a monopolist, start with the demand curve perceived by the monopolist. Table 2 shows quantities along the demand curve and the price at each quantity demanded, and then calculates total revenue by multiplying price times quantity at each level of output. (In this example, the output is given as 1, 2, 3, 4, and so on, for the sake of simplicity. If you prefer a dash of greater realism, you can imagine that these output levels and the corresponding prices are measured per 1,000 or 10,000 pills.) As the figure illustrates, total revenue for a monopolist rises, flattens out, and then falls. In this example, total revenue is highest at a quantity of 6 or 7.

Clearly, the total revenue for a monopolist is not a straight upward-sloping line, in the way that total revenue was for a perfectly competitive firm. The different total revenue pattern for a monopolist occurs because the quantity that a monopolist chooses to produce affects the market price, which was not true for a perfectly competitive firm. If the monopolist charges a very high price, then quantity demanded drops, and so total revenue is very low. If the monopolist charges a very low price, then, even if quantity demanded is very high, total revenue will not add up to much. At some intermediate level, total revenue will be highest.

However, the monopolist is not seeking to maximize revenue, but instead to earn the highest possible profit. Profits are calculated in the final row of the table. In the HealthPill example in Figure 2, the highest profit will occur at the quantity where total revenue is the farthest above total cost. Of the choices given in the table, the highest profits occur at an output of 4, where profit is 800.

MARGINAL REVENUE AND MARGINAL COST FOR A MONOPOLIST

In the real world, a monopolist often does not have enough information to analyze its entire total revenues or total costs curves; after all, the firm does not know exactly what would happen if it were to alter production dramatically. But a monopolist often has fairly reliable information about how changing output by small or moderate amounts will affect its marginal revenues and marginal costs, because it has had experience with such changes over time and because modest changes are easier to extrapolate from current experience. A monopolist can use information on **marginal revenue** and **marginal cost** to seek out the profit-maximizing combination of quantity and price.

The first four columns of Table 3 use the numbers on total cost from the HealthPill example in the previous exhibit and calculate marginal cost and average cost. This monopoly faces a typical upward-sloping marginal cost curve, as shown in Figure 3. The second four columns of Table 3 use the total revenue information from the previous exhibit and calculate marginal revenue.

Notice that marginal revenue is zero at a quantity of 7, and turns negative at quantities higher than 7. It may seem counterintuitive that marginal revenue could ever be zero or negative: after all, does an increase in quantity sold not always mean more revenue? For a perfect competitor, each additional unit sold brought a positive marginal revenue, because marginal revenue was equal to the given market price. But a monopolist can sell a larger quantity and see a decline in **total revenue**. When a monopolist increases sales by one unit, it gains some marginal revenue from selling that extra unit, but also loses some marginal revenue because every other unit must now be sold at a lower price. As the quantity sold becomes higher, the drop in price affects a greater quantity of sales, eventually causing a situation where more sales cause marginal revenue to be negative.

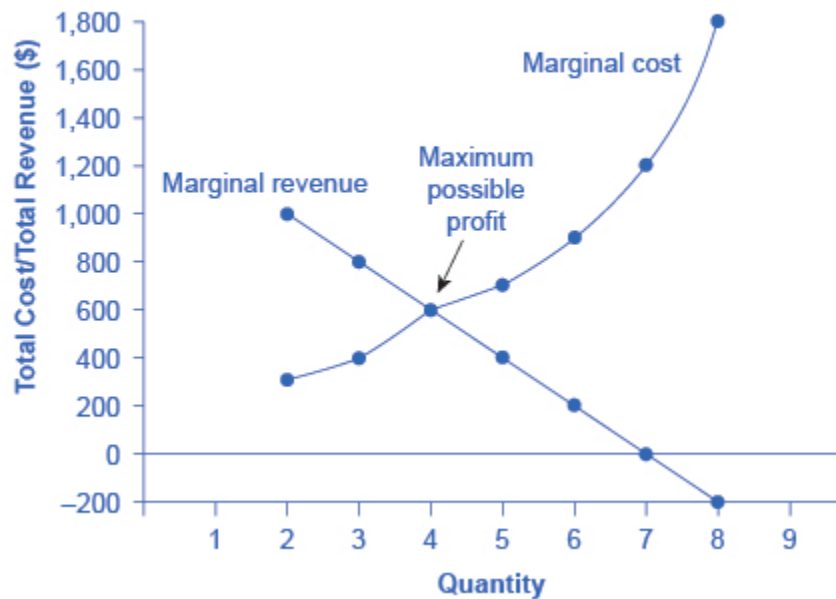


Figure 3. Marginal Revenue and Marginal Cost for the HealthPill Monopoly. For a monopoly like HealthPill, marginal revenue decreases as additional units are sold. The marginal cost curve is upward-sloping. The profit-maximizing choice for the monopoly will be to produce at the quantity where marginal revenue is equal to marginal cost: that is, $MR = MC$. If the monopoly produces a lower quantity, then $MR > MC$ at those levels of output, and the firm can make higher profits by expanding output. If the firm produces at a greater quantity, then $MC > MR$, and the firm can make higher profits by reducing its quantity of output.

Cost Information				Revenue Information			
Quantity	Total Cost	Marginal Cost	Average Cost	Quantity	Price	Total Revenue	Marginal Revenue
1	1,500	1,500	1,500	1	1,200	1,200	1,200
2	1,800	300	900	2	1,100	2,200	1,000
3	2,200	400	733	3	1,000	3,000	800
4	2,800	600	700	4	900	3,600	600
5	3,500	700	700	5	800	4,000	400
6	4,200	700	700	6	700	4,200	200
7	5,600	1,400	800	7	600	4,200	0
8	7,400	1,800	925	8	500	4,000	-200

Table 3. Costs and Revenues of HealthPill

A monopolist can determine its profit-maximizing price and quantity by analyzing the marginal revenue and marginal costs of producing an extra unit. If the marginal revenue exceeds the marginal cost, then the firm should produce the extra unit.

For example, at an output of 3 in Figure 3, marginal revenue is 800 and marginal cost is 400, so producing this unit will clearly add to overall profits. At an output of 4, marginal revenue is 600 and marginal cost is 600, so producing this unit still means overall profits are unchanged. However, expanding output from 4 to 5 would involve a marginal revenue of 400 and a marginal cost of 700, so that fifth unit would actually reduce profits. Thus, the monopoly can tell from the marginal revenue and marginal cost that of the choices given in the table, the profit-maximizing level of output is 4.

Indeed, the monopoly could seek out the profit-maximizing level of output by increasing quantity by a small amount, calculating marginal revenue and marginal cost, and then either increasing output as long as marginal revenue exceeds marginal cost or reducing output if marginal cost exceeds marginal revenue. This process works without any need to calculate total revenue and total cost. Thus, a profit-maximizing monopoly should follow the rule of producing up to the quantity where marginal revenue is equal to marginal cost—that is, $MR = MC$.

MAXIMIZING PROFITS

If you find it counterintuitive that producing where marginal revenue equals marginal cost will maximize profits, working through the numbers will help.

Step 1. Remember that marginal cost is defined as the change in total cost from producing a small amount of additional output.

$$MC = \frac{\text{change in total cost}}{\text{change in quantity produced}}$$

Step 2. Note that in Table 3, as output increases from 1 to 2 units, total cost increases from \$1500 to \$1800. As a result, the marginal cost of the second unit will be:

$$\begin{aligned} MC &= \frac{\$1800 - \$1500}{1} \\ &= \$300 \end{aligned}$$

Step 3. Remember that, similarly, marginal revenue is the change in total revenue from selling a small amount of additional output.

$$MR = \frac{\text{change in total revenue}}{\text{change in quantity sold}}$$

Step 4. Note that in Table 3, as output increases from 1 to 2 units, total revenue increases from \$1200 to \$2200. As a result, the marginal revenue of the second unit will be:

$$\begin{aligned} MR &= \frac{\$2200 - \$1200}{1} \\ &= \$1000 \end{aligned}$$

Quantity	Marginal Revenue	Marginal Cost	Marginal Profit	Total Profit
1	1,200	1,500	-300	-300
2	1,000	300	700	400
3	800	400	400	800
4	600	600	0	800
5	400	700	-300	500
6	200	700	-500	0
7	0	1,400	-1,400	-1,400

Table 4. Marginal Revenue, Marginal Cost, Marginal and Total Profit

Table 4 repeats the marginal cost and marginal revenue data from Table 3, and adds two more columns: **Marginal profit** is the profitability of each additional unit sold. It is defined as marginal revenue minus marginal cost. Finally, total profit is the sum of marginal profits. As long as marginal profit is positive, producing more output will increase total profits. When marginal profit turns negative, producing more output will decrease total profits. Total profit is maximized where marginal revenue equals marginal cost. In this example, maximum profit occurs at 4 units of output.

A perfectly competitive firm will also find its profit-maximizing level of output where $MR = MC$. The key difference with a perfectly competitive firm is that in the case of perfect competition, marginal revenue is equal to price ($MR = P$), while for a monopolist, marginal revenue is not equal to the price, because changes in quantity of output affect the price.

ILLUSTRATING MONOPOLY PROFITS

It is straightforward to calculate profits of given numbers for total revenue and total cost. However, the size of monopoly profits can also be illustrated graphically with Figure 4, which takes the marginal cost and marginal revenue curves from the previous exhibit and adds an average cost curve and the monopolist's perceived demand curve.

Figure 5 illustrates the three-step process where a monopolist: selects the profit-maximizing quantity to produce; decides what price to charge; determines total revenue, total cost, and profit.

Step 1: The Monopolist Determines Its Profit-Maximizing Level of Output

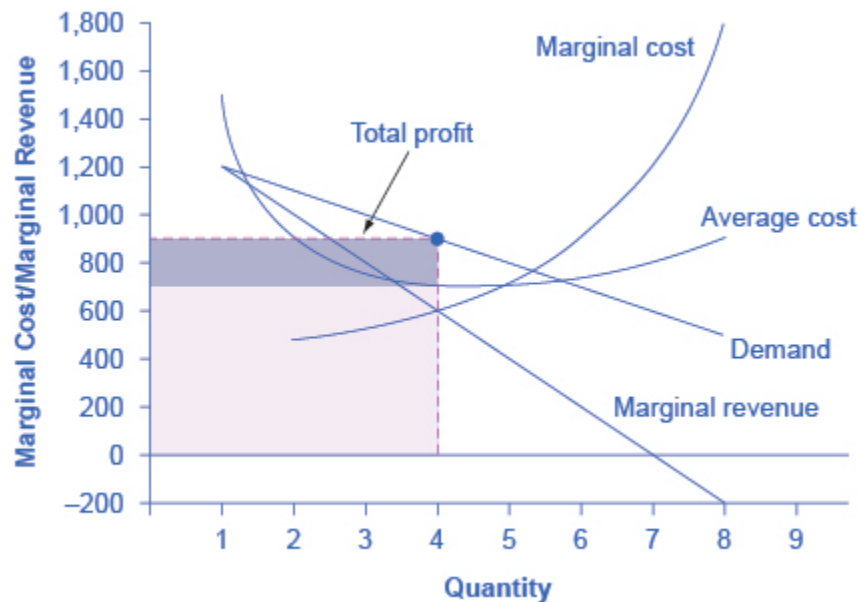


Figure 4. Illustrating Profits at the HealthPill Monopoly. This figure begins with the same marginal revenue and marginal cost curves from the HealthPill monopoly presented in Figure 3. It then adds an average cost curve and the demand curve faced by the monopolist. The HealthPill firm first chooses the quantity where $MR = MC$; in this example, the quantity is 4. The monopolist then decides what price to charge by looking at the demand curve it faces. The large box, with quantity on the horizontal axis and marginal revenue on the vertical axis, shows total revenue for the firm. Total costs for the firm are shown by the lighter-shaded box, which is quantity on the horizontal axis and marginal cost of production on the vertical axis. The large total revenue box minus the smaller total cost box leaves the darkly shaded box that shows total profits. Since the price charged is above average cost, the firm is earning positive profits.

The firm can use the points on the demand curve D to calculate total revenue, and then, based on total revenue, calculate its marginal revenue curve. The profit-maximizing quantity will occur where $MR = MC$ —or at the last possible point before marginal costs start exceeding marginal revenue. On Figure 4, $MR = MC$ occurs at an output of 4.

Step 2: The Monopolist Decides What Price to Charge

The monopolist will charge what the market is willing to pay. A dotted line drawn straight up from the profit-maximizing quantity to the demand curve shows the profit-maximizing price. This price is above the average cost curve, which shows that the firm is earning profits.

Step 3: Calculate Total Revenue, Total Cost, and Profit

Total revenue is the overall shaded box, where the width of the box is the quantity being sold and the height is the price. In Figure 4, the bottom part of the shaded box, which is shaded more lightly, shows total costs; that is, quantity on the horizontal axis multiplied by average cost on the vertical axis. The larger box of total revenues minus the smaller box of total costs will equal profits, which is shown by the darkly shaded box. In a perfectly competitive market, the forces of entry would erode this profit in the long run. But a monopolist is protected by barriers to entry. In fact, one telltale sign of a pos-

sible monopoly is when a firm earns profits year after year, while doing more or less the same thing, without ever seeing those profits eroded by increased competition.

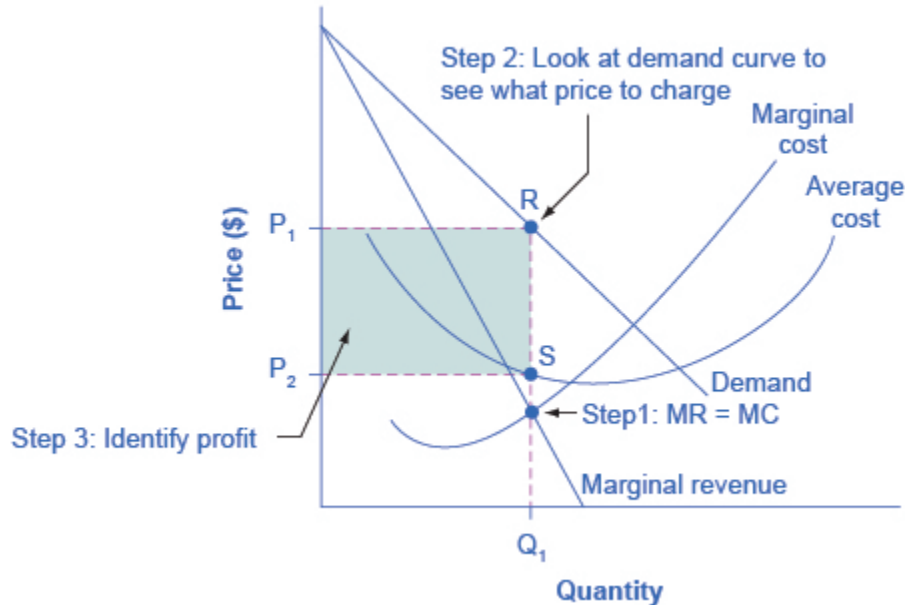


Figure 5. How a Profit-Maximizing Monopoly Decides Price In Step 1, the monopoly chooses the profit-maximizing level of output Q_1 , by choosing the quantity where $MR = MC$. In Step 2, the monopoly decides how much to charge for output level 1 by drawing a line straight up from Q_1 to point R on its perceived demand curve. Thus, the monopoly will charge a price (P_1). In Step 3, the monopoly identifies its profit. Total revenue will be Q_1 multiplied by P_1 . Total cost will be Q_1 multiplied by the average cost of producing Q_1 , which is shown by point S on the average cost curve to be P_2 . Profits will be the total revenue rectangle minus the total cost rectangle, shown by the shaded zone in the figure.

WHY IS A MONOPOLIST'S MARGINAL REVENUE ALWAYS LESS THAN THE PRICE?

The marginal revenue curve for a monopolist always lies beneath the market demand curve. To understand why, think about increasing the quantity along the demand curve by one unit, so that you take one step down the demand curve to a slightly higher quantity but a slightly lower price. A demand curve is not sequential: It is not that first we sell Q_1 at a higher price, and then we sell Q_2 at a lower price. Rather, a demand curve is conditional: If we charge the higher price, we would sell Q_1 . If, instead, we charge a lower price (on all the units that we sell), we would sell Q_2 .

So when we think about increasing the quantity sold by one unit, marginal revenue is affected in two ways. First, we sell one additional unit at the new market price. Second, all the previous units, which could have been sold at the higher price, now sell for less. Because of the lower price on all units sold, the marginal revenue of selling a unit is less than the price of that unit—and the marginal revenue curve is below the demand curve. *Tip:* For a straight-line demand curve, MR and demand have the same vertical intercept. As output increases, marginal revenue decreases twice as fast as demand, so that the horizontal intercept of MR is halfway to the horizontal intercept of demand. You can see this in the Figure 6.

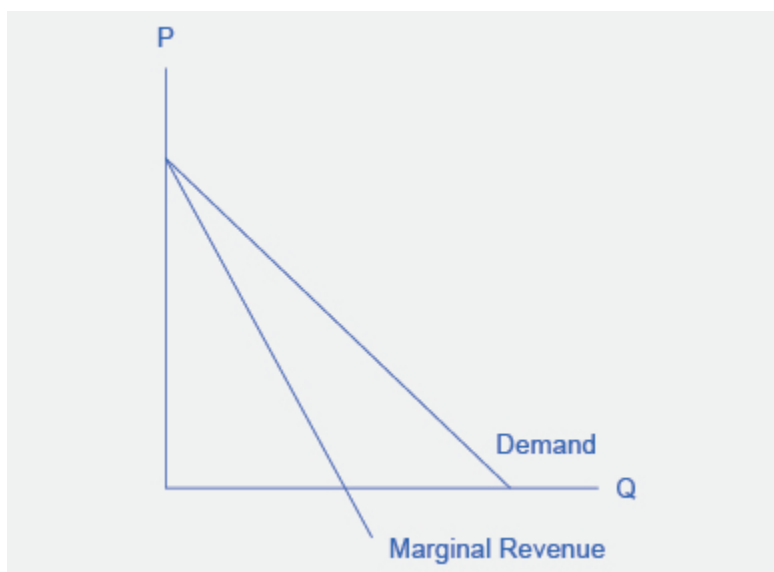


Figure 6. The Monopolist's Marginal Revenue Curve versus Demand Curve.

Because the market demand curve is conditional, the marginal revenue curve for a monopolist lies beneath the demand curve.

THE INEFFICIENCY OF MONOPOLY

Most people criticize monopolies because they charge too high a price, but what economists object to is that monopolies do not supply enough output to be allocatively efficient. To understand why a monopoly is inefficient, it is useful to compare it with the benchmark model of perfect competition.

Allocative efficiency is a social concept. It refers to producing the optimal quantity of some output, the quantity where the marginal benefit to society of one more unit just equals the marginal cost. The rule of profit maximization in a world of perfect competition was for each firm to produce the quantity of output where $P = MC$, where the price (P) is a measure of how much buyers value the good and the marginal cost (MC) is a measure of what marginal units cost society to produce. Following this rule assures allocative efficiency. If $P > MC$, then the marginal benefit to society (as measured by P) is greater than the marginal cost to society of producing additional units, and a greater quantity should be produced. But in the case of monopoly, price is always greater than marginal cost at the profit-maximizing level of output, as can be seen by looking back at Figure 4. Thus, consumers will suffer from a monopoly because a lower quantity will be sold in the market, at a higher price, than would have been the case in a perfectly competitive market.

The problem of inefficiency for monopolies often runs even deeper than these issues, and also involves incentives for efficiency over longer periods of time. There are counterbalancing incentives here. On one side, firms may strive for new inventions and new intellectual property because they want to become monopolies and earn high profits—at least for a few years until the competition catches up. In this way, monopolies may come to exist because of competitive pressures on firms. However, once a barrier to entry is in place, a monopoly that does not need to fear competition can just produce the same old products in the same old way—while still ringing up a healthy rate of profit. John Hicks, who won the Nobel Prize for economics in 1972, wrote in 1935: “The best of all monopoly profits is a quiet life.” He did not mean the comment in a complimentary way. He meant that monopolies may bank their profits and slack off on trying to please their customers.

When AT&T provided all of the local and long-distance phone service in the United States, along with manufacturing most of the phone equipment, the payment plans and types of phones did not change much. The old joke was that you could have any color phone you wanted, as long as it was black. But in 1982, AT&T was split up by government litigation into a number of local phone companies, a long-distance phone company, and a phone equipment manufacturer. An explosion of innovation followed. Services like call waiting, caller ID, three-way calling, voice mail through the phone company, mobile phones, and wireless connections to the Internet all became available. A wide range of payment plans was offered, as well. It was no longer true that all phones were black; instead, phones came in a wide variety of shapes and colors. The end of the telephone monopoly brought lower prices, a greater quantity of services, and also a wave of innovation aimed at attracting and pleasing customers.

THE REST IS HISTORY

In the opening case, the East India Company and the Confederate States were presented as a monopoly or near monopoly provider of a good. Nearly every American schoolchild knows the result of the ‘unwelcome visit’ the ‘Mohawks’ bestowed upon Boston Harbor’s tea-bearing ships—the Boston Tea Party. Regarding the cotton industry, we also know Great Britain remained neutral during the Civil War, taking neither side during the conflict.

Did the monopoly nature of these business have unintended and historical consequences? Might the American Revolution have been deterred, if the East India Company had sailed the tea-bearing ships back to England? Might the southern states have made different decisions had they not been so confident “King Cotton” would force diplomatic recognition of the Confederate States of America? Of course, it is not possible to definitively answer these questions; after all we cannot roll back the clock and try a different scenario. We can, however, consider the monopoly nature of these businesses and the roles they played and hypothesize about what might have occurred under different circumstances.

Perhaps if there had been legal free tea trade, the colonists would have seen things differently; there was smuggled Dutch tea in the colonial market. If the colonists had been able to freely purchase Dutch tea, they would have paid lower prices and avoided the tax.

What about the cotton monopoly? With one in five jobs in Great Britain depending on Southern cotton and the Confederate States nearly the sole provider of that cotton, why did Great Britain remain neutral during the Civil War? At the beginning of the war, Britain simply drew down massive stores of cotton. These stockpiles lasted until near the end of 1862. Why did Britain not recognize the Confederacy at that point? Two reasons: The Emancipation Proclamation and new sources of cotton. Having outlawed slavery throughout the United Kingdom in 1833, it was politically impossible for Great Britain, empty cotton warehouses or not, to recognize, diplomatically, the Confederate States. In addition, during the two years it took to draw down the stockpiles, Britain expanded cotton imports from India, Egypt, and Brazil.

Monopoly sellers often see no threats to their superior marketplace position. In these examples did the power of the monopoly blind the decision makers to other possibilities? Perhaps. But, as they say, the rest is history.

KEY CONCEPTS AND SUMMARY

A monopolist is not a price taker, because when it decides what quantity to produce, it also determines the market price. For a monopolist, total revenue is relatively low at low quantities of output, because not much is being sold. Total revenue is also relatively low at very high quantities of output, because a very high quantity will sell only at a low price. Thus, total revenue for a monopolist will start low, rise, and then decline. The marginal revenue for a monopolist from selling additional units will decline. Each additional unit sold by a monopolist will push down the overall market price, and as more units are sold, this lower price applies to more and more units.

The monopolist will select the profit-maximizing level of output where $MR = MC$, and then charge

the price for that quantity of output as determined by the market demand curve. If that price is above average cost, the monopolist earns positive profits.

Monopolists are not productively efficient, because they do not produce at the minimum of the average cost curve. Monopolists are not allocatively efficient, because they do not produce at the quantity where $P = MC$. As a result, monopolists produce less, at a higher average cost, and charge a higher price than would a combination of firms in a perfectly competitive industry. Monopolists also may lack incentives for innovation, because they need not fear entry.

SELF-CHECK QUESTIONS

1. Suppose demand for a monopoly's product falls so that its profit-maximizing price is below average variable cost. How much output should the firm supply? *Hint: Draw the graph.*
2. Imagine a monopolist could charge a different price to every customer based on how much he or she were willing to pay. How would this affect monopoly profits?

REVIEW QUESTIONS

1. How is the demand curve perceived by a perfectly competitive firm different from the demand curve perceived by a monopolist?
2. How does the demand curve perceived by a monopolist compare with the market demand curve?
3. Is a monopolist a price taker? Explain briefly.
4. What is the usual shape of a total revenue curve for a monopolist? Why?
5. What is the usual shape of a marginal revenue curve for a monopolist? Why?
6. How can a monopolist identify the profit-maximizing level of output if it knows its total revenue and total cost curves?
7. How can a monopolist identify the profit-maximizing level of output if it knows its marginal revenue and marginal costs?
8. When a monopolist identifies its profit-maximizing quantity of output, how does it decide what price to charge?
9. Is a monopolist allocatively efficient? Why or why not?
10. How does the quantity produced and price charged by a monopolist compare to that of a perfectly competitive firm?

CRITICAL THINKING QUESTIONS

1. Imagine that you are managing a small firm and thinking about entering the market of a monopolist. The monopolist is currently charging a high price, and you have calculated that you can make a nice profit charging 10% less than the monopolist. Before you go ahead and challenge the monopolist, what possibility should you consider for how the monopolist might react?

2. If a monopoly firm is earning profits, how much would you expect these profits to be diminished by entry in the long run?

PROBLEMS

1. Draw the demand curve, marginal revenue, and marginal cost curves from Figure 4, and identify the quantity of output the monopoly wishes to supply and the price it will charge. Suppose demand for the monopoly's product increases dramatically. Draw the new demand curve. What happens to the marginal revenue as a result of the increase in demand? What happens to the marginal cost curve? Identify the new profit-maximizing quantity and price. Does the answer make sense to you?
2. Draw a monopolist's demand curve, marginal revenue, and marginal cost curves. Identify the monopolist's profit-maximizing output level. Now, think about a slightly higher level of output (say $Q_0 + 1$). According to the graph, is there any consumer willing to pay more than the marginal cost of that new level of output? If so, what does this mean?

REFERENCES

Aboukhadijeh, Feross. "Chapter 20: Girding for War – The North and the South, 1861-1865." *StudyNotes, Inc.* Accessed July 7, 2013. <http://www.apstudynotes.org/us-history/outlines/chapter-20-girding-for-war-the-north-and-the-south-1861-1865/>.

British Parliament. "(28 August 1833). Slavery Abolition Act 1833; Section LXIV." Accessed July 2013. http://www.pdavis.nl/Legis_07.htm.

Dattel, E. (nd). "Cotton and the Civil War." *Mississippi Historical Society*. Accessed July 2013. <http://mshistorynow.mdah.state.ms.us/articles/291/cotton-and-the-civil-war>.

Gartner. 2015. "Gartner Says Tablet Sales Continue to Be Slow in 2015." Accessed March 12, 2015. <http://www.gartner.com/newsroom/id/2954317>.

Grogan, David. 2015. "Federal Judge Finds AmEx's Anti-Steering Rule Violates Antitrust Law." *American Booksellers Association*. Accessed March 12, 2015. <http://www.bookweb.org/news/federal-judge-finds-amex%E2%80%99s-anti-steering-rule-violates-antitrust-law>.

Massachusetts Historical Society. "The Coming of the American Revolution 1764-1776: The Boston Tea Party." Retrieved from <http://www.masshist.org/revolution/teaparty.php>.

Massachusetts Historical Society. "Whereas our Nation." *The Massachusetts Gazette*, p. 2. Accessed July 2013 http://www.masshist.org/revolution/image-viewer.php?old=1&item_id=457&img_step=1&nmask=1&mode=large.

Pelegrin, William. 2015. "Judge Overrules Antitrust Case Against Google , Says Setting Default Search Engines Is Fair." *Digital Trends*. Accessed March 12, 2015. <http://www.digitaltrends.com/mobile/judge-tosses-out-google-antitrust-lawsuit/>.

GLOSSARY

allocative efficiency producing the optimal quantity of some output; the quantity where the marginal benefit to society of one more unit just equals the marginal cost

marginal profit profit of one more unit of output, computed as marginal revenue minus marginal cost

SOLUTIONS

Answers to Self-Check Questions

1. If price falls below AVC, the firm will not be able to earn enough revenues even to cover its variable costs. In such a case, it will suffer a smaller loss if it shuts down and produces no output. By contrast, if it stayed in operation and produced the level of output where $MR = MC$, it would lose all of its fixed costs plus some variable costs. If it shuts down, it only loses its fixed costs.
2. This scenario is called “perfect price discrimination.” The result would be that the monopolist would produce more output, the same amount in fact as would be produced by a perfectly competitive industry. However, there would be no consumer surplus since each buyer is paying exactly what they think the product is worth. Therefore, the monopolist would be earning the maximum possible profits.

CHAPTER 13. MONOPOLISTIC COMPETITION AND OLIGOPOLY

INTRODUCTION TO MONOPOLISTIC COMPETITION AND OLIGOPOLY



Figure 1. Competing Brands? The laundry detergent market is one that is characterized neither as perfect competition nor monopoly. (Credit: modification of work by Pixel Drip/Flickr Creative Commons)

THE TEMPTATION TO DEFY THE LAW

Laundry detergent and bags of ice—products of industries that seem pretty mundane, maybe even boring. Hardly! Both have been the center of clandestine meetings and secret deals worthy of a spy novel. In France, between 1997 and 2004, the top four laundry detergent producers (Proctor & Gamble, Henkel, Unilever, and Colgate-Palmolive) controlled about 90 percent of the French soap market. Officials from the soap firms were meeting secretly, in out-of-the-way, small cafés around Paris. Their goals: Stamp out competition and set prices.

Around the same time, the top five Midwest ice makers (Home City Ice, Lang Ice, Tinley Ice, Sisler's Dairy, and Products of Ohio) had similar goals in mind when they secretly agreed to divide up the bagged ice market.

If both groups could meet their goals, it would enable each to act as though they were a single firm—in essence, a monop-

oly—and enjoy monopoly-size profits. The problem? In many parts of the world, including the European Union and the United States, it is illegal for firms to divide up markets and set prices collaboratively.

These two cases provide examples of markets that are characterized neither as perfect competition nor monopoly. Instead, these firms are competing in market structures that lie between the extremes of monopoly and perfect competition. How do they behave? Why do they exist? We will revisit this case later, to find out what happened.

CHAPTER OBJECTIVES

Introduction to Monopolistic Competition and Oligopoly

In this chapter, you will learn about:

- Monopolistic Competition
- Oligopoly

Perfect competition and monopoly are at opposite ends of the competition spectrum. A perfectly competitive market has many firms selling identical products, who all act as price takers in the face of the competition. If you recall, **price takers** are firms that have no market power. They simply have to take the market price as given.

Monopoly arises when a single firm sells a product for which there are no close substitutes. Microsoft, for instance, has been considered a monopoly because of its domination of the operating systems market.

What about the vast majority of real world firms and organizations that fall between these extremes, firms that could be described as **imperfectly competitive**? What determines their behavior? They have more influence over the price they charge than perfectly competitive firms, but not as much as a monopoly would. What will they do?

One type of imperfectly competitive market is called **monopolistic competition**. Monopolistically competitive markets feature a large number of competing firms, but the products that they sell are not identical. Consider, as an example, the Mall of America in Minnesota, the largest shopping mall in the United States. In 2010, the Mall of America had 24 stores that sold women's "ready-to-wear" clothing (like Ann Taylor and Urban Outfitters), another 50 stores that sold clothing for both men and women (like Banana Republic, J. Crew, and Nordstrom's), plus 14 more stores that sold women's specialty clothing (like Motherhood Maternity and Victoria's Secret). Most of the markets that consumers encounter at the retail level are monopolistically competitive.

The other type of imperfectly competitive market is **oligopoly**. Oligopolistic markets are those dominated by a small number of firms. Commercial aircraft provides a good example: Boeing and Airbus each produce slightly less than 50% of the large commercial aircraft in the world. Another example is the U.S. soft drink industry, which is dominated by Coca-Cola and Pepsi. Oligopolies are characterized by high barriers to entry with firms choosing output, pricing, and other decisions strategically based on the decisions of the other firms in the market. In this chapter, we first explore how monopolistically competitive firms will choose their profit-maximizing level of output. We will then discuss oligopolistic firms, which face two conflicting temptations: to collaborate as if they were a single

monopoly, or to individually compete to gain profits by expanding output levels and cutting prices. Oligopolistic markets and firms can also take on elements of monopoly and of perfect competition.

13.1 MONOPOLISTIC COMPETITION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the significance of differentiated products
- Describe how a monopolistic competitor chooses price and quantity
- Discuss entry, exit, and efficiency as they pertain to monopolistic competition
- Analyze how advertising can impact monopolistic competition

Monopolistic competition involves many **firms** competing against each other, but selling products that are distinctive in some way. Examples include stores that sell different styles of clothing; restaurants or grocery stores that sell different kinds of food; and even products like golf balls or beer that may be at least somewhat similar but differ in public perception because of advertising and brand names. There are over 600,000 restaurants in the United States. When products are distinctive, each firm has a mini-monopoly on its particular style or flavor or brand name. However, firms producing such products must also compete with other styles and flavors and brand names. The term “monopolistic competition” captures this mixture of mini-monopoly and tough competition, and the following Clear It Up feature introduces its derivation.

WHO INVENTED THE THEORY OF IMPERFECT COMPETITION?

The theory of imperfect competition was developed by two economists independently but simultaneously in 1933. The first was Edward Chamberlin of Harvard University who published *The Economics of Monopolistic Competition*. The second was Joan Robinson of Cambridge University who published *The Economics of Imperfect Competition*. Robinson subsequently became interested in macroeconomics where she became a prominent Keynesian, and later a post-Keynesian economist. (See the Welcome to Economics! and The Keynesian Perspective chapters for more on Keynes.)

DIFFERENTIATED PRODUCTS

A firm can try to make its products different from those of its competitors in several ways: physical aspects of the product, location from which the product is sold, intangible aspects of the product, and perceptions of the product. Products that are distinctive in one of these ways are called **differentiated products**.

Physical aspects of a product include all the phrases you hear in advertisements: unbreakable bottle,

nonstick surface, freezer-to-microwave, non-shrink, extra spicy, newly redesigned for your comfort. The location of a firm can also create a difference between producers. For example, a gas station located at a heavily traveled intersection can probably sell more gas, because more cars drive by that corner. A supplier to an automobile manufacturer may find that it is an advantage to locate close to the car factory.

Intangible aspects can differentiate a product, too. Some intangible aspects may be promises like a guarantee of satisfaction or money back, a reputation for high quality, services like free delivery, or offering a loan to purchase the product. Finally, **product differentiation** may occur in the minds of buyers. For example, many people could not tell the difference in taste between common varieties of beer or cigarettes if they were blindfolded but, because of past habits and advertising, they have strong preferences for certain brands. Advertising can play a role in shaping these intangible preferences.

The concept of differentiated products is closely related to the degree of variety that is available. If everyone in the economy wore only blue jeans, ate only white bread, and drank only tap water, then the markets for clothing, food, and drink would be much closer to perfectly competitive. The variety of styles, flavors, locations, and characteristics creates product differentiation and monopolistic competition.

PERCEIVED DEMAND FOR A MONOPOLISTIC COMPETITOR

A monopolistically competitive firm perceives a demand for its goods that is an intermediate case between monopoly and competition. Figure 1 offers a reminder that the **demand curve** as faced by a perfectly competitive firm is **perfectly elastic** or flat, because the perfectly competitive firm can sell any quantity it wishes at the prevailing **market price**. In contrast, the demand curve, as faced by a monopolist, is the market demand curve, since a monopolist is the only firm in the market, and hence is downward sloping.

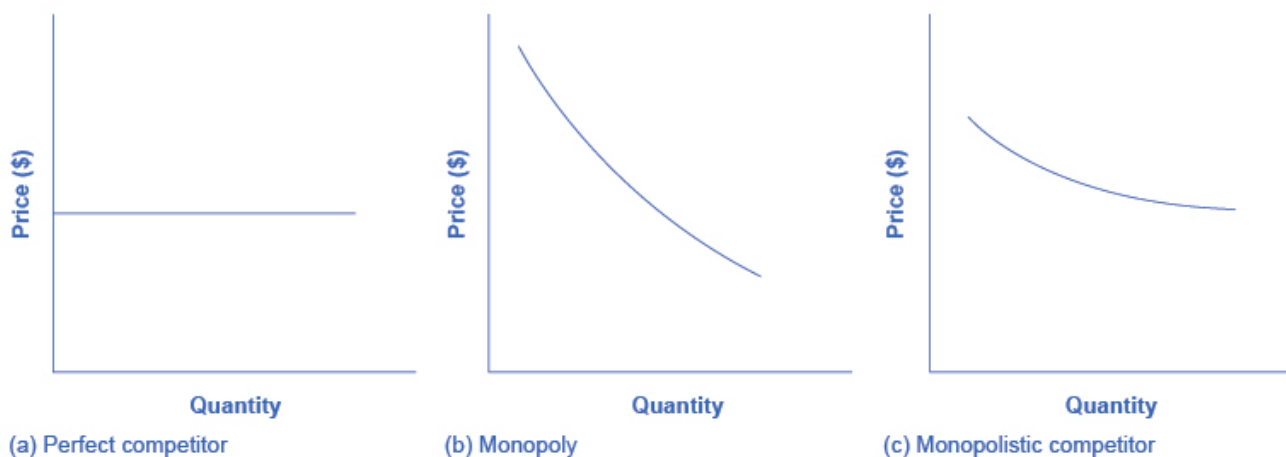


Figure 1. Perceived Demand for Firms in Different Competitive Settings. The demand curve faced by a perfectly competitive firm is perfectly elastic, meaning it can sell all the output it wishes at the prevailing market price. The demand curve faced by a monopoly is the market demand. It can sell more output only by decreasing the price it charges. The demand curve faced by a monopolistically competitive firm falls in between.

The demand curve as faced by a monopolistic competitor is not flat, but rather downward-sloping, which means that the monopolistic competitor can raise its price without losing all of its customers

or lower the price and gain more customers. Since there are substitutes, the demand curve facing a monopolistically competitive firm is more elastic than that of a monopoly where there are no close substitutes. If a monopolist raises its price, some consumers will choose not to purchase its product—but they will then need to buy a completely different product. However, when a monopolistic competitor raises its price, some consumers will choose not to purchase the product at all, but others will choose to buy a similar product from another firm. If a monopolistic competitor raises its price, it will not lose as many customers as would a perfectly competitive firm, but it will lose more customers than would a monopoly that raised its prices.

At a glance, the demand curves faced by a monopoly and by a monopolistic competitor look similar—that is, they both slope down. But the underlying economic meaning of these perceived demand curves is different, because a monopolist faces the market demand curve and a monopolistic competitor does not. Rather, a monopolistically competitive firm's demand curve is but one of many firms that make up the “before” market demand curve. Are you following? If so, how would you categorize the market for golf balls? Take a swing, then see the following Clear It Up feature.

ARE GOLF BALLS REALLY DIFFERENTIATED PRODUCTS?

Monopolistic competition refers to an industry that has more than a few firms, each offering a product which, from the consumer's perspective, is different from its competitors. The U.S. Golf Association runs a laboratory that tests 20,000 golf balls a year. There are strict rules for what makes a golf ball legal. The weight of a golf ball cannot exceed 1.620 ounces and its diameter cannot be less than 1.680 inches (which is a weight of 45.93 grams and a diameter of 42.67 millimeters, in case you were wondering). The balls are also tested by being hit at different speeds. For example, the distance test involves having a mechanical golfer hit the ball with a titanium driver and a swing speed of 120 miles per hour. As the testing center explains: “The USGA system then uses an array of sensors that accurately measure the flight of a golf ball during a short, indoor trajectory from a ball launcher. From this flight data, a computer calculates the lift and drag forces that are generated by the speed, spin, and dimple pattern of the ball. ... The distance limit is 317 yards.”

Over 1800 golf balls made by more than 100 companies meet the USGA standards. The balls do differ in various ways, like the pattern of dimples on the ball, the types of plastic used on the cover and in the cores, and so on. Since all balls need to conform to the USGA tests, they are much more alike than different. In other words, golf ball manufacturers are monopolistically competitive.

However, retail sales of golf balls are about \$500 million per year, which means that a lot of large companies have a powerful incentive to persuade players that golf balls are highly differentiated and that it makes a huge difference which one you choose. Sure, Tiger Woods can tell the difference. For the average duffer (golf-speak for a “mediocre player”) who plays a few times a summer—and who loses a lot of golf balls to the woods and lake and needs to buy new ones—most golf balls are pretty much indistinguishable.

HOW A MONOPOLISTIC COMPETITOR CHOOSES PRICE AND QUANTITY

The monopolistically competitive firm decides on its profit-maximizing quantity and price in much the same way as a monopolist. A monopolistic competitor, like a monopolist, faces a downward-sloping demand curve, and so it will choose some combination of price and quantity along its perceived demand curve.

As an example of a **profit-maximizing** monopolistic competitor, consider the Authentic Chinese Pizza store, which serves pizza with cheese, sweet and sour sauce, and your choice of vegetables and meats. Although Authentic Chinese Pizza must compete against other pizza businesses and restau-

rants, it has a differentiated product. The firm's perceived demand curve is downward sloping, as shown in Figure 2 and the first two columns of Table 1.

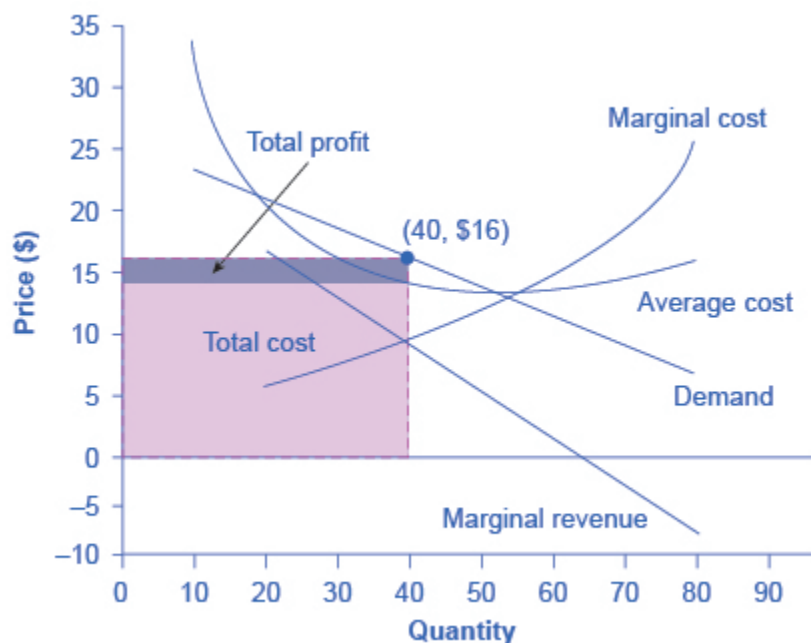


Figure 2. How a Monopolistic Competitor Chooses its Profit Maximizing Output and Price. To maximize profits, the Authentic Chinese Pizza shop would choose a quantity where marginal revenue equals marginal cost, or Q where $MR = MC$. Here it would choose a quantity of 40 and a price of \$16.

Quantity	Price	Total Revenue	Marginal Revenue	Total Cost	Marginal Cost	Average Cost
10	\$23	\$230	–	\$340	–	\$34
20	\$20	\$400	\$17	\$400	\$6	\$20
30	\$18	\$540	\$14	\$480	\$8	\$16
40	\$16	\$640	\$10	\$580	\$10	\$14.50
50	\$14	\$700	\$6	\$700	\$12	\$14
60	\$12	\$720	\$2	\$840	\$14	\$14
70	\$10	\$700	–\$2	\$1,020	\$18	\$14.57
80	\$8	\$640	–\$6	\$1,280	\$26	\$16

Table 1. Revenue and Cost Schedule

The combinations of price and quantity at each point on the demand curve can be multiplied to calculate the total revenue that the firm would receive, which is shown in the third column of Table 1. The fourth column, marginal revenue, is calculated as the change in total revenue divided by the change in quantity. The final columns of Table 1 show total cost, marginal cost, and average cost. As always, marginal cost is calculated by dividing the change in total cost by the change in quantity, while average cost is calculated by dividing total cost by quantity. The following Work It Out feature shows how these firms calculate how much of its product to supply at what price.

HOW A MONOPOLISTIC COMPETITOR DETERMINES HOW MUCH TO PRODUCE AND AT WHAT PRICE

The process by which a monopolistic competitor chooses its profit-maximizing quantity and price resembles closely how a monopoly makes these decisions process. First, the firm selects the profit-maximizing quantity to produce. Then the firm decides what price to charge for that quantity.

Step 1. The monopolistic competitor determines its profit-maximizing level of output. In this case, the Authentic Chinese Pizza company will determine the profit-maximizing quantity to produce by considering its marginal revenues and marginal costs. Two scenarios are possible:

- If the firm is producing at a quantity of output where marginal revenue exceeds marginal cost, then the firm should keep expanding production, because each marginal unit is adding to profit by bringing in more revenue than its cost. In this way, the firm will produce up to the quantity where $MR = MC$.
- If the firm is producing at a quantity where marginal costs exceed marginal revenue, then each marginal unit is costing more than the revenue it brings in, and the firm will increase its profits by reducing the quantity of output until $MR = MC$.

In this example, MR and MC intersect at a quantity of 40, which is the profit-maximizing level of output for the firm.

Step 2. The monopolistic competitor decides what price to charge. When the firm has determined its profit-maximizing quantity of output, it can then look to its perceived demand curve to find out what it can charge for that quantity of output. On the graph, this process can be shown as a vertical line reaching up through the profit-maximizing quantity until it hits the firm's perceived demand curve. For Authentic Chinese Pizza, it should charge a price of \$16 per pizza for a quantity of 40.

Once the firm has chosen price and quantity, it's in a position to calculate total revenue, total cost, and profit. At a quantity of 40, the price of \$16 lies above the average cost curve, so the firm is making economic profits. From Table 1 we can see that, at an output of 40, the firm's total revenue is \$640 and its total cost is \$580, so profits are \$60. In Figure 2, the firm's total revenues are the rectangle with the quantity of 40 on the horizontal axis and the price of \$16 on the vertical axis. The firm's total costs are the light shaded rectangle with the same quantity of 40 on the horizontal axis but the average cost of \$14.50 on the vertical axis. Profits are total revenues minus total costs, which is the shaded area above the average cost curve.

Although the process by which a monopolistic competitor makes decisions about quantity and price is similar to the way in which a monopolist makes such decisions, two differences are worth remembering. First, although both a monopolist and a monopolistic competitor face downward-sloping demand curves, the monopolist's perceived demand curve is the market demand curve, while the perceived **demand curve** for a monopolistic competitor is based on the extent of its product differentiation and how many competitors it faces. Second, a monopolist is surrounded by barriers to entry and need not fear entry, but a monopolistic competitor who earns profits must expect the entry of firms with similar, but differentiated, products.

MONOPOLISTIC COMPETITORS AND ENTRY

If one monopolistic competitor earns positive economic profits, other firms will be tempted to enter the market. A gas station with a great location must worry that other gas stations might open across the street or down the road—and perhaps the new gas stations will sell coffee or have a carwash or some other attraction to lure customers. A successful restaurant with a unique barbecue sauce must be concerned that other restaurants will try to copy the sauce or offer their own unique recipes. A

laundry detergent with a great reputation for quality must be concerned that other competitors may seek to build their own reputations.

The entry of other firms into the same general market (like gas, restaurants, or detergent) shifts the demand curve faced by a monopolistically competitive firm. As more firms enter the market, the quantity demanded at a given price for any particular firm will decline, and the firm's perceived demand curve will shift to the left. As a firm's perceived demand curve shifts to the left, its marginal revenue curve will shift to the left, too. The shift in marginal revenue will change the profit-maximizing quantity that the firm chooses to produce, since marginal revenue will then equal marginal cost at a lower quantity.

Figure 3 (a) shows a situation in which a monopolistic competitor was earning a profit with its original perceived demand curve (D_0). The intersection of the marginal revenue curve (MR_0) and marginal cost curve (MC) occurs at point S, corresponding to quantity Q_0 , which is associated on the demand curve at point T with price P_0 . The combination of price P_0 and quantity Q_0 lies above the average cost curve, which shows that the firm is earning positive economic profits.

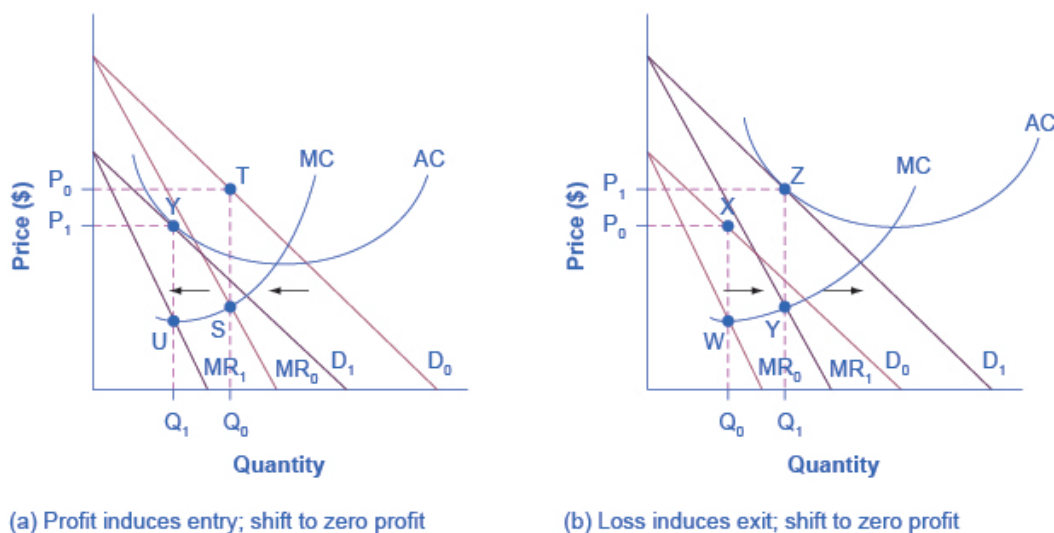


Figure 3. Monopolistic Competition, Entry, and Exit. (a) At P_0 and Q_0 , the monopolistically competitive firm shown in this figure is making a positive economic profit. This is clear because if you follow the dotted line above Q_0 , you can see that price is above average cost. Positive economic profits attract competing firms to the industry, driving the original firm's demand down to D_1 . At the new equilibrium quantity (P_1 , Q_1), the original firm is earning zero economic profits, and entry into the industry ceases. In (b) the opposite occurs. At P_0 and Q_0 , the firm is losing money. If you follow the dotted line above Q_0 , you can see that average cost is above price. Losses induce firms to leave the industry. When they do, demand for the original firm rises to D_1 , where once again the firm is earning zero economic profit.

Unlike a monopoly, with its high barriers to entry, a monopolistically competitive firm with positive economic profits will attract competition. When another competitor enters the market, the original firm's perceived demand curve shifts to the left, from D_0 to D_1 , and the associated marginal revenue curve shifts from MR_0 to MR_1 . The new profit-maximizing output is Q_1 , because the intersection of

the MR_1 and MC now occurs at point U. Moving vertically up from that quantity on the new demand curve, the optimal price is at P_1 .

As long as the **firm** is earning positive economic profits, new competitors will continue to enter the market, reducing the original firm's demand and marginal revenue curves. The long-run **equilibrium** is shown in the figure at point Y, where the firm's perceived demand curve touches the average cost curve. When price is equal to average cost, economic profits are zero. Thus, although a monopolistically competitive firm may earn positive economic profits in the short term, the process of new entry will drive down economic profits to zero in the long run. Remember that zero economic profit is not equivalent to zero **accounting profit**. A zero economic profit means the firm's accounting profit is equal to what its resources could earn in their next best use. Figure 3 (b) shows the reverse situation, where a monopolistically competitive firm is originally losing money. The adjustment to long-run equilibrium is analogous to the previous example. The economic losses lead to firms exiting, which will result in increased demand for this particular firm, and consequently lower losses. Firms exit up to the point where there are no more losses in this market, for example when the demand curve touches the average cost curve, as in point Z.

Monopolistic competitors can make an economic profit or loss in the short run, but in the long run, entry and exit will drive these firms toward a zero economic profit outcome. However, the zero economic profit outcome in monopolistic competition looks different from the zero economic profit outcome in perfect competition in several ways relating both to efficiency and to variety in the market.

MONOPOLISTIC COMPETITION AND EFFICIENCY

The long-term result of entry and exit in a perfectly competitive market is that all firms end up selling at the price level determined by the lowest point on the average cost curve. This outcome is why perfect competition displays **productive efficiency**: goods are being produced at the lowest possible average cost. However, in monopolistic competition, the end result of entry and exit is that firms end up with a price that lies on the downward-sloping portion of the average cost curve, not at the very bottom of the AC curve. Thus, monopolistic competition will not be productively efficient.

In a perfectly competitive market, each firm produces at a quantity where price is set equal to marginal cost, both in the short run and in the long run. This outcome is why perfect competition displays allocative efficiency: the social benefits of additional production, as measured by the marginal benefit, which is the same as the price, equal the marginal costs to society of that production. In a monopolistically competitive market, the rule for maximizing profit is to set $MR = MC$ —and price is higher than marginal revenue, not equal to it because the demand curve is downward sloping. When $P > MC$, which is the outcome in a monopolistically competitive market, the benefits to society of providing additional quantity, as measured by the price that people are willing to pay, exceed the marginal costs to society of producing those units. A monopolistically competitive firm does not produce more, which means that society loses the net benefit of those extra units. This is the same argument we made about monopoly, but in this case to a lesser degree. Thus, a monopolistically competitive industry will produce a lower quantity of a good and charge a higher price for it than would a perfectly competitive industry. See the following Clear It Up feature for more detail on the impact of demand shifts.

WHY DOES A SHIFT IN PERCEIVED DEMAND CAUSE A SHIFT IN MARGINAL REVENUE?

The combinations of price and quantity at each point on a firm's perceived demand curve are used to calculate total revenue for each combination of price and quantity. This information on total revenue is then used to calculate marginal revenue, which is the change in total revenue divided by the change in quantity. A change in perceived demand will change total revenue at every quantity of output and in turn, the change in total revenue will shift marginal revenue at each quantity of output. Thus, when entry occurs in a monopolistically competitive industry, the perceived demand curve for each firm will shift to the left, because a smaller quantity will be demanded at any given price. Another way of interpreting this shift in demand is to notice that, for each quantity sold, a lower price will be charged. Consequently, the marginal revenue will be lower for each quantity sold—and the marginal revenue curve will shift to the left as well. Conversely, exit causes the perceived demand curve for a monopolistically competitive firm to shift to the right and the corresponding marginal revenue curve to shift right, too.

A monopolistically competitive industry does not display productive and allocative efficiency in either the short run, when firms are making economic profits and losses, nor in the long run, when firms are earning zero profits.

THE BENEFITS OF VARIETY AND PRODUCT DIFFERENTIATION

Even though monopolistic competition does not provide productive efficiency or allocative efficiency, it does have benefits of its own. Product differentiation is based on variety and innovation. Many people would prefer to live in an economy with many kinds of clothes, foods, and car styles; not in a world of perfect competition where everyone will always wear blue jeans and white shirts, eat only spaghetti with plain red sauce, and drive an identical model of car. Many people would prefer to live in an economy where firms are struggling to figure out ways of attracting customers by methods like friendlier service, free delivery, guarantees of quality, variations on existing products, and a better shopping experience.

Economists have struggled, with only partial success, to address the question of whether a market-oriented economy produces the optimal amount of variety. Critics of market-oriented economies argue that society does not really need dozens of different athletic shoes or breakfast cereals or automobiles. They argue that much of the cost of creating such a high degree of product differentiation, and then of advertising and marketing this differentiation, is socially wasteful—that is, most people would be just as happy with a smaller range of **differentiated products** produced and sold at a lower price. Defenders of a market-oriented economy respond that if people do not want to buy differentiated products or highly advertised brand names, no one is forcing them to do so. Moreover, they argue that consumers benefit substantially when firms seek short-term profits by providing differentiated products. This controversy may never be fully resolved, in part because deciding on the optimal amount of variety is very difficult, and in part because the two sides often place different values on what variety means for consumers. Read the following Clear It Up feature for a discussion on the role that advertising plays in monopolistic competition.

HOW DOES ADVERTISING IMPACT MONOPOLISTIC COMPETITION?

The U.S. economy spent about \$180.12 billion on advertising in 2014, according to eMarketer.com. Roughly one third of this was television advertising, and another third was divided roughly equally between Internet, newspapers, and radio.

The remaining third was divided up between direct mail, magazines, telephone directory yellow pages, and billboards. Mobile devices are increasing the opportunities for advertisers.

Advertising is all about explaining to people, or making people believe, that the products of one firm are differentiated from the products of another firm. In the framework of monopolistic competition, there are two ways to conceive of how advertising works: either advertising causes a firm's perceived demand curve to become more inelastic (that is, it causes the perceived demand curve to become steeper); or advertising causes demand for the firm's product to increase (that is, it causes the firm's perceived demand curve to shift to the right). In either case, a successful advertising campaign may allow a firm to sell either a greater quantity or to charge a higher price, or both, and thus increase its profits.

However, economists and business owners have also long suspected that much of the advertising may only offset other advertising. Economist A. C. Pigou wrote the following back in 1920 in his book, *The Economics of Welfare*:

It may happen that expenditures on advertisement made by competing monopolists [that is, what we now call monopolistic competitors] will simply neutralise one another, and leave the industrial position exactly as it would have been if neither had expended anything. For, clearly, if each of two rivals makes equal efforts to attract the favour of the public away from the other, the total result is the same as it would have been if neither had made any effort at all.

KEY CONCEPTS AND SUMMARY

Monopolistic competition refers to a market where many firms sell differentiated products. Differentiated products can arise from characteristics of the good or service, location from which the product is sold, intangible aspects of the product, and perceptions of the product.

The perceived demand curve for a monopolistically competitive firm is downward-sloping, which shows that it is a price maker and chooses a combination of price and quantity. However, the perceived demand curve for a monopolistic competitor is more elastic than the perceived demand curve for a monopolist, because the monopolistic competitor has direct competition, unlike the pure monopolist. A profit-maximizing monopolistic competitor will seek out the quantity where marginal revenue is equal to marginal cost. The monopolistic competitor will produce that level of output and charge the price that is indicated by the firm's demand curve.

If the firms in a monopolistically competitive industry are earning economic profits, the industry will attract entry until profits are driven down to zero in the long run. If the firms in a monopolistically competitive industry are suffering economic losses, then the industry will experience exit of firms until economic profits are driven up to zero in the long run.

A monopolistically competitive firm is not productively efficient because it does not produce at the minimum of its average cost curve. A monopolistically competitive firm is not allocatively efficient because it does not produce where $P = MC$, but instead produces where $P > MC$. Thus, a monopolistically competitive firm will tend to produce a lower quantity at a higher cost and to charge a higher price than a perfectly competitive firm.

Monopolistically competitive industries do offer benefits to consumers in the form of greater variety and incentives for improved products and services. There is some controversy over whether a market-oriented economy generates too much variety.

SELF-CHECK QUESTIONS

1. Suppose that, due to a successful advertising campaign, a monopolistic competitor experiences an increase in demand for its product. How will that affect the price it charges and the quantity it supplies?
2. Continuing with the scenario outlined in question 1, in the long run, the positive economic profits earned by the monopolistic competitor will attract a response either from existing firms in the industry or firms outside. As those firms capture the original firm's profit, what will happen to the original firm's profit-maximizing price and output levels?

REVIEW QUESTIONS

1. What is the relationship between product differentiation and monopolistic competition?
2. How is the perceived demand curve for a monopolistically competitive firm different from the perceived demand curve for a monopoly or a perfectly competitive firm?
3. How does a monopolistic competitor choose its profit-maximizing quantity of output and price?
4. How can a monopolistic competitor tell whether the price it is charging will cause the firm to earn profits or experience losses?
5. If the firms in a monopolistically competitive market are earning economic profits or losses in the short run, would you expect them to continue doing so in the long run? Why?
6. Is a monopolistically competitive firm productively efficient? Is it allocatively efficient? Why or why not?

CRITICAL THINKING QUESTIONS

1. Aside from advertising, how can monopolistically competitive firms increase demand for their products?
2. Make a case for why monopolistically competitive industries never reach long-run equilibrium.
3. Would you rather have efficiency or variety? That is, one opportunity cost of the variety of products we have is that each product costs more per unit than if there were only one kind of product of a given type, like shoes. Perhaps a better question is, "What is the right amount of variety? Can there be too many varieties of shoes, for example?"

PROBLEMS

Andrea's Day Spa began to offer a relaxing aromatherapy treatment. The firm asks you how much to charge to maximize profits. The demand curve for the treatments is given by the first two columns in Table 2; its total costs are given in the third column. For each level of output, calculate total revenue, marginal revenue, average cost, and marginal cost. What is the profit-maximizing level of output for the treatments and how much will the firm earn in profits?

Price	Quantity	TC
\$25.00	0	\$130
\$24.00	10	\$275
\$23.00	20	\$435
\$22.50	30	\$610
\$22.00	40	\$800
\$21.60	50	\$1,005
\$21.20	60	\$1,225

Table 2.

REFERENCES

Kantar Media. “Our Insights: Infographic—U.S. Advertising Year End Trends Report 2012.” Accessed October 17, 2013. <http://kantarmedia.us/insight-center/reports/infographic-us-advertising-year-end-trends-report-2012>.

Statistica.com. 2015. “Number of Restaurants in the United States from 2011 to 2014.” Accessed March 27, 2015. <http://www.statista.com/statistics/244616/number-of-qsr-fsr-chain-independent-restaurants-in-the-us/>.

GLOSSARY

differentiated product a product that is perceived by consumers as distinctive in some way
imperfectly competitive firms and organizations that fall between the extremes of monopoly and perfect competition

monopolistic competition many firms competing to sell similar but differentiated products

oligopoly when a few large firms have all or most of the sales in an industry

SOLUTIONS

Answers to Self-Check Questions

1. An increase in demand will manifest itself as a rightward shift in the demand curve, and a rightward shift in marginal revenue. The shift in marginal revenue will cause a movement up the marginal cost curve to the new intersection between MR and MC at a higher level of output. The new price can be read by drawing a line up from the new output level to the new demand curve, and then over to the vertical axis. The new price should be higher. The increase in quantity will cause a movement along the average cost curve to a possibly higher level of average cost. The price, though, will increase more, causing an increase in total profits.
2. As long as the original firm is earning positive economic profits, other firms will respond in ways that take away the original firm's profits. This will manifest itself as a decrease in demand for the original firm's product, a decrease in the firm's profit-maximizing price and a decrease in the firm's profit-maximizing

level of output, essentially unwinding the process described in the answer to question 1. In the long-run equilibrium, all firms in monopolistically competitive markets will earn zero economic profits.

13.2 OLIGOPOLY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain why and how oligopolies exist
- Contrast collusion and competition
- Interpret and analyze the prisoner's dilemma diagram
- Evaluate the tradeoffs of imperfect competition

Many purchases that individuals make at the retail level are produced in markets that are neither perfectly competitive, monopolies, nor monopolistically competitive. Rather, they are oligopolies. **Oligopoly** arises when a small number of large firms have all or most of the sales in an industry. Examples of oligopoly abound and include the auto industry, cable television, and commercial air travel. Oligopolistic firms are like cats in a bag. They can either scratch each other to pieces or cuddle up and get comfortable with one another. If oligopolists compete hard, they may end up acting very much like perfect competitors, driving down costs and leading to zero profits for all. If oligopolists collude with each other, they may effectively act like a monopoly and succeed in pushing up prices and earning consistently high levels of profit. Oligopolies are typically characterized by mutual interdependence where various decisions such as output, price, advertising, and so on, depend on the decisions of the other firm(s). Analyzing the choices of oligopolistic firms about pricing and quantity produced involves considering the pros and cons of competition versus collusion at a given point in time.

WHY DO OLIGOPOLIES EXIST?

A combination of the barriers to entry that create monopolies and the product differentiation that characterizes monopolistic competition can create the setting for an oligopoly. For example, when a government grants a patent for an invention to one firm, it may create a monopoly. When the government grants patents to, for example, three different pharmaceutical companies that each has its own drug for reducing high blood pressure, those three firms may become an oligopoly.

Similarly, a **natural monopoly** will arise when the quantity demanded in a market is only large enough for a single firm to operate at the minimum of the long-run average cost curve. In such a setting, the market has room for only one firm, because no smaller firm can operate at a low enough average cost to compete, and no larger firm could sell what it produced given the quantity demanded in the market.

Quantity demanded in the market may also be two or three times the quantity needed to produce at the minimum of the average cost curve—which means that the market would have room for only two or three oligopoly firms (and they need not produce differentiated products). Again, smaller firms would have higher average costs and be unable to compete, while additional large firms would produce such a high quantity that they would not be able to sell it at a profitable price. This combination of economies of scale and market demand creates the barrier to entry, which led to the Boeing-Airbus oligopoly for large passenger aircraft.

The product differentiation at the heart of monopolistic competition can also play a role in creating oligopoly. For example, firms may need to reach a certain minimum size before they are able to spend enough on advertising and marketing to create a recognizable brand name. The problem in competing with, say, Coca-Cola or Pepsi is not that producing fizzy drinks is technologically difficult, but rather that creating a brand name and marketing effort to equal Coke or Pepsi is an enormous task.

COLLUSION OR COMPETITION?

When oligopoly firms in a certain market decide what quantity to produce and what price to charge, they face a temptation to act as if they were a monopoly. By acting together, oligopolistic firms can hold down industry output, charge a higher price, and divide up the profit among themselves. When firms act together in this way to reduce output and keep prices high, it is called **collusion**. A group of firms that have a formal agreement to collude to produce the monopoly output and sell at the monopoly price is called a **cartel**. See the following Clear It Up feature for a more in-depth analysis of the difference between the two.

COLLUSION VERSUS CARTELS: HOW CAN I TELL WHICH IS WHICH?

In the United States, as well as many other countries, it is illegal for firms to collude since collusion is anti-competitive behavior, which is a violation of antitrust law. Both the **Antitrust Division of the Justice Department** and the **Federal Trade Commission** have responsibilities for preventing collusion in the United States.

The problem of enforcement is finding hard evidence of collusion. Cartels are formal agreements to collude. Because cartel agreements provide evidence of collusion, they are rare in the United States. Instead, most collusion is tacit, where firms implicitly reach an understanding that competition is bad for profits.

The desire of businesses to avoid competing so that they can instead raise the prices that they charge and earn higher profits has been well understood by economists. **Adam Smith** wrote in *Wealth of Nations* in 1776: “People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices.”

Even when oligopolists recognize that they would benefit as a group by acting like a monopoly, each individual oligopoly faces a private temptation to produce just a slightly higher quantity and earn slightly higher profit—while still counting on the other oligopolists to hold down their production and keep prices high. If at least some oligopolists give in to this temptation and start producing more, then the market price will fall. Indeed, a small handful of oligopoly firms may end up competing so fiercely that they all end up earning zero economic profits—as if they were perfect competitors.

THE PRISONER'S DILEMMA

Because of the complexity of oligopoly, which is the result of mutual interdependence among firms, there is no single, generally-accepted theory of how oligopolies behave, in the same way that we have theories for all the other market structures. Instead, economists use **game theory**, a branch of mathematics that analyzes situations in which players must make decisions and then receive payoffs based on what other players decide to do. Game theory has found widespread applications in the social sciences, as well as in business, law, and military strategy.

The **prisoner's dilemma** is a scenario in which the gains from cooperation are larger than the rewards from pursuing self-interest. It applies well to oligopoly. The story behind the prisoner's dilemma goes like this:

Two co-conspiratorial criminals are arrested. When they are taken to the police station, they refuse to say anything and are put in separate interrogation rooms. Eventually, a police officer enters the room where Prisoner A is being held and says: "You know what? Your partner in the other room is confessing. So your partner is going to get a light prison sentence of just one year, and because you're remaining silent, the judge is going to stick you with eight years in prison. Why don't you get smart? If you confess, too, we'll cut your jail time down to five years, and your partner will get five years, also." Over in the next room, another police officer is giving exactly the same speech to Prisoner B. What the police officers do not say is that if both prisoners remain silent, the evidence against them is not especially strong, and the prisoners will end up with only two years in jail each.

The game theory situation facing the two prisoners is shown in Table 3. To understand the dilemma, first consider the choices from Prisoner A's point of view. If A believes that B will confess, then A ought to confess, too, so as to not get stuck with the eight years in prison. But if A believes that B will not confess, then A will be tempted to act selfishly and confess, so as to serve only one year. The key point is that A has an incentive to confess regardless of what choice B makes! B faces the same set of choices, and thus will have an incentive to confess regardless of what choice A makes. Confess is considered the dominant strategy or the strategy an individual (or firm) will pursue regardless of the other individual's (or firm's) decision. The result is that if prisoners pursue their own self-interest, both are likely to confess, and end up doing a total of 10 years of jail time between them.

		Prisoner B	
		Remain Silent (cooperate with other prisoner)	Confess (do not cooperate with other prisoner)
Prisoner A	Remain Silent (cooperate with other prisoner)	A gets 2 years, B gets 2 years	A gets 8 years, B gets 1 year
	Confess (do not cooperate with other prisoner)	A gets 1 year, B gets 8 years	A gets 5 years B gets 5 years

Table 3. The Prisoner's Dilemma Problem

The game is called a dilemma because if the two prisoners had cooperated by both remaining silent, they would only have had to serve a total of four years of jail time between them. If the two prisoners can work out some way of cooperating so that neither one will confess, they will both be better off than if they each follow their own individual self-interest, which in this case leads straight into longer jail terms.

THE OLIGOPOLY VERSION OF THE PRISONER'S DILEMMA

The members of an oligopoly can face a prisoner's dilemma, also. If each of the oligopolists cooperates in holding down output, then high monopoly profits are possible. Each oligopolist, however, must worry that while it is holding down output, other firms are taking advantage of the high price by raising output and earning higher profits. Table 4 shows the prisoner's dilemma for a two-firm oligopoly—known as a **duopoly**. If Firms A and B both agree to hold down output, they are acting together as a monopoly and will each earn \$1,000 in profits. However, both firms' dominant strategy is to increase output, in which case each will earn \$400 in profits.

		Firm B	
		Hold Down Output (cooperate with other firm)	Increase Output (do not cooperate with other firm)
Firm A	Hold Down Output (cooperate with other firm)	A gets \$1,000, B gets \$1,000	A gets \$200, B gets \$1,500
	Increase Output (do not cooperate with other firm)	A gets \$1,500, B gets \$200	A gets \$400, B gets \$400

Table 4. A Prisoner's Dilemma for Oligopolists

Can the two firms trust each other? Consider the situation of Firm A:

- If A thinks that B will cheat on their agreement and increase output, then A will increase output, too, because for A the profit of \$400 when both firms increase output (the bottom right-hand choice in Table 4) is better than a profit of only \$200 if A keeps output low and B raises output (the upper right-hand choice in the table).
- If A thinks that B will cooperate by holding down output, then A may seize the opportunity to earn higher profits by raising output. After all, if B is going to hold down output, then A can earn \$1,500 in profits by expanding output (the bottom left-hand choice in the table) compared with only \$1,000 by holding down output as well (the upper left-hand choice in the table).

Thus, firm A will reason that it makes sense to expand output if B holds down output and that it also makes sense to expand output if B raises output. Again, B faces a parallel set of decisions.

The result of this prisoner's dilemma is often that even though A and B could make the highest combined profits by cooperating in producing a lower level of output and acting like a monopolist, the two firms may well end up in a situation where they each increase **output** and earn only \$400 each in **profits**. The following Clear It Up feature discusses one cartel scandal in particular.

WHAT IS THE LYSINE CARTEL?

Lysine, a \$600 million-a-year industry, is an amino acid used by farmers as a feed additive to ensure the proper growth of swine and poultry. The primary U.S. producer of lysine is Archer Daniels Midland (ADM), but several other large European and Japanese firms are also in this market. For a time in the first half of the 1990s, the world's major lysine producers met together in hotel conference rooms and decided exactly how much each firm would sell and what it would charge. The U.S. Federal Bureau of Investigation (FBI), however, had learned of the cartel and placed wire taps on a number of their phone calls and meetings.

From FBI surveillance tapes, following is a comment that Terry Wilson, president of the corn processing division at ADM, made to the other lysine producers at a 1994 meeting in Mona, Hawaii:

I wanna go back and I wanna say something very simple. If we're going to trust each other, okay, and if I'm assured that I'm gonna get 67,000 tons by the year's end, we're gonna sell it at the prices we agreed to . . . The only thing we need to talk about there because we are gonna get manipulated by these [expletive] buyers—they can be smarter than us if we let them be smarter. . . . They [the customers] are not your friend. They are not my friend. And we gotta have 'em, but they are not my friends. You are my friend. I wanna be closer to you than I am to any customer. Cause you can make us . . . money. . . . And all I wanna tell you again is let's—let's put the prices on the board. Let's all agree that's what we're gonna do and then walk out of here and do it.

The price of lysine doubled while the cartel was in effect. Confronted by the FBI tapes, Archer Daniels Midland pled guilty in 1996 and paid a fine of \$100 million. A number of top executives, both at ADM and other firms, later paid fines of up to \$350,000 and were sentenced to 24–30 months in prison.

In another one of the FBI recordings, the president of Archer Daniels Midland told an executive from another competing firm that ADM had a slogan that, in his words, had “penetrated the whole company.” The company president stated the slogan this way: “Our competitors are our friends. Our customers are the enemy.” That slogan could stand as the motto of cartels everywhere.

HOW TO ENFORCE COOPERATION

How can parties who find themselves in a prisoner's dilemma situation avoid the undesired outcome and cooperate with each other? The way out of a prisoner's dilemma is to find a way to penalize those who do not cooperate.

Perhaps the easiest approach for colluding oligopolists, as you might imagine, would be to sign a contract with each other that they will hold output low and keep prices high. If a group of U.S. companies signed such a contract, however, it would be illegal. Certain international organizations, like the nations that are members of the **Organization of Petroleum Exporting Countries (OPEC)**, have signed international agreements to act like a monopoly, hold down output, and keep prices high so that all of the countries can make high profits from oil exports. Such agreements, however, because they fall in a gray area of international law, are not legally enforceable. If Nigeria, for example, decides to start cutting prices and selling more oil, Saudi Arabia cannot sue Nigeria in court and force it to stop.

Visit the Organization of the Petroleum Exporting Countries website and learn more about its history and how it defines itself.



Because oligopolists cannot sign a legally enforceable contract to act like a monopoly, the firms may instead keep close tabs on what other firms are producing and charging. Alternatively, oligopolists

may choose to act in a way that generates pressure on each firm to stick to its agreed quantity of output.

One example of the pressure these firms can exert on one another is the **kinked demand curve**, in which competing oligopoly firms commit to match price cuts, but not price increases. This situation is shown in Figure 1. Say that an oligopoly airline has agreed with the rest of a cartel to provide a quantity of 10,000 seats on the New York to Los Angeles route, at a price of \$500. This choice defines the kink in the firm's perceived demand curve. The reason that the firm faces a kink in its demand curve is because of how the other oligopolists react to changes in the firm's price. If the oligopoly decides to produce more and cut its price, the other members of the cartel will immediately match any price cuts—and therefore, a lower price brings very little increase in quantity sold.

If one **firm** cuts its price to \$300, it will be able to sell only 11,000 seats. However, if the airline seeks to raise prices, the other oligopolists will not raise their prices, and so the firm that raised prices will lose a considerable share of sales. For example, if the firm raises its price to \$550, its sales drop to 5,000 seats sold. Thus, if oligopolists always match price cuts by other firms in the cartel, but do not match price increases, then none of the oligopolists will have a strong incentive to change prices, since the potential gains are minimal. This strategy can work like a silent form of cooperation, in which the cartel successfully manages to hold down output, increase **price**, and share a monopoly level of profits even without any legally enforceable agreement.

Many real-world oligopolies, prodded by economic changes, legal and political pressures, and the egos of their top executives, go through episodes of cooperation and competition. If oligopolies could sustain cooperation with each other on output and pricing, they could earn profits as if they were a single monopoly. However, each firm in an oligopoly has an incentive to produce more and grab a bigger share of the overall market; when firms start behaving in this way, the market outcome in terms of prices and quantity can be similar to that of a highly competitive market.

TRADEOFFS OF IMPERFECT COMPETITION

Monopolistic competition is probably the single most common market structure in the U.S. economy. It provides powerful incentives for innovation, as firms seek to earn profits in the short run, while entry assures that firms do not earn economic profits in the long run. However, monopolistically competitive firms do not produce at the lowest point on their average cost curves. In addition, the endless search to impress consumers through product differentiation may lead to excessive social expenses on advertising and marketing.

Oligopoly is probably the second most common market structure. When oligopolies result from patented innovations or from taking advantage of economies of scale to produce at low average cost, they may provide considerable benefit to consumers. Oligopolies are often buffeted by significant barriers to entry, which enable the oligopolists to earn sustained profits over long periods of time. Oligopolists also do not typically produce at the minimum of their average cost curves. When they lack vibrant competition, they may lack incentives to provide innovative products and high-quality service.

The task of public policy with regard to competition is to sort through these multiple realities, attempting to encourage behavior that is beneficial to the broader society and to discourage behavior

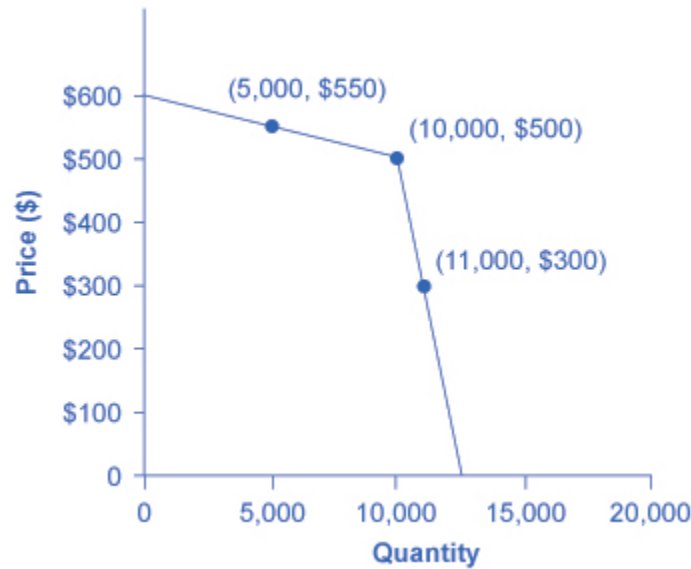


Figure 1. A Kinked Demand Curve. Consider a member firm in an oligopoly cartel that is supposed to produce a quantity of 10,000 and sell at a price of \$500. The other members of the cartel can encourage this firm to honor its commitments by acting so that the firm faces a kinked demand curve. If the oligopolist attempts to expand output and reduce price slightly, other firms also cut prices immediately—so if the firm expands output to 11,000, the price per unit falls dramatically, to \$300. On the other side, if the oligopoly attempts to raise its price, other firms will not do so, so if the firm raises its price to \$550, its sales decline sharply to 5,000. Thus, the members of a cartel can discipline each other to stick to the pre-agreed levels of quantity and price through a strategy of matching all price cuts but not matching any price increases.

that only adds to the profits of a few large companies, with no corresponding benefit to consumers. Monopoly and Antitrust Policy discusses the delicate judgments that go into this task.

THE TEMPTATION TO DEFY THE LAW

Oligopolistic firms have been called “cats in a bag,” as this chapter mentioned. The French detergent makers chose to “cozy up” with each other. The result? An uneasy and tenuous relationship. When the *Wall Street Journal* reported on the matter, it wrote: “According to a statement a Henkel manager made to the [French anti-trust] commission, the detergent makers wanted ‘to limit the intensity of the competition between them and clean up the market.’ Nevertheless, by the early 1990s, a price war had broken out among them.” During the soap executives’ meetings, which sometimes lasted more than four hours, complex pricing structures were established. “One [soap] executive recalled ‘chaotic’ meetings as each side tried to work out how the other had bent the rules.” Like many cartels, the soap cartel disintegrated due to the very strong temptation for each member to maximize its own individual profits.

How did this soap opera end? After an investigation, French antitrust authorities fined Colgate-Palmolive, Henkel, and Procter & Gamble a total of €361 million (\$484 million). A similar fate befell the icemakers. Bagged ice is a commodity, a

perfect substitute, generally sold in 7- or 22-pound bags. No one cares what label is on the bag. By agreeing to carve up the ice market, control broad geographic swaths of territory, and set prices, the icemakers moved from perfect competition to a monopoly model. After the agreements, each firm was the sole supplier of bagged ice to a region; there were profits in both the long run and the short run. According to the courts: “These companies illegally conspired to manipulate the marketplace.” Fines totaled about \$600,000—a steep fine considering a bag of ice sells for under \$3 in most parts of the United States.

Even though it is illegal in many parts of the world for firms to set prices and carve up a market, the temptation to earn higher profits makes it extremely tempting to defy the law.

KEY CONCEPTS AND SUMMARY

An oligopoly is a situation where a few firms sell most or all of the goods in a market. Oligopolists earn their highest profits if they can band together as a cartel and act like a monopolist by reducing output and raising price. Since each member of the oligopoly can benefit individually from expanding output, such collusion often breaks down—especially since explicit collusion is illegal.

The prisoner’s dilemma is an example of game theory. It shows how, in certain situations, all sides can benefit from cooperative behavior rather than self-interested behavior. However, the challenge for the parties is to find ways to encourage cooperative behavior.

SELF-CHECK QUESTIONS

1. Consider the curve shown in Figure 2, which shows the market demand, marginal cost, and marginal revenue curve for firms in an oligopolistic industry. In this example, we assume firms have zero fixed costs.

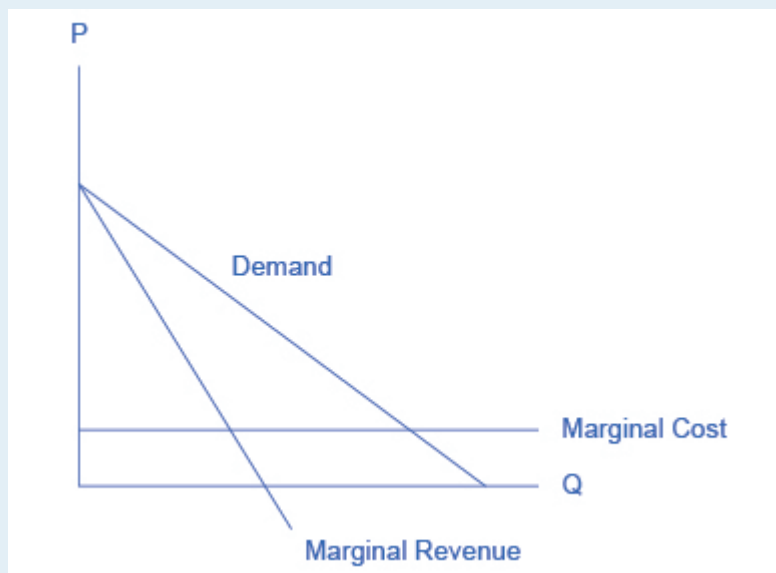


Figure 2.

- a. Suppose the firms collude to form a cartel. What price will the cartel charge? What quantity will the cartel supply? How much profit will the cartel earn?
- b. Suppose now that the cartel breaks up and the oligopolistic firms compete as vigorously as

possible by cutting the price and increasing sales. What will the industry quantity and price be? What will the collective profits be of all firms in the industry?

- c. Compare the equilibrium price, quantity, and profit for the cartel and cutthroat competition outcomes.
2. Sometimes oligopolies in the same industry are very different in size. Suppose we have a duopoly where one firm (Firm A) is large and the other firm (Firm B) is small, as shown in the prisoner's dilemma box in Table 5.

	Firm B colludes with Firm A	Firm B cheats by selling more output
Firm A colludes with Firm B	A gets \$1,000, B gets \$100	A gets \$800, B gets \$200
Firm A cheats by selling more output	A gets \$1,050, B gets \$50	A gets \$500, B gets \$20

Table 5.

Assuming that the payoffs are known to both firms, what is the likely outcome in this case?

REVIEW QUESTIONS

1. Will the firms in an oligopoly act more like a monopoly or more like competitors? Briefly explain.
2. Does each individual in a prisoner's dilemma benefit more from cooperation or from pursuing self-interest? Explain briefly.
3. What stops oligopolists from acting together as a monopolist and earning the highest possible level of profits?

CRITICAL THINKING QUESTIONS

1. Would you expect the kinked demand curve to be more extreme (like a right angle) or less extreme (like a normal demand curve) if each firm in the cartel produces a near-identical product like OPEC and petroleum? What if each firm produces a somewhat different product? Explain your reasoning.
2. When OPEC raised the price of oil dramatically in the mid-1970s, experts said it was unlikely that the cartel could stay together over the long term—that the incentives for individual members to cheat would become too strong. More than forty years later, OPEC still exists. Why do you think OPEC has been able to beat the odds and continue to collude? *Hint:* You may wish to consider non-economic reasons.

PROBLEMS

1. Mary and Raj are the only two growers who provide organically grown corn to a local grocery store. They know that if they cooperated and produced less corn, they could raise the price of the corn. If they work independently, they will each earn \$100. If they decide to work together and both lower their output, they

can each earn \$150. If one person lowers output and the other does not, the person who lowers output will earn \$0 and the other person will capture the entire market and will earn \$200. Table 6 represents the choices available to Mary and Raj. What is the best choice for Raj if he is sure that Mary will cooperate? If Mary thinks Raj will cheat, what should Mary do and why? What is the prisoner's dilemma result? What is the preferred choice if they could ensure cooperation? A = Work independently; B = Cooperate and Lower Output. (Each results entry lists Raj's earnings first, and Mary's earnings second.)

2. Jane and Bill are apprehended for a bank robbery. They are taken into separate rooms and questioned by the police about their involvement in the crime. The police tell them each that if they confess and turn the other person in, they will receive a lighter sentence. If they both confess, they will be each be sentenced to 30 years. If neither confesses, they will each receive a 20-year sentence. If only one confesses, the confessor will receive 15 years and the one who stayed silent will receive 35 years. Table 7 below represents the choices available to Jane and Bill. If Jane trusts Bill to stay silent, what should she do? If Jane thinks that Bill will confess, what should she do? Does Jane have a dominant strategy? Does Bill have a dominant strategy? A = Confess; B = Stay Silent. (Each results entry lists Jane's sentence first (in years), and Bill's sentence second.)

		Jane	
		A	B
Bill	A	(30, 30)	(15, 35)
	B	(35, 15)	(20, 20)

Table 7.

REFERENCES

The United States Department of Justice. "Antitrust Division." Accessed October 17, 2013. <http://www.justice.gov/atr/>.

eMarketer.com. 2014. "Total US Ad Spending to See Largest Increase Since 2004: Mobile advertising leads growth; will surpass radio, magazines and newspapers this year. Accessed March 12, 2015. <http://www.emarketer.com/Article/Total-US-Ad-Spending-See-Largest-Increase-Since-2004/1010982>.

Federal Trade Commission. "About the Federal Trade Commission." Accessed October 17, 2013. <http://www.ftc.gov/ftc/about.shtm>.

GLOSSARY

cartel a group of firms that collude to produce the monopoly output and sell at the monopoly price

collusion when firms act together to reduce output and keep prices high

duopoly an oligopoly with only two firms

game theory a branch of mathematics often used by economists that analyzes situations in which players must make decisions and then receive payoffs based on what decisions the other players make

kinked demand curve a perceived demand curve that arises when competing oligopoly firms commit to match price cuts, but not price increases

prisoner's dilemma a game in which the gains from cooperation are larger than the rewards from pursuing self-interest

SOLUTIONS

Answers to Self-Check Questions

1.
 - a. If the firms form a cartel, they will act like a monopoly, choosing the quantity of output where $MR = MC$. Drawing a line from the monopoly quantity up to the demand curve shows the monopoly price. Assuming that fixed costs are zero, and with an understanding of cost and profit, we can infer that when the marginal cost curve is horizontal, average cost is the same as marginal cost. Thus, the cartel will earn positive economic profits equal to the area of the rectangle, with a base equal to the monopoly quantity and a height equal to the difference between price (on the demand above the monopoly quantity) and average cost, as shown in the following figure.

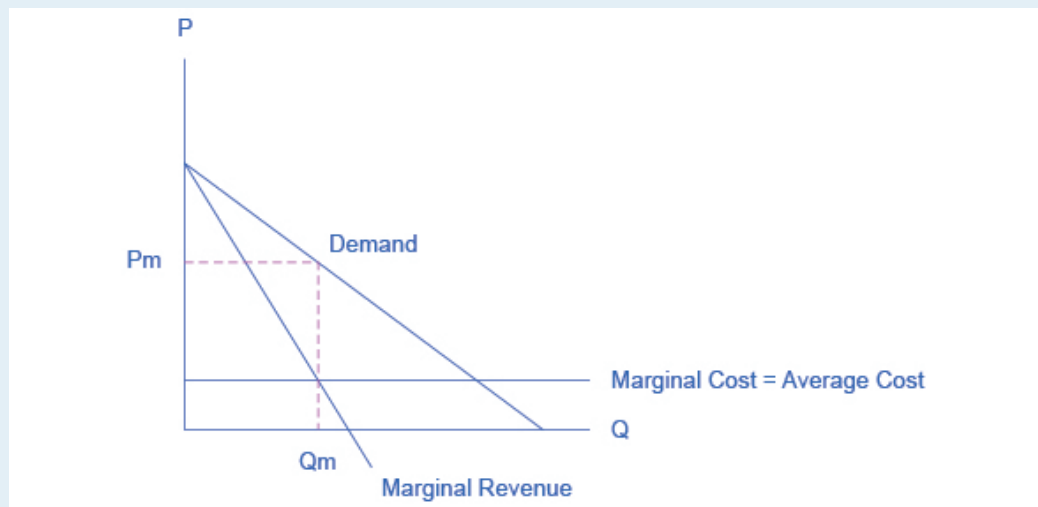


Figure 3.

- b. The firms will expand output and cut price as long as there are profits remaining. The long-run equilibrium will occur at the point where average cost equals demand. As a result, the oligopoly will earn zero economic profits due to “cutthroat competition,” as shown in the next figure.
 - c. $P_c > P_{cc}$. $Q_c < Q_{cc}$. Profit for the cartel is positive and large. Profit for cutthroat competition is zero.
2. Firm B reasons that if it cheats and Firm A does not notice, it will double its money. Since Firm A's profits will decline substantially, however, it is likely that Firm A will notice and if so, Firm A will cheat also, with the result that Firm B will lose 90% of what it gained by cheating. Firm A will reason that Firm B is unlikely to risk cheating. If neither firm cheats, Firm A earns \$1000. If Firm A cheats, assuming Firm B does not cheat, A can boost its profits only a little, since Firm B is so small. If both firms cheat, then Firm A loses at least 50% of what it could have earned. The possibility of a small gain (\$50) is probably not enough to induce Firm A to cheat, so in this case it is likely that both firms will collude.

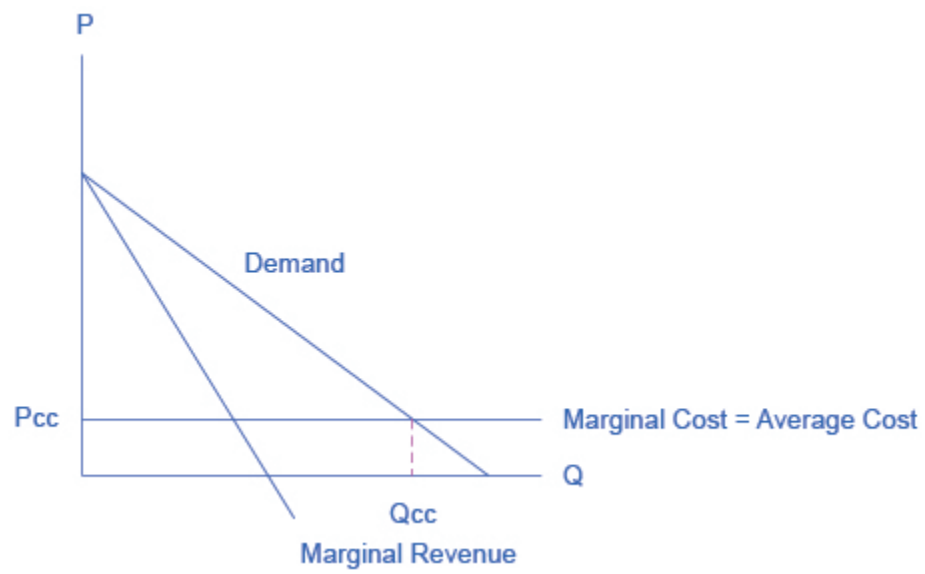


Figure 4.

CHAPTER 14. THE RISE OF BIG BUSINESS

INTRODUCTION TO THE RISE OF BIG BUSINESS

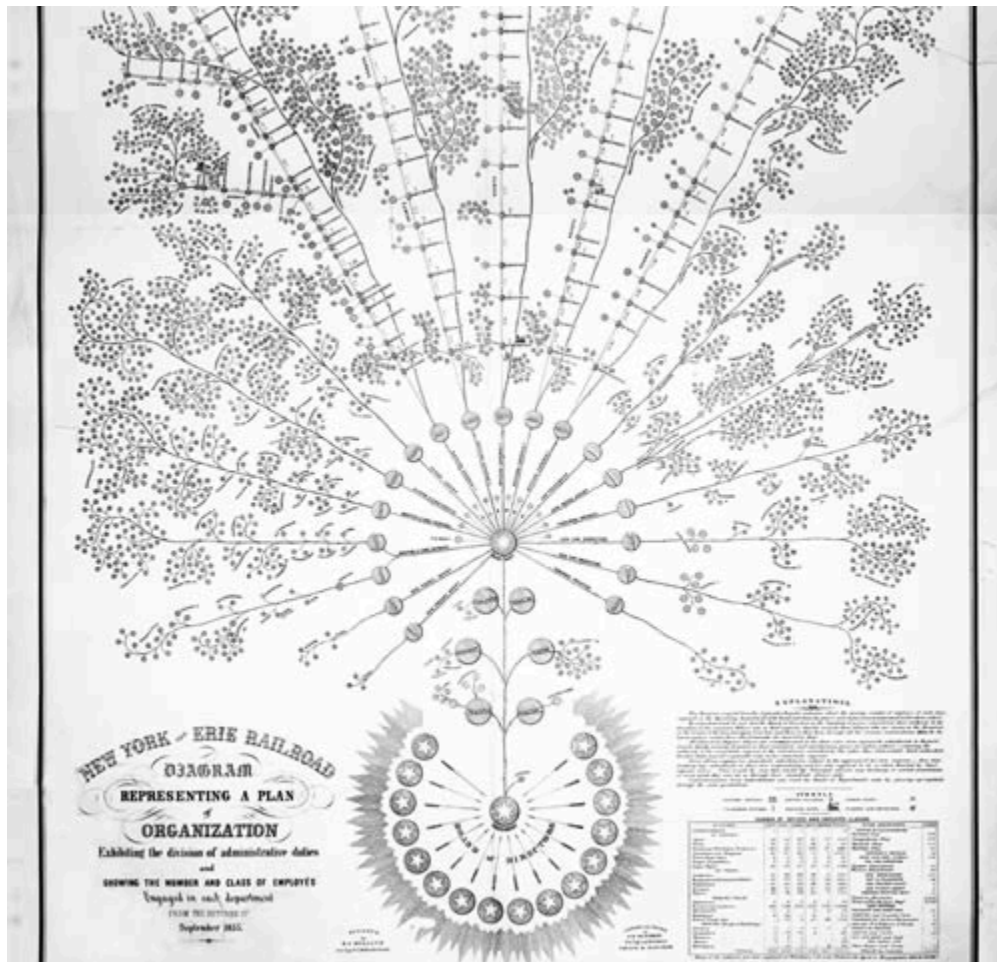


Figure 1. Organizational diagram of the New York and Erie Railroad, 1855. Daniel McCallum & George Holt Henshaw. Public Domain.

LEARNING OBJECTIVES

Introduction to the Rise of Big Business

In this chapter, you will learn about:

- Big Business in American History
- Industrialization and the Factory

- Big Business and Organized Labor

As you will see in the next few chapters, heterodox economists take a very different approach to studying the modern business enterprise. In contrast to focusing on the neoclassical ideal of relatively small firms in perfectly competitive markets, heterodox economists start with the recognition that the modern business enterprise tends to be large. We may think of large firms as possessing differential advantage over their competitors and the ability to exert control over key moments throughout the production process. Finally, we choose to place big business at the center of our economic inquiry, because the large corporation possesses the power to act in ways that other agents cannot. In this regard, the Megacorp—as Post Keynesian economist Alfred Eichner call it—may be viewed as a social actor—that is, one who may act with institutional capacity.

What do we mean when we say that some persons act with institutional capacity? It is helpful to return briefly to the problem of institutions. In an earlier chapter on consumption we defined institutions as collectively shared habits of thought—of knowing, doing, and valuing. The American institutional economist John R. Commons suggests that institutions result in shaping individual behavior by controlling, liberating and expanding individual action. Commons' view of the relational effects that institutions exert on individual action facilitates examination of history as a process of cumulative causation. We wish to know how individuals interact with society as a whole, and how they may affect the systems as a whole in varying degrees of efficacy. And in turn, we wish to know how social structures constrain, shape and focus individual action.

It is important that we keep these concepts in mind as we study the history of big business. Without a clear view as to how individuals interact with one another and the system as a whole, it is difficult to situate the megacorp in history, instead relegating its emergence as one of technical inevitability. With these tools let us proceed with an overview of the history of big business.

THE JOINT-STOCK CORPORATION AND LONG DISTANCE TRADE

The modern business enterprise has its roots in the early trade regimes of European nation states during the 17th and 18th centuries.

Corporations like the East India Company were granted special privileges by the political ruling establishment in order to guard against the uncertainties involved with long-distance trade – an inherently risky proposition. Such privileges are referred to in economic history as Crown Monopolies, by reference to their relationship with the monarchy. Crown monopolies belong to a class of restrictive trade practices that may be considered to be common to all capitalist markets. However, it is important to stress that the actual flavor that such market governance mechanisms embody depends upon the historical context in which we examine them. It is useful to draw comparisons between the joint-stock corporations that enjoyed crown monopolies, such as the East India Company, and modern corporations that benefit from other protections in accordance with modern market governance institutions, but we should exercise caution as some modern developments do not fit with early capitalist institutions.



Figure 2. Coat of Arms of the British East India Company. Wikimedia Commons. CC BY-SA 3.0

MERCANTILISM

Mercantilism is a term that economists use to describe a social provisioning process peculiar to Western European nation-states during the 17th and 18th centuries. Philosophers who thought and wrote about economic ideas believed that a strong nation-state was one that was able to accumulate more gold and silver than their contemporaries. Hence, the orientation for the economies of these monarchies was that of external trade. To effect this end, the monarch had to enable a fleet of merchant vessels that were capable of engaging in long-distance trade, often with costly and uncertain outcomes at the end of a long voyage. Merchant vessels and crew are costly endeavors – far beyond the capacity of any given individual to finance on a sustainable basis. Therefore, it was necessary to pool the funds of many investors vested in a single shipping enterprise. But, capitalists are easily spooked by the prospect of losses (early, proto-capitalists were no exception) and so the practice of granting special protections and privileges to investors pooling funds for shipping concerns facilitated and enabled the investment. It is important to note, also, that capitalists vested in joint-stock corporations for the purposes of long-distance trade were not random citizens. They were connected via kinship networks to the aristocracy.

Adam Smith's magnum opus, *An Inquiry into the Nature and Causes of the Wealth of Nations* [1776], may be viewed as an attack on the political economy of mercantilism. Smith's critique lies on the theory of value embodied in mercantilism. Smith observes that mercantilism was predicated upon a zero sum game. Enriching the welfare of a given nation-state meant that another had to lose relative shares of total economic value. Smith argues that production, not trade, was the wellspring of value. Productivity, not special privileges to extended members of the aristocracy, should be the end in view of any nation wishing to enrich itself. For Smith, productivity is enhanced through a greater elaboration of the division of labor and extension of the market in society. Restrictive trade practices only stymied this development, Smith believed.

Despite Smith's protestations market governance institutions that result in creating special privileges for specific firms remain a commonplace in capitalist society.

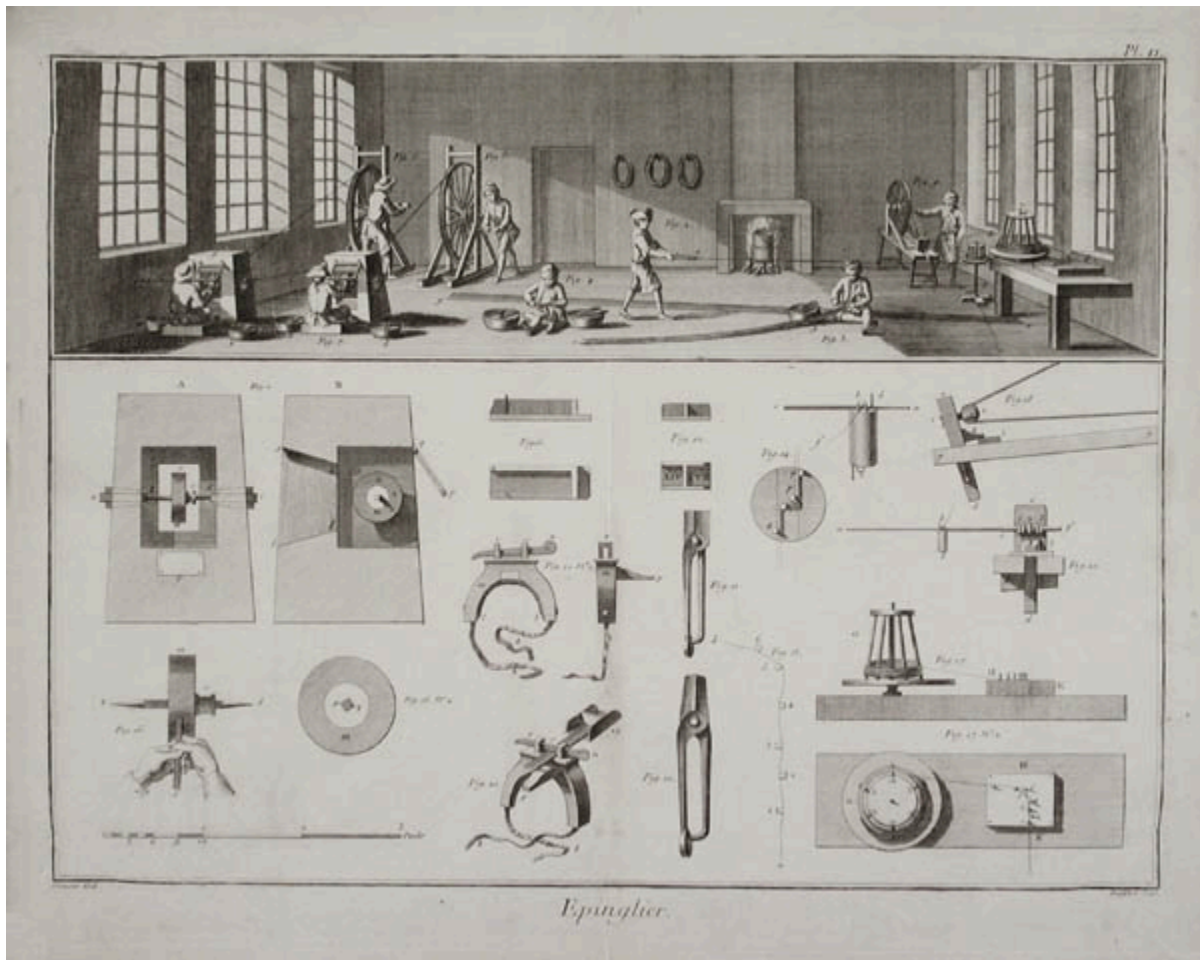


Figure 3. Pin maker's factory, indicating the division of labor in making a simple pin. Depicted in Diderot's *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers* 1762: Epinglier, Plate II. This image is in the public domain.

14.1 BIG BUSINESS IN AMERICAN HISTORY

The history of the American experience tends to be viewed as exceptional relative to Europe. The tradition has generally been to view the course of American development as unique and separate from the antecedent European history that America originates from. One consequence of this ideological tradition has been to view corporate power as an anomaly to an otherwise well-functioning system, in which unfettered market forces generally produce highly competitive, small firms that do not possess effective market power. The actual history is quite different.

From the very foundation of the American colonies big business has been the rule, not the exception. The Virginia Company, a joint-stock company that enjoyed an exclusive patent on very large swathes of land (see figure 1 below), established the first British colony at Jamestown in 1607. A patent is taken here to mean an exclusive grant by the sovereign with respect to a particular land tract, and establishes a fundamental claim to ownership and disposal of that land.

The managerial class vested with the power of the charter administered all aspects of colonial life. Virtually all of the land that makes up the area that is described as the original thirteen colonies was held in ownership by a small set of individuals vested with property rights derived from monarchical legal traditions, granted with an exclusive patent for private development.

BUSINESS ENTERPRISE AND PROPERTY RIGHTS

The corporate firm and the legal structure of property rights coevolved in the course of US economic history. When we examine the going concern with an emphasis on its proprietary relationships with other aspects of the economy this codevelopment appears all the more transparent. That is, the firm is the site of a great many transactions in the economic system. Monetary transactions in which the firm makes a claim to resources generated in the economy involve rules governing the proprietary relationships between the transacting party and the resources in question. In order to secure exclusive and steady claim to a set of resources the firm must be contractually entitled to the resources as private property. Hence, it is important to study the codevelopment between firms and the legal framework on the matter of property.

These transactions provide the foundation for the intangible value of the firm as property. Intangible value emerges as the firm is able to maintain differential advantage with respect to other participants in the market, as well as maintain control over stages in the production process in which the firm is engaged, while exerting its influence over matters of market governance. Precisely how society chooses to assign rights to such property is a matter of primary importance from the standpoint of determining the status of vestiture. Hence, insofar as firms embody the creation of property and affect existing claims to property, it is necessary to adjust the legal framework to establish a coherent system of property rights. By setting up the problem as essentially a relationship between the law and the



Figure 1. Map of the Colony of Virginia — during 17th and early 18th centuries, by Knusser. CC BY SA 2.5

corporation, we can appreciate the importance of the corporation in the development of American capitalism.

COLONIAL LEGAL INSTITUTIONS

Colonial America borrowed its legal framework from the English common law system. The common law tradition allows for judges to freely interpret new laws in light of old ones, suggesting that judges take an active role in shaping the subsequent development of legal institutions. This legal tradition of judicial instrumentalism will be important as we examine the changing nature of property and the firm.

REAL PROPERTY RIGHTS

The colonists derived their property rights as privileges of the king. Since colonists were subjects of the king, they enjoyed all the rights to property that citizens in London enjoyed. The Crown permitted the **free and common socage** system of land tenure in the American colonies, as opposed to other land tenure systems that reflected the myriad feudal relations and systems of mutual obligation. Free and common socage is described by the following features:

- Perpetual ownership

- Transmissible to heirs.
- Free disposal of property
- Alienable

The fee simple system, as Americans would come to know it, embodies a set of rights. Property is often thought of as a “bundle of rights” for this reason. Accompanying these bundles of rights are a set of rents due the donor of the land, the king. Quit rents, then, represent the debt the holder of real property owes to the sovereign for the right to remain at liberty with the property. If the landholder defaults on this obligation by failing to make the quit rent payment, ownership reverts to the sovereign. Modern observers may look to a city like Kansas City, Missouri which possesses a large inventory of housing whose landlords are not current in their tax liability to the local government. Eventually, the state takes possession of the real property and all rights associated with ownership are revoked, until the property is disposed of again as private property.

POLICE POWER

One of the principal legal developments in American economic history is that the state holds broad powers to establish and define markets, or otherwise exercise its police powers toward some economic end. This power allows the state to establish special privileges to private entities so that it may collect rents by controlling access to resources. For example, suppose a local municipality wished to transform a waterfall into a lock & canal system in order to facilitate inland navigation via steamship. The state may grant a franchise to a firm for exclusive use of the public right of way, thereby entitling the owner of the concern to a rent. Alternatively, the state may attempt to directly develop the falls to remove the impediment to navigation. In either case, the extent to which income is transferred from the community as a result of the establishment of a lock & canal depends entirely upon the discretion of those in control of the property, who must decide whether and to what extent they will charge rents. In general, the firm must earn revenues from its ongoing marketing activity in order to remain a going concern.

EMERGENCE OF PRIORITY RIGHTS

One peculiarity of English common law was the institution of granting existing property holders protections against a future neighbor who may impose costs upon them in their own pursuits regarding their property. The doctrine of ancient lights, as it was known, is the underlying concept of the practice of establishing a basis for prescriptive rights in common law. This common law principle suggests that the long tenured holders of property rights are protected from new developments that would diminish or impede upon their property, in an absolute sense. For example, a property holder who enjoys the benefit of, say, unimpeded sunlight would have the reasonable expectation that rival parties who wish to construct a home adjacent to their property would not be able to do so if it resulted in blocking such daylight. But, the doctrine of ancient lights was not restricted solely to the issue of sunlight, but more broadly as referring to general benefits to property ownership. In effect, the doctrine of ancient lights imputes a conservative bias in the course of development in such a way that favors temporal or generation advantage to prior property holders.

As the American economy developed, this doctrine was tested in court and found to obstruct the economic development of urban areas. Prescriptive rights were found to be inconsistent with the social conditions in which American economic development was embedded. In 1838, a case in New York state, *Parker v. Foote*, established precedent for overturning the institution of prescriptive rights,

wherein the judge ruled that the doctrine of ancient lights cannot be applied without inhibiting the growth of cities and internal improvements. Hence, the law was adjusted by judicial instrumentalism to accommodate and legitimate negative externalities as a byproduct of economic development. When economists refer to negative externalities they are describing costs that impact third parties as a result of some economic process not born by the agents principally engaged in such activity.

THE VALIDITY OF CONTRACTS BETWEEN UNEQUALS

Under English contract law, contracts were held as valid only when struck on an equitable basis. That contract validity rested upon notions of equity was vestigial to feudal institutions. Contracts were not considered valid if they were struck between two parties that were viewed as too unequal. For example, contracts between someone of sound mind and one that is mentally impaired would be considered invalid on the premise that the two parties were not equal in their contracting ability. It follows that based upon this equity consideration in the common-law tradition contracts between, say, a factory owner and an individual laborer would not be valid. In American capitalism, such an idea is anachronistic and incompatible with the manner in which value is extracted, appropriated and capitalized. While the equity tradition remained central to American contract law throughout the 18th century, four important changes to contract law would emerge in the course of American development.

First, in the 19th century the courts interpreted contracts struck between free persons without compulsions as valid, on the principle that mutual assent generally implied validity irrespective of equity. For example, if a factory owner made the terms of employment known to the public and workers accepted a labor contract in exchange for wages, then the contract is valid whether or not the individual worker stands on equitable terms with her employer. In this circumstance equity refers to the balance of bargaining power between parties in transaction with each other. It would be absurd to think that a factory owner possesses the same bargaining power as an individual worker, for the owner controls and exercises their discretion over the the production process. Despite this lack of equity, the labor contract is considered valid so long as the worker knows in advance what she signing up for.

Second, the rule embodied above, that of mutual assent or “meeting of the wills” as it is referred to in contract law, applied to transactions in markets. Provided the buyer is free to inspect the quality of the goods, then any asymmetry of knowledge regarding its quality or serviceability does not impair the validity of the contract in exchange.

Third, contracts were struck in terms of market prices, as opposed to customary prices that ensure equity through time. That is to say, if the prices of the goods changed from the time the contract is struck and the buyer takes delivery, then the contracted price remains valid. This institutional change facilitated the development of futures contracts. A futures contract is nothing more than a promise to pay at a later date for delivery of some good, where the contracted price is settled in current market prices. The following breakout box provides a simply example to illustrate how they work.

HOW FUTURES CONTRACTS WORK

Suppose I am in the wheat milling business. I know that I need 100 bushels of wheat so that I can mill all of the meal I expect to sell next month. I have a feeling that next month the price of wheat is going to increase, so I decide to lock in

today's prices on the wheat that I'll receive next month. I can do this by buying a futures contract from a wheat farmer for delivery of her wheat in a month, but at today's price.

Futures contracts are the foundation of a whole class of financial instruments called options, which are ubiquitous in modern American capitalism as they are used to provide risk hedging positions for firms that have an interest in the commodities that undergird the futures contract in question. They are also a means of speculating in financial markets, but that is a topic we shall avoid for now. The main point to remember here, is that changing from customary to market prices in contract law opened up space for the financialized firm to thrive and grow, essential to the development of the modern corporation.

Fourth, the courts began to interpret contract law such that employer liability for hazards incurred by workers was limited. The presumption that, with the principle of mutual assent, dangers and risks associated with the workplace were encapsulated in the wage, implied that workers that agreed to the terms of employment absolved employers' liability of injury. Additionally, the "fellow servant doctrine" suggested that employers were further limited in liability where injuries to an employee resulted from negligence on the part of his coworkers. Given the nature of production, it is common to expect that contributory negligence plays some part in the hazards, making it difficult for workers to establish negligence and liability on the part of their employer.

THE SUPREME COURT

Since John Marshall's tenure as Chief Justice (1801- 1835), the Supreme Court has exercised considerable power in the shaping of American economic legal institutions. In the 1819 case *McCullough v. Maryland*, Chief Justice Marshall established the **implied powers** of the court in engaging in judicial review and revision of the legal framework as the court sees fit. The basic premise was that the constitution was intended to be robust through the ages to changes in the technical and social character of the nation. Accordingly, the court reserved the right and asserted its power to interpret its scope broadly; subsequent rulings would be made in light of new economic and social developments, while preserving the legitimacy of the constitution. In *McCullough v. Maryland*, the constitutional legitimacy of the Second Bank of the United States was under question. Marshall ruled that since the bank was chartered with the purpose of effecting the ends of the constitution itself, namely regulation of interstate commerce, then it was legitimate despite no specific provision in the constitution for a central bank.

Several antebellum cases before the Supreme Court helped establish the power of the federal government to regulate interstate commerce via the Commerce Clause of the Constitution, superseding state and local desires for the same. In *Gibbons v. Ogden* (1824), the court overturned a monopoly granted by the state of New York to a steamship concern because its business involved traffic in New Jersey, ruling that individual states do not have the authority to regulate interstate commerce. While the powers of the port of Philadelphia were upheld in the service of regulating commerce for its municipal needs, Chief Justice Taney in *Cooley v. Board of Wardens* (1851) held that where such regulation touched upon issues of national importance, particularly involving its seaport, those regulatory powers were reserved for Congress. *Brown v. Maryland* (1827) ruled that no state had the power to license and tax importers.

While the Supreme Court was laying the foundation for federal regulation of business enterprise, the immediate impact following *Gibbons v. Ogden* was the liberation of private corporations from state regulation. Not until 1887 with the establishment of the Interstate Commerce Commission, which emerged in response to problems associated with market governance in railroads, did the federal government assert itself on matters of regulation. As a consequence, corporations were free to grow and pursue national markets without the interference of local intervention.

Recall, this discussion of the history of changing legal institutions is in regards to the proprietary nature of corporations. Perhaps the most important case governing the subsequent development of the legal foundations of corporate property and futurity in America economic is *Dartmouth College v. Woodward* (1819). Again, Justice Marshall plays a pivotal role in the coevolution of property law and corporations. The direct issue at stake was the legitimacy of Dartmouth's corporate charter as it rested upon a grant from King George III, prior to the establishment of the United States. Indirectly, all patents granted by the King prior to the revolution were called into question if the court ruled the proprietary basis for Dartmouth's charter to be invalid following the American Revolution. Marshall ruled that revolutions do not inherently undermine vested property rights. This granted legitimacy by judicial review to all original patents by the King, including the original Virginia Company patent, which accounted for a considerable portion of all land in the United State. Nearly all of the areas affected by the Northwest Ordinances were derived from the original crown patents, suggesting that overturning the Dartmouth charter could undermine the legitimacy of all American real property rights.

CANALS, STEAMBOATS AND RAILROADS

The canalization of the America's inland waterways, a period spanning 1815 – 1843, marks an important moment in the development of the corporation. Particularly, the canal building era laid the foundations for a newfound importance for corporations and the ability for private citizens to control the development of the social provisioning process. In short, society's relationship with corporations changed as it used them as institutions for the development of large-scale public works projects.

As the US grew westward, institutions were established to facilitate inter-regional trade, and bring to eastern markets the product of the hinterland. State governments issued franchises to concerns that incorporated on the premise that they would improve waterways in the interior watersheds of the nation. In return, the corporation was entitled to revenues associated with control over the canals. Similarly, firms were incorporated to engage in inland navigation along the canalized waterways and establish rates sufficient to cover their costs and in accordance with the principle of *charging what the market may bear*. That is to say, firms set prices and administer them to the market, allowing the firm to capture quasi-rents and accumulate monetary claims on its balance sheet. Firms are limited in the extent to which such rents may be appropriated from the public, which includes factors such as the ability to pay from participants in the market as well as rules, formal and informal, that may provide a governance structure to the market. For instance, an association may exist between navigation concerns that facilitate the orderly division of the market, which may impose limits on the pricing discretion of its members. In other situations, the market may be divided between firms with different cost structures. In the event that the firms with lowest cost structure establishes a market price, less efficient firms will be limited in their ability to capture quasi-rents from the public due to pressure to retain market share.

The enterprises that were incorporated for the purposes of building and operating the canals were

large. The engineering challenges before them required the organizational capacity to administer the provisioning of resources and labor that spanned whole regions. The firm was faced with planning challenges that exceeded the scope and capacity of small-scale proprietors. Consequently, most enterprises engaged in the canal business were cooperative ventures between private interests and the state. Private interests included merchants who wished to lower transport costs and grow the market for their trade, capitalists vested in navigation companies, as well as land speculators, who believed that improvements following canalization of waterways would increase the salable value of land held adjacent to the improvement. The state became the financial partner in the venture, providing institutional means for generating revenue through tax levies and bond issues, and remaining ultimately responsible for financing its ongoing maintenance. Notably, the Erie Canal, constructed between 1817 and 1825, was a landmark achievement of the cooperation between private interests and the State of New York toward the end of transforming economic relationships, such that corporations become a central site of social provisioning in the development of the American economy.



Figure 2. Workers operating a lock at Lockport, New York ca. 1839 along the Erie Canal. Engraving by W. H. Bartlett. Public Domain.

To better understand how large scale public works projects, like the Erie Canal, were so transformative consider the financial and organizational implications of its development. First, the Erie Canal was financed largely through the issuance of bonds. Such bonds were held as wealth by private citizens who became vested in the interests of the canal as a going concern. To remain a going concern, the canal required the establishment of rates sufficient to cover its costs as well as sufficient traffic to provide revenues to support its financial liabilities. As a result, those financially vested in the canal as bondholders were compelled to take an interest in the development of the economic space as a whole such that the canal remained viable, to include promotion of settlement in regions adjacent to the canal and its feeder lines. The grandness of the venture required a further elaboration of the large-scale, economic interdependence in which national markets became increasingly important. Second, given the extent of private, monied interests vested in the canal, an administrative structure was required in order to ensure the orderly operation of the canal and its finances. In the case of the Erie Canal, such an administrative framework was established by legislature and structured as a commission, whose leadership were drawn from the business communities affected by the existence of the canal.

As America's inland waterways were improved through lock and canal systems, the steamship became increasingly important to the social provisioning process. For some communities, navigation companies engaged in the steamer trade emerged as its first large, powerful corporations. For example, consider the Oregon Steam Navigation Company (OSN) whose business was centered in Portland, Oregon in the mid to late 19th century. While small compared to East Coast metropolises, Portland was the most important city on the West Coast after San Francisco, until the 1890s when Seattle was connected with the transcontinental rail system via the Northern Pacific. The OSN was Oregon's first, great monopoly whose **intangible value** provided the basis for wealth for many of the region's "founding fathers." The intangible value of the corporation lied in:

1. the the exclusive rights of way that the firm enjoyed in Portland
2. ownership and control over key portages along the Columbia River, and
3. access to the state legislature via the social networks of its owners and agents.

In 1866 efforts to extend a regional trunk of the nascent transcontinental railway system into Portland from the Sacramento Valley in California involved courting the owners of the Oregon Steamship Navigation Company (OSN) on the presumption that their political connections would ensure special acts of incorporation from the Oregon legislature favorable to the Californian interests. Initially, such owners attempted to block the advance of the railroad promoters for threat of the viability of the going business of the OSN. However, later that year, perhaps hedging against the threat of eventual encroachment on their business, members of the OSN leadership sought to secure the franchise for the Oregon section of the regional trunk of the transcontinental, thereby securing associated land grants and privileges from the state, which set off a race between two competing concerns to build a railroad through the Willamette Valley in Oregon that would connect Portland with a road from California. The ultimate objective lied in establishing legitimate claim to the exclusive charter, ostensibly by demonstrating commitment to actually build the road and fulfill the spirit of the charter, while undermining the legitimacy of the rival concern through the courts and in the public mind through influence of the press. The outcome was mutual insolvency between the two parties, who had each incorporated variants on the name Oregon Central Railroad.

In the late 1860s, the stagecoach magnate Ben Holladay resolved the conflict by purchasing control of one of the firms, persuading the legislature to reassign the charter to the concern he had acquired, and acquiring controlling stake in his rivals. Holladay reorganized his newly acquired railway assets under a new name, California & Oregon Railroad, together with some steamship ventures that he had established or acquired in order to provide a revenue source while he developed his railway aspirations. To finance this consolidation scheme Holladay floated a considerable amount of bonds to counterparties in Europe, who were falsely led to believe that the market in the Willamette Valley in Oregon was capable of supporting the revenue requirements of his steamship and railway concerns. The balance sheet of the Oregon and California Railroad relied upon lines of credit extended from speculative financiers to remain solvent.

The Panic of 1873 undermined the solvency of the Oregon and California, as Holladay did not earn enough through revenues from his properties to service his debt obligations without the ability to issue new debt to refinance old. During a financial panic, creditors tend to stop lending, preferring instead to hold more liquid assets on their balance sheet for fear of becoming insolvent themselves, resulting in a vicious cycle of asset value deflation. The response by Holladay's creditor's was to install an agent to take receivership of Holladay's properties and reorganize his properties on a sounder

basis. The surest way to achieve this end was to buy out Holladay's interests, thereby preventing him from exerting any control over the enterprises.

Over the next decade, Henry Villard, agent for German bondholders would proceed to establish new market governance institutions in the Pacific Northwest such that economic activity adjacent to the railroads would be harmonized or rationalized, with the ultimate end of servicing the debt obligations to foreign creditors. Villard's aspirations were grander than the German bondholders: Villard wanted to engage in thoroughgoing economic development and population growth. Villard wanted to build an empire on the Columbia River. A central problem facing Villard was maintaining control over the market for railroad traffic in the face of the encroaching Northern Pacific, the main northern line of the transcontinental system. The directors of the Northern Pacific wanted to run the line to Tacoma or Seattle, where the natural deepwater harbors would facilitate better access to international markets. Yet, Villard was vested in properties centered on Portland and its hinterlands. Villard organized a holding company in 1879, the Oregon Railway & Navigation Company, that consolidated his railroads and steamships in the Willamette with his acquisition of controlling stake in the OSN. Initially, Villard sought order in the market along the Columbia River via traffic sharing agreements with the directors of the Northern Pacific. However, the Northern Pacific was determined to build toward the Puget Sound. In order to protect the intangible value embodied in the corporations that he had come to control, he engineered a financial takeover of the Northern Pacific, through the use of another holding company, the Oregon and Transcontinental Company incorporated in 1881, which allowed Villard to pool funds to acquire control in the Northern Pacific. Once in control of the Northern Pacific, Villard connected the Northern Pacific with the Oregon Railway and Navigation Company, ensuring that any traffic destined for the Pacific would pass through Portland.

It is important to note that holding companies are essentially financial corporations chartered for the purpose of acquiring majority ownership in another corporation, with a minimum outlay of cash. Holding companies can be layered, such that several holding companies stand between a given set of capitalists and the actual underlying corporations that the holding companies are designed to facilitate control. This can be illustrated with a simple example: suppose that a simple majority of stock ownership is sufficient to control a railroad corporation. If there are 200 shares of stock outstanding in the corporation, then it would be necessary to acquire 101 shares to ensure control. Now, suppose that a holding company can be incorporated and capitalized by issuing 101 shares of stocks and assuming parity value between the two stocks. If the holding company is sufficiently capitalized and has access to credit, it is possible that by owning a simple majority in the holding company allows one to effectively control the first corporation if the holding company is able to acquire a simple majority of its stock, for half the ownership requirement (see Figure 5 below). With each layer of holding companies between the capitalist and the firm he wishes to control, the ownership stake required for control is halved. In this way, great corporations may be brought under control of a single individual with relatively little personal outlay involved. This has been a common theme in the area of corporate control since the second half of the 19th century, and continued well into the 20th century.

GLOSSARY

free and common socage

a system of land tenure in the American colonies which included perpetual ownership, transmissible to heirs, free disposal of property, alienability

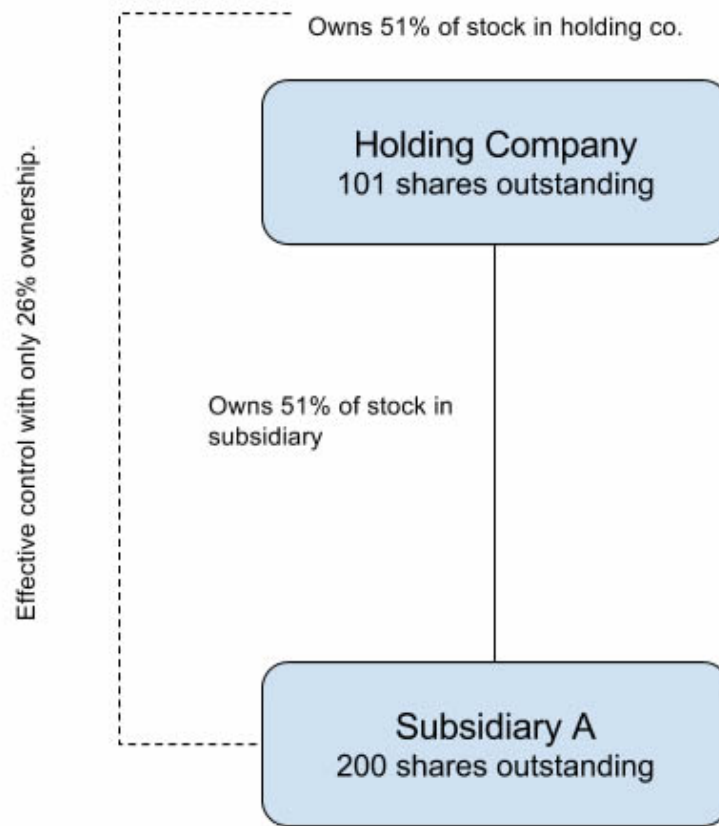


Figure 3.

intangible value

the value of a business enterprise over and above the value of its tangible assets

14.2 INDUSTRIALIZATION AND THE FACTORY

After the Civil War, the United States economy passed into an era of rapid industrialization. The term industrialization is used here to describe a process of development that reorganizes productive activities such that they conform to a machine logic, as opposed to a handicraft logic. That is, the economy becomes increasingly constituted by a series of interconnected machine processes. When we think of the United States undergoing industrialization we are really envisioning its economy becoming a large-scale machine, capable of churning out all of the goods required to sustain the community efficiently and to sustain the machine itself, through production of intermediate goods. When we think of the economy as industrializing we are actually considering a qualitative change in the economic system. To better understand what a qualitative change in the economic system entails, consider the following key features of an industrialized, machine process economy.

LARGE-SCALE TECHNOLOGIES THAT MAKE UP THE CORE OF THE ECONOMIC SYSTEM

Particular combinations of technology incorporated into the economic system were massive in scale. For example, consider again the case of the railroads. We have previously discussed why railroads required administration from the managerial bodies of corporate entities capable of organizing the business over the scale and scope of its activities. The technology itself was massive – establishment of steam powered with a rail and car transportation system over vast geographical areas was quite literally the greatest engineering marvel that humans had achieved to date. With the railroads came the telegraph, which created the possibility for transnational and interregional rapid communication. Other hallmark technologies of an industrializing 19th century American economy include: electric light and power, streetcars, improvements in steel production, and petroleum refining.

INTEGRATED CHAINS OF PRODUCTION THAT LINK MARKETS AND INDUSTRIES

The largeness of the business enterprise implies a high degree of interconnectedness between different stages in the production of the social product. To illustrate our point, consider the generation and provisioning of electric power. We can envision a simplified model of electric generation by breaking the whole production process into four stages:

1. Production of fuel source. In this example, we will assume we're talking about a coal fired power plant. This is a good choice, because coal served as the primary fuel source for electric generating during its advent in the United States. Coal is produced by mining and is subject to geological processes that have resulted in an uneven spatial allocation of coal in the earth's crust. Therefore, coal can only be mined in certain areas. The story of coal in America begins in Pennsylvania.
2. Energy conversion. The chemical energy embodied in coal must be converted into mechanical energy before it may be used to generate electricity. This conversion process is accomplished by burning the coal to transform water into steam, which creates the

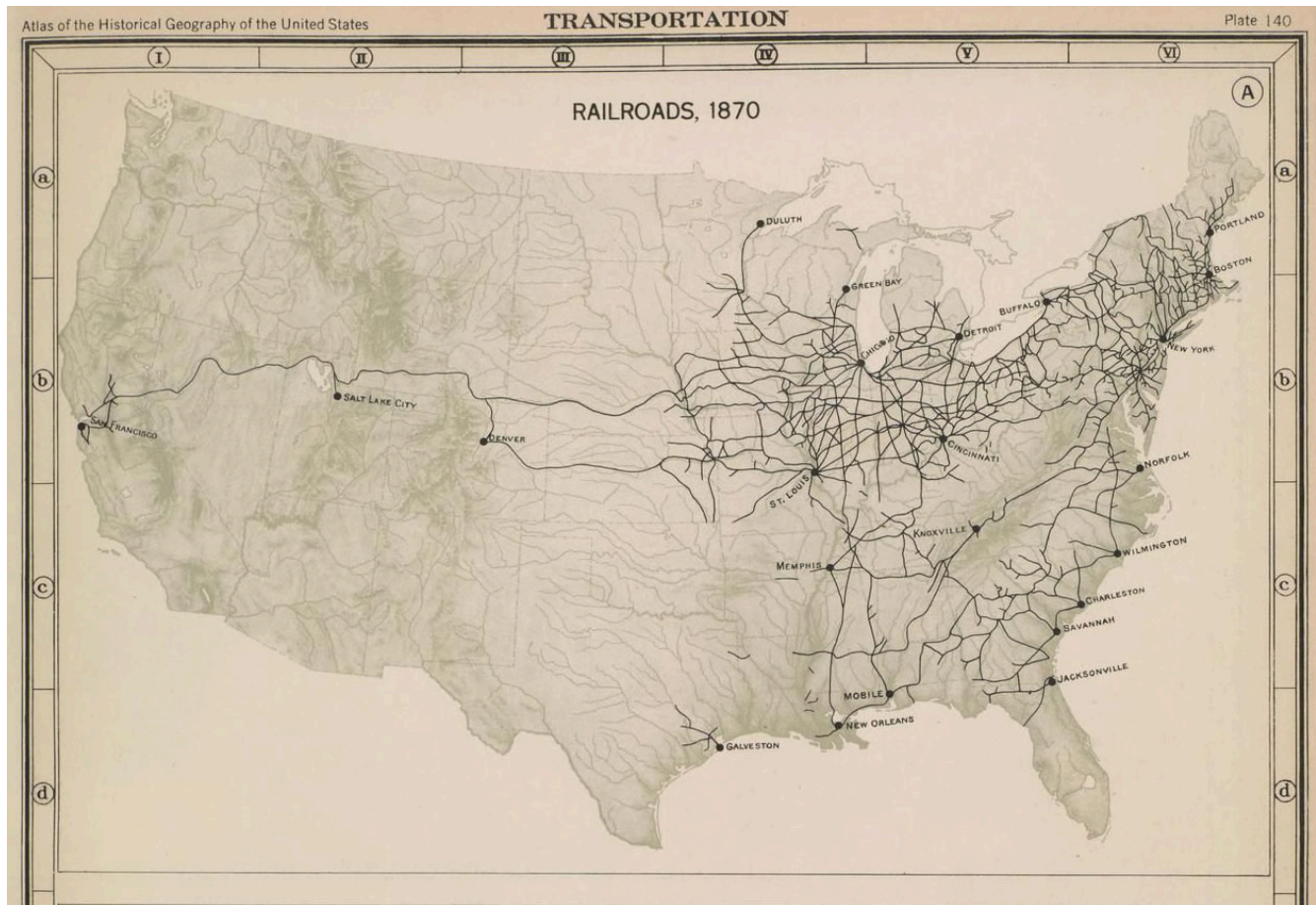


Figure 4. Atlas of the Historical Geography of the United States, Plate 140 by Charles O. Paullin. Curated by the Carnegie Institution of Science

necessary pressure to power an engine. The engine now embodies some of the energy that existed in the coal, less losses due to inefficiencies in the conversion process. The steam engine itself is the product of a set of interconnected production processes.

3. Electric Generation. The mechanical energy of the steam engine is then used to power an electric generator. Electric generators require steel, copper, and most importantly, a patent to produce.
4. Distribution. Once generated, the electricity is distributed to the end user via cables that conduct electricity. Edison's first system in New York used copper as its conductor.

If you think about each of these stages of production occurring as a production process that occupies a different space geographically, then you begin to see how something like the provisioning of electricity is really a network of interdependent production processes that establishes connections between different regions. Coal is mined in Pennsylvania, and then shipped to New York for use as fuel in the production of steam. A boiler and steam-powered electric generator is purchased from a firm in New York that is licensed to sell Edison patented systems. Edison's generators are produced in upstate New York and require a steady flow of steel from Pennsylvania and copper from Michigan for ongoing production. And so on. Every single economic activity is the product of an orchestra of interconnected production processes. To function efficiently those interdependencies must be managed so

that they they function as a whole machine. This is what we mean when we describe the economy as industrialized.

14.3 BIG BUSINESS AND ORGANIZED LABOR

As the business enterprise becomes large it operates on a scale of production that requires the labor of a proportionally large number of workers. The managers of the firm must ensure that they reliably and efficiently meet their production goals. To achieve this end, managers established methods for controlling their labor force so that rate of production remains within their full discretion. Introduction of mechanization into the production process serves as one historical example of the methods of controlling labor. Mechanization establishes the pace of production, by powering machine technology. For example, an assembly line sets the rate at which workers must labor. In this sense, workers on an assembly line must work at roughly the same pace, harmonizing their collected labor. By adjusting the speed of the machine, worker productivity may be adjusted across the board as desired, subject to physical limits of the labor force. Similarly, the machine process results in a “deskilling” process, which tends to homogenize the labor requirements for production. Deskilling transforms the labor process by removing the power that workers enjoy via their specialized knowledge of craft production. Machine production depends, to a much lesser degree, on the specialized knowledge of workers. Instead, this knowledge is transferred to the managerial class, who may possess knowledge of the machine and its technical requirements or retain the expertise of an engineer.

Loss of worker control over the production process results in the ability for the managerial class to retain a larger share of the monetary value of the output of the labor process. Managerial focus on cost efficiency and meeting its production targets, coupled with control over the labor process, resulted in dangerous and poor working conditions. The large business enterprise, by exercising its control over the labor process in order to exploit its workers, created the conditions for the emergence for an organized labor movement.

An exhaustive review of the labor history in the United States is beyond our reach, however, we may highlight some of the major themes and events that helped shape its development.

THE KNIGHTS OF LABOR

Organized in 1869 by tailors in Pennsylvania, the secret organization named the Noble Order of the Knights of Labor would serve as an organizational body for workers seeking collective action following the Panic of 1873. Coal miners helped drive membership growth in the Knights of Labor through the depression of the 1870s. By 1884 there were 70,000 members of the Knights of Labor. By 1886 this figure exploded to 700,000.

The Knights of Labor participated in strikes in order to achieve their primary goal of an 8 hour work-day. Most notably, the Knights of Labor won a struggle against Jay Gould over a dispute concerning the Wabash Railroad in 1885.

THE GREAT RAILWAY STRIKE OF 1877

Railroads figure prominently in the course of American economic history for they have left lasting and transformative imprints on the course of development. It is natural, then, to link a major moment in the history of the labor movement to consequences of railroad speculation and administration. The economic depression of the 1870's was global in nature and was caused largely by a financial crisis related to the value of railroad securities. The Panic of 1873 emerged from a financial system that had become increasingly fragile during the speculative episode in railroad securities, mostly bonds, following the Civil War. Jay Cooke, a major US banker during the latter part of the 19th century, was heavily vested in the market for Northern Pacific bonds to interests in Europe. During the 1870s Europeans were less keen to accept the risk associated with holding American railroad bonds. In turn, this led to the inability for Cooke to continue issuing debt and ultimately resulted in his firm becoming insolvent. The failure of Jay Cooke & Company set off a wave of bank failures and a closure of the New York Stock Exchange. The result of this financial panic was a nearly decade long depression.

As a result, railroads were under pressure to cut their labor costs. Notable railroads such as the Pennsylvania and Baltimore & Ohio Railroads had administered waves of wage cuts. Railroad workers organized themselves in the summer of 1877 and succeeded in bringing the railroad system in the eastern part of the United States to a grinding halt. Striking railroad workers were joined by sympathetic workers in other mining and manufacturing sectors, which resulted in fears on the part of the state and business interests that labor was growing into a coherent and powerful countervailing force. To quell the strikers, initially local militias were deployed. However, local militias were ineffective due to their propensity to defect and join the strikers, as they found themselves in common interest with their fellow members of the community. Consequently, National Guard and federal troops were brought into cities under the control of the strikers and were far more effective, because the soldiers were not particularly vested with the interests of the community and more likely to follow the orders of their commanding officers.

The strikes of 1877 serve as an important moment in the development of the political economy of the labor situation in the United States. Following the strikes, membership in labor organizations increased substantially. Likewise, more national guard units were established with an eye toward checking the power of labor going forward. In the realm of economic theory, the struggle between big business and big labor was taken as a serious matter of consideration. The development of neoclassical economics owes much of its theory of distribution to the international labor struggle that reached a fever pitch in the 1870s. For example, John Bates Clark, the father of the marginal product theory of distribution, was working out a theory of distribution that sought to move consideration of the distribution of the social product away from conflict between social classes, toward one based upon marginal contributions of each *factor* of production (land, labor, capital) in the production process. The result, is a theoretical system that finds fairness in an unequal distribution of income and wealth. Bates' theory was a direct response to the class conflict embodied in events like the Great Strikes of 1877.

While it is not possible to give the history of the labor movement in the United States a thorough treatment here, we can summarize some of central implications and consequences:

- Ongoing coherence and emphasize around the demand for an 8 hour working day.
- Development of systems of administration and organization capable of exerting a countervailing force against big business.



Figure 5. Maryland National Guard Sixth Regiment fighting to suppress strikers in Baltimore. Harper's Weekly, Journal of Civilization, Vol XXL, No. 1076, New York. Public domain.

- Formation of institutions capable of allowing workers to regain some of the control over production that was lost with the demise of handicraft production as a result of industrialization.

14.4 REGULATION OF BIG BUSINESS

A central problem facing the business enterprise is maintaining the orderly governance of the markets in which they operate. The going concern wishes to form stable expectations over the viability of its business through time. An industrialized economics system means that firms are embedded in networks of interdependent financial flows, which suggests that firms must find mechanisms to coordinate their operations and establish norms governing how other participants in the markets will behave, especially via their pricing policy. Formal and informal cartel systems were frequently established in an attempt to bring harmony to markets. Earlier we have referred to these arrangements as **restrictive trade practices**, due to their objective of avoiding price competition between firms over market share. Recall, the business enterprise must ensure that it is capable of administering a price to the market that embodies a sufficient markup that recovers their costs over some defined period. Prices must also provide a flow of income to its shareholders as profits, as well as a means of accumulating funds for investment purposes. If firms engage in a price war to capture a larger share of the market, they will undermine their revenue streams and have difficulty achieving their desired cash flows. The solution is to institute controls on individual firms so that they are properly incentivized to cooperate.

Market governance institutions are best viewed as efforts on the part of the business community to socialize the risks and rewards to individual firms. However, voluntary and privately organized market governance institutions were not terribly stable. There was always the risk that some participants would defect and undermine the cooperative agreement, resulting in a newfound round of ruinous competition.

A solution to the instability of private market governance structures was found in the federal government. The Interstate Commerce Commission (ICC) was established in 1887 in order to provide regulation over railroad rates, and emerged largely as a result of lobbying efforts by the railroads themselves. Government regulation over markets was not new to the American economic history. However, establishment of the ICC resulted in the federal government exercising its discretion over the development of economic affairs, by instituting a new form of market governance by regulatory commission. With a standing commission now responsible for authorizing proposed rate schedules the problem of coordinating and enforcing voluntary rate design policies solved by outsourcing this function to the federal government.

The Interstate Commerce Commission emerged as a result of two political processes. On one hand, agricultural interests organized through the Grange, Patrons of Husbandry, and the like, viewed railroads as monopolies that extracted unfair monetary flows from them by charging exorbitant rates, and administering differential rate schedules that had the effect of advantaging some ratepayers at the expense of others, and undermining the ability for some farmers to bring their goods to market. These agricultural interests tended to view their monopoly power as the primary concern, and sought to compel railroads to act as if they were competitive and thereby undermine their ability to cap-

ture rents. Businessmen in the railroad and finance industry sought regulation as a means to ensure coordination between competitors, but held different views about how such regulation should be structured. Ultimately, the railway interests embraced the ICC for it helped to quell anti-monopolist agitation while preserving private control and ownership of the railway system.

THE SHERMAN ACT

The 1890s mark a turning point in American economic history. Historians generally mark 1890 as the close of the frontier, as the transcontinental railways had linked the Pacific with the Atlantic and allowed for rapid and low cost transportation and communication. In the 1890s cities in the United States began electrifying. By the 1890s, big business had become a powerful and central organizing institution in the American economy. Passage of the Sherman Act in 1890 serves as an important landmark in the history of big business and provides a useful bookend for delineating a new phase in the American economic development.

The Sherman Act of 1890 sought to prohibit behavior that may be construed as acting in restraint of trade. The anti-trust act makes illegal cartels and conspiracies to engage in price fixing. Consequently, some mechanisms that firms relied upon previously to govern markets were no longer legal, such as pooling arrangements and cartels. While the Sherman Act was seldom applied in earnest as a means to restrain corporate power, it did encourage the practice of firms attempting to consolidate control through mergers and acquisitions of their competitors.

The first great merger wave occurs between 1895 and 1904 and represents the effect of the Sherman Act on the methods by which firms attempted to control markets in the wake of the Panic of 1893. Efforts by firms to capture more market share in order to spread their fixed costs more widely, led to ruinous price wars. Since previous forms of market governance had been largely prohibited by law in the Sherman Act, firms responded by consolidating in order to eliminate the costs associated with redundant capacity in light of the depressed economic conditions of the early 1890s. Between 1895 and 1904 roughly 300 firms disappeared on an annual average basis. From 1898 to 1902 this wave of mergers was particularly acute, with 1028 firms lost to consolidation in 1899 alone. The 1890s merger movement produced some of the more notable monopolies in American history, such as U.S. Steel, American Tobacco, Du Pont, and International Harvester.

The Sherman Act, when measured by its efficacy in preventing the consolidation of corporate control, was a total failure. However, it is important to note its foundational importance to subsequent processes of coevolution between the going concern as big business and the legal framework. Extensive revision of the laws governing corporate behavior would occur in the 20th century. Nevertheless, the Sherman Act serves as an excellent reference point for a turning point in the history of the corporation in American and closing our discussion of the history of big business.

14.5 CONCLUSION

This chapter has sought to provide a broad outline for understanding some of the more important themes that contextualize the big business in American history. The history is expansive, and so we have held a fairly narrow view of the going concern in a limited period of American history. The value in this approach, however, has been to focus on a foundational period in American economic history, the 19th century, and explore how the interaction between corporations under the control of members of the business community and legal frameworks established a course of development in which big business becomes a normal feature of capitalism in the United States. By tracing the corporation from its roots in colonial Virginia to the landmark Sherman Act in 1890, we have taken care to view the firm as a vehicle for institutional change and the locus of cumulative causation. Therefore, we walk away from this reading of American economic history with the understanding that thinking about the firm as a site of institutional capacity facilitates a more dynamic understanding of economic development. We see that power, control and influence matter in formulating a rich understanding of why the United States economy developed as it did, rather than following some other technical possibility.

CHAPTER 15. COSTS AND PRICES: THE EVIDENCE

INTRODUCTION TO COSTS AND PRICES

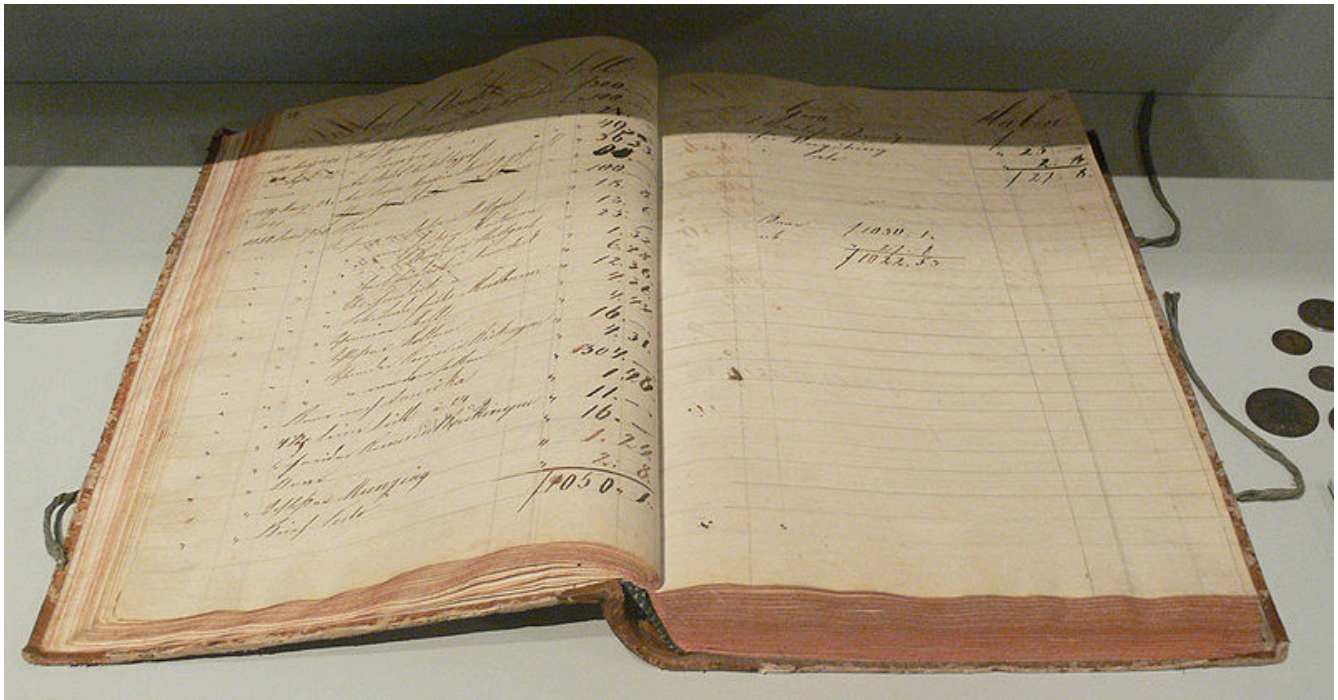


Figure 1. Ledger from a German general store, 1828. (Praefcke, Public Domain).

CHAPTER OBJECTIVES

Introduction to Prices and Costs

In this chapter, you will learn about:

- Testing the Neoclassical Theory of the Firm
- Costing and Pricing: A Heterodox Alternative

15.1 TESTING THE NEOCLASSICAL THEORY OF THE FIRM

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Construct testable hypotheses from the neoclassical theory of the firm
- Test these hypotheses against the given data

Let's review the basic properties of the neoclassical theory of the firm as given in previous chapters ("Cost and Industry Structure," "Perfect Competition," and "Monopoly"). To keep things simple, we'll focus on the short run behavior of a single firm, and we'll look to the similarities of this behavior regardless of whether the market is competitive or monopolistic. First, the firm is portrayed as a functional relationship between inputs (factors) and outputs (products). Because we're considering the short run, there are variable and fixed inputs, and therefore variable and fixed costs.

Second, the firm accepts the demand for its product as given and determines how much it should produce—its quantity of output (Q)—so that its profits are maximized where the cost of producing the last unit equals the revenue from selling it—that is, where marginal cost (MC) equals marginal revenue (MR). We assume that the firm is subject to diminishing marginal returns. Therefore marginal costs will always eventually rise to meet marginal revenue and will cause average total cost (ATC) to eventually rise (giving it its 'U' shape). Hence, the very simple **Hypothesis 1**: over some short-run period, managers are aware of their firms' marginal costs associated with increasing (or decreasing) their production, and they generally find that marginal cost increases over a relevant range of production.

Now, consider the conditions presented in the chapters "Cost and Industry Structure" and "Perfect Competition":

1. If $MC < ATC$, then ATC is decreasing (marginal cost is 'dragging down' average total cost the same way a low test score would drag down your course grade). Therefore, if ATC is decreasing, it must be the case that $MC < ATC$.
2. Profits = $Q(P - ATC)$, where P represents price. Therefore, firms make losses when $P < ATC$, break even when $P = ATC$, and make profits when $P > ATC$.
3. Profit maximizing firms under perfect competition produce where $MC = MR = P$.

We can read these three conditions in reverse to establish a necessary condition for firms to avoid making a loss. Plugging (3) into (2), it should be clear that, since the firm will produce where $P = MC$, the profit equation can be restated as Profits = $Q(MC - ATC)$ under perfect competition. Hence, if the firm is going to break even or make a profit then MC must be $\geq ATC$. Taking this back to (1),

then, for MC to be equal to or greater than ATC, ATC must either be constant or increasing at some reasonable level of output.

Hence, **Hypothesis 2**: if firms maximize profits by producing a quantity where marginal cost equals marginal revenue, then under competitive conditions firms must normally be producing in a situation in which average total cost is either constant or increasing. Really, this is just an extension of hypothesis one which allows us to test our theory by looking at average total costs rather than marginal costs.

TESTING THE THEORY

It is generally held that a developed science should be capable of expressing ideas about how the world works in hypotheses, and that these hypotheses should be compared with actual observations to determine how well the theory explains what we see in the real world. Fortunately (and perhaps surprisingly), a considerable number of studies on both costs and prices have been produced going back over 100 years. Saving the reader a laborious trek through all of them, we'll just look at a couple to test the validity of our hypotheses.

In 1998 Alan Blinder and colleagues published a book, *Asking About Prices*, which was not actually concerned with microeconomics (it was concerned with testing theories of macroeconomic price 'stickiness'), but just the same can help us test our hypotheses [1]. Their approach was somewhat uncommon to mainstream economics at the end of the 20th century: they simply asked business people about how they did things. For our purposes, their work (see p. 103) sheds light on hypothesis 1. Asking business executives about the shape of their marginal cost curve was apparently a bit tricky. Evidently, most do not normally consider the concept of marginal cost, so the surveyors had to repeat, rephrase, and further explain the question before a response could be given.

And the responses were not encouraging for the neoclassical theory of the firm. Firms representing only 11% of GDP were believed to have upward sloping marginal costs curves. Nearly half reported constant marginal costs, and, almost shockingly, 41% reported decreasing marginal costs (suggesting increasing returns). Evidently, the 'law' of diminishing marginal returns is more like a local custom, appropriate to only a small part (about 11%) of the US economy.

In fact, similar findings had been reported over 50 years ago. In an article published in the *American Economic Review* in 1952, Wilford Eitemen and Glenn Guthrie reported their findings from a survey of manufacturing companies across the US. Specifically, they found that only 5% of the goods produced conformed to the standard 'U'-shaped depiction of average total costs. The majority (60%) of products were reported as having average total cost curves which simply declined as output increased up to maximum capacity. (Almost all remaining products were reported as similar but with a slight rise in average total costs just before capacity was reached.)

DESIGNING THE AVERAGE TOTAL COST CURVE

Eitemen offered a simple, intuitive reason why we should expect that most average total cost curves would always decline with increased output. One way to look at the matter is in terms of the percentage of designed maximum capacity output at which you would want your production process to operate most efficiently. To give an example, imagine you're an engineer and you have been tasked with designing a machine (or a factory, or the layout of an office) that can produce some maximum number of units in a given day. In designing the machine, you find that it will have to operate more

or less efficiently depending on how heavily it's being used. Would you design it to operate most efficiently when it's only running at 20% capacity? Maybe 60%? Perhaps 99%?

Most would say that a production process should work best when it's operating somewhere near the level of output it was designed to be capable of producing—otherwise, why should it be made to produce so much output in the first place? Translate 'work best' into 'operate at lowest average total cost' and the implication should be clear: a 'U'-shaped average total cost curve indicates a production process that was designed to be capable of producing up to some maximum amount inefficiently. In contrast, a declining average total cost curve indicates that the production process was made to work best when it's producing as much as it was designed to be capable of producing—the most efficient, lowest average total cost point being at or near 100% of maximum capacity output.

Returning to the three postulates we used to develop hypothesis 2, Eitemen and Guthrie's findings indicate a clear departure from the standard model of firm behavior. If most firms believe their average total costs decline as output increases, then mathematically their marginal cost curves must be below ('dragging down') their average total costs. (Perhaps in many cases, the firm's marginal costs are constant as Blinder et al. indicate is common throughout the economy.) Now, if marginal costs are always less than average total costs, then there are only two possible conclusions to be drawn:

1. Firms produce a quantity where $MC = MR$. But, since MC is always $< ATC$, and under perfect competition $\text{profits} = Q(MC - ATC)$, then firms in perfectly competitive markets will always make a loss. This doesn't seem like an attractive conclusion—surely we can find firms in competitive markets that manage to make a profit. So it's worth considering the alternative:
2. Firms do not produce a quantity where $MC = MR$. That is, under the cost structure that most firms report, the standard profit maximization behavior postulated in the neoclassical theory of the firm is not only unrealistic, it's impossible.

Clearly, the implications of these findings do not bode well for the neoclassical theory of the firm presented in earlier chapters. The preponderance of evidence may not wholly refute the neoclassical model, but it does expose severe limitations. In the breakout box below, two additional hypotheses are examined, giving greater reason to abandon the neoclassical theory of the firm. The section that follows then introduces a few basic concepts from an alternative, heterodox viewpoint.

TWO MORE HYPOTHESES

From the neoclassical perspective, we can imagine a firm incrementally increasing its, say, weekly output, and continuing to do so as long as its marginal revenue is greater than its marginal cost. Once the additional revenue of one additional unit of output just covers the cost of that unit's production, the firm settles on that level of output—profit maximizing equilibrium has been reached where $MC = MR$. Assuming that marginal cost must be positive (most workers won't pay their boss to let them work more hours, but perhaps there are exceptions), then marginal revenue must also be positive.

Now, refer back to chapter "Elasticity," specifically the section asking "Does Raising Price Bring in More Revenue." There you learned that a firm that is able to set its own price can increase revenues by cutting prices if demand is elastic. If on the other hand demand is inelastic, cutting your price will increase your sales (that's the law of demand after all), but the lower price will weigh heavier on your revenues than the gained sales, and your revenues will actually decline. Mathematically, this works out to a simple rule: if demand is elastic then selling more will increase revenues—which is to say that marginal

revenue is positive when the elasticity of demand > 1 (and MR is zero when the elasticity of demand $= 1$, and negative when elasticity of demand is < 1). Hence, **hypothesis 3**: If firms produce where $MC = MR$, then MR must be positive, and therefore demand must be elastic.

Blinder et al. (1998 p. 99) found some support for this—but not much. While noting many businesses do not actually calculate or even think along these lines, the authors found that, among those that could give an answer, 84% believed demand to be inelastic (and, by implication, negative revenues). In fact, 41% believed that the elasticity of demand was zero—that is, a price cut would not increase sales at all.

Now consider a more general issue. In chapter “Supply and Demand” you were introduced to the centerpiece of neoclassical economics: the market model. The chief argument of that model is that prices in the market will adjust to bring supply and demand into an equilibrium in which the quantity people wish to sell is equal to the quantity people wish to purchase—that is, the price mechanism acts to clear markets. This model fits into the broader narrative of neoclassical thought which holds that capitalist economies are organized by price adjustments in the complex network of markets that make up these economies. Therefore, **Hypothesis 4**: unless the world changes very seldom or very slowly, leaving most markets in their previously established equilibria for long stretches of time, prices should be changing rapidly with frequent bidding, as in an auction market.



Figure 2 – Left: Auctioneer does brisk business at the Hickman Saturday Auction, 15 miles south of Lincoln, NE. (O’Rear, Public Domain). Right: Stock brokers on the New York Stock Exchange, 1963. (US News & World Report, Public Domain). Price bidding is essential to auction markets like the ones depicted above. But, of the things you’ve purchased in the last month, for how many of them did you bid—or even negotiate—a price?

By now you could probably predict what the evidence was going to tell us about this important part of standard neoclassical theory: throughout most of the modern US economy, prices change relatively infrequently. Blinder et al. (1998, p. 84) find that 49% of responding firms change their prices no more than once in a year or even longer. Indeed, only 22% reprice daily, weekly, or monthly. The auction market model at the core of neoclassical economics exists, indeed; but, it would appear that it lies somewhere between minor and insignificant in today’s economy.

The evidence reviewed above suggests that basic ideas from neoclassical microeconomics about how businesses, markets, and capitalist economies function have, at best, very limited application to the real world. While the evidence doesn't necessarily deliver a fatal blow to the neoclassical theory of the firm, it should be disconcerting to find so many businesses viewing their world and their work in a fundamentally contradictory way. Likewise, the prevalence of yearly price changes doesn't invalidate the auction model of markets you learned in an earlier chapter; but it does raise questions as to its usefulness, and to whether an alternative model should take priority in our understanding of modern capitalist economies.

15.2 COSTING AND PRICING: A HETERODOX ALTERNATIVE

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the concepts of costing, depreciation, going concerns, and pricing
- Identify the basic pricing methods of full cost pricing and target rate of return pricing
- Calculate prices according to these methods
- Discuss the implications of planning, costing, and pricing for the existence of a supply curve

The empirical findings presented in the previous section suggest that our standard models of how firms behave and what determines prices are not appropriate for understanding today's economies. Choosing the profit maximizing level of output does not appear to be relevant to firm behavior in the short run. Likewise, frequent price adjustments 'in the market' are not characteristic of most real-world markets, in which prices are clearly determined by producers and maintained over relatively long periods of time. All of this suggests that a deeper look into the actual nature of firms' costs and the actual manner in which prices are determined is necessary. Fortunately, ample information about these processes is available, and it comes from the costing practices of accountants and the pricing practices of management.

Costing is the process of estimating the costs of production before production actually takes place—and, hence, before the actual costs of production are known with certainty. To do this, of course, it would be necessary to have some idea of how much output the business will be producing and the direct and indirect expenses that will be involved at that level of production. Making such calculations may involve a simple, educated guess or a sophisticated process of research, experiment, and forecast. What is important, from a theoretical standpoint, is that it is a fundamentally uncertain task which takes place before exchanges occur in the market. This view is consistent with what was suggested above, that firms plan their production processes before, a topic discussed further in chapter "The Megacorp."

DEPRECIATION AND THE GOING CONCERN

A complete review of cost accounting isn't necessary here, but one particular type of cost, depreciation, is of particular historical and conceptual significance. **Depreciation** is a way of accounting for the expense of an asset—say, a machine press—over the life of the asset. For instance, suppose your machine shop purchases a press (a machine that does exactly what it sounds like it does) for \$20,000.

If you expect that the press will be in use for the next 10 years you might account for a depreciation expense of \$2,000 per year for the next decade.

To understand the significance of depreciation, consider how businesses usually calculated income before the late 1800s. Before then, business enterprise was often treated as a terminal venture, having a clearly defined beginning and end date. Investors would pool money to start a business, purchase materials and capital (say, a ship and local goods to be traded abroad), and hire workers (the ship's crew). At some predetermined date (perhaps when the ship returned to port) the business would be liquidated: its crew would be paid and any remaining assets would be sold off. The resulting profit to the investors would, in essence, simply be whatever money was leftover. In this approach, what was the relationship between the productive asset (the ship) and profit?

Profits were simply decreased by the cost of the ship, but increased by whatever price could be fetched by selling the ship at the end of the venture.

But, accountants in the late 1800s asked, is that an appropriate way to think of, say, a railroad company? Will a railroad lay thousands of miles of line, run trains across it for some predetermined number of years, and then pull up the line to be sold for scrap metal? Clearly not. Instead, fixed assets like a railroad's line, or an airline's planes, or a law firm's office building are depreciated over the course of their useful life. Conceptually, this is simply the recognition of the role those assets play in allowing these businesses to generate an income into the foreseeable future. Most importantly, this accounting method reflects the fact that these businesses are not treated as terminal ventures. Rather, they, as nearly all businesses today, are considered **going concerns**: organizations which are expected to continue to exist into the foreseeable future.

PRICES FROM PRICING

Similar to costing, pricing refers to procedures businesses use to determine, beforehand, the price at which they will sell their product once production is up and running and sales can be made. While modern pricing procedures can be complex and will vary widely across different businesses and industries, two basic methods should be understood: full cost pricing and target rate of return pricing. Both are instances of markup (or cost-plus) pricing: setting the price of a business enterprise's product by adding some dollar amount over and above average costs of production.

Full cost pricing (sometimes called normal cost pricing) is the simpler of the two methods. It can be written as

$$P = (ATC)(1 + r)$$

Where:

P is the price at which the business plans to sell its product

ATC is the average total (or per-unit) cost determined in the costing process

r is the predetermined markup

Target rate of return pricing is similar, but a bit more complicated. Here the price is being set, not to

achieve a particular percentage profit above costs, but to earn a desired return on the money invested into the business. The formula can be written as

$$P = ATC + \frac{ROI(IC)}{Q}$$

Where:

ROI is desired return on invested capital

IC is invested capital—that is, money invested into producing the product

Q is the expected quantity of output sold

CALCULATING PRICES

To illustrate both approaches, consider a business that invests \$10 million into a plant designed to manufacture inexpensive steak knives. It expects that over some relevant period it will be able to produce and sell 2 million knives; and, at that level of production, its per-unit costs will be \$1.80 per knife. The calculated prices using our two pricing procedures are given below (assuming that in the first case the desired markup is 10% (or 0.1), and in the second the desired return on invested capital is also 10%).

Full cost price: $P = (1.80)(1 + 0.1) = \$1.98$

Target rate of return price:
$$P = (1.80) + \frac{(0.1)(10,000,000)}{2,000,000} = \$2.30$$

Notice also that, even though the markup and desired return on invested capital are both 0.1 (10%), the resulting markups and hence the prices are not the same. This is because, although both procedures are essentially marking the price up over costs, the treatment of the costs being marked up are different.

It is worth reflecting on the significance of these insights into cost accounting and markup pricing, as they represent important general concepts in heterodox economics which are usually neglected in standard neoclassical theory. First, they suggest that business enterprises are making decisions before anything is even produced, let alone 'brought to market'. In particular, pricing practices (and the intended quantity of output and corresponding cost estimates on which pricing is based) are a component of the planning process which takes place within the business enterprise. Contrary to the axiom that firms cannot recover fixed costs in the short run and therefore should ignore them in making short run decisions, it is long run planning driving short run behavior that is most important for understanding what determines prices.

Second, to acknowledge costing and pricing as it actually occurs is to acknowledge that the future is fundamentally unknowable. While standard (neoclassical) models assume that firms know their production costs and, typically, also the amount they can sell and the resulting revenues they can expect to take in, actual firms face uncertainty in how their plans will work out. A particular implication of

this reality: since firms set prices based on estimated average total costs at an expected level of output, a change in the actual quantity of production/sales is unlikely to affect the predetermined price. This suggests that price and quantity supplied are determined completely separately, which in turn means that there is no such thing as a supply curve.

THE ILLUSORY SUPPLY CURVE

Recall from the basics of the neoclassical market model that a supply schedule (and its corresponding supply curve) simply shows the relationship between how much a firm (or firms) would be willing to supply at various market prices. That is, supply simply refers to the functional relationship between quantity supplied and the market price, with the market price determining the quantity supplied. If, however, the two are determined separately then there's no way around the implication: quantity supplied is not functionally related to the market price—that is, there is no supply curve.

The astute reader may have already realized the impossibility of supply curves under certain conditions from the failed hypotheses discussed earlier in this chapter. Referring back to chapter “Perfect Competition,” specifically the section titled “Marginal Cost and the Firm’s Supply Curve,” you’ll recall that a firm’s marginal cost curve (above minimum average variable cost) is its supply curve. (This is because quantity supplied is determined where $MC = MR$ and, under competitive conditions, $MR = P$. Hence, quantity supplied is determined by $P = MC$.) Now, as was shown with the test of hypothesis 2 above, firms simply couldn’t determine their quantity supplied this way—at least not under competitive conditions and having the empirically typical average total cost curves. This, of course, means that the neoclassical theory of supply must be rejected for these cases.

This doesn’t mean that the basic ideas of supply—higher prices leading to higher output and vice versa, for instance—are completely absent in the real world. Some industries—particularly, those related to mining and agriculture—do in fact see diminishing returns. In these (albeit limited) parts of modern economies upward-sloping supply curves may be found. However, as our examination of the cost structures of actual firms suggested earlier in this chapter, this relegates what is considered the normal case in neoclassical economics to a special—and pretty rare—case.

Third and finally, a review of the evidence and history of actual businesses reveals an anachronism within the neoclassical theory of the firm. As you learned in chapter “Perfect Competition,” the firm chooses the most profitable line of business (and appropriate production technique) in the long run, and the profit maximizing quantity to produce in the short run. If, in the short run, the firm is making a loss it will choose to shut down (if its fixed cost losses would be lower than the losses on continuing production). In an abstract, but important way the business enterprise this theory is describing is a terminal venture. Yet, beyond the halls and offices of economics departments, firms are generally seen as going concerns. This is reflected, for instance, in the accounting practices discussed above, as well as in the relationships firms maintain with customers. Blinder et al. (1998, pp. 96-7) found that 85% of all sales in the economy are made to regular customers whom the business expects to sell to in the future. In manufacturing and wholesale trade that number is over 90%.

As will be explored in more depth in the chapter “The Megacorp,” the idea that businesses are organized and run as going concerns is a significant theoretical innovation over the standard neoclassical theory of the firm. For now, we only need to consider what it means for prices. The role of the price mechanism—the ‘invisible hand of the market’—in neoclassical economics cannot be overstated. It is the process by which self-interested people (consumers, workers, entrepreneurs, landlords, and all the rest) are brought together in exchange for their mutual benefit. It is the mechanism that allows econ-

omists to believe in a (potentially) optimal equilibrium state—in an individual market, and in a capitalist economy as a whole.

In contrast, what is being argued in this section is that prices—at least those prices not actually determined through an auction—are set by businesses themselves as part of their planning process. The reader may have noticed that in the markup pricing introduced above a glaring question was ignored: namely, what determines the markup? A succinct, if incomplete answer can now be given: if the firm is to be a going concern, the markup, as well as the procedures that determined costs, will reflect the needs of the firm to continue to do business into the foreseeable future. For most firms there will also be plans to grow. Hence, from this view, prices are not exchange-based, market clearing values at all. They are, rather, reproduction prices—allowing the firm to reproduce itself through time—and, typically, also growth prices—ensuring the firm brings in the earnings necessary to expand. To use a now-familiar term, the vast majority of the prices we see in actual capitalist economies today might best be called **going concern prices**.

15.3 COMPARING NEOCLASSICAL AND HETERODOX THEORY

For quick reference, a table contrasting general views from neoclassical and heterodox economics on firms, costs, and prices is provided below. The student should be aware that we are painting in broad strokes here. Both neoclassical and heterodox economics include a variety of sophisticated models and theories, some of which would contradict the characterizations made below. However, as an introduction to microeconomics this table should suffice as a reflection of the core positions and standard ideas within these two camps.

	Neoclassical	Heterodox
Prices are determined in...	the market	the planning processes of the business enterprise
Prices function to...	clear markets	ensure the reproduction and growth of business enterprises
As production increases, average total costs...	first decline to some minimum (or efficient) point and then start to rise again (that is, ATC is 'U'-shaped)	typically decline up to (or very near) the maximum output at which the production process was designed to operate
Marginal costs...	eventually rise with higher levels of production due to the law of diminishing returns	are not generally contemplated by business executives and are not relevant to microeconomic theory, in part because firms don't (and couldn't) typically produce according marginal costs and revenues.
Price and quantity supplied...	are functionally related—that is, when price changes quantity will generally change	are, in most instances, determined completely separately
Firms are viewed as...	terminal ventures, always and everywhere maximizing profits	going concerns, organized and operated to (try to) stay in business into the foreseeable future
Firms make short-run decisions...	ignoring fixed costs, because by definition these cannot be changed	typically in accordance with the long-run strategies determined through their long-run planning process

CHAPTER 16. THE MEGACORP

INTRODUCTION TO THE MEGACORP

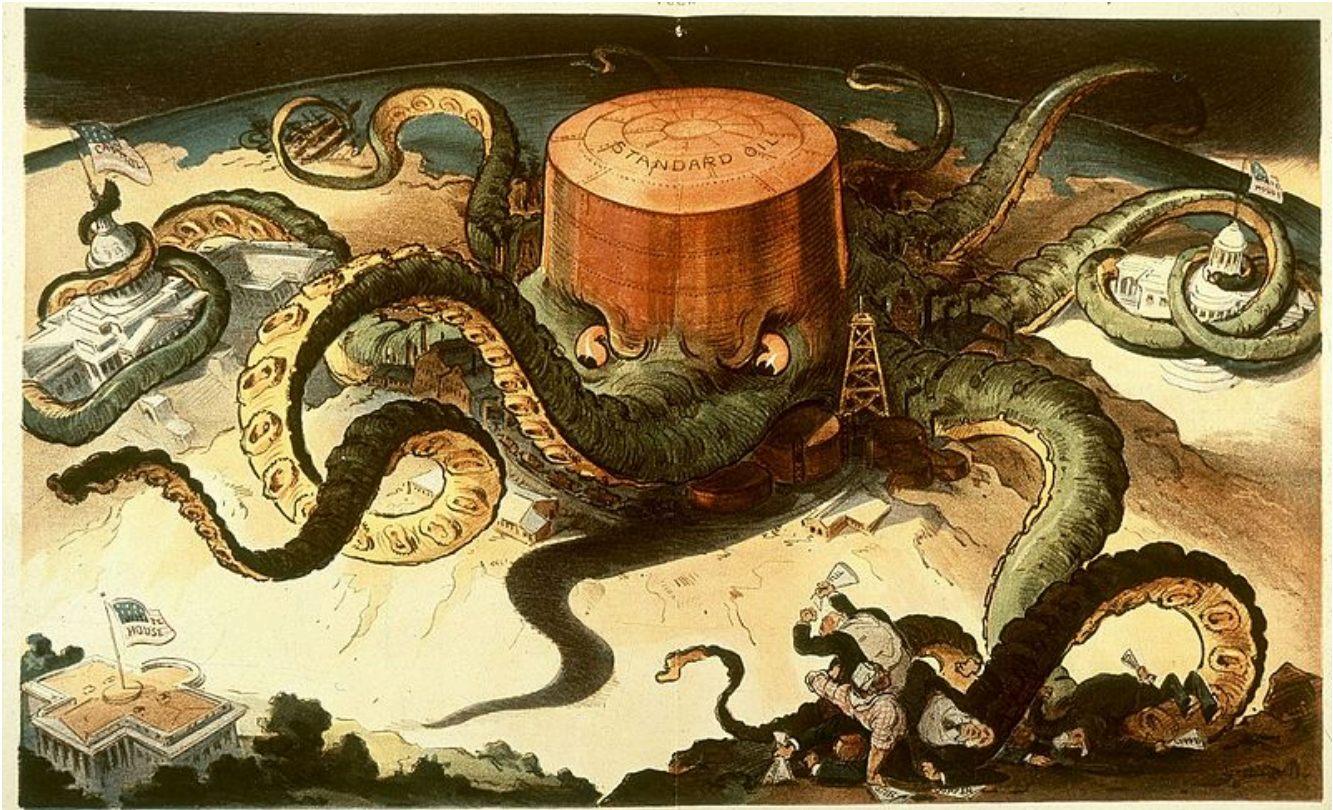


Figure 1. The power of the Standard Oil trust. Udo Keppler's political cartoon depicting the corporation which, circa 1900, had effectively monopolized petroleum distribution. The tentacles of this gargantuan octopus represented the power Standard Oil exercised over other industries (copper, steel, and shipping) as well as politics. (Wikimedia, Public Domain)

CHAPTER OBJECTIVES

Introduction

In this chapter, you will learn about:

- The imperatives of technology
- Business models, plural: Aims and methods of the megacorp
- Stabilizing unstable markets

- A case in point: The high prices of college textbooks

The importance of businesses in modern capitalist economies cannot be overstated. They are the local barber shops and factories, as well as the multinational oil giants and online retail enterprises. We do business with them when we purchase nearly everything we buy as well as when we set out to earn a living. While economists prefer simply to say ‘firm’, everyday conversations will often include terms like ‘big business’, ‘corporations’, ‘entrepreneurs’, and ‘job creators’. They may reference particular firms, known by name to everyone, or commonly understood terms for whole industries—for instance, ‘the media’ or ‘Wall Street’. Certainly, businesses are an important and complex part of the modern world—not just the economy, but society more broadly. They have been applauded and derided, praised and demonized since before the Sons of Liberty cast crates of tea belonging to the East India Company, a large British corporation, into Boston Harbor in 1773.

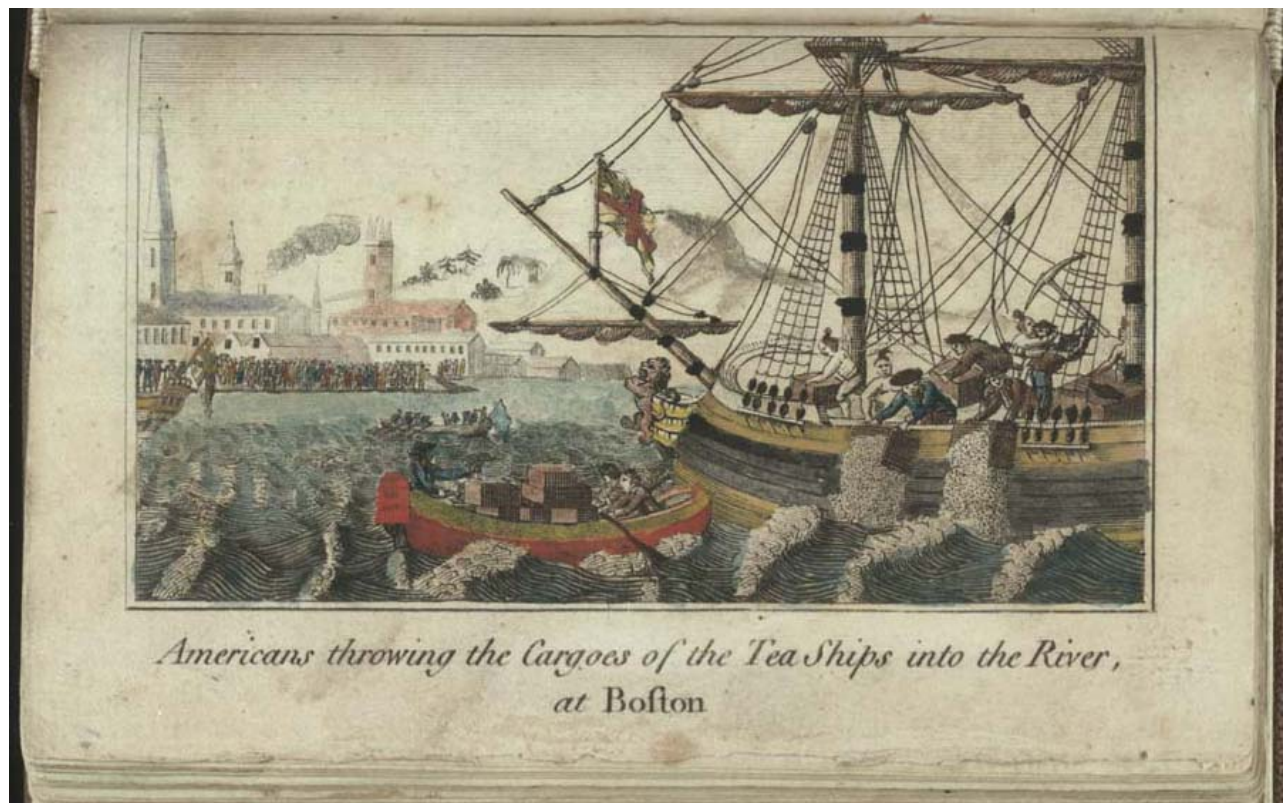


Figure 2. “Boston Tea Party,” by W.D. Cooper, *The History of North America*. London: E. Newbury, 1789. Engraving. Plate opposite p. 58. Rare Book and Special Collections Division, Library of Congress (40). (Wikimedia, Public Domain)

Yet, neoclassical economics treats ‘the firm’ with typical abstraction: it is simply a production function, translated into a set of cost curves, with a singular objective to maximize profits. That model of the firm, almost regardless of the actual business or businesses it might represent, has a ‘U’-shaped average total cost curve, and increasing marginal costs due to diminishing marginal returns (see chapter “Cost and Industry Structure”). It pursues, always and everywhere, the greatest possible profits by deciding what, how, and how much to produce, taking technological possibilities as given. As you

have seen in previous chapters, the market structure—and therefore the outcome—may differ from industry to industry, location to location, and so forth, but the firm remains essentially the same.

Many heterodox economists will note that abstraction is not, in itself, a bad thing. A complex reality has to be simplified in our understanding of it, if we have any hope of understanding it at all. But, it is important to get it right—to simplify reality without departing too far from it. The neoclassical depiction of the firm is not only fundamentally problematic, heterodox economists would argue (see chapter “Costs and Prices: The Evidence”), it is part of a view of capitalism that may no longer be appropriate to our modern economy. Specifically, the plot of the mainstream economics narrative casts the market as the star of the show—and, at least where that market is competitive, really consumers play the roles of king and queen. In contrast, heterodox economists tell a story in which modern capitalist economies have largely replaced consumer sovereignty with producer sovereignty.

In this alternative story, the true importance of the producer—specifically the large, economically, politically, and culturally influential corporation—is acknowledged. The business enterprise (a term many heterodox economists prefer to ‘firm’) is a socially recognized, legitimate organization. It has certain requirements if it is to be maintained as a coherent organization. And, especially in its ‘big business’ form—what Post Keynesian economist Alfred Eichner termed the Megacorp—those requirements take priority in the way modern capitalist economies are structured.

DOES MONOPOLY POWER CAPTURE THE REAL BREADTH OF CORPORATE POWER?

The starting point and focus of most of neoclassical theory has traditionally been voluntary exchange in a competitive market—that is, individuals choosing to trade one good for another, and in doing so making each other better off. This has colored the notion of power within this paradigm, limiting it somewhat to a failure to meet the standards of the competitive ideal. Such a narrow treatment of power, in the realm of big business, is evident in the monopoly model presented in a previous chapter. There, the reader will notice, the notion of power is explicitly treated; but the focus is on the ability to raise prices and restrict output compared to the outcome that would exist under perfect competition. Treatment of barriers to entry is secondary, showing the practical means by which business enterprises may achieve higher prices through lower competition. (Likewise, monopolistic competition and oligopoly are typically defined as some mix of the two extremes of perfect competition and monopoly, rather than fundamentally different situations.) The quantity, moreover, is still determined by a comparison of marginal costs and revenues, and price is still set with reference to demand—the same basic tenets of the model of perfect competition remain in tact.

Yet, consider the forms of power which the neoclassical abstraction sets aside. The power to dictate both prices and quantities, for instance through the billions of dollars of advertising spent each year in the US, is ignored—or, at best, treated as a special case. So, too, is the power to create new technologies, as well as to limit access to those technologies through things like digital rights management. Likewise, the political and cultural influences of business are almost completely neglected in an analysis that focuses only on prices and quantities.

If the reader is willing to accept this alternative story, setting big business at the center of the economic landscape, as a plausible way to look at how our economy works, a host of new questions—and new answers—will open up. The following sections will explore some of those questions while introducing a number of important concepts and theories from heterodox economics. To illustrate these ideas, the chapter will conclude with a look at the college textbook industry and a question you’ve probably asked yourself: why do textbooks costs so much?

16.1 THE IMPERATIVES OF TECHNOLOGY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify the important consequences of technological change for the organization of business
- Explain the implications for the relationship between business enterprises and market structure

In chapter “Costs and Prices” you learned (or will learn) that modern businesses are treated as going concerns—that is, organizations which are expected to continue to exist into the foreseeable future.

This idea, reflected in accounting practices developed in the second half of the 19th century, is central to how heterodox economists understand the nature of the modern business enterprise. Its historical roots are explored in more detail in chapter “The Rise of the Big Business,” but here we want to focus on a particular cause, technological advance, and its implications.

In his classic, *The New Industrial State* (1967), economist John Kenneth Galbraith compared the Ford Motor Company in the year of its founding, 1903, to the same business at the introduction of the first Mustang in 1964. In 1903 the company employed around 125 people, worked with approximately \$150,000 in capital, and required only months of negotiations, preparation, and production to bring an automobile to the market. It would be an understatement to say that, six decades later, the process was more involved. The Mustang was the result of several years of preparation, millions of dollars on engineering and ‘styling’, and tens of millions on tooling for production. At the time, Ford employed over 300,000 people.

Galbraith argued that the essential cause of this massive increase in preparation, capital, and labor was the tremendous advance in **technology** (which he defined as “the application of scientific or other organized knowledge to practical tasks”). Increasingly sophisticated technology requires specialization—of machinery and the materials it will work on, as well as of workers. All of which, of course, requires more capital (that is, money). It also means that increasingly specialized workers, machines, and so on must be more carefully managed; and that a commitment to produce something must be made much further in advance of actually having something to sell. As Galbraith explained, the Dodge brothers’ machine shop, responsible for machining the engine and chassis of the original Ford automobiles, could have hypothetically been asked to accommodate significant changes in the car’s design with only a few hours delay. The factories which built the Mustang, on the other hand, were effectively locked into making Mustangs for well over a year.

Finally, and most importantly, there is the technological imperative of planning. As Galbraith argued, the large financial and material commitments necessary to design a product and prepare for its pro-



Figure 3. A 1964 Ford Mustang in Wimbledon White, photographed at the 50th anniversary of the Mustang at the Charlotte Motor Speedway. (Source: Sicnag, Wikimedia, CC-BY-2.0)

duction require that each part of the process works today so that, in the distant future, when all the parts are brought together, the job is done correctly. This means that the business enterprise will require specialists in predicting (and shaping) future market conditions, specialists in materials acquisition, specialists in education to ensure a competent workforce, and so on—all to ensure that what the organization set out to do some months or years or decades earlier actually gets done and, with some luck, proves successful. Without that capacity, the confidence to set out on such a sophisticated enterprise would never have existed in the first place.

What bearing, then, does this have on our theories of how businesses behave and how markets operate? We'll start with the general points: first, the small firm of the perfect competition model will not do in an age of increasingly sophisticated technology. Both the complexity of specialized knowledge and its management through a large organization suggest that a large business enterprise will be the norm. Second, that complexity of knowledge and organization suggests that the modern business enterprise ought to be considered in more dimensions than simply its cost curves and a goal of maximizing profits. The politics and culture of organization become relevant to understanding what Galbraith called the **technostructure**—that is, the whole of those people who participate in the group decision-making that directs the business enterprise through time.

viewpoint.

DOES THE INTERNET MOVE US CLOSER TO PERFECT COMPETITION?

A common refrain in recent years argues that information technology—in particular, the internet—contradicts the tendency toward big business. Recall that the model of perfect competition doesn't suggest that firms must be absolutely tiny—say, having no more than a handful of employees. Rather, it simply considers situations in which firms are small *relative to the size of the market*. Many have argued that the internet has made global markets out of local markets; and, as a result, firms that used to be considered large are now small *relative to* the whole world of competitors.

Yet, within the tech sector, as in most other industries, concentration and big business appear to be the norm. Large business like IBM have long dominated traditional computing, before the World Wide Web, and relatively newer firms like Google and Facebook continue to expand control over much of the newer, web-based markets. Of course, Amazon, PayPal, eBay, AirBnB, and Uber have allowed for a proliferation of small merchants, taxi drivers, and hotels; but with these large corporations looming over the markets they've created, have we really returned to the traditional competitive product markets of past centuries—or are we looking at the competitive (one might say, 'divide and conquer') *labor* markets that characterized the rise of big business in the late 1800's?

Third, if planning is the central concern of the modern business enterprise, then we have to reconsider how we see firms relating to their markets. In previous chapters, you have studied the basic neoclassical models of market structure and how firms respond to particular market conditions. The implicit assumption in these models is that firms react to markets. Yet, if long term planning is to be done effectively it would seem obvious that businesses cannot take the markets as given and respond accordingly. Instead, they will have to actively shape those markets to ensure stable, predictable conditions into the future.

This third point has a particular bearing on our theories of prices and competition. History shows that businesses often try to avoid competition in terms of price. You can probably imagine why this is, given the complexity of modern technologies and the magnitude of investments necessary to get a business off the ground. Where starting a new business or a new product line for an existing business requires huge sums of money and years of planning, it is simply a must that the investors and management are reasonably sure that it will be able to sell the product profitably in the relatively distant future. The business enterprise's problem in this scenario is that competitors may be making the same investments; and, in an attempt to make sure those investments don't go to waste due to weak sales, they may try to grab market share by lowering their prices. If everyone in the market pursues the same strategy, what do you think the result would be?

Stiff competition based on prices is relatively rare in established markets for a simple reason: it tends to be mutually destructive to most or all competitors who engage in it. In fact, when prominent businesses engage in that sort of scenario—as, for instance, the fast food giants occasionally have in recent years—it is often considered newsworthy, usually showing up in the headlines of the business press as a 'price war'. A century ago, the term 'cut throat competition,' which includes aggressive price cuts and other efforts to drive competitors out of the market, was common. Clearly, both of the terms suggest that price competition can be a violent thing—something, one would hope at least, that should be avoided if at all possible.

In summary, then, heterodox economists typically see the modern business enterprise as a going concern, engaged in long term planning which includes managing and shaping both technological and market conditions. In this view firms don't maximize profits so much as they seek to survive through time. If it is appropriate to say that businesses tend to maximize anything, it would be growth through

time rather than profits specifically. Likewise, outright price competition is a relatively rare occurrence in mature markets today. Though most businesses will take competition into consideration when setting their prices, it would be a better description of the real world to say that competition occurs mainly through other practices, like advertising and branding, product innovations, or more cost-effective production practices. That is to say that businesses compete chiefly through their investments, not their prices.

16.2 BUSINESS MODELS, PLURAL: AIMS AND METHODS OF THE MEGACORP

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify different general methods by which businesses can pursue profits
- Analyze the nature and significance of advertizing
- Apply heterodox concepts to the analysis of the pharmaceutical industry

As a professor of mine, James Sturgeon, is fond of saying, there's more than one way to make money: You could...

- Steal
- Extort
- Accept a bribe
- Speculate on the financial or real estate markets
- Inherit
- Con
- Or, perhaps failing all of the above, you could earn it

For our purposes here, this can be interpreted as a rather cheeky way of saying that different businesses have different **business models**—that is, different ways of making money in a market (or several markets). And, while they might all be treated with the utmost abstraction as combining inputs to produce outputs of greater value, heterodox economists are inclined to believe that not every way of making money is the same.

Breaking down all of the means by which modern businesses make money is far beyond the scope of this chapter (though the odds are good that the reader is taking this course to fulfill a requirement for a business degree for which she or he will be spending plenty of time doing just that). Instead, we'll borrow from Karl Marx's extensive work on how capitalist economies function to create a simplified picture, Figure 4, of what a business enterprise does. And from there, we'll look to institutional economics to understand qualitative difference in how businesses generate their earnings.



Figure 4. The Business Enterprise in a Monetary Production Economy.

In Figure 4, M represents an amount of money and C represents a commodity (for instance the lumber and other building materials a construction company may use to build a house). P represents the production process that converts the commodity C (building materials) into some other commodity C' (for instance, a house) to be sold for some amount of money M' . This process can be treated as a shorthand depiction of what any business does, with the requirement that if the business is to remain in business M' must be greater than M . Take a moment to review the chapter “Cost and Industry Structure,” specifically the definitions of firm and production therein. How does Figure 4 differ?

The difference may not be obvious, but it is important. In the neoclassical tradition in economics—indeed, in the classical tradition which Marx was critiquing—production by businesses is treated as a $C \rightarrow C'$ process. That is, all business activity is part of converting commodity inputs into commodity outputs which are of greater value, ultimately, to consumers. Clearly, Figure 4 is more than that: it treats money and commodities as distinct things. This allows us to look at the step-by-step process by which money is converted into commodities, production creates new commodities, and money is created by sale of those new commodities. The whole process of turning money into commodity inputs and ultimately selling commodity outputs for money we’ll call **monetary production**. The full implications of these distinctions for Marxian (or radical political) economics would require its own course. For now, we can use this shorthand description of how a business enterprise works to think about different types of business models.

The three numbered arrows in Figure 4 represent three different ways that managers, engineers, marketers, and others within a business enterprise can influence the process of monetary production. The first arrow, following M to P , indicates something close to the traditional way of thinking of a business: money is invested into a production process that turns inputs into outputs for sale. This could include building a new factory or setting up a research and development (R&D) team to design new products and more efficient production practices. What is significant here is that the focus is on producing something—turning stuff C into stuff C' , presumably with the hope that people will be willing to pay more for C' than was paid for the C necessary to produce it.

The second arrow indicates a different business model—that is, another way the monetary production process could be influenced to make money. Here, you will notice, the initial money invested is going not to production, P , but to the produced commodity, C' . This is meant to indicate investments made to change the *perceived* value of the product the business sells. Of course, one way to do this is to improve the product itself, which would be indicated by the first arrow. With arrow 2, however, we’re

dealing with changing the perceived value of the product without actually changing the product itself; and you've probably already guessed the most salient way this can be done: advertising.

Here, the founder of institutional economics, Thorstein Veblen, is worth quoting at length:

The end sought by the systematic advertising of the larger business concerns is...a *monopoly of custom and prestige*.... The great end of consistent advertising is to establish such differential monopolies resting on popular conviction.... The cost, as well as the pecuniary value and the magnitude, of this *organized fabrication of popular convictions* is indicated by such statements as that the proprietors of a certain well-known household remedy, reputed among medical authorities to be of entirely dubious value, have for a series of years found their profits in spending several million dollars annually in advertisements. This case is by no means unique.

It has been said, no doubt in good faith and certainly with some reason, that advertising as currently carried on gives the body of consumers valuable information and guidance as to the ways and means whereby their wants can be satisfied and their purchasing power can be best utilized. To the extent to which this holds true, advertising is a service to the community. But there is a large reservation to be made on this head. Advertising is competitive; the greater part of it aims to divert purchases, etc., from one channel to another channel of the same general class. And to the extent to which the efforts of advertising in all its branches are spent on this *competitive disturbance of trade*, they are, on the whole, of *slight if any immediate service to the community*. (Veblen 1904, pp. 55-7, Public Domain, emphasis added)

Veblen's inimitable prose may be a bit irksome to read, but reference to the italicized bits above should be enough to understand the argument. 'Custom and prestige,' and the 'fabrication of popular conviction' suggest that businesses are competing through persuasion rather than production of something of value. The second paragraph, then, considers whether these activities are actually of value to society (or perhaps, in neoclassical terms, whether or not they're 'welfare enhancing'). For the most part, they are decidedly not. They may create a pecuniary (that is, money) value for the business—the *perception* of C' is improved and M' increases as a result. But, to society in general they amount to little more than a needless shifting of buying habits and consumption patterns. As Professor Sturgeon notes, there's more than one way to make money.

We should pause to consider the usefulness of this model. Separating money from commodities in the monetary production process has allowed us to consider different business activities as essentially different things. The abstract firm in standard neoclassical theory might lead us to treat anything that makes a business more money as part of the production process ($C...P...C'$), suggesting that advertising was just another commodity in the production process improving the value of the product to be sold. By implication, then, the advertising added value to society. Yet, treating production and advertising as two different activities allowed room for the argument that, in fact, the advertising was generally wasteful.

DISTINGUISHING ARROW 1 FROM ARROW 2: THE PHARMACEUTICAL INDUSTRY

A contemporary example should indicate the importance of distinguishing between arrows 1 and 2. The price of pharmaceutical drugs continues to be a controversial topic in the United States, and it's no surprise why: according to the OECD,

pharmaceutical spending has risen from an average of \$159 per person (or 0.87% of all spending in the US) in 1985 to \$1,112 per person (or 2.04% of total spending) in 2014. Moreover, a 2016 Reuters article found that the prices of 4 of the top 10 most widely used drugs more than doubled in just the previous 5 years.

A number of factors are important to explaining these rapid price increases, including an aging population in the US and an exceedingly complex distribution and insurance system for pharmaceuticals. But, let's focus on the common explanation given by the producers themselves: high prices are necessary to recover the high costs of research and development (much of which leads to dead-end investments that don't produce effective drugs that can be brought to market) and the clinical trials necessary for regulatory approval. Clearly, this is in line with the general framework of heterodox economic theory. Pharmaceutical companies are engaging in long-run planning, including significant investments into better products (that is, investments into P in Figure 1), and prices are a reflection of those investments.

But, what's more, our monetary production model can help us find a fuller explanation of the nature and extent of those investments. Consider estimates from Gagnon and Lexchin (2008) on promotional expenditures by the industry. The authors found that, for 2004, the industry devoted about 24% of its sales revenue to promotion, versus about 13% to research and development. Looking specifically to promotion directed at physicians, the authors estimate that approximately \$61,000 was spent *per physician* in the US. To be sure, there are other estimates lower than those made by Gagnon and Lexchin. Just the same, it is clear that pharmaceutical enterprises are investing heavily not only in developing new drugs, but in promoting their use as well.

Moreover, the pharmaceutical industry provides us with a somewhat unique opportunity to clearly distinguish investments in C' (arrow 2 in figure 1) from investments in P (arrow 1). Because these drugs are systematically evaluated for their effectiveness in treating specific illnesses, data exist to indicate how well supposedly-arrow 1 investments are improving the product. On this matter, Gagnon (2013) finds that fewer and fewer new molecular entities (truly new drugs, without precedent in drugs already in use) are being developed, and only a small minority of new drugs represent substantial therapeutic advances over existing treatments. The proliferation of 'me-too' drugs, which are new products to be marketed if not actually better treatments for illnesses, suggests that the business model in this industry is focused less on improvements in P and more on inflating C'. That is, arrow 2 appears to dominate, with arrow 1 a secondary concern, in the normal business models of this industry.

Sources: <https://data.oecd.org/healthres/pharmaceutical-spending.htm>

<http://www.reuters.com/article/us-usa-healthcare-drugpricing-idUSKCN0X10TH>

Gagnon, M. A. and J. Lexchin. (2008), "The Cost of Pushing Pills: A New Estimate of Pharmaceutical Promotion Expenditures in the United States," *PLoS Medicine*, 5(1), 29-33.

Gagnon, M. A. (2013). Corruption of pharmaceutical markets: Addressing the misalignment of financial incentives and public health. *The Journal of Law, Medicine & Ethics*, 41(3), 571-580.

Finally, there is arrow 3, a direct line from an initial sum of money to a greater amount of money, bypassing all the troublesome in-between steps of producing and selling a good or service. This sort of business model takes many forms and is becoming increasingly popular in the capitalism of today. The traditional exemplar is the financial firm—or, say, the stock broker—which has no direct ties to actually producing something, but rather only to trading in the ownership rights of those producers.

The full role and consequences of financial firms and markets cannot be covered here. What's of interest presently is business activity that can turn a profit without being concerned with producing something for sale, or even simply with persuading people to buy a product. You may have heard

about the causes of the Great Recession of 2008-09 being chiefly linked to business models in finance, insurance, and real estate (the so-called FIRE sector). The problem, many believe, is in the opacity and complexity of a system in which mortgage contracts representing an agreement to pay for a house over time were packaged together into derivative contracts to be sold to investors, then resold, repackaged, and so on. Indeed to make a system even more complex, many financial firms were making side bets that the contracts they themselves were selling would in fact not pay out. Many articles, books, and documentaries have been produced retelling that story, and *The Big Short* is an effective and accurate dramatization of some of those events. For now, it is enough to say that the businesses engaged in those activities were largely concerned with turning money into more money by persuading others that these contracts built on contracts (built on contracts...) were worth more than they were. These were arrow 3 business activities, having little to do with the actual business of housing people.

A century earlier, Veblen had described similar practices in the stock markets, where the goal is to,

induce a discrepancy between the putative and the actual earning-capacity, by expedients well known and approved for the purpose. Partial information, as well as misinformation, sagaciously given out at a critical juncture, will go far toward producing a favorable temporary discrepancy of this kind. (Veblen, 1904, p. 156, Public Domain)

It was, more recently, precisely this type of behavior which ultimately caused the deepest recession since the Great Depression (which, incidentally, had also been caused by this type of behavior). As a result, many of the largest banks and other financial institutions in the US and abroad have been made to pay fines in the 10's of billions of dollars. Whether these fines will be enough to make this business model unattractive in the future is debatable—but it seems unlikely.

16.3 STABILIZING UNSTABLE MARKETS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Situate businesses and their business models within their market governance structures
- Define infrastructure, corporate governance, and market governance

The previous sections dealt mainly with the nature of the modern business enterprise as an individual organization. But, of course, no business is created in a vacuum, and no business can operate in complete isolation. This section will look at the economic, social, and political nature of markets to better understand how real businesses fit into a heterodox understanding of the economy.

Any business model—including those discussed above and beyond—is a plan for how to successfully operate a business within one or more markets. A good business plan will have to consider, among other things, the competition, the potential customer base, rules and regulations, and the necessary infrastructure to produce and distribute the product, whatever it may be. Established businesses will have worked out their business models over time, will have built (or had built for them) the necessary infrastructure (for instance, roads or communications protocols), and will typically have helped define the rules and regulations that dictate which individual and competitive activities are permissible and which are not. This is to say that markets themselves are defined by (and the way business enterprises behave in these markets is guided by) their infrastructure and their corporate and market governance structures.

DEFINITIONS

Infrastructure: the common structures, facilities, and systems necessary for organizations to operate. These may include physical infrastructure like highways, bridges, electrical grids, and sewage systems; as well as systems like the complex networks of computers tied together through communications lines and protocols we know as the internet.

Corporate governance: the rules and practices defining how and by whom a business is directed as well as how members of the business will interact with each other and those outside of the business. These are both formal rules like accounting practices or the corporate hierarchy of who reports to whom, as well as informal norms, traditions, and relationships.

Market governance: the rules and practices defining how business enterprises will behave in a market, including especially

how they will interact with their competitors. These may include formal agreements—for instance, a consent decree by which a business agrees not to engage in an activity the government considers anticompetitive, or a joint venture operation between two energy companies to explore a potential source of crude oil. Probably more important, though, are the various informal arrangements by which certain types of competition and cooperation between businesses are allowed while others are not.

A side note: these shared rules and infrastructure are not fixed in time—they evolve, whether by unintended consequence or intentional change. Moreover, they are socially constructed, which means there is a political element to all of them.

Although we're focusing for now on the private, business side of the matter, it should be remembered that there is almost always a government role in the development, regulation, and sometimes prohibition of these structures. Finally, a fourth component, **property rights**, is worth adding. Property rights are legal norms defining who can own what, what can be done with that property, and therefore how businesses can generate earnings and who has claims on those earnings. While property is often considered simply a natural right, the actual content of property rights is a complex, perpetually evolving, and highly contested subject.

These attributes of the organization of businesses and the markets in which they operate are all geared toward essentially the same thing: stability. As an earlier section explained, most parts of the modern economy involve sophisticated and usually large-scale technologies. For this and other reasons long term planning is necessary for production to go forward; and long term planning requires predictable outcomes. Hence, markets and businesses must be organized to promote stability. In particular, as a previous section explains, businesses require predictability in prices.

Now, consider the neoclassical models of previous chapters. In each of these some, perhaps natural, equilibration process is used—that is, a process by which firm's, consumers, and ultimately markets move toward an equilibrium, toward stability. The utility maximizing consumer seeks an optimal combination of consumers goods, the profit maximizing firm seeks an optimal level of production, and the market, through competition and price bidding, moves toward the equilibrium price and output.

Many heterodox economists reject each of these models, in part because of the unrealistic mechanisms by which equilibrium is reached. For instance, as chapter “Costs and Prices” demonstrates, few markets in the modern US economy are characterized by the sort of price adjustments that the standard market model relies on to reach a stable equilibrium. Moreover, considerable evidence suggests that competition—especially price competition—actually promotes instability. And it is for this reason that the concept of market governance is so important. Without a workable set of norms concerning acceptable and unacceptable forms of competition and cooperation, most markets would never reach equilibrium. Instead, price competition and chicanery would wreak havoc on businesses and consumers alike.

16.4 THE HIGH PRICE OF COLLEGE TEXTBOOKS

In this concluding section of the chapter we'll look at a question you've probably asked yourself: why do college textbooks cost so much? The question is particularly interesting for economic theory. Nationally, textbooks prices have risen more than three times the prices of other goods and services in the economy—an increase from the 1977 to 2015 of 1,041%, reports ABC News. As of September 2016, the average undergraduate student will be spending just shy of \$1,300 a year on textbooks and supplies. That's no small sum.

Standard analysis would look for low elasticities of demand and lack of competition to explain high prices. Question: as a student, is your demand for college textbooks elastic or inelastic? Why? Likely, you answered 'very inelastic' since you have little choice in buying a textbook that is required for a course. (In fact a 2014 Student PIRG survey found that two thirds of students had foregone buying at least one textbook due to cost. Not surprisingly, almost all of those students indicated concern that the decision would impact their grades negatively.)

As for competition in the textbook publishing business, it may not surprise you to learn that, like in most other markets, there are only a handful of large corporations controlling the bulk of the college textbook supply. International corporations, which in many cases have existed for over a century, dominate the global market for these products, bringing in billions of dollars in revenue each year. While not a pure monopoly, it would seem that this industry is closer to our neoclassical monopoly model than to perfect competition. The outcome of this captive-market situation, as you've seen in a previous chapter, is clear: fewer students purchase textbooks than would like, and they pay higher prices than would exist under more competitive conditions.

But, then, is competition the solution? Certainly, the industry meets the definition of high concentration, suggesting that textbook publishing is far from the ideal model of perfect competition. Publishers appear to enjoy significant market power, especially in setting their own prices, rather than taking the competitive market price as given. But, if we accept that these large corporations are not price takers in the markets for their books, we still have to answer: how—that is, for what reasons—do they determine the prices they will charge? And this is where the business model becomes important.

The mainstream assumption that firms, always and everywhere, maximize profits suggests that output is being set according to marginal costs and marginal revenues, and that the high inelasticity of demand allows publishers to raise prices well above the actual costs of production. Firms produce to maximize profits, and a lack of competition allows them to take monopoly rents through higher prices.

But, consider the heterodox position—that business enterprises generally want to survive and grow, and that competition occurs chiefly through investments, not prices. This view suggests that it's not so much that publishers wish to get every last dollar of profit to be had (that is, to maximize profits),

nor that they will raise prices to whatever the students (that is, demanders) are willing to pay. Rather, these producers are seeking to make the investments necessary to survive (to make a profit) in their markets. The prices they charge, then, may seem like price gouging, but may really just be the results of the markups necessary to cover the costs of those investments. (However, the more recent focus among publicly traded companies on keeping stock prices up may drive them toward price gouging over simply pricing to cover full costs.)

And this is where we come back to the original question: is competition the solution? From the perspective of heterodox economics, the high price of college textbooks reflects the high cost of investing in new textbook editions—in computer software tools to complement the text (and ensure the student pays for the online access), in test banks and slideshows to aid the instructor using the text, and so on. And what compels publishers to make these investments? Largely, competition. If a publisher neglected to develop new editions, it could not compete with the used textbook markets. If a publisher fell behind in developing software and supplements for instructors, the other publishers would have a competitive advantage in pitching their texts to instructors. Perhaps surprisingly, the conclusion is that prices are high, not so much from a lack of price competition, but because of a particular dynamic in investment competition. It is—or so one could argue—a lack of appropriate planning across the industry as a whole that keeps publishers grasping at market share by investing in anything and everything that might help sell their textbooks...and the student pays the price.

THE COURTS AND PROPERTY RIGHTS

In a previous breakout box it was mentioned that a proper understanding of market structure would require, among other things, an understanding of the specific property rights relevant to production, exchange, and consumption processes of the market. A recent US Supreme Court case offers a good illustration of that very point.

John Wiley & Sons, Inc. is a global publishing company with thousands of employees and a market cap (total value of outstanding shares) in the billions of dollars. In 2008 Wiley sued Supap Kirtsaeng for copyright infringement. Kirtsaeng, a Thailand native who had studied in the US, had discovered that foreign edition textbooks could be imported into the US to be sold at lower prices than the texts' American-edition prices. Wiley believed that this was a violation of their property rights which included the right to prevent importation of their copyrighted works.

While two lower courts sided with the publisher's position on the control that copyright afforded them, the US Supreme Court did not. They argued in their ruling that the 'first-sale doctrine', allowing those who obtained a textbook legally to re-sell it, limited Wiley's right to control the importation of their books. As an economist-in-training, how do you think things would be different if the Supreme Court had ruled in favor of Wiley? Would prices continue to rise, plateau, or come down? Would this change the direction of investment in the publishing industry? How would students, faculties, and higher education more generally be affected?

CHAPTER 17. MONOPOLY AND ANTITRUST POLICY

INTRODUCTION TO MONOPOLY AND ANTITRUST POLICY



Figure 1. Oligopoly versus Competitors in the Marketplace. Large corporations, such as the natural gas producer Kinder Morgan, can bring economies of scale to the marketplace. Will that benefit consumers? Or is more competition better for consumers? (Credit: modification of work by Derrick Coetzee/Flickr Creative Commons)

MORE THAN COOKING, HEATING, AND COOLING

If you live in the United States, there is a slightly better than 50–50 chance your home is heated and cooled using natural gas. You may even use natural gas for cooking. However, those uses are not the primary uses of natural gas in the U.S. In 2012, according to the U.S. Energy Information Administration, home heating, cooling, and cooking accounted for just 18% of natural gas usage. What accounts for the rest? The greatest uses for natural gas are the generation of electric power (39%) and in industry (30%). Together these three uses for natural gas touch many areas of our lives, so why would there be any opposition to a merger of two natural gas firms? After all, a merger could mean increased efficiencies and reduced costs to people like you and me.

In October 2011, Kinder Morgan and El Paso Corporation, two natural gas firms, announced they were merging. The announcement stated the combined firm would link “nearly every major production region with markets,” cut costs by “eliminating duplication in pipelines and other assets,” and that “the savings could be passed on to consumers.”

The objection? The \$21.1 billion deal would give Kinder Morgan control of more than 80,000 miles of pipeline, making the new firm the third largest energy producer in North America. As the third largest energy producer, policymakers and the public wondered whether the cost savings really would be passed on to consumers, or would the merger give Kinder Morgan a strong oligopoly position in the natural gas marketplace?

That brings us to the central question this chapter poses: What should the balance be between corporate size and a larger number of competitors in a marketplace? We will also consider what role the government should play in this balancing act.

CHAPTER OBJECTIVES

Introduction to Monopoly and Antitrust Policy

In this chapter, you will learn about:

- Corporate Mergers
- Regulating Anticompetitive Behavior
- Regulating Natural Monopolies
- The Great Deregulation Experiment

The previous chapters on the theory of the firm identified three important lessons: First, that competition, by providing consumers with lower prices and a variety of innovative products, is a good thing; second, that large-scale production can dramatically lower average costs; and third, that markets in the real world are rarely perfectly competitive. As a consequence, government policymakers must determine how much to intervene to balance the potential benefits of large-scale production against the potential loss of competition that can occur when businesses grow in size, especially through mergers.

For example, in 2011, AT&T and T-Mobile proposed a merger. At the time, there were only four major mobile phone service providers. The proposal was blocked by both the Justice Department and the FCC.

The two companies argued that the merger would benefit consumers, who would be able to purchase better telecommunications services at a cheaper price because the newly created firm would be able to produce more efficiently by taking advantage of economies of scale and eliminating duplicate investments. However, a number of activist groups like the Consumer Federation of America and Public Knowledge expressed fears that the merger would reduce competition and lead to higher prices for consumers for decades to come. In December 2006, the federal government allowed the merger to proceed. By 2009, the new post-merger AT&T was the eighth largest company by revenues in the United States, and by that measure the largest telecommunications company in the world. Economists have spent – and will still spend – years trying to determine whether the merger of AT&T and BellSouth, as well as other smaller mergers of telecommunications companies at about this same time, helped consumers, hurt them, or did not make much difference.

This chapter discusses public policy issues about competition. How can economists and governments determine when mergers of large companies like AT&T and BellSouth should be allowed and when they should be blocked? The government also plays a role in policing anticompetitive behavior other than mergers, like prohibiting certain kinds of contracts that might restrict competition. In the case of natural monopoly, however, trying to preserve competition probably will not work very well, and so government will often resort to regulation of price and/or quantity of output. In recent decades,

there has been a global trend toward less government intervention in the price and output decisions of businesses.

17.1 CORPORATE MERGERS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

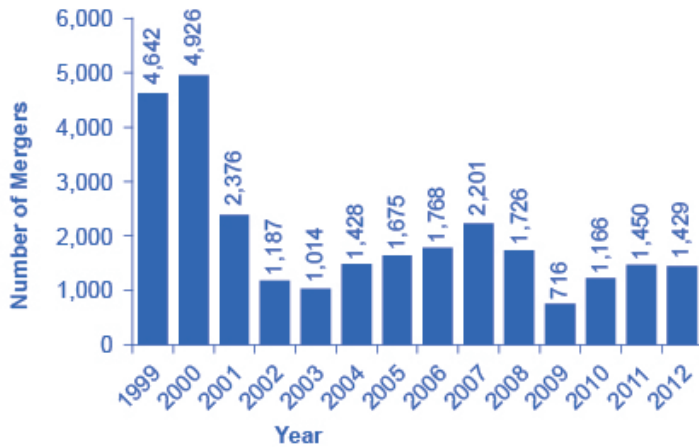
- Explain antitrust law and its significance
- Calculate concentration ratios
- Calculate the Herfindahl-Herschman Index (HHI)
- Evaluate methods of antitrust regulation

A corporate **merger** occurs when two formerly separate firms combine to become a single firm. When one firm purchases another, it is called an **acquisition**. An acquisition may not look just like a merger, since the newly purchased firm may continue to be operated under its former company name. Mergers can also be lateral, where two firms of similar sizes combine to become one. However, both mergers and acquisitions lead to two formerly separate firms being under common ownership, and so they are commonly grouped together.

REGULATIONS FOR APPROVING MERGERS

Since a merger combines two firms into one, it can reduce the extent of competition between firms. Therefore, when two U.S. firms announce a merger or acquisition where at least one of the firms is above a minimum size of sales (a threshold that moves up gradually over time, and was at \$70.9 million in 2013), or certain other conditions are met, they are required under law to notify the U.S. Federal Trade Commission (FTC). The left-hand panel of Figure 1 (a) shows the number of mergers submitted for review to the FTC each year from 1999 to 2012. Mergers were very high in the late 1990s, diminished in the early 2000s, and then rebounded somewhat in a cyclical fashion. The right-hand panel of Figure 1 (b) shows the distribution of those mergers submitted for review in 2012 as measured by the size of the transaction. It is important to remember that this total leaves out many small mergers under \$50 million, which only need to be reported in certain limited circumstances. About a quarter of all reported merger and acquisition transactions in 2012 exceeded \$500 million, while about 11 percent exceeded \$1 billion. In 2014, the FTC took action against mergers likely to stifle competition in markets worth 18.6 billion in sales.

The laws that give government the power to block certain mergers, and even in some cases to break up large firms into smaller ones, are called **antitrust laws**. Before a large merger happens, the antitrust regulators at the FTC and the U.S. Department of Justice can allow the merger, prohibit it, or allow it if certain conditions are met. One common condition is that the merger will be allowed if the firm



(a) Number of mergers submitted for review by the Federal Trade Commission, 1999-2012



(b) Size of transaction for mergers submitted for review in 2012 (in millions of dollars)

Figure 1. Number and Size of Mergers. (a) The number of mergers in 1999 and 2000 were relatively high compared to the annual numbers seen from 2001–2012. While 2001 and 2007 saw a high number of mergers, these were still only about half the number of mergers in 1999 and 2000. (b) In 2012, the greatest number of mergers submitted for review was for transactions between \$100 and \$150 million.

agrees to sell off certain parts. For example, in 2006, Johnson & Johnson bought the Pfizer’s “consumer health” division, which included well-known brands like Listerine mouthwash and Sudafed cold medicine. As a condition of allowing the merger, Johnson & Johnson was required to sell off six brands to other firms, including Zantac® heartburn relief medication, Cortizone anti-itch cream, and Balmex diaper rash medication, to preserve a greater degree of competition in these markets.

The U.S. government approves most proposed mergers. In a market-oriented economy, firms have the freedom to make their own choices. Private firms generally have the freedom to:

- expand or reduce production
- set the price they choose
- open new factories or sales facilities or close them
- hire workers or to lay them off
- start selling new products or stop selling existing ones

If the owners want to acquire a firm or be acquired, or to merge with another firm, this decision is just one of many that firms are free to make. In these conditions, the managers of private firms will sometimes make mistakes. They may close down a factory which, it later turns out, would have been profitable. They may start selling a product that ends up losing money. A merger between two companies can sometimes lead to a clash of corporate personalities that makes both firms worse off. But the fundamental belief behind a market-oriented economy is that firms, not governments, are in the best position to know if their actions will lead to attracting more customers or producing more efficiently.

Indeed, government regulators agree that most mergers are beneficial to consumers. As the Federal Trade Commission has noted on its website (as of November, 2013): “Most mergers actually benefit

competition and consumers by allowing firms to operate more efficiently.” At the same time, the FTC recognizes, “Some [mergers] are likely to lessen competition. That, in turn, can lead to higher prices, reduced availability of goods or services, lower quality of products, and less innovation. Indeed, some mergers create a concentrated market, while others enable a single firm to raise prices.” The challenge for the antitrust regulators at the FTC and the U.S. Department of Justice is to figure out when a merger may hinder competition. This decision involves both numerical tools and some judgments that are difficult to quantify. The following Clear it Up helps explain how antitrust laws came about.

WHAT IS U.S. ANTITRUST LAW?

In the closing decades of the 1800s, many industries in the U.S. economy were dominated by a single firm that had most of the sales for the entire country. Supporters of these large firms argued that they could take advantage of economies of scale and careful planning to provide consumers with products at low prices. However, critics pointed out that when competition was reduced, these firms were free to charge more and make permanently higher profits, and that without the goading of competition, it was not clear that they were as efficient or innovative as they could be.

In many cases, these large firms were organized in the legal form of a “trust,” in which a group of formerly independent firms were consolidated together by mergers and purchases, and a group of “trustees” then ran the companies as if they were a single firm. Thus, when the U.S. government passed the **Sherman Antitrust Act** in 1890 to limit the power of these trusts, it was called an antitrust law. In an early demonstration of the law’s power, the U.S. Supreme Court in 1911 upheld the government’s right to break up Standard Oil, which had controlled about 90% of the country’s oil refining, into 34 independent firms, including Exxon, Mobil, Amoco, and Chevron. In 1914, the **Clayton Antitrust Act** outlawed mergers and acquisitions (where the outcome would be to “substantially lessen competition” in an industry), price discrimination (where different customers are charged different prices for the same product), and tied sales (where purchase of one product commits the buyer to purchase some other product). Also in 1914, the Federal Trade Commission (FTC) was created to define more specifically what competition was unfair. In 1950, the **Celler-Kefauver Act** extended the Clayton Act by restricting vertical and conglomerate mergers. In the twenty-first century, the FTC and the U.S. Department of Justice continue to enforce antitrust laws.

THE FOUR-FIRM CONCENTRATION RATIO

Regulators have struggled for decades to measure the degree of monopoly power in an industry. An early tool was the **concentration ratio**, which measures what share of the total sales in the industry are accounted for by the largest firms, typically the top four to eight firms. For an explanation of how high market concentrations can create inefficiencies in an economy, refer to Monopoly.

Say that the market for replacing broken automobile windshields in a certain city has 18 firms with the market shares shown in Table 1, where the **market share** is each firm’s proportion of total sales in that market. The four-firm concentration ratio is calculated by adding the market shares of the four largest firms: in this case, $16 + 10 + 8 + 6 = 40$. This concentration ratio would not be considered especially high, because the largest four firms have less than half the market.

If the market shares in the market for replacing automobile windshields are:

Smooth as Glass Repair Company	16% of the market
The Auto Glass Doctor Company	10% of the market
Your Car Shield Company	8% of the market
Seven firms that each have 6% of the market	42% of the market, combined
Eight firms that each have 3% of the market	24% of the market, combined

Then the four-firm concentration ratio is $16 + 10 + 8 + 6 = 40$.

Table 1. Calculating Concentration Ratios from Market Shares

The concentration ratio approach can help to clarify some of the fuzziness over deciding when a merger might affect competition. For instance, if two of the smallest firms in the hypothetical market for repairing automobile windshields merged, the four-firm concentration ratio would not change—which implies that there is not much worry that the degree of competition in the market has notably diminished. However, if the top two firms merged, then the four-firm concentration ratio would become 46 (that is, $26 + 8 + 6 + 6$). While this concentration ratio is modestly higher, the four-firm concentration ratio would still be less than half, so such a proposed merger might barely raise an eyebrow among antitrust regulators.

Visit this website to read an article about Google's run-in with the FTC.



THE HERFINDAHL-HIRSHMAN INDEX

A four-firm concentration ratio is a simple tool, which may reveal only part of the story. For example, consider two industries that both have a four-firm concentration ratio of 80. However, in one industry five firms each control 20% of the market, while in the other industry, the top firm holds 77% of the market and all the other firms have 1% each. Although the four-firm concentration ratios are identical, it would be reasonable to worry more about the extent of competition in the second case—where the largest firm is nearly a monopoly—than in the first.

Another approach to measuring industry concentration that can distinguish between these two cases is called the **Herfindahl-Hirschman Index (HHI)**. The HHI, as it is often called, is calculated by summing the squares of the market share of each firm in the industry, as the following Work it Out shows.

CALCULATING HHI

Step 1. Calculate the HHI for a monopoly with a market share of 100%. Because there is only one firm, it has 100% market share. The HHI is $100^2 = 10,000$.

Step 2. For an extremely competitive industry, with dozens or hundreds of extremely small competitors, the value of the HHI might drop as low as 100 or even less. Calculate the HHI for an industry with 100 firms that each have 1% of the market. In this case, the HHI is $100(1^2) = 100$.

Step 3. Calculate the HHI for the industry shown in Table 1. In this case, the HHI is $16^2 + 10^2 + 8^2 + 7(6^2) + 8(3^2) = 744$.

Step 4. Note that the HHI gives greater weight to large firms.

Step 5. Consider the example given earlier, comparing one industry where five firms each have 20% of the market with an industry where one firm has 77% and the other 23 firms have 1% each. The two industries have the same four-firm concentration ratio of 80. But the HHI for the first industry is $5(20^2) = 2,000$, while the HHI for the second industry is much higher at $77^2 + 23(1^2) = 5,952$.

Step 6. Note that the near-monopolist in the second industry drives up the HHI measure of industrial concentration.

Step 7. Review Table 2 which gives some examples of the four-firm concentration ratio and the HHI in various U.S. industries in 2009. (You can find market share data from multiple industry sources. Data in the table are from: Verizon (for wireless), *The Wall Street Journal* (for automobiles), IDC Worldwide (for computers) and the U.S. Bureau of Transportation Statistics (for airlines).)

U.S. Industry	Four-Firm Ratio	HHI
Wireless	91	2,311
Largest five: Verizon, AT&T, Sprint, T-Mobile, MetroPCS		
Automobiles	63	1,121
Largest five: GM, Toyota, Ford, Honda, Chrysler		
Computers	74	1,737
Largest five: HP, Dell, Acer, Apple, Toshiba		
Airlines	44	536
Largest five: Southwest, American, Delta, United, U.S. Airways		

Table 2. Examples of Concentration Ratios and HHIs in the U.S. Economy, 2009

In the 1980s, the FTC followed these guidelines: If a merger would result in an HHI of less than 1,000, the FTC would probably approve it. If a merger would result in an HHI of more than 1,800, the FTC would probably challenge it. If a merger would result in an HHI between 1,000 and 1,800, then the FTC would scrutinize the plan and make a case-by-case decision. However, in the last several decades, the antitrust enforcement authorities have moved away from relying as heavily on measures of concentration ratios and HHIs to determine whether a merger will be allowed, and instead carried out more case-by-case analysis on the extent of competition in different industries.

NEW DIRECTIONS FOR ANTITRUST

Both the four-firm concentration ratio and the Herfindahl-Hirschman index share some weaknesses. First, they begin from the assumption that the “market” under discussion is well-defined, and the only question is measuring how sales are divided in that market. Second, they are based on an implicit assumption that competitive conditions across industries are similar enough that a broad measure of concentration in the market is enough to make a decision about the effects of a merger. These assump-

tions, however, are not always correct. In response to these two problems, the antitrust regulators have been changing their approach in the last decade or two.

Defining a **market** is often controversial. For example, Microsoft in the early 2000s had a dominant share of the software for computer operating systems. However, in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 14% in 2014. A narrowly defined market will tend to make concentration appear higher, while a broadly defined market will tend to make it appear smaller.

There are two especially important shifts affecting how markets are defined in recent decades: one centers on technology and the other centers on **globalization**. In addition, these two shifts are interconnected. With the vast improvement in communications technologies, including the development of the Internet, a consumer can order books or pet supplies from all over the country or the world. As a result, the degree of competition many local retail businesses face has increased. The same effect may operate even more strongly in markets for business supplies, where so-called “business-to-business” websites can allow buyers and suppliers from anywhere in the world to find each other.

Globalization has changed the boundaries of markets. As recently as the 1970s, it was common for measurements of concentration ratios and HHIs to stop at national borders. Now, many industries find that their competition comes from the global market. A few decades ago, three companies, General Motors, Ford, and Chrysler, dominated the U.S. auto market. By 2014, however, these three firms were making less than half of U.S. auto sales, and facing competition from well-known car manufacturers such as Toyota, Honda, Nissan, Volkswagen, Mitsubishi, and Mazda. When HHIs are calculated with a global perspective, concentration in most major industries—including cars—is lower than in a purely domestic context.

Because attempting to define a particular market can be difficult and controversial, the Federal Trade Commission has begun to look less at market share and more at the data on actual competition between businesses. For example, in February 2007, Whole Foods Market and Wild Oats Market announced that they wished to merge. These were the two largest companies in the market that the government defined as “premium natural and organic supermarket chains.” However, one could also argue that they were two relatively small companies in the broader market for all stores that sell groceries or specialty food products.

Rather than relying on a market definition, the government antitrust regulators looked at detailed evidence on profits and prices for specific stores in different cities, both before and after other competitive stores entered or exited. Based on that evidence, the Federal Trade Commission decided to block the merger. After two years of legal battles, the merger was eventually allowed in 2009 under the conditions that Whole Foods sell off the Wild Oats brand name and a number of individual stores, to preserve competition in certain local markets. For more on the difficulties of defining markets, refer to Monopoly.

This new approach to antitrust regulation involves detailed analysis of specific markets and companies, instead of defining a market and counting up total sales. A common starting point is for antitrust regulators to use statistical tools and real-world evidence to estimate the **demand curves** and **supply curves** faced by the firms that are proposing the merger. A second step is to specify how competition occurs in this specific industry. Some possibilities include competing to cut prices, to raise output, to build a brand name through advertising, and to build a reputation for good service or high qual-

ity. With these pieces of the puzzle in place, it is then possible to build a statistical model that estimates the likely outcome for consumers if the two firms are allowed to merge. Of course, these models do require some degree of subjective judgment, and so they can become the subject of legal disputes between the antitrust authorities and the companies that wish to merge.

KEY CONCEPTS AND SUMMARY

A corporate merger involves two private firms joining together. An acquisition refers to one firm buying another firm. In either case, two formerly independent firms become one firm. Antitrust laws seek to ensure active competition in markets, sometimes by preventing large firms from forming through mergers and acquisitions, sometimes by regulating business practices that might restrict competition, and sometimes by breaking up large firms into smaller competitors.

A four-firm concentration ratio is one way of measuring the extent of competition in a market. It is calculated by adding the market shares—that is, the percentage of total sales—of the four largest firms in the market. A Herfindahl-Hirschman Index (HHI) is another way of measuring the extent of competition in a market. It is calculated by taking the market shares of all firms in the market, squaring them, and then summing the total.

The forces of globalization and new communications and information technology have increased the level of competition faced by many firms by increasing the amount of competition from other regions and countries.

SELF-CHECK QUESTIONS

1. Is it true that both the four-firm concentration ratio and the Herfindahl-Hirschman Index can be affected by a merger between two firms that are not already in the top four by size? Explain briefly.
2. Is it true that the four-firm concentration ratio puts more emphasis on one or two very large firms, while the Herfindahl-Hirschman Index puts more emphasis on all the firms in the entire market? Explain briefly.
3. Some years ago, two intercity bus companies, Greyhound Lines, Inc. and Trailways Transportation System, wanted to merge. One possible definition of the market in this case was “the market for intercity bus service.” Another possible definition was “the market for intercity transportation, including personal cars, car rentals, passenger trains, and commuter air flights.” Which definition do you think the bus companies preferred, and why?
4. As a result of globalization and new information and communications technology, would you expect that the definitions of markets used by antitrust authorities will become broader or narrower?

REVIEW QUESTIONS

1. What is a corporate merger? What is an acquisition?
2. What is the goal of antitrust policies?
3. How is a four-firm concentration ratio measured? What does a high measure mean about the extent of competition?

4. How is a Herfindahl-Hirshman Index measured? What does a low measure mean about the extent of competition?
5. Why can it be difficult to decide what a “market” is for purposes of measuring competition?

CRITICAL THINKING QUESTIONS

1. Does either the four-firm concentration ratio or the HHI directly measure the amount of competition in an industry? Why or why not?
2. What would be evidence of serious competition between firms in an industry? Can you identify two highly competitive industries?

PROBLEMS

1. Use Table 3 to calculate the four-firm concentration ratio for the U.S. auto market. Does this indicate a concentrated market or not?

GM	19%
Ford	17%
Toyota	14%
Chrysler	11%

Table 3. Global Auto Manufacturers with Top Four U.S. Market Share, June 2013. (Source: <http://www.zacks.com/commentary/27690/auto-industry-stock-outlook-june-2013>)

2. Use Table 3 and Table 4 to calculate the Herfindal-Hirschman Index for the U.S. auto market. Would the FTC approve a merger between GM and Ford?

Honda	10%
Nissan	7%
Hyundai	5%
Kia	4%
Subaru	3%
Volkswagen	3%

Table 4. Global Auto Manufacturers with additional U.S. Market Share, June 2013. (Source: <http://www.zacks.com/commentary/27690/auto-industry-stock-outlook-june-2013>)

GLOSSARY

acquisition when one firm purchases another

antitrust laws laws that give government the power to block certain mergers, and even in some cases to break up large firms into smaller ones

concentration ratio an early tool to measure the degree of monopoly power in an industry;

measures what share of the total sales in the industry are accounted for by the largest firms, typically the top four to eight firms

four-firm concentration ratio the percentage of the total sales in the industry that are accounted for by the largest four firms

Herfindahl-Hirschman Index (HHI) approach to measuring market concentration by adding the square of the market share of each firm in the industry

market share the percentage of total sales in the market

merger when two formerly separate firms combine to become a single firm

SOLUTIONS

Answers to Self-Check Questions

1. Yes, it is true. The HHI example is easy enough: since the market shares of all firms are included in the HHI calculation, a merger between two of the firms will change the HHI. For the four-firm concentration ratio, it is quite possible that a merger between, say, the fifth and sixth largest firms in the market could create a new firm that is then ranked in the top four in the market. In this case, a merger of two firms, neither in the top four, would still change the four-firm concentration ratio.
2. No, it is not true. The HHI includes the market shares of all firms in its calculation, but the squaring of the market shares has the effect of making the impact of the largest firms relatively bigger than in the 4-firm or 8-firm ratio.
3. The bus companies wanted the broader market definition (i.e., the second definition). If the narrow definition had been used, the combined bus companies would have had a near-monopoly on the market for intercity bus service. But they had only a sliver of the market for intercity transportation when everything else was included. The merger was allowed.
4. The common expectation is that the definition of markets will become broader because of greater competition from faraway places. However, this broadening doesn't necessarily mean that antitrust authorities can relax. There is also a fear that companies with a local or national monopoly may use the new opportunities to extend their reach across national borders, and that it will be difficult for national authorities to respond.

17.2 REGULATING ANTICOMPETITIVE BEHAVIOR

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Analyze restrictive practices
- Explain tying sales, bundling, and predatory pricing
- Evaluate a real-world situation of possible anticompetitive and restrictive practices

The U.S. antitrust laws reach beyond blocking mergers that would reduce competition to include a wide array of anticompetitive practices. For example, it is illegal for competitors to form a cartel to collude to make pricing and output decisions, as if they were a monopoly firm. The Federal Trade Commission and the U.S. Department of Justice prohibit firms from agreeing to fix prices or output, rigging bids, or sharing or dividing markets by allocating customers, suppliers, territories, or lines of commerce.

In the late 1990s, for example, the antitrust regulators prosecuted an international cartel of vitamin manufacturers, including the Swiss firm Hoffman-La Roche, the German firm BASF, and the French firm Rhone-Poulenc. These firms reached agreements on how much to produce, how much to charge, and which firm would sell to which customers. The high-priced vitamins were then bought by firms like General Mills, Kellogg, Purina-Mills, and Proctor and Gamble, which pushed up the prices more. Hoffman-La Roche pleaded guilty in May 1999 and agreed both to pay a fine of \$500 million and to have at least one top executive serve four months of jail time.

Under U.S. antitrust laws, monopoly itself is not illegal. If a firm has a monopoly because of a newly patented invention, for example, the law explicitly allows a firm to earn higher-than-normal profits for a time as a reward for innovation. If a firm achieves a large share of the market by producing a better product at a lower price, such behavior is not prohibited by antitrust law.

RESTRICTIVE PRACTICES

Antitrust law includes rules against **restrictive practices**—practices that do not involve outright agreements to raise price or to reduce the quantity produced, but that might have the effect of reducing competition. Antitrust cases involving restrictive practices are often controversial, because they delve into specific contracts or agreements between firms that are allowed in some cases but not in others.

For example, if a product manufacturer is selling to a group of dealers who then sell to the general public it is illegal for the manufacturer to demand a **minimum resale price maintenance agreement**, which would require the dealers to sell for at least a certain minimum price. A minimum price contract is illegal because it would restrict competition among dealers. However, the manufacturer is legally allowed to “suggest” minimum prices and to stop selling to dealers who regularly undercut the suggested price. If you think this rule sounds like a fairly subtle distinction, you are right.

An **exclusive dealing** agreement between a manufacturer and a dealer can be legal or illegal. It is legal if the purpose of the contract is to encourage competition between dealers. For example, it is legal for the Ford Motor Company to sell its cars to only Ford dealers, for General Motors to sell to only GM dealers, and so on. However, exclusive deals may also limit competition. If one large retailer obtained the exclusive rights to be the sole distributor of televisions, computers, and audio equipment made by a number of companies, then this exclusive contract would have an anticompetitive effect on other retailers.

Tying sales happen when a customer is required to buy one product only if the customer also buys a second product. Tying sales are controversial because they force consumers to purchase a product that they may not actually want or need. Further, the additional, required products are not necessarily advantageous to the customer. Suppose that to purchase a popular DVD, the store required that you also purchase a portable TV of a certain model. These products are only loosely related, thus there is no reason to make the purchase of one contingent on the other. Even if a customer was interested in a portable TV, the tying to a particular model prevents the customer from having the option of selecting one from the numerous types available in the market. A related, but not identical, concept is called **bundling**, where two or more products are sold as one. Bundling typically offers an advantage for the consumer by allowing them to acquire multiple products or services for a better price. For example, several cable companies allow customers to buy products like cable, internet, and a phone line through a special price available through bundling. Customers are also welcome to purchase these products separately, but the price of bundling is usually more appealing.

In some cases, tying sales and bundling can be viewed as anticompetitive. However, in other cases they may be legal and even common. It is common for people to purchase season tickets to a sports team or a set of concerts so that they can be guaranteed tickets to the few contests or shows that are most popular and likely to sell out. Computer software manufacturers may often bundle together a number of different programs, even when the buyer wants only a few of the programs. Think about the software that is included in a new computer purchase, for example.

Recall from the chapter on Monopoly that **predatory pricing** occurs when the existing firm (or firms) reacts to a new firm by dropping prices very low, until the new firm is driven out of the market, at which point the existing firm raises prices again. This pattern of pricing is aimed at deterring the entry of new firms into the market. But in practice, it can be hard to figure out when pricing should be considered predatory. Say that American Airlines is flying between two cities, and a new airline starts flying between the same two cities, at a lower price. If American Airlines cuts its price to match the new entrant, is this predatory pricing? Or is it just market competition at work? A commonly proposed rule is that if a firm is selling for less than its average variable cost—that is, at a price where it should be shutting down—then there is evidence for predatory pricing. But calculating in the real world what costs are variable and what costs are fixed is often not obvious, either.

The Microsoft antitrust case embodies many of these gray areas in restrictive practices, as the next Clear it Up shows.

DID MICROSOFT® ENGAGE IN ANTICOMPETITIVE AND RESTRICTIVE PRACTICES?

The most famous restrictive practices case of recent years was a series of lawsuits by the U.S. government against Microsoft—lawsuits that were encouraged by some of Microsoft’s competitors. All sides admitted that Microsoft’s Windows program had a near-monopoly position in the market for the software used in general computer operating systems. All sides agreed that the software had many satisfied customers. All sides agreed that the capabilities of computer software that was compatible with Windows—both software produced by Microsoft and that produced by other companies—had expanded dramatically in the 1990s. Having a **monopoly** or a near-monopoly is not necessarily illegal in and of itself, but in cases where one company controls a great deal of the market, antitrust regulators look at any allegations of restrictive practices with special care.

The antitrust regulators argued that Microsoft had gone beyond profiting from its software innovations and its dominant position in the software market for operating systems, and had tried to use its market power in operating systems software to take over other parts of the software industry. For example, the government argued that Microsoft had engaged in an anticompetitive form of exclusive dealing by threatening computer makers that, if they did not leave another firm’s software off their machines (specifically, Netscape’s Internet browser), then Microsoft would not sell them its operating system software. Microsoft was accused by the government antitrust regulators of tying together its Windows operating system software, where it had a monopoly, with its Internet Explorer browser software, where it did not have a monopoly, and thus using this bundling as an anticompetitive tool. Microsoft was also accused of a form of predatory pricing; namely, giving away certain additional software products for free as part of Windows, as a way of driving out the competition from other makers of software.

In April 2000, a federal court held that Microsoft’s behavior had crossed the line into unfair competition, and recommended that the company be broken into two competing firms. However, that penalty was overturned on appeal, and in November 2002 Microsoft reached a settlement with the government that it would end its restrictive practices.

The concept of restrictive practices is continually evolving, as firms seek new ways to earn profits and government regulators define what is permissible and what is not. A situation where the law is evolving and changing is always somewhat troublesome, since laws are most useful and fair when firms know what they are in advance. In addition, since the law is open to interpretation, competitors who are losing out in the market can accuse successful firms of anticompetitive restrictive practices, and try to win through government regulation what they have failed to accomplish in the market. Officials at the Federal Trade Commission and the Department of Justice are, of course, aware of these issues, but there is no easy way to resolve them.

KEY CONCEPTS AND SUMMARY

Firms are blocked by antitrust authorities from openly colluding to form a cartel that will reduce output and raise prices. Companies sometimes attempt to find other ways around these restrictions and, consequently, many antitrust cases involve restrictive practices that can reduce competition in certain circumstances, like tie-in sales, bundling, and predatory pricing.

SELF-CHECK QUESTIONS

Why would a firm choose to use one or more of the anticompetitive practices described in Regulating Anticompetitive Behavior?

REVIEW QUESTIONS

1. What is a minimum resale price maintenance agreement? How might it reduce competition and when might it be acceptable?
2. What is exclusive dealing? How might it reduce competition and when might it be acceptable?
3. What is a tie-in sale? How might it reduce competition and when might it be acceptable?
4. What is predatory pricing? How might it reduce competition, and why might it be difficult to tell when it should be illegal?

CRITICAL THINKING QUESTIONS

1. Can you think of any examples of successful predatory pricing in the real world?
2. If you were developing a product (like a web browser) for a market with significant barriers to entry, how would you try to get your product into the market successfully?

GLOSSARY

bundling a situation in which multiple products are sold as one

exclusive dealing an agreement that a dealer will sell only products from one manufacturer

minimum resale price maintenance agreement an agreement that requires a dealer who buys from a manufacturer to sell for at least a certain minimum price

restrictive practices practices that reduce competition but that do not involve outright agreements between firms to raise prices or to reduce the quantity produced

tying sales a situation where a customer is allowed to buy one product only if the customer also buys another product

SOLUTIONS

Answers to Self-Check Questions

Because outright collusion to raise profits is illegal and because existing regulations include gray areas which firms may be able to exploit.

17.3 REGULATING NATURAL MONOPOLIES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Evaluate the appropriate competition policy for a natural monopoly
- Interpret a graph of regulatory choices
- Contrast cost-plus and price cap regulation

Most true monopolies today in the U.S. are regulated, natural monopolies. A natural monopoly poses a difficult challenge for competition policy, because the structure of costs and demand seems to make competition unlikely or costly. A **natural monopoly** arises when average costs are declining over the range of production that satisfies market demand. This typically happens when fixed costs are large relative to variable costs. As a result, one firm is able to supply the total quantity demanded in the market at lower cost than two or more firms—so splitting up the natural monopoly would raise the average cost of production and force customers to pay more.

Public utilities, the companies that have traditionally provided water and electrical service across much of the United States, are leading examples of natural monopoly. It would make little sense to argue that a local water company should be broken up into several competing companies, each with its own separate set of pipes and water supplies. Installing four or five identical sets of pipes under a city, one for each water company, so that each household could choose its own water provider, would be terribly costly. The same argument applies to the idea of having many competing companies for delivering electricity to homes, each with its own set of wires. Before the advent of wireless phones, the argument also applied to the idea of many different phone companies, each with its own set of phone wires running through the neighborhood.

THE CHOICES IN REGULATING A NATURAL MONOPOLY

So what then is the appropriate competition policy for a natural monopoly? Figure 1 illustrates the case of natural monopoly, with a market demand curve that cuts through the downward-sloping portion of the **average cost curve**. Points A, B, C, and F illustrate four of the main choices for regulation. Table 5 outlines the regulatory choices for dealing with a natural monopoly.

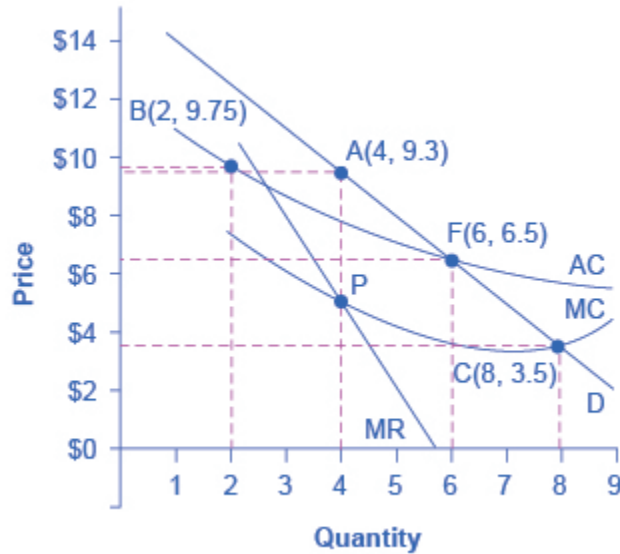


Figure 1. Regulatory Choices in Dealing with Natural Monopoly.

A natural monopoly will maximize profits by producing at the quantity where marginal revenue (MR) equals marginal costs (MC) and by then looking to the market demand curve to see what price to charge for this quantity. This monopoly will produce at point A, with a quantity of 4 and a price of 9.3. If antitrust regulators split this company exactly in half, then each half would produce at point B, with average costs of 9.75 and output of 2. The regulators might require the firm to produce where marginal cost crosses the market demand curve at point C. However, if the firm is required to produce at a quantity of 8 and sell at a price of 3.5, the firm will suffer from losses. The most likely choice is point F, where the firm is required to produce a quantity of 6 and charge a price of 6.5.

Quantity	Price	Total Revenue*	Marginal Revenue	Total Cost	Marginal Cost	Average Cost
1	14.7	14.7	—	11.0	—	11.00
2	12.4	24.7	10.0	19.5	8.5	9.75
3	10.6	31.7	7.0	25.5	6.0	8.50
4	9.3	37.2	5.5	31.0	5.5	7.75
5	8.0	40.0	2.8	35.0	4.0	7.00
6	6.5	39.0	−1.0	39.0	4.0	6.50
7	5.0	35.0	−4.0	42.0	3.0	6.00
8	3.5	28.0	−7.0	45.5	3.5	5.70
9	2.0	18.0	−10.0	49.5	4.0	5.5

Table 5. Regulatory Choices in Dealing with Natural Monopoly. (*Total Revenue is given by multiplying price and quantity. However, some of the price values in this table have been rounded for ease of presentation.)

The first possibility is to leave the natural monopoly alone. In this case, the monopoly will follow its normal approach to maximizing profits. It determines the quantity where $MR = MC$, which happens at point P at a quantity of 4. The firm then looks to point A on the demand curve to find that it can charge a price of 9.3 for that profit-maximizing quantity. Since the price is above the average cost curve, the natural monopoly would earn economic profits.

A second outcome arises if antitrust authorities decide to divide the company, so that the new firms can compete. As a simple example, imagine that the company is cut in half. Thus, instead of one large firm producing a quantity of 4, two half-size firms each produce a quantity of 2. Because of the declining average cost curve (AC), the average cost of production for each of the half-size companies each producing 2, as shown at point B, would be 9.75, while the average cost of production for a larger firm producing 4 would only be 7.75. Thus, the economy would become less productively efficient, since the good is being produced at a higher average cost. In a situation with a downward-sloping average cost curve, two smaller firms will always have higher average costs of production than one larger firm for any quantity of total output. In addition, the antitrust authorities must worry that splitting the natural monopoly into pieces may be only the start of their problems. If one of the two firms grows larger than the other, it will have lower average costs and may be able to drive its competitor out of the market. Alternatively, two firms in a market may discover subtle ways of coordinating their behavior and keeping prices high. Either way, the result will not be the greater competition that was desired.

A third alternative is that regulators may decide to set prices and quantities produced for this industry. The regulators will try to choose a point along the market demand curve that benefits both consumers and the broader social interest. Point C illustrates one tempting choice: the regulator requires that the firm produce the quantity of output where marginal cost crosses the demand curve at an output of 8, and charge the price of 3.5, which is equal to **marginal cost** at that point. This rule is appealing because it requires price to be set equal to marginal cost, which is what would occur in a perfectly competitive market, and it would assure consumers a higher quantity and lower price than at the monopoly choice A. In fact, efficient allocation of resources would occur at point C, since the value to the consumers of the last unit bought and sold in this market is equal to the marginal cost of producing it.

Attempting to bring about point C through force of regulation, however, runs into a severe difficulty. At point C, with an output of 8, a price of 3.5 is below the average cost of production, which is 5.7, and so if the firm charges a price of 3.5, it will be suffering losses. Unless the regulators or the government offer the firm an ongoing public subsidy (and there are numerous political problems with that option), the firm will lose money and go out of business.

Perhaps the most plausible option for the regulator is point F; that is, to set the price where AC crosses the demand curve at an output of 6 and a price of 6.5. This plan makes some sense at an intuitive level: let the natural monopoly charge enough to cover its average costs and earn a normal rate of profit, so that it can continue operating, but prevent the firm from raising prices and earning abnormally high monopoly profits, as it would at the monopoly choice A. Of course, determining this level of output and price with the political pressures, time constraints, and limited information of the real world is much harder than identifying the point on a graph. For more on the problems that can arise from a centrally determined price, see the discussion of price floors and price ceilings in Demand and Supply.

COST-PLUS VERSUS PRICE CAP REGULATION

Indeed, regulators of public utilities for many decades followed the general approach of attempting to choose a point like F in Figure 1. They calculated the average cost of production for the water or electricity companies, added in an amount for the normal rate of profit the firm should expect to earn, and set the price for consumers accordingly. This method was known as **cost-plus regulation**.

Cost-plus regulation raises difficulties of its own. If producers are reimbursed for their costs, plus a bit more, then at a minimum, producers have less reason to be concerned with high costs—because they can just pass them along in higher prices. Worse, firms under cost-plus regulation even have an incentive to generate high costs by building huge factories or employing lots of staff, because what they can charge is linked to the costs they incur.

Thus, in the 1980s and 1990s, some regulators of public utilities began to use **price cap regulation**, where the regulator sets a price that the firm can charge over the next few years. A common pattern was to require a price that declined slightly over time. If the firm can find ways of reducing its costs more quickly than the price caps, it can make a high level of profits. However, if the firm cannot keep up with the price caps or suffers bad luck in the market, it may suffer losses. A few years down the road, the regulators will then set a new series of price caps based on the firm's performance.

Price cap regulation requires delicacy. It will not work if the price regulators set the price cap unrealistically low. It may not work if the market changes dramatically so that the firm is doomed to incurring losses no matter what it does—say, if energy prices rise dramatically on world markets, then the company selling natural gas or heating oil to homes may not be able to meet price caps that seemed reasonable a year or two ago. But if the regulators compare the prices with producers of the same good in other areas, they can, in effect, pressure a natural monopoly in one area to compete with the prices being charged in other areas. Moreover, the possibility of earning greater profits or experiencing losses—instead of having an average rate of profit locked in every year by cost-plus regulation—can provide the natural monopoly with incentives for efficiency and innovation.

With natural monopoly, market competition is unlikely to take root, so if consumers are not to suffer the high prices and restricted output of an unrestricted monopoly, government regulation will need to play a role. In attempting to design a system of price cap regulation with flexibility and incentive, government regulators do not have an easy task.

KEY CONCEPTS AND SUMMARY

In the case of a natural monopoly, market competition will not work well and so, rather than allowing an unregulated monopoly to raise price and reduce output, the government may wish to regulate price and/or output. Common examples of regulation are public utilities, the regulated firms that often provide electricity and water service.

Cost-plus regulation refers to government regulation of a firm which sets the price that a firm can charge over a period of time by looking at the firm's accounting costs and then adding a normal rate of profit. Price cap regulation refers to government regulation of a firm where the government sets a price level several years in advance. In this case, the firm can either make high profits if it manages to produce at lower costs or sell a higher quantity than expected or suffer low profits or losses if costs are high or it sells less than expected.

SELF-CHECK QUESTIONS

1. Urban transit systems, especially those with rail systems, typically experience significant economies of scale in operation. Consider the transit system whose data is given in the Table 6. Note that the quantity is in millions of riders.

Demand:	Quantity	1	2	3	4	5	6	7	8	9	10
	Price	10	9	8	7	6	5	4	3	2	1
	Marginal Revenue	10	8	6	4	2	0	-2	-4	-6	-8
Costs:	Marginal Cost	9	6	5	3	2	3	4	5	7	10
	Average Cost	9	7.5	6.7	5.8	5	4.7	4.6	4.6	4.9	5.4

Table 6.

- Draw the demand, marginal revenue, marginal cost, and average cost curves. Do they have the normal shapes?
2. From the graph you drew to answer Self-Check Question 1, would you say this transit system is a natural monopoly? Justify.

REVIEW QUESTIONS

1. If public utilities are a natural monopoly, what would be the danger in deregulating them?
2. If public utilities are a natural monopoly, what would be the danger in splitting them up into a number of separate competing firms?
3. What is cost-plus regulation?
4. What is price cap regulation?

CRITICAL THINKING QUESTIONS

1. In the middle of the twentieth century, major U.S. cities had multiple competing city bus companies. Today, there is usually only one and it runs as a subsidized, regulated monopoly. What do you suppose caused the change?
2. Why are urban areas willing to subsidize urban transit systems? Does the argument for subsidies make sense to you?

PROBLEM

Use Table 6 to answer the following questions.

1. If the transit system was allowed to operate as an unregulated monopoly, what output would it supply and what price would it charge?
2. If the transit system was regulated to operate with no subsidy (i.e., at zero economic profit), what approximate output would it supply and what approximate price would it charge?
3. If the transit system was regulated to provide the most allocatively efficient quantity of output, what output would it supply and what price would it charge? What subsidy would be necessary to insure this efficient provision of transit services?

GLOSSARY

cost-plus regulation when regulators permit a regulated firm to cover its costs and to make a normal level of profit

price cap regulation when the regulator sets a price that a firm cannot exceed over the next few years

SOLUTIONS

Answers to Self-Check Questions

1. Yes, all curves have normal shapes.

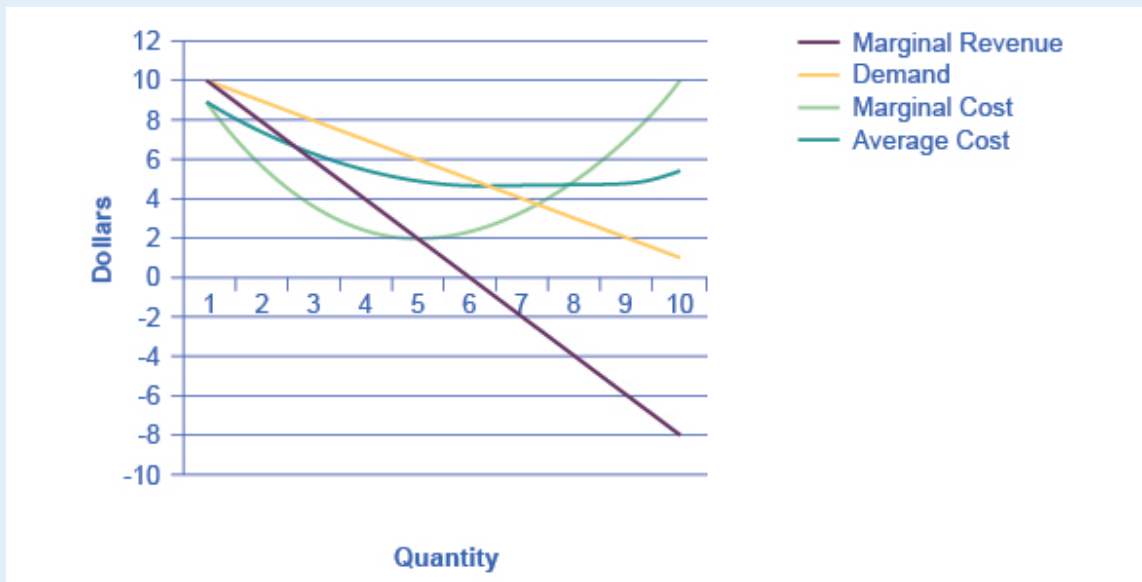


Figure 2.

2. Yes it is a natural monopoly because average costs decline over the range that satisfies the market demand. For example, at the point where the demand curve and the average cost curve meet, there are economies of scale.

17.4 THE GREAT DEREGULATION EXPERIMENT

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Evaluate the effectiveness of price regulation and antitrust policy
- Explain regulatory capture and its significance

Governments at all levels across the United States have regulated prices in a wide range of industries. In some cases, like water and electricity that have natural monopoly characteristics, there is some room in economic theory for such regulation. But once politicians are given a basis to intervene in markets and to choose prices and quantities, it is hard to know where to stop.

DOUBTS ABOUT REGULATION OF PRICES AND QUANTITIES

Beginning in the 1970s, it became clear to policymakers of all political leanings that the existing price regulation was not working well. The United States carried out a great policy experiment—the **deregulation** discussed in Monopoly—removing government controls over prices and quantities produced in airlines, railroads, trucking, intercity bus travel, natural gas, and bank interest rates. The Clear it Up discusses the outcome of deregulation in one industry in particular—airlines.

WHAT ARE THE RESULTS OF AIRLINE DEREGULATION?

Why did the pendulum swing in favor of deregulation? Consider the airline industry. In the early days of air travel, no airline could make a profit just by flying passengers. Airlines needed something else to carry and the Postal Service provided that something with airmail. And so the first U.S. government regulation of the airline industry happened through the Postal Service, when in 1926 the Postmaster General began giving airlines permission to fly certain routes based on the needs of mail delivery—and the airlines took some passengers along for the ride. In 1934, the Postmaster General was charged by the antitrust authorities with colluding with the major airlines of that day to monopolize the nation's airways. In 1938, the Civil Aeronautics Board (CAB) was created to regulate airfares and routes instead. For 40 years, from 1938 to 1978, the CAB approved all fares, controlled all entry and exit, and specified which airlines could fly which routes. There was zero entry of new airlines on the main routes across the country for 40 years, because the CAB did not think it was necessary.

In 1978, the Airline Deregulation Act took the government out of the business of determining airfares and schedules. The new law shook up the industry. Famous old airlines like Pan American, Eastern, and Braniff went bankrupt and disappeared. Some new airlines like People Express were created—and then vanished.

The greater competition from deregulation reduced airfares by about one-third over the next two decades, saving consumers billions of dollars a year. The average flight used to take off with just half its seats full; now it is two-thirds full, which is far more efficient. Airlines have also developed hub-and-spoke systems, where planes all fly into a central hub city at a certain time and then depart. As a result, one can fly between any of the spoke cities with just one connection—and there is greater service to more cities than before deregulation. With lower fares and more service, the number of air passengers doubled from the late 1970s to the start of the 2000s—an increase that, in turn, doubled the number of jobs in the airline industry. Meanwhile, with the watchful oversight of government safety inspectors, commercial air travel has continued to get safer over time.

The U.S. airline industry is far from perfect. For example, a string of mergers in recent years has raised concerns over how competition might be compromised.

One difficulty with government price regulation is what economists call **regulatory capture**, in which the firms supposedly being regulated end up playing a large role in setting the regulations that they will follow. When the airline industry was being regulated, for example, it suggested appointees to the regulatory board, sent lobbyists to argue with the board, provided most of the information on which the board made decisions, and offered well-paid jobs to at least some of the people leaving the board. In this situation, consumers can easily end up being not very well represented by the regulators. The result of regulatory capture is that government price regulation can often become a way for existing competitors to work together to reduce output, keep prices high, and limit competition.

THE EFFECTS OF DEREGULATION

Deregulation, both of airlines and of other industries, has its negatives. The greater pressure of competition led to entry and exit. When firms went bankrupt or contracted substantially in size, they laid off workers who had to find other jobs. Market competition is, after all, a full-contact sport.

A number of major accounting scandals involving prominent corporations such as Enron, Tyco International, and WorldCom led to the **Sarbanes-Oxley Act** in 2002. Sarbanes-Oxley was designed to increase confidence in financial information provided by public corporations to protect investors from accounting fraud.

The Great Recession which began in late 2007 and which the U.S. economy is still struggling to recover from was caused at least in part by a global financial crisis, which began in the United States. The key component of the crisis was the creation and subsequent failure of several types of unregulated financial assets, such as collateralized mortgage obligations (CMOs, a type of mortgage-backed security), and credit default swaps (CDSs, insurance contracts on assets like CMOs that provided a payoff even if the holder of the CDS did not own the CMO). Many of these assets were rated very safe by private credit rating agencies such as Standard & Poors, Moody's, and Fitch.

The collapse of the markets for these assets precipitated the financial crisis and led to the failure of Lehman Brothers, a major investment bank, numerous large commercial banks, such as Wachovia, and even the Federal National Mortgage Corporation (Fannie Mae), which had to be nationalized—that is, taken over by the federal government. One response to the financial crisis was the **Dodd-Frank Act**, which attempted major reforms of the financial system. The legislation's purpose, as noted on dodd-frank.com is:

To promote the financial stability of the United States by improving accountability and transparency in the

financial system, to end “too big to fail,” to protect the American taxpayer by ending bailouts, [and] to protect consumers from abusive financial services practices. . .

We will explore the financial crisis and the Great Recession in more detail in the macroeconomic chapters of this book, but for now it should be clear that many Americans have grown disenchanted with deregulation, at least of financial markets.

All market-based economies operate against a background of laws and regulations, including laws about enforcing contracts, collecting taxes, and protecting health and the environment. The government policies discussed in this chapter—like blocking certain anticompetitive mergers, ending restrictive practices, imposing price cap regulation on natural monopolies, and deregulation—demonstrate the role of government to strengthen the incentives that come with a greater degree of competition.

MORE THAN COOKING, HEATING, AND COOLING

What did the Federal Trade Commission (FTC) decide on the Kinder Morgan / El Paso Corporation merger? After careful examination, federal officials decided there was only one area of significant overlap that might provide the merged firm with strong market power. The FTC approved the merger, provided Kinder Morgan divest itself of the overlap area. Tallgrass purchased Kinder Morgan Interstate Gas Transmission, Trailblazer Pipeline Co. LLC, two processing facilities in Wyoming, and Kinder Morgan’s 50 percent interest in the Rockies Express Pipeline to meet the FTC requirements. The FTC was attempting to strike a balance between potential cost reductions resulting from economies of scale and concentration of market power.

Did the price of natural gas decrease? Yes, rather significantly. In 2010, the wellhead price of natural gas was \$4.48 per thousand cubic foot; in 2012 the price had fallen to just \$2.66. Was the merger responsible for the large drop in price? The answer is uncertain. The larger contributor to the sharp drop in price was the overall increase in the supply of natural gas. More and more natural gas was able to be recovered by fracturing shale deposits, a process called fracking. Fracking, which is controversial for environmental reasons, enabled the recovery of known reserves of natural gas that previously were not economically feasible to tap. Kinder Morgan’s control of 80,000-plus miles of pipeline likely made moving the gas from wellheads to end users smoother and allowed for an even greater benefit from the increased supply.

KEY CONCEPTS AND SUMMARY

The U.S. economy experienced a wave of deregulation in the late 1970s and early 1980s, when a number of government regulations that had set prices and quantities produced in a number of industries were eliminated. Major accounting scandals in the early 2000s and, more recently, the Great Recession have spurred new regulation to prevent similar occurrences in the future. Regulatory capture occurs when the industries being regulated end up having a strong influence over what regulations exist.

SELF-CHECK QUESTIONS

Use the following information to answer the next three questions. In the years before wireless phones, when telephone technology required having a wire running to every home, it seemed plausible that telephone service had diminishing average costs and might need to be regulated like a natural monopoly. For most of the twentieth century, the national U.S. phone company was AT&T, and the company functioned as a regulated monopoly. Think about the deregulation of

the U.S. telecommunications industry that has happened over the last few decades. (This is not a research assignment, but a thought assignment based on what you have learned in this chapter.)

1. What real world changes made the deregulation possible?
2. What are some of the benefits of the deregulation?
3. What might some of the negatives of deregulation be?

REVIEW QUESTIONS

1. What is deregulation? Name some industries that have been deregulated in the United States.
2. What is regulatory capture?
3. Why does regulatory capture reduce the persuasiveness of the case for regulating industries for the benefit of consumers?

CRITICAL THINKING QUESTIONS

1. Deregulation, like all changes in government policy, always has pluses and minuses. What do you think some of the minuses might be for airline deregulation?
2. Do you think it is possible for government to outlaw everything that businesses could do wrong? If so, why does government not do that? If not, how can regulation stay ahead of rogue businesses that push the limits of the system until it breaks?

REFERENCES

Bishop, Todd. 2014. "Microsoft Exec Admits New Reality: Market Share No Longer 90% — It's 14%." *GeekWire*. Accessed March 27, 2015. <http://www.geekwire.com/2014/microsoft-exec-admits-new-reality-market-share-longer-90-14/>.

Catan, T., & Dezember, R. "Kinder-El Paso Merger to Face Antitrust Scrutiny," *Wall Street Journal*. October 19, 2011.

Collins, A. "Tallgrass Energy to Acquire Kinder Morgan Assets for \$1.8B." *The Middle Market*, Accessed August 2013. <http://www.themiddlemarket.com/news/tallgrass-energy-to-acquire-kinder-morgan-assets-for-1-point-8-billion-232862-1.html>.

Conoco Phillips. "Why Natural Gas." Accessed August 2013. <http://www.powerincooperation.com/en/pages/useful.html?gclid=COXg7rH3hrgCFWlp7AodtkgA3g>.

De la Merced, M. (2012, August 20). "Kinder Morgan to Sell Assets to Tallgrass for \$1.8 Billion." *The New York Times*. August 20, 2012.

The Federal Trade Commission. n.d. "The Federal Trade Commission's (FTC) Mission." Accessed

March 27, 2015. <https://www.ftc.gov/system/files/documents/reports/2014-one-page-ftc-performance-snapshot/150218fy14snapshot.pdf>.

Kahn, C. “Kinder Morgan to Buy El Paso Corp. for \$20.7B.” *USA Today*. October 17, 2011.

Kinder Morgan. (2013). “Investor Information.” Accessed August 2013. <http://www.kindermorgan.com/investor/>.

Rogowsky, Mark. 2014. “There’d Be No Wireless Wars Without The Blocked T-Mobile Merger, So Where Does That Leave Comcast-TWC?” *Forbes.com*. Accessed March 12, 2015. <http://www.forbes.com/sites/markrogowsky/2014/08/27/t-mobile-and-sprint-continue-to-battle-thanks-to-the-government/>.

U.S. Energy Information Administration (a). “Natural Gas Consumption by End User.” Accessed May 31, 2013. http://www.eia.gov/dnav/ng/ng_cons_sum_dcu_nus_a.htm.

U.S. Energy Information Administration (b). “Natural Gas Prices.” Accessed June 28, 2013. http://www.eia.gov/dnav/ng/ng_pri_sum_dcu_nus_a.htm.

The Wall Street Journal. 2015. “Auto Sales.” Accessed April 10, 2015. http://online.wsj.com/mdc/public/page/2_3022-autosales.html.

GLOSSARY

regulatory capture when the firms supposedly being regulated end up playing a large role in setting the regulations that they will follow and as a result, they “capture” the people doing the regulation, usually through the promise of a job in that “regulated” industry once their term in government has ended.

SOLUTIONS

Answers to Self-Check Questions

1. Improvements in technology that allowed phone calls to be made via microwave transmission, communications satellites, and other wireless technologies.
2. More consumer choice. Cheaper phone calls, especially long distance. Better-quality phone service in many cases. Cheaper, faster, and better-quality data transmission. Spin-off technologies like free Internet-based calling and video calling.
3. More choice can sometimes make for difficult decisions—not knowing if you got the best plan for your situation, for example. Some phone service providers are less reliable than AT&T used to be.

CHAPTER 18. ENVIRONMENTAL PROTECTION AND NEGATIVE EXTERNALITIES

INTRODUCTION TO ENVIRONMENTAL PROTECTION AND NEGATIVE EXTERNALITIES



Figure 1. Environmental Debate. Across the country, countless people have protested, even risking arrest, against the Keystone XL Pipeline. (Credit: modification of image by “NoKXL”/Flickr Creative Commons)

KEYSTONE XL

You might have heard about Keystone XL in the news. It is a pipeline system designed to bring oil from Canada to the refineries near the Gulf of Mexico, as well as to boost crude oil production in the United States. While a private company, TransCanada, will own the pipeline, U.S. government approval is required because of its size and location. The pipeline is being built in four phases, with the first two currently in operation, bringing oil from Alberta, Canada, east across Canada, south through the United States into Nebraska and Oklahoma, and northeast again to Illinois. The third and fourth phases of the project, known as Keystone XL, would create a pipeline southeast from Alberta straight to Nebraska, and then from Oklahoma to the Gulf of Mexico.

Sounds like a great idea, right? A pipeline that would move much needed crude oil to the Gulf refineries would increase oil production for manufacturing needs, reduce price pressure at the gas pump, and increase overall economic growth. Supporters argue that the pipeline is one of the safest pipelines built yet, and would reduce America's dependence on politically vulnerable Middle Eastern oil imports.

Not so fast, say its critics. The Keystone XL would be constructed over an enormous aquifer (one of the largest in the world) in the Midwest, and through an environmentally fragile area in Nebraska, causing great concern among environmentalists about possible destruction to the natural surroundings. They argue that leaks could taint valuable water sources and construction of the pipeline could disrupt and even harm indigenous species. Environmentalist groups have fought government approval of the proposed construction of the pipeline, and as of press time the pipeline projects remain stalled.

Of course, environmental concerns matter when discussing issues related to economic growth. But how much should they factor in? In the case of the pipeline, how do we know how much damage it would cause when we do not know how to put a value on the environment? Would the benefits of the pipeline outweigh the opportunity cost? The issue of how to balance economic progress with unintended effects on our planet is the subject of this chapter.

CHAPTER OBJECTIVES

Introduction to Environmental Protection and Negative Externalities

In this chapter, you will learn about:

- The Economics of Pollution
- Command-and-Control Regulation
- Market-Oriented Environmental Tools
- The Benefits and Costs of U.S. Environmental Laws
- International Environmental Issues
- The Tradeoff between Economic Output and Environmental Protection

In 1969, the Cuyahoga River in Ohio was so polluted that it spontaneously burst into flame. Air pollution was so bad at that time that Chattanooga, Tennessee was a city where, as an article from Sports Illustrated put it: “the death rate from tuberculosis was double that of the rest of Tennessee and triple that of the rest of the United States, a city in which the filth in the air was so bad it melted nylon stockings off women’s legs, in which executives kept supplies of clean white shirts in their offices so they could change when a shirt became too gray to be presentable, in which headlights were turned on at high noon because the sun was eclipsed by the gunk in the sky.”

The problem of pollution arises for every economy in the world, whether high-income or low-income, and whether market-oriented or command-oriented. Every country needs to strike some balance between production and environmental quality. This chapter begins by discussing how firms may fail to take certain social costs, like pollution, into their planning if they do not need to pay these costs. Traditionally, policies for environmental protection have focused on governmental limits on how much of each pollutant could be emitted. While this approach has had some success, economists have suggested a range of more flexible, market-oriented policies that reduce pollution at a lower cost. We will consider both approaches, but first let’s see how economists frame and analyze these issues.

18.1 THE ECONOMICS OF POLLUTION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain and give examples of positive and negative externalities
- Identify equilibrium price and quantity
- Evaluate how firms can contribute to market failure

From 1970 to 2012, the U.S. population increased by one-third and the size of the U.S. economy more than doubled. Since the 1970s, however, the United States, using a variety of anti-pollution policies, has made genuine progress against a number of pollutants. Table 1 lists the change in carbon dioxide emissions by users of energy (from residential to industrial) according to the U.S. Energy Information Administration (EIA). The table shows that emissions of certain key air pollutants declined substantially from 2007 to 2012; they dropped 730 million metric tons (MMT) a year—a 12% reduction. This seems to indicate that progress has been made in the United States in reducing overall carbon dioxide emissions, which cause greenhouse gases.

End-use Sector	Primary Fossil Fuels			Purchased Electric Power	Total Primary Fossil Fuels
	Coal	Petroleum	Natural Gas		
Residential	(0)	(14)	(31)	(134)	(179)
Commercial	(2)	(2)	(7)	(126)	(136)
Industrial	(40)	(62)	31	(118)	(191)
Transportation	0	(228)	5	(1)	(224)
Power	(464)	(36)	(122)	–	–
Change 2007–2012	(508)	(342)	121	(378)	(730)

Table 1. U.S. Carbon Dioxide (CO₂) Emissions from Fossil Fuels Consumed 2007–2012, Million Metric Tons (MMT) per Year. (Source: EIA Monthly Energy Review)

Despite the gradual reduction in emissions from fossil fuels, many important environmental issues remain. Along with the still high levels of air and water pollution, other issues include hazardous waste disposal, destruction of wetlands and other wildlife habitats, and the impact on human health from pollution.

EXTERNALITIES

Private markets, such as the cell phone industry, offer an efficient way to put buyers and sellers together and determine what goods are produced, how they are produced, and who gets them. The principle that voluntary exchange benefits both buyers and sellers is a fundamental building block of the economic way of thinking. But what happens when a voluntary exchange affects a third party who is neither the buyer nor the seller?

As an example, consider a concert producer who wants to build an outdoor arena that will host country music concerts a half-mile from your neighborhood. You will be able to hear these outdoor concerts while sitting on your back porch—or perhaps even in your dining room. In this case, the sellers and buyers of concert tickets may both be quite satisfied with their voluntary exchange, but you have no voice in their market transaction. The effect of a market exchange on a third party who is outside or “external” to the exchange is called an **externality**. Because externalities that occur in market transactions affect other parties beyond those involved, they are sometimes called **spillovers**.

Externalities can be negative or positive. If you hate country music, then having it waft into your house every night would be a **negative externality**. If you love country music, then what amounts to a series of free concerts would be a **positive externality**.

POLLUTION AS A NEGATIVE EXTERNALITY

Pollution is a negative externality. Economists illustrate the **social costs** of production with a demand and supply diagram. The social costs include the private costs of production incurred by the company and the external costs of pollution that are passed on to society. Figure 1 shows the demand and supply for manufacturing refrigerators. The demand curve (D) shows the quantity demanded at each price. The supply curve (S_{private}) shows the quantity of refrigerators supplied by all the firms at each price if they are taking only their private costs into account and they are allowed to emit pollution at zero cost. The market equilibrium (E_0), where quantity supplied and quantity demanded are equal, is at a price of \$650 and a quantity of 45,000. This information is also reflected in the first three columns of Table 2.

Price	Quantity Demanded	Quantity Supplied before Considering Pollution Cost	Quantity Supplied after Considering Pollution Cost
\$600	50,000	40,000	30,000
\$650	45,000	45,000	35,000
\$700	40,000	50,000	40,000
\$750	35,000	55,000	45,000
\$800	30,000	60,000	50,000
\$850	25,000	65,000	55,000
\$900	20,000	70,000	60,000

Table 2. A Supply Shift Caused by Pollution Costs

However, as a by-product of the metals, plastics, chemicals and energy that are used in manufacturing refrigerators, some pollution is created. Let’s say that, if these pollutants were emitted into the air and water, they would create costs of \$100 per refrigerator produced. These costs might occur because of injuries to human health, property values, wildlife habitat, reduction of recreation possibilities, or

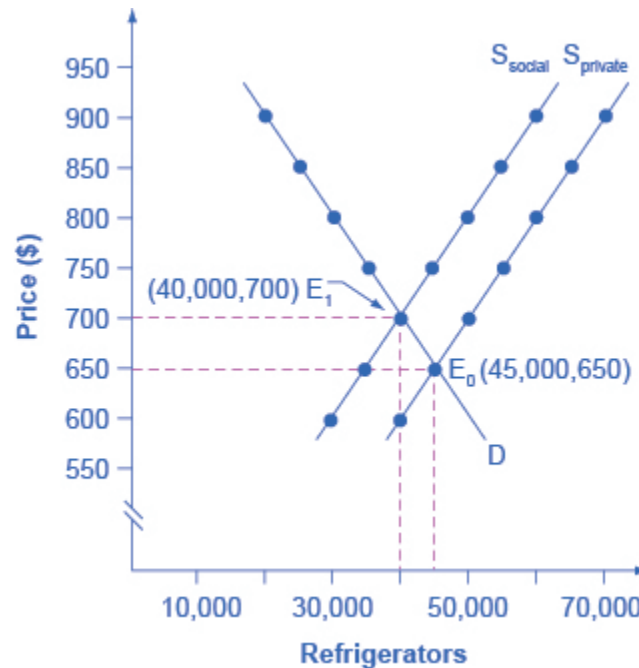


Figure 1. Taking Social Costs into Account: A Supply Shift. If the firm takes only its own costs of production into account, then its supply curve will be S_{private} , and the market equilibrium will occur at E_0 . Accounting for additional external costs of \$100 for every unit produced, the firm's supply curve will be S_{social} . The new equilibrium will occur at E_1 .

because of other negative impacts. In a market with no anti-pollution restrictions, firms can dispose of certain wastes absolutely free. Now imagine that firms which produce refrigerators must factor in these **external costs** of pollution—that is, the firms have to consider not only the costs of labor and materials needed to make a refrigerator, but also the broader costs to society of injuries to health and other values caused by pollution. If the firm is required to pay \$100 for the **additional external costs** of pollution each time it produces a refrigerator, production becomes more costly and the entire supply curve shifts up by \$100.

As illustrated in the fourth column of Table 2 and in Figure 1, the firm will need to receive a price of \$700 per refrigerator and produce a quantity of 40,000—and the firm's new supply curve will be S_{social} . The new equilibrium will occur at E_1 , taking the additional external costs of pollution into account results in a higher price, a lower quantity of production, and a lower quantity of pollution. The following Work It Out feature will walk you through an example, this time with musical accompaniment.

IDENTIFYING THE EQUILIBRIUM PRICE AND QUANTITY

Table 3 shows the supply and demand conditions for a firm that will play trumpets on the streets when requested. Output is measured as the number of songs played.

Price	Quantity Demanded	Quantity Supplied without paying the costs of the externality	Quantity Supplied after paying the costs of the externality
\$20	0	10	8
\$18	1	9	7
\$15	2.5	7.5	5.5
\$12	4	6	4
\$10	5	5	3
\$5	7.5	2.5	0.5

Table 3. Supply and Demand Conditions for a Trumpet-Playing Firm

Step 1. Determine the negative externality in this situation. To do this, you must think about the situation described and consider all parties that might be impacted. A negative externality might be the increase in noise pollution in the area where the firm is playing.

Step 2. Identify the equilibrium price and quantity when only private costs are taken into account, and then when social costs are taken into account. Remember that equilibrium is where the quantity demanded is equal to the quantity supplied.

Step 3. Look down the columns to where the quantity demanded (the second column) is equal to the “quantity supplied without paying the costs of the externality” (the third column). Then refer to the first column of that row to determine the equilibrium price. In this case, the equilibrium price and quantity when only private costs are taken into account would be at a price of \$10 and a quantity of five.

Step 4. Identify the equilibrium price and quantity when the additional external costs are taken into account. Look down the columns of quantity demanded (the second column) and the “quantity supplied after paying the costs of the externality” (the fourth column) then refer to the first column of that row to determine the equilibrium price. In this case, the equilibrium will be at a price of \$12 and a quantity of four.

Step 5. Consider how taking the externality into account affects the equilibrium price and quantity. Do this by comparing the two equilibrium situations. If the firm is forced to pay its additional external costs, then production of trumpet songs becomes more costly, and the supply curve will shift up.

Remember that the supply curve is based on choices about production that firms make while looking at their marginal costs, while the demand curve is based on the benefits that individuals perceive while maximizing utility. If no externalities existed, private costs would be the same as the costs to society as a whole, and private benefits would be the same as the benefits to society as a whole. Thus, if no externalities existed, the interaction of demand and supply will coordinate social costs and benefits.

However, when the externality of pollution exists, the supply curve no longer represents all social costs. Because externalities represent a case where markets no longer consider all social costs, but only some of them, economists commonly refer to externalities as an example of **market failure**. When there is market failure, the private market fails to achieve efficient output, because either firms do not account for all costs incurred in the production of output and/or consumers do not account for all benefits obtained (a positive externality). In the case of pollution, at the market output, social costs of production exceed social benefits to consumers, and the market produces too much of the product.

We can see a general lesson here. If firms were required to pay the social costs of pollution, they would create less pollution but produce less of the product and charge a higher price. In the next module, we will explore how governments require firms to take the social costs of pollution into account.

KEY CONCEPTS AND SUMMARY

Economic production can cause environmental damage. This tradeoff arises for all countries, whether high-income or low-income, and whether their economies are market-oriented or command-oriented.

An externality occurs when an exchange between a buyer and seller has an impact on a third party who is not part of the exchange. An externality, which is sometimes also called a spillover, can have a negative or a positive impact on the third party. If those parties imposing a negative externality on others had to take the broader social cost of their behavior into account, they would have an incentive to reduce the production of whatever is causing the negative externality. In the case of a positive externality, the third party is obtaining benefits from the exchange between a buyer and a seller, but they are not paying for these benefits. If this is the case, then markets would tend to under produce output because suppliers are not aware of the additional demand from others. If the parties that are generating benefits to others would be somehow compensated for these external benefits, they would have an incentive to increase production of whatever is causing the positive externality.

SELF-CHECK QUESTIONS

1. Identify the following situations as an example of a negative or a positive externality:
 - a. You are a birder (bird watcher), and your neighbor has put up several birdhouses in the yard as well as planting trees and flowers that attract birds.
 - b. Your neighbor paints his house a hideous color.
 - c. Investments in private education raise your country's standard of living.
 - d. Trash dumped upstream flows downstream right past your home.
 - e. Your roommate is a smoker, but you are a nonsmoker.
2. Identify whether the market supply curve will shift right or left or will stay the same for the following:
 - a. Firms in an industry are required to pay a fine for their emissions of carbon dioxide.
 - b. Companies are sued for polluting the water in a river.
 - c. Power plants in a specific city are not required to address the impact of their emissions on the quality of air.
 - d. Companies that use fracking to remove oil and gas from rock are required to clean up the damage.
3. For each of your answers to Self-Check Question 2, will equilibrium price rise or fall or stay the same?
4. The supply and demand conditions for a manufacturing firm are given in Table 4. The third column represents a supply curve without taking the social cost of pollution into account. The fourth column represents the supply curve when the firm is required to take the social cost of pollution into account. Identify the equilibrium before the social cost of production is included and after the social cost of production is included.

Price	Quantity Demanded	Quantity Supplied without paying the cost of the pollution	Quantity Supplied after paying the cost of the pollution
\$10	450	400	250
\$15	440	440	290
\$20	430	480	330
\$25	420	520	370
\$30	410	560	410

Table 4.

REVIEW QUESTIONS

1. What is an externality?
2. Give an example of a positive externality and an example of a negative externality.
3. What is the difference between private costs and social costs?
4. In a market without environmental regulations, will the supply curve for a firm take into account private costs, external costs, both, or neither? Explain.

CRITICAL THINKING QUESTIONS

Suppose you want to put a dollar value on the external costs of carbon emissions from a power plant. What information or data would you obtain to measure the external [not social] cost?

PROBLEMS

1. Show the market for cigarettes in equilibrium, assuming that there are no laws banning smoking in public. Label the equilibrium private market price and quantity as P_m and Q_m . Add whatever is needed to the model to show the impact of the negative externality from second-hand smoking. (Hint: In this case it is the consumers, not the sellers, who are creating the negative externality.) Label the social optimal output and price as P_e and Q_e . On the graph, shade in the deadweight loss at the market output.
2. Refer to Table 2. The externality created by the production of refrigerators was \$100. However, once both the private and additional external costs were taken into consideration, the market price increased by only \$50. If the external costs were \$100 why did the price only increase by \$50 when all costs were taken into account?
3. Table 5, shows the supply and demand conditions for a firm that will play trumpets on the streets when requested. Q_{s1} is the quantity supplied without social costs. Q_{s2} is the quantity supplied with social costs. What is the negative externality in this situation? Identify the equilibrium price and quantity when only private costs are taken into account, and then when social costs are taken into account. How does taking the externality into account affect the equilibrium price and quantity?

P	Qd	Qs1	Qs2
\$20	0	10	8
\$18	1	9	7
\$15	2.5	7.5	5.5
\$12	4	6	4
\$10	5	5	3
\$5	7.5	2.5	0.5

Table 5.

REFERENCES

Johnson, Oscar William. "Back on Track: Earth Day Success Story; The Chattanooga Choo-Choo No Longer Spews Foul Air." *Sports Illustrated*. April 30, 1990. <http://www.si.com/vault/1990/04/30/121923/back-on-track-earth-day-success-story-the-chattanooga-choo-choo-no-longer-spews-foul-air>.

U.S. Energy Information Administration. "Total Energy: Monthly Energy Review." *U.S. Department of Energy*. Accessed December 19, 2013. <http://www.eia.gov/totalenergy/data/monthly/>.

GLOSSARY

additional external cost additional costs incurred by third parties outside the production process when a unit of output is produced

externality a market exchange that affects a third party who is outside or "external" to the exchange; sometimes called a "spillover"

market failure When the market on its own does not allocate resources efficiently in a way that balances social costs and benefits; externalities are one example of a market failure

negative externality a situation where a third party, outside the transaction, suffers from a market transaction by others

positive externality a situation where a third party, outside the transaction, benefits from a market transaction by others

social costs costs that include both the private costs incurred by firms and also additional costs incurred by third parties outside the production process, like costs of pollution

spillover see externality

SOLUTIONS

Answers to Self-Check Questions

1.
 - a. positive externality
 - b. negative externality
 - c. positive externality
 - d. negative externality

- e. negative externality
2.
 - a. supply shifts left
 - b. supply shifts left
 - c. supply stays the same
 - d. supply shifts left
 3.
 - a. price will rise
 - b. price will rise
 - c. price stays the same
 - d. price will rise.
 4. The original equilibrium (before the external social cost of pollution is taken into account) is where the private supply curve crosses the demand curve. This original equilibrium is at a price of \$15 and a quantity of 440. After taking into account the additional external cost of pollution, the production becomes more costly, and the supply curve shifts up. The new equilibrium will be at a price of \$30 and a quantity of 410.

18.2 COMMAND-AND-CONTROL REGULATION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain command-and-control regulation
- Evaluate the effectiveness of command-and-control regulation

When the United States started passing comprehensive environmental laws in the late 1960s and early 1970s, a typical law specified how much pollution could be emitted out of a smoke-stack or a drainpipe and imposed penalties if that limit was exceeded. Other laws required the installation of certain equipment—for example, on automobile tailpipes or on smokestacks—to reduce pollution. These types of laws, which specify allowable quantities of pollution and which also may detail which pollution-control technologies must be used, fall under the category of **command-and-control regulation**. In effect, command-and-control regulation requires that firms increase their costs by installing anti-pollution equipment; firms are thus required to take the social costs of pollution into account.

Command-and-control regulation has been highly successful in protecting and cleaning up the U.S. environment. In 1970, the Environmental Protection Agency (EPA) was created to oversee all environmental laws. In the same year, the **Clean Air Act** was enacted to address air pollution. Just two years later, in 1972, Congress passed and the president signed the far-reaching **Clean Water Act**. These command-and-control environmental laws, and their amendments and updates, have been largely responsible for America's cleaner air and water in recent decades. However, economists have pointed out three difficulties with command-and-control environmental regulation.

First, command-and-control regulation offers no incentive to improve the quality of the environment beyond the standard set by a particular law. Once the command-and-control regulation has been satisfied, polluters have zero incentive to do better.

Second, command-and-control regulation is inflexible. It usually requires the same standard for all polluters, and often the same pollution-control technology as well. This means that command-and-control regulation draws no distinctions between firms that would find it easy and inexpensive to meet the pollution standard—or to reduce pollution even further—and firms that might find it difficult and costly to meet the standard. Firms have no reason to rethink their production methods in fundamental ways that might reduce pollution even more and at lower cost.

Third, command-and-control regulations are written by legislators and the EPA, and so they are

subject to compromises in the political process. Existing firms often argue (and lobby) that stricter environmental standards should not apply to them, only to new firms that wish to start production. Consequently, real-world environmental laws are full of fine print, loopholes, and exceptions.

Although critics accept the goal of reducing pollution, they question whether command-and-control regulation is the best way to design policy tools for accomplishing that goal. A different approach is the use of market-oriented tools, which are discussed in the next section.

KEY CONCEPTS AND SUMMARY

Command-and-control regulation sets specific limits for pollution emissions and/or specific pollution-control technologies that must be used. Although such regulations have helped to protect the environment, they have three shortcomings: they provide no incentive for going beyond the limits they set; they offer limited flexibility on where and how to reduce pollution; and they often have politically-motivated loopholes.

SELF-CHECK QUESTIONS

Consider two approaches to reducing emissions of CO₂ into the environment from manufacturing industries in the United States. In the first approach, the U.S. government makes it a policy to use only predetermined technologies. In the second approach, the U.S. government determines which technologies are cleaner and subsidizes their use. Of the two approaches, which is the command-and-control policy?

REVIEW QUESTIONS

1. What is command-and-control environmental regulation?
2. What are the three problems that economists have noted with regard to command-and-control regulation?

CRITICAL THINKING QUESTIONS

1. Would environmentalists favor command-and-control policies as a way to reduce pollution? Why or why not?
2. Consider two ways of protecting elephants from poachers in African countries. In one approach, the government sets up enormous national parks that have sufficient habitat for elephants to thrive and forbids all local people to enter the parks or to injure either the elephants or their habitat in any way. In a second approach, the government sets up national parks and designates 10 villages around the edges of the park as official tourist centers that become places where tourists can stay and bases for guided tours inside the national park. Consider the different incentives of local villagers—who often are very poor—in each of these plans. Which plan seems more likely to help the elephant population?

GLOSSARY

command-and-control regulation laws that specify allowable quantities of pollution and that also may detail which pollution-control technologies must be used

SOLUTIONS**Answers to Self-Check Questions**

The first policy is command-and-control because it is a requirement that applies to all producers.

18.3 MARKET-ORIENTED ENVIRONMENTAL TOOLS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Show how pollution charges impact firm decisions
- Suggest other laws and regulations that could fall under pollution charges
- Explain the significance of marketable permits and property rights
- Evaluate which policies are most appropriate for various situations

Market-oriented environmental policies create incentives to allow firms some flexibility in reducing pollution. The three main categories of market-oriented approaches to pollution control are pollution charges, marketable permits, and better-defined property rights. All of these policy tools, discussed below, address the shortcomings of command-and-control regulation—albeit in different ways.

POLLUTION CHARGES

A **pollution charge** is a tax imposed on the quantity of pollution that a firm emits. A pollution charge gives a profit-maximizing firm an incentive to figure out ways to reduce its emissions—as long as the marginal cost of reducing the emissions is less than the tax.

For example, consider a small firm that emits 50 pounds per year of small particles, such as soot, into the air. Particulate matter, as it is called, causes respiratory illnesses and also imposes costs on firms and individuals.

Figure 1 illustrates the marginal costs that a firm faces in reducing pollution. The marginal cost of pollution reduction, like most **marginal cost curves** increases with output, at least in the short run. Reducing the first 10 pounds of particulate emissions costs the firm \$300. Reducing the second 10 pounds would cost \$500; reducing the third ten pounds would cost \$900; reducing the fourth 10 pounds would cost \$1,500; and the fifth 10 pounds would cost \$2,500. This pattern for the costs of reducing pollution is common, because the firm can use the cheapest and easiest method to make initial reductions in pollution, but additional reductions in pollution become more expensive.

Imagine the firm now faces a pollution tax of \$1,000 for every 10 pounds of particulates emitted. The firm has the choice of either polluting and paying the tax, or reducing the amount of particulates they emit and paying the cost of abatement as shown in the figure. How much will the firm pollute and

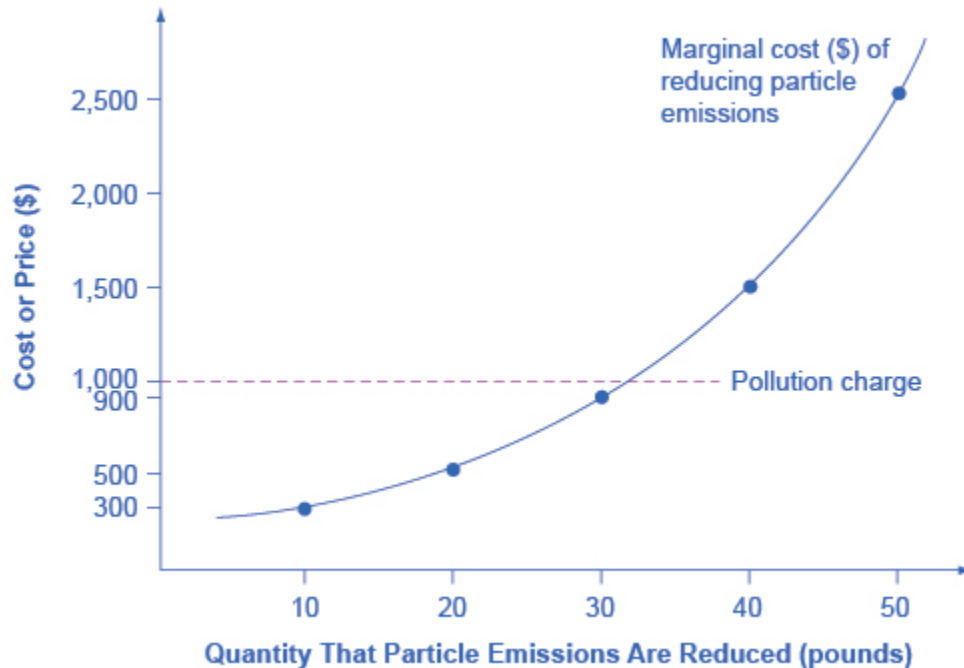


Figure 1. A Pollution Charge. If a pollution charge is set equal to \$1,000, then the firm will have an incentive to reduce pollution by 30 pounds because the \$900 cost of these reductions would be less than the cost of paying the pollution charge.

how much will the firm abate? The first 10 pounds would cost the firm \$300 to abate. This is substantially less than the \$1,000 tax, so they will choose to abate. The second 10 pounds would cost \$500 to abate, which is still less than the tax, so they will choose to abate. The third 10 pounds would cost \$900 to abate, which is slightly less than the \$1,000 tax. The fourth 10 pounds would cost \$1,500, which is much more costly than paying the tax. As a result, the firm will decide to reduce pollutants by 30 pounds, because the marginal cost of reducing pollution by this amount is less than the pollution tax. With a tax of \$1,000, the firm has no incentive to reduce pollution more than 30 pounds.

A firm that has to pay a pollution tax will have an incentive to figure out the least expensive technologies for reducing pollution. Firms that can reduce pollution cheaply and easily will do so to minimize their pollution taxes, whereas firms that will incur high costs for reducing pollution will end up paying the pollution tax instead. If the pollution tax applies to every source of pollution, then no special favoritism or loopholes are created for politically well-connected producers.

For an example of a pollution charge at the household level, consider two ways of charging for garbage collection. One method is to have a flat fee per household, no matter how much garbage a household produces. An alternative approach is to have several levels of fees, depending on how much garbage the household produces—and to offer lower or free charges for recyclable materials. As of 2006 (latest statistics available), the EPA had recorded over 7,000 communities that have implemented “pay as you throw” programs. When people have a financial incentive to put out less garbage and to increase recycling, they find ways of doing so.

Visit this website to learn more about pay-as-you-throw programs, including viewing a map and a table that shows the number of communities using this program in each state.



A number of environmental policies are really pollution charges, although they often do not travel under that name. For example, the federal government and many state governments impose taxes on gasoline. We can view this tax as a charge on the air pollution that cars generate as well as a source of funding for maintaining roads. Indeed, gasoline taxes are far higher in most other countries than in the United States.

Similarly, the refundable charge of five or 10 cents that only 10 states have for returning recyclable cans and bottles works like a pollution tax that provides an incentive to avoid littering or throwing bottles in the trash. Compared with command-and-control regulation, a pollution tax reduces pollution in a more flexible and cost-effective way.

Visit this website to see the current U.S. states with bottle bills and the states that have active campaigns for new bottle bills. You can also view current and proposed bills in Canada and other countries around the world.



MARKETABLE PERMITS

When a city or state government sets up a **marketable permit program** (e.g. cap-and-trade), it must start by determining the overall quantity of pollution it will allow as it tries to meet national pollution standards. Then, a number of permits allowing only this quantity of pollution are divided among the firms that emit that pollutant. These permits to pollute can be sold or given to firms free.

Now, add two more conditions. Imagine that these permits are designed to reduce total emissions over time. For example, a permit may allow emission of 10 units of pollution one year, but only nine units the next year, then eight units the year after that, and so on down to some lower level. In addition, imagine that these are marketable permits, meaning that firms can buy and sell them.

To see how marketable permits can work to reduce pollution, consider the four firms listed in Table 6. The table shows current emissions of lead from each firm. At the start of the marketable permit program, each firm receives permits to allow this level of pollution. However, these permits are shrinkable, and next year the permits allow the firms to emit only half as much pollution. Let's say that in a year, Firm Gamma finds it easy and cheap to reduce emissions from 600 tons of lead to 200 tons, which means that it has permits that it is not using that allow emitting 100 tons of lead. Firm Beta reduces its lead pollution from 400 tons to 200 tons, so it does not need to buy any permits, and it does not have any extra permits to sell. However, although Firm Alpha can easily reduce pollution from 200 tons to 150 tons, it finds that it is cheaper to purchase permits from Gamma rather than to reduce its own emissions to 100. Meanwhile, Firm Delta did not even exist in the first period, so the only way it can start production is to purchase permits to emit 50 tons of lead.

The total quantity of pollution will decline. But the buying and selling of the marketable permits will determine exactly which firms reduce pollution and by how much. With a system of marketable permits, the firms that find it least expensive to do so will reduce pollution the most.

	Firm Alpha	Firm Beta	Firm Gamma	Firm Delta
Current emissions—permits distributed free for this amount	200 tons	400 tons	600 tons	0 tons
How much pollution will these permits allow in one year?	100 tons	200 tons	300 tons	0 tons
Actual emissions one year in the future	150 tons	200 tons	200 tons	50 tons
Buyer or seller of marketable permit?	Buys permits for 50 tons	Doesn't buy or sell permits	Sells permits for 100 tons	Buys permits for 50 tons

Table 6. How Marketable Permits Work

Another application of marketable permits occurred when the **Clean Air Act** was amended in 1990. The revised law sought to reduce sulfur dioxide emissions from electric power plants to half of the 1980 levels out of concern that sulfur dioxide was causing acid rain, which harms forests as well as buildings. In this case, the marketable permits the federal government issued were free of charge (no pun intended) to electricity-generating plants across the country, especially those that were burning coal (which produces sulfur dioxide). These permits were of the “shrinkable” type; that is, the amount of pollution allowed by a given permit declined with time.

BETTER-DEFINED PROPERTY RIGHTS

A clarified and strengthened idea of property rights can also strike a balance between economic activity and pollution. Ronald Coase (1910–2013), who won the 1991 Nobel Prize in economics, offered a vivid illustration of an externality: a railroad track running beside a farmer's field where the railroad locomotive sometimes gives off sparks and sets the field ablaze. Coase asked whose responsibility it was to address this spillover. Should the farmer be required to build a tall fence alongside the field to block the sparks? Or should the railroad be required to put some gadget on the locomotive's smokestack to reduce the number of sparks?

Coase pointed out that this issue cannot be resolved until **property rights** are clearly defined—that is, the legal rights of ownership on which others are not allowed to infringe without paying compensation. Does the farmer have a property right not to have a field burned? Does the railroad have a property right to run its own trains on its own tracks? If neither party has a property right, then the two

sides may squabble endlessly, nothing will be done, and sparks will continue to set the field aflame. However, if either the farmer or the railroad has a well-defined legal responsibility, then that party will seek out and pay for the least costly method of reducing the risk that sparks will hit the field. The property right determines whether the farmer or the railroad pays the bills.

The property rights approach is highly relevant in cases involving endangered species. The U.S. government's endangered species list includes about 1,000 plants and animals, and about 90% of these species live on privately owned land. The protection of these endangered species requires careful thinking about incentives and property rights. The discovery of an endangered species on private land has often triggered an automatic reaction from the government to prohibit the landowner from using that land for any purpose that might disturb the imperiled creatures. Consider the incentives of that policy: If you admit to the government that you have an endangered species, the government effectively prohibits you from using your land. As a result, rumors abounded of landowners who followed a policy of "shoot, shovel, and shut up" when they found an endangered animal on their land. Other landowners have deliberately cut trees or managed land in a way that they knew would discourage endangered animals from locating there.

HOW EFFECTIVE ARE MARKET-ORIENTED ENVIRONMENTAL POLICY TOOLS?

Environmentalists sometimes fear that market-oriented environmental tools are an excuse to weaken or eliminate strict limits on pollution emissions and instead to allow more pollution. It is true that if pollution charges are set very low or if marketable permits do not reduce pollution by very much then market-oriented tools will not work well. But command-and-control environmental laws can also be full of loopholes or have exemptions that do not reduce pollution by much, either. The advantage of market-oriented environmental tools is not that they reduce pollution by more or less, but because of their incentives and flexibility, they can achieve any desired reduction in pollution at a lower cost to society.

A more productive policy would consider how to provide private landowners with an incentive to protect the endangered species that they find and to provide a habitat for additional endangered species. For example, the government might pay landowners who provide and maintain suitable habitats for endangered species or who restrict the use of their land to protect an endangered species. Again, an environmental law built on incentives and flexibility offers greater promise than a command-and-control approach, which tries to oversee millions of acres of privately owned land.

APPLYING MARKET-ORIENTED ENVIRONMENTAL TOOLS

Market-oriented environmental policies are a tool kit. Specific policy tools will work better in some situations than in others. For example, marketable permits work best when a few dozen or a few hundred parties are highly interested in trading, as in the cases of oil refineries that trade lead permits or electrical utilities that trade sulfur dioxide permits. However, for cases in which millions of users emit small amounts of pollution—such as emissions from car engines or unrecycled soda cans—and have no strong interest in trading, pollution charges will typically offer a better choice. Market-oriented environmental tools can also be combined. Marketable permits can be viewed as a form of improved property rights. Or the government could combine marketable permits with a pollution tax on any emissions not covered by a permit.

KEY CONCEPTS AND SUMMARY

Examples of market-oriented environmental policies, also referred to as cap and trade programs, include pollution charges, marketable permits, and better-defined property rights. Market-oriented environmental policies include taxes, markets, and property rights so that those who impose negative externalities must face the social cost.

SELF-CHECK QUESTIONS

SELF-CHECK QUESTIONS

- Classify the following pollution-control policies as command-and-control or market incentive based.
 - A state emissions tax on the quantity of carbon emitted by each firm.
 - The federal government requires domestic auto companies to improve car emissions by 2020.
 - The EPA sets national standards for water quality.
 - A city sells permits to firms that allow them to emit a specified quantity of pollution.
 - The federal government pays fishermen to preserve salmon.
- An emissions tax on a quantity of emissions from a firm is not a command-and-control approach to reducing pollution. Why?
- Four firms called Elm, Maple, Oak, and Cherry, produce wooden chairs. However, they also produce a great deal of garbage (a mixture of glue, varnish, sandpaper, and wood scraps). The first row of Table 7 shows the total amount of garbage (in tons) currently produced by each firm. The other rows of the table show the cost of reducing garbage produced by the first five tons, the second five tons, and so on. First, calculate the cost of requiring each firm to reduce the weight of its garbage by one-fourth. Now, imagine that marketable permits are issued for the current level of garbage, but the permits will shrink the weight of allowable garbage for each firm by one-fourth. What will be the result of this alternative approach to reducing pollution?

	Elm	Maple	Oak	Cherry
Current production of garbage (in tons)	20	40	60	80
Cost of reducing garbage by first five tons	\$5,500	\$6,300	\$7,200	\$3,000
Cost of reducing garbage by second five tons	\$6,000	\$7,200	\$7,500	\$4,000
Cost of reducing garbage by third five tons	\$6,500	\$8,100	\$7,800	\$5,000
Cost of reducing garbage by fourth five tons	\$7,000	\$9,000	\$8,100	\$6,000
Cost of reducing garbage by fifth five tons	\$0	\$9,900	\$8,400	\$7,000

Table 7.

- The rows in Table 8 show three market-oriented tools for reducing pollution. The columns of the table show three complaints about command-and-control regulation. Fill in the table by stating briefly how each market-oriented tool addresses each of the three concerns.

	Incentives to Go Beyond	Flexibility about Where and How Pollution Will Be Reduced	Political Process Creates Loopholes and Exceptions
Pollution Charges			
Marketable Permits			
Property Rights			
Table 8.			

REVIEW QUESTIONS

1. What is a pollution charge and what incentive does it provide for a firm to take external costs into account?
2. What is a marketable permit and what incentive does it provide for a firm to take external costs into account?
3. What are better-defined property rights and what incentive do they provide to take external costs into account?

CRITICAL THINKING QUESTIONS

1. Will a system of marketable permits work with thousands of firms? Why or why not?
2. Is zero pollution possible under a marketable permits system? Why or why not?
3. Is zero pollution an optimal goal? Why or why not?

REFERENCES

Environmental Protection Agency. "2006 Pay-As-You-Throw Programs." Accessed December 20, 2013. <http://www.epa.gov/epawaste/conservation/tools/payt/states/06comm.htm>.

GLOSSARY

marketable permit program a permit that allows a firm to emit a certain amount of pollution; firms with more permits than pollution can sell the remaining permits to other firms

pollution charge a tax imposed on the quantity of pollution that a firm emits; also called a pollution tax

property rights the legal rights of ownership on which others are not allowed to infringe without paying compensation

SOLUTIONS

Answers to Self-Check Questions

1.
- a. market-based

b. command-and-control

c. command-and-control

d. market-based

e. market-based
2. Even though state or local governments impose these taxes, a company has the flexibility to adopt technologies that will help it avoid the tax.
3. First, if each firm is required to reduce its garbage output by one-fourth, then Elm will reduce five tons at a cost of \$5,500; Maple will reduce 10 tons at a cost of \$13,500; Oak will reduce three tons at a cost of \$22,500; and Cherry will reduce four tons at a cost of \$18,000. Total cost of this approach: \$59,500. If the system of marketable permits is put in place, and those permits shrink the weight of allowable garbage by one-quarter, then pollution must still be reduced by the same overall amount. However, now the reduction in pollution will take place where it is least expensive.

Reductions in Garbage	Who does the reducing?	At what cost?
First 5 tons	Cherry	\$3,000
Second 5 tons	Cherry	\$4,000
Third 5 tons	Cherry	\$5,000
Fourth 5 tons	Elm	\$5,500
Fifth and sixth 5 tons	Elm and Cherry	\$6,000 each
Seventh 5 tons	Maple	\$6,300
Eighth 5 tons	Elm	\$6,500
Ninth and tenth 5 tons	Elm and Cherry	\$7,000 each

Table 9.

Thus, the overall pattern of reductions here will be that Elm reduces garbage by 20 tons and has 15 tons of permits to sell. Maple reduces by five tons and needs to buy five tons of permits. Oak does not reduce garbage at all, and needs to buy 15 tons of permits. Cherry reduces garbage by 25 tons, which leaves it with five tons of permits to sell. The total cost of these reductions would be \$56,300, a definite reduction in costs from the \$59,500 cost of the command-and-control option.

4.

	Incentives to Go Beyond	Flexibility about Where and How Pollution Will Be Reduced	Political Process Creates Loopholes and Exceptions
Pollution Charges	If you keep reducing pollution you reduce your charge	Reducing pollution by any method is fine	If charge applies to all emissions of pollution then no loopholes
Marketable Permits	If you reduce your pollution you can sell your extra pollution permits	Reductions of pollution will happen at firms where it is cheapest to do so, by the least expensive methods	If all polluters are required to have permits then there are no loopholes
Property Rights	The party that has to pay for the pollution has incentive to do so in a cost effect way	Reducing pollution by any method is fine	If the property rights are clearly defined, then it is not legally possible to avoid cleanup

Table 10.

18.4 THE BENEFITS AND COSTS OF U.S. ENVIRONMENTAL LAWS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Evaluate the benefits and costs of environmental protection
- Explain the effects of ecotourism
- Apply marginal analysis to illustrate the marginal costs and marginal benefits of reducing pollution

Government economists have estimated that U.S. firms may pay more than \$200 billion per year to comply with federal environmental laws. That is big bucks. Is the money well spent?

BENEFITS AND COSTS OF CLEAN AIR AND CLEAN WATER

The benefits of a cleaner environment can be divided into four areas: (1) people may stay healthier and live longer; (2) certain industries that rely on clean air and water, such as farming, fishing, and tourism, may benefit; (3) property values may be higher; and (4) people may simply enjoy a cleaner environment in a way that does not need to involve a market transaction. Some of these benefits, such as gains to tourism or farming, are relatively easy to value in economic terms. It is harder to assign a monetary value to others, such as the value of clean air for someone with asthma. It seems impossible to put a clear-cut monetary value on still others, such as the satisfaction you might feel from knowing that the air is clear over the Grand Canyon, even if you have never visited the Grand Canyon.

Although estimates of environmental benefits are not precise, they can still be revealing. For example, a study by the Environmental Protection Agency looked at the costs and benefits of the Clean Air Act from 1970 to 1990. It found that total costs over that time period were roughly \$500 billion—a huge amount. However, it also found that a middle-range estimate of the health and other benefits from cleaner air was \$22 trillion—about 44 times higher than the costs. A more recent study by the EPA estimated that the environmental benefits to Americans from the **Clean Air Act** will exceed their costs by a margin of four to one. The EPA estimated that “in 2010 the benefits of Clean Air Act programs will total about \$110 billion. This estimate represents the value of avoiding increases in illness and premature death which would have prevailed.” Saying that overall benefits of environmental regulation have exceeded costs in the past, however, is very different from saying that every environmental regulation makes sense. For example, studies suggest that when breaking down emission reductions by type of contaminants, the benefits of air pollution control outweigh the costs primarily for particulates and lead, but when looking at other air pollutants, the costs of reducing them may be comparable

to or greater than the benefits. Just because some environmental regulations have had benefits much higher than costs does not prove that every individual regulation is a sensible idea.

ECOTOURISM: MAKING ENVIRONMENTALISM PAY

The definition of **ecotourism** is a little vague. Does it mean sleeping on the ground, eating roots, and getting close to wild animals? Does it mean flying in a helicopter to shoot anesthetic darts at African wildlife? Or a little of both? The definition may be fuzzy, but tourists who hope to appreciate the ecology of their destination—“eco tourists”—are the impetus to a big and growing business. The International Ecotourism Society estimates that international tourists interested in seeing nature or wildlife will take 1.56 billion trips by 2020.

Visit The International Ecotourism Society’s website to learn more about The International Ecotourism Society, its programs, and tourism’s role in sustainable community development.



Realizing the attraction of ecotourism, the residents of low-income countries may come to see that preserving wildlife habitats is more lucrative than, say, cutting down forests or grazing livestock to survive. In South Africa, Namibia, and Zimbabwe, for example, a substantial expansion of both rhinoceros and elephant populations is broadly credited to ecotourism, which has given local communities an economic interest in protecting them. Some of the leading ecotourism destinations include: Costa Rica and Panama in Central America; the Caribbean; Malaysia, and other South Pacific destinations; New Zealand; the Serengeti in Tanzania; the Amazon rain forests; and the Galapagos Islands. In many of these countries and regions, governments have enacted policies whereby revenues from ecotourism are shared with local communities, to give people in those local communities a kind of property right that encourages them to conserve their local environment.

Ecotourism needs careful management, so that the combination of eager tourists and local entrepreneurs does not destroy what the visitors are coming to see. But whatever one’s qualms are about certain kinds of ecotourism—such as the occasional practice of rich tourists shooting elderly lions with high-powered rifles—it is worth remembering that the alternative is often that low-income people in poor countries will damage their local environment in their effort to survive.

MARGINAL BENEFITS AND MARGINAL COSTS

We can use the tools of marginal analysis to illustrate the marginal costs and the marginal benefits of reducing pollution. Figure 1 illustrates a theoretical model of this situation. When the quantity of environmental protection is low so that pollution is extensive—for example, at quantity Q_a —there are usually a lot of relatively cheap and easy ways to reduce pollution, and the **marginal benefits** of doing

so are quite high. At Q_a , it makes sense to allocate more resources to fight pollution. However, as the extent of environmental protection increases, the cheap and easy ways of reducing pollution begin to decrease, and more costly methods must be used. The **marginal cost curve** rises. Also, as environmental protection increases, the largest marginal benefits are achieved first, followed by reduced marginal benefits. As the quantity of environmental protection increases to, say, Q_b , the gap between marginal benefits and marginal costs narrows. At point Q_c the marginal costs will exceed the marginal benefits. At this level of environmental protection, society is not allocating resources efficiently, because too many resources are being given up to reduce pollution.

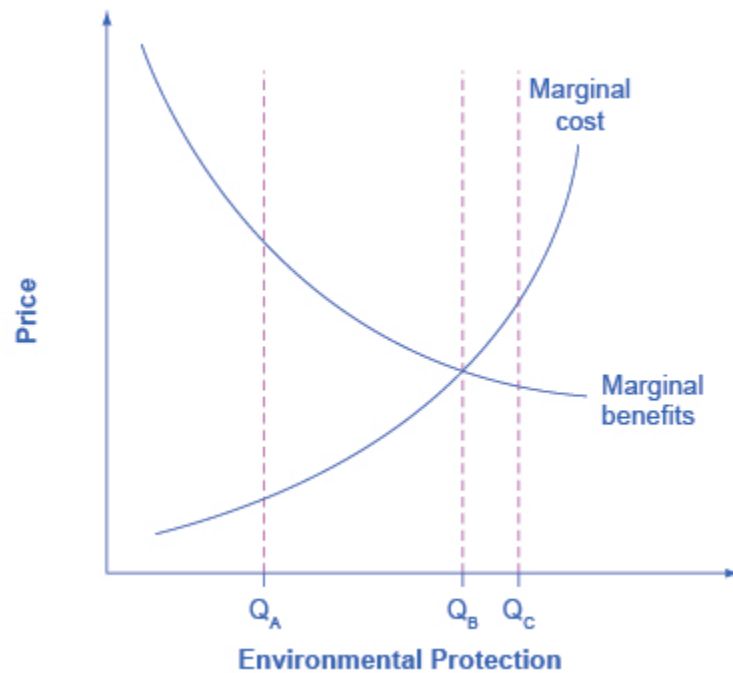


Figure 1. Marginal Costs and Marginal Benefits of Environmental Protection. Reducing pollution is costly—resources must be sacrificed. The marginal costs of reducing pollution are generally increasing, because the least expensive and easiest reductions can be made first, leaving the more expensive methods for later. The marginal benefits of reducing pollution are generally declining, because the steps that provide the greatest benefit can be taken first, and steps that provide less benefit can wait until later.

As society draws closer to Q_b , some might argue that it becomes more important to use market-oriented environmental tools to hold down the costs of reducing pollution. Their objective would be to avoid environmental rules that would provide the quantity of environmental protection at Q_c , where marginal costs exceed marginal benefits. The following Clear It Up feature delves into how the EPA measures its policies – and the monetary value of our lives.

WHAT'S A LIFE WORTH?

The U.S. Environmental Protection Agency (EPA) must estimate the value of saving lives by reducing pollution against the

additional costs. In measuring the benefits of government environmental policies, the EPA’s National Center for Environmental Economics (NCEE) values a statistical human life at \$7.4 million (in 2006 U.S. dollars). Economists value a human life on the basis of studies of the value that people actually place on human lives in their own decisions. For example, some jobs have a higher probability of death than others, and these jobs typically pay more to compensate for the risk. Examples are ocean fishery as opposed to fish farming, and ice trucking in Alaska as opposed to truck driving in the “lower forty-eight” states.

Government regulators use estimates such as these when deciding what proposed regulations are “reasonable,” which means deciding which proposals have high enough benefits to justify their cost. For example, when the U.S. Department of Transportation makes decisions about what safety systems should be required in cars or airplanes, it will approve rules only where the estimated cost per life saved is \$3 million or less.

Resources spent on life-saving regulations create tradeoff. A study by W. Kip Viscusi of Vanderbilt University estimated that when a regulation costs \$50 million, it diverts enough spending in the rest of the economy from health care and safety expenditures that it costs a life. This finding suggests that any regulation that costs more than \$50 million per life saved actually costs lives, rather than saving them.

KEY CONCEPTS AND SUMMARY

We can make a strong case, taken as a whole, that the benefits of U.S. environmental regulation have outweighed the costs. As the extent of environment regulation increases, additional expenditures on environmental protection will probably have increasing marginal costs and decreasing marginal benefits. This pattern suggests that the flexibility and cost savings of market-oriented environmental policies will become more important.

SELF-CHECK QUESTIONS

1. Suppose a city releases 16 million gallons of raw sewage into a nearby lake. Table 11 shows the total costs of cleaning up the sewage to different levels, together with the total benefits of doing so. (Benefits include environmental, recreational, health, and industrial benefits.)

	Total Cost (in thousands of dollars)	Total Benefits (in thousands of dollars)
16 million gallons	Current situation	Current situation
12 million gallons	50	800
8 million gallons	150	1300
4 million gallons	500	1650
0 gallons	1200	1900

Table 11.

- a. Using the information in Table 11, calculate the marginal costs and marginal benefits of reducing sewage emissions for this city. See Cost and Industry Structure if you need a refresher on how to calculate marginal costs.
- b. What is the optimal level of sewage for this city?
- c. Why not just pass a law that zero sewage can be emitted? After all, the total benefits of zero emissions exceed the total costs.
2. The state of Colorado requires oil and gas companies who use fracking techniques to return the land to its

original condition after the oil and gas extractions. Table 12 shows the total cost and total benefits (in dollars) of this policy.

Land Restored (in acres)	Total Cost	Total Benefit
0	\$0	\$0
100	\$20	\$140
200	\$80	\$240
300	\$160	\$320
400	\$280	\$380

Table 12.

- Calculate the marginal cost and the marginal benefit at each quantity (acre) of land restored. See Cost and Industry Structure if you need a refresher on how to calculate marginal costs and benefits.
- If we apply marginal analysis, what is the optimal amount of land to be restored?

REVIEW QUESTIONS

- As the extent of environmental protection expands, would you expect marginal costs of environmental protection to rise or fall? Why or why not?
- As the extent of environmental protection expands, would you expect the marginal benefits of environmental protection to rise or fall? Why or why not?

CRITICAL THINKING QUESTIONS

- From an economic perspective, is it sound policy to pursue a goal of zero pollution? Why or why not?
- Recycling is a relatively inexpensive solution to much of the environmental contamination from plastics, glass, and other waste materials. Is it a sound policy to make it mandatory for everybody to recycle?

PROBLEMS

A city currently emits 16 million gallons (MG) of raw sewage into a lake that is beside the city. Table 13 shows the total costs (TC) in thousands of dollars of cleaning up the sewage to different levels, together with the total benefits (TB) of doing so. Benefits include environmental, recreational, health, and industrial benefits.

	TC	TB
16 MG	Current	Current
12 MG	50	800
8 MG	150	1300
4 MG	500	1850
0 MG	1200	2000

Table 13.

- a. Using the information in Table 13 calculate the marginal costs and marginal benefits of reducing sewage emissions for this city.
- b. What is the optimal level of sewage for this city? How can you tell?

REFERENCES

Ryan, Dave. "New Report Shows Benefits of 1990 Clean Air Amendments Outweigh Costs by Four-to-One Margin," press release, November 16, 1999. *United States Environmental Protection Agency*. Accessed December 19, 2013. <http://www.epa.gov/oar/sect812/r-140.html>.

National Center for Environmental Economics (NCEE). "Frequently Asked Questions on Mortality Risk Valuation." *United States Environmental Protection Agency*. Accessed December 19, 2013. <http://yosemite.epa.gov/ee/epa/eed.nsf/pages/MortalityRiskValuation.html#whatvalue> World Tourism Organization, "Tourism 2020 Vision." Accessed December 19, 2013. http://www.world-tourism.org/market_research/facts/market_trends.htm.

Viscusi, Kip W. *Fatal Tradeoffs: Public and Private Responsibilities for Risk*. New York: Oxford University Press, 1995.

SOLUTIONS

Answers to Self-Check Questions

1.
 - a. See the answers in the following table. The marginal cost is calculated as the change in total cost divided by the change in quantity.

	Total Cost (in thousands of dollars) [marginal cost]	Total Benefits (in thousands of dollars) [marginal benefit]
16 million gallons	Current situation	Current situation
12 million gallons	50 [50]	800 [800]
8 million gallons	150 [100]	1,300 [500]
4 million gallons	500 [350]	1,850 [350]
0 gallons	1,200 [700]	2,000 [150]

Table 14.

- b. The “optimal” level of pollution is where the marginal benefits of reducing it are equal to the marginal cost. This is at four million gallons.
 - c. Marginal analysis tells us if the marginal costs of cleanup are greater than the marginal benefit, society could use those resources more efficiently elsewhere in the economy.
- 2.
- a. See the next table for the answers, which were calculated using the traditional calculation of marginal cost equal to change in total cost divided by change in quantity.

Land Restored (in acres)	Total Cost [marginal cost]	Total Benefit [marginal benefit]
0	\$0	\$0
100	\$20 [0.2]	\$140 [1.4]
200	\$80 [0.6]	\$240 [1]
300	\$160 [0.8]	\$320 [0.8]
400	\$280 [1.2]	\$480 [0.6]

Table 15.

- b. The optimal amount of restored land is 300 acres. Beyond this quantity the marginal costs are greater than the marginal benefits.

18.5 INTERNATIONAL ENVIRONMENTAL ISSUES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain biodiversity
- Analyze the partnership of high-income and low-income countries in efforts to address international externalities

Many countries around the world have become more aware of the benefits of environmental protection. Yet even if most nations individually took steps to address their environmental issues, no nation acting alone can solve certain environmental problems which spill over national borders. No nation by itself can reduce emissions of carbon dioxide and other gases by enough to solve the problem of global warming—not without the cooperation of other nations. Another issue is the challenge of preserving **biodiversity**, which includes the full spectrum of animal and plant genetic material. Although a nation can protect biodiversity within its own borders, no nation acting alone can protect biodiversity around the world. Global warming and biodiversity are examples of **international externalities**.

Bringing the nations of the world together to address environmental issues requires a difficult set of negotiations between countries with different income levels and different sets of priorities. If nations such as China, India, Brazil, Mexico, and others are developing their economies by burning vast amounts of fossil fuels or by stripping their forest and wildlife habitats, then the world's high-income countries acting alone will not be able to reduce greenhouse gases. However, low-income countries, with some understandable exasperation, point out that high-income countries do not have much moral standing to lecture them on the necessities of putting environmental protection ahead of economic growth. After all, high-income countries have historically been the primary contributors to greenhouse warming by burning **fossil fuels**—and still are today. It is hard to tell people who are living in a low-income country, where adequate diet, health care, and education are lacking, that they should sacrifice an improved quality of life for a cleaner environment.

Can rich and poor countries come together to address global environmental spillovers? At the initiative of the European Union and the most vulnerable developing nations, the Durban climate conference in December 2011 launched negotiations to develop a new international climate change agreement that covers all countries. The agreement will take the form of an agreed upon outcome with legal force applicable to all parties. According to the EU, the goal is to adopt the plan in

2015 and implement it in 2020. For the agreement to work, the two biggest emitters of greenhouse gases—China and the United States—will have to sign on.

Visit this website to learn more about the European Commission.



If high-income countries want low-income countries to reduce their emissions of greenhouse gases, then the high-income countries may need to pay some of the costs. Perhaps some of these payments will happen through private markets; for example, some tourists from rich countries will pay handsomely to vacation near the natural treasures of low-income countries. Perhaps some of the transfer of resources can happen through making modern pollution-control technology available to poorer countries.

The practical details of what such an international system might look like and how it would operate across international borders are forbiddingly complex. But it seems highly unlikely that some form of world government will impose a detailed system of environmental command-and-control regulation around the world. As a result, a decentralized and market-oriented approach may be the only practical way to address international issues such as global warming and biodiversity.

KEY CONCEPTS AND SUMMARY

Certain global environmental issues, such as global warming and biodiversity, spill over national borders and will need to be addressed with some form of international agreement.

SELF-CHECK QUESTIONS

Consider the case of global environmental problems that spill across international borders as a prisoner's dilemma of the sort studied in Monopolistic Competition and Oligopoly. Say that there are two countries, A and B. Each country can choose whether to protect the environment, at a cost of 10, or not to protect it, at a cost of zero. If one country decides to protect the environment, there is a benefit of 16, but the benefit is divided equally between the two countries. If both countries decide to protect the environment, there is a benefit of 32, which is divided equally between the two countries.

- a. In Table 16, fill in the costs, benefits, and total payoffs to the countries of the following decisions. Explain why, without some international agreement, they are likely to end up with neither country acting to protect the environment.

		Country B	
		Protect	Not Protect
Country A	Protect		
	Not Protect		

Table 16.

REVIEW QUESTIONS

1. What are the economic tradeoffs between low-income and high-income countries in international conferences on global environmental damage?
2. What arguments do low-income countries make in international discussions of global environmental clean-up?

CRITICAL THINKING QUESTIONS

1. Can extreme levels of pollution hurt the economic development of a high-income country? Why or why not?
2. How can high-income countries benefit from covering much of the cost of reducing pollution created by low-income countries?

REFERENCES

European Union. "Durban Conference Delivers Breakthrough for Climate." Press release. December 11, 2012. Accessed December 19, 2013. http://europa.eu/rapid/press-release_MEMO-11-895_en.htm?locale=en.

GLOSSARY

biodiversity the full spectrum of animal and plant genetic material

international externalities externalities that cross national borders and that cannot be resolved by a single nation acting alone

SOLUTIONS

Answers to Self-Check Questions

		Country B	
		Protect	Not Protect
Country A	Protect	Both A and B have a cost of 10 and a benefit of 16; each country has net = 6	A has a cost of 10 and a benefit of 8 (net = -2); B has a cost of 0 and a benefit of 8 (net = 8)
	Not Protect	A has a cost of 0 and a benefit of 8 (net = 8); B has a cost of 10 and a benefit of 8 (net = -2)	Both A and B have a zero cost and a zero benefit; each country has net = 0

Table 17.

Country B will reason this way: If A protects the environment, then we will have benefits of 6 if we act to protect the environment, but 8 if we do not, so we will not protect it. If A is not protecting the environment, we will have losses of 2 if we protect, but have zero if we do not protect, so again, we will not protect it. Country A will reason in a similar manner. The result is that both countries choose to not protect, even though they will achieve the largest social benefits—a combined benefit of 12 for the two countries—if they both choose to protect. Environmental treaties can be viewed as a way for countries to try to extricate themselves from this situation.

18.6 THE TRADEOFF BETWEEN ECONOMIC OUTPUT AND ENVIRONMENTAL PROTECTION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Apply the production possibility frontier to evaluate the tradeoff between economic output and the environment
- Interpret a graphic representation of the tradeoff between economic output and environmental protection

The tradeoff between economic output and the environment can be analyzed with a **production possibility frontier (PPF)** such as the one shown in Figure 1. At one extreme, at a choice like P, a country would be selecting a high level of economic output but very little environmental protection. At the other extreme, at a choice like T, a country would be selecting a high level of environmental protection but little economic output. According to the graph, an increase in environmental protection involves an opportunity cost of less economic output. No matter what their preferences, all societies should wish to avoid choices like M, which are productively inefficient. Efficiency requires that the choice should be on the production possibility frontier.

Economists do not have a great deal to say about the choice between P, Q, R, S and T in Figure 1, all of which lie along the production possibility frontier. Countries with low per capita **gross domestic product (GDP)**, such as China, place a greater emphasis on economic output—which in turn helps to produce nutrition, shelter, health, education, and desirable consumer goods. Countries with higher income levels, where a greater share of people have access to the basic necessities of life, may be willing to place a relatively greater emphasis on environmental protection.

However, economists are united in their belief that an inefficient choice such as M is undesirable. Rather than choosing M, a nation could achieve either greater economic output with the same environmental protection, as at point Q, or greater environmental protection with the same level of output, as at point S. The problem with command-and-control environmental laws is that they sometimes involve a choice like M. Market-oriented environmental tools offer a mechanism either for providing either the same environmental protection at lower cost, or providing a greater degree of environmental protection for the same cost.

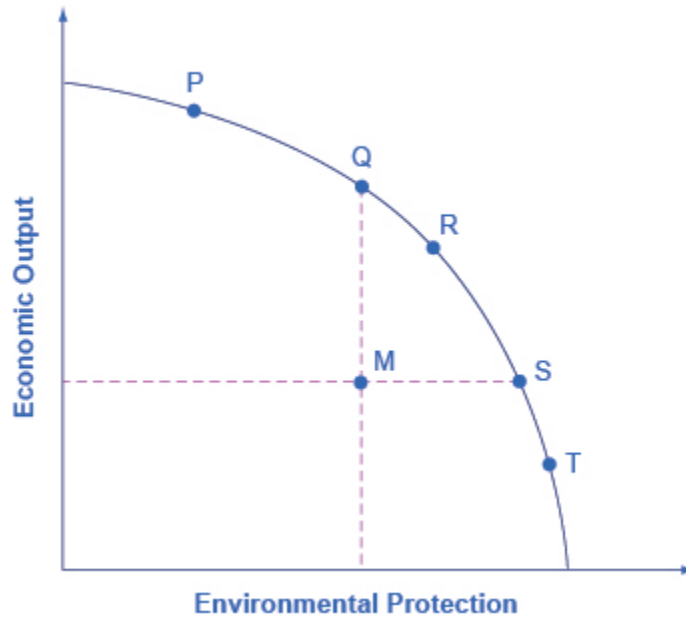


Figure 1. The Tradeoff between Economic Output and Environmental Protection. Each society will have to weigh its own values and decide whether it prefers a choice like P with more economic output and less environmental protection, or a choice like T with more environmental protection and less economic output.

KEYSTONE XL

So how would an economist respond to claims of environmental damage caused by the Keystone XL project? Clearly the environmental cost of oil spills would be considered a negative externality, but how many external costs would arise? And are these costs “too high” when measured against any potential for economic benefit?

As this chapter indicates, in deciding whether construction of the pipeline is a good idea, an economist would want to know not only about the marginal benefits resulting from the additional pipeline construction, but also the potential marginal costs—and especially the marginal external costs of the pipeline. Typically these come in the form of environmental impact statements, which are usually required for these kinds of projects. The most recent impact statement, released in March 2013 by the Nebraska Department of State, considered the possibility of fewer miles of pipeline going over the aquifer system and avoiding completely environmentally fragile areas; it indicated that “most resources” would not be harmed by construction of the pipeline.

As of press time, the Obama Administration has not approved construction of the Keystone XL project. While the economic benefits of additional oil in the United States may be fairly easily quantified, the social costs are not. It seems that, in a period of economic expansion, people want to err on the side of caution and estimate the marginal costs to be greater than the marginal benefits of additional oil generation. Those estimates may change, however, if the price of gasoline continues to rise.

KEY CONCEPTS AND SUMMARY

Depending on their different income levels and political preferences, countries are likely to make different choices about **allocative efficiency**—that is, the choice between economic output and environ-

mental protection along the production possibility frontier. However, all countries should prefer to make a choice that shows **productive efficiency**—that is, the choice is somewhere on the production possibility frontier rather than inside it. Revisit Choice in a World of Scarcity for more on these terms.

SELF-CHECK QUESTIONS

A country called Sherwood is very heavily covered with a forest of 50,000 trees. There are proposals to clear some of Sherwood's forest and grow corn, but obtaining this additional economic output will have an environmental cost from reducing the number of trees. Table 18 shows possible combinations of economic output and environmental protection.

<i>Combos</i>	Corn Bushels (thousands)	Number of Trees (thousands)
P	9	5
Q	2	30
R	7	20
S	2	40
T	6	10

Table 18.

- Sketch a graph of a production possibility frontier with environmental quality on the horizontal axis, measured by the number of trees, and the quantity of economic output, measured in corn, on the vertical axis.
- Which choices display productive efficiency? How can you tell?
- Which choices show allocative efficiency? How can you tell?
- In the choice between T and R, decide which one is better. Why?
- In the choice between T and S, can you say which one is better, and why?
- If you had to guess, which choice would you think is more likely to represent a command-and-control environmental policy and which choice is more likely to represent a market-oriented environmental policy, choice Q or S? Why?

REVIEW QUESTIONS

- In the tradeoff between economic output and environmental protection, what do the combinations on the production possibility curve represent?
- What does a point inside the production possibility frontier represent?

CRITICAL THINKING QUESTIONS

Technological innovations shift the production possibility curve. Look at graph you sketched for Self-Check Question 1 Which types of technologies should a country promote? Should “clean” technologies be promoted over other technologies? Why or why not?

PROBLEMS

In the Land of Purity, there is only one form of pollution, called “gunk.” Table 19 shows possible combinations of economic output and reduction of gunk, depending on what kinds of environmental regulations are chosen.

<i>Combos</i>	Eco Output	Gunk Cleaned Up
J	800	10%
K	500	30%
L	600	40%
M	400	40%
N	100	90%

Table 19.

- Sketch a graph of a production possibility frontier with environmental quality on the horizontal axis, measured by the percentage reduction of gunk, and with the quantity of economic output on the vertical axis.
- Which choices display productive efficiency? How can you tell?
- Which choices show allocative efficiency? How can you tell?
- In the choice between K and L, can you say which one is better and why?
- In the choice between K and N, can you say which one is better, and why?
- If you had to guess, which choice would you think is more likely to represent a command-and-control environmental policy and which choice is more likely to represent a market-oriented environmental policy, choice L or M? Why?

SOLUTIONS

Answers to Self-Check Questions

- Of the choices provided, P, R, and S demonstrate productive efficiency. These are the choices on the production possibility frontier.
- Allocative efficiency is determined by the preferences—in this case by the preferences of society as expressed through government and other social institutions. Because you do not have information about these preferences, you really cannot say much about allocative efficiency.
- In the choice between T and R, R should clearly be preferred, because it has both more corn and more trees. This answer illustrates why productive efficiency is beneficial. Compared with choices inside the PPF, it means more of one or both goods.
- In the choice between T and S, it is not possible to say which choice is better. True, S is on the PPF and T is not—but that only addresses the issue of productive efficiency. If a society has a strong preference for economic output and places a lower value on trees, then allocative efficiency may lead to a choice of T over S. Of course, the reverse could also be true, leading to a choice of S. Without information on society’s preferences to judge allocative efficiency, this question cannot be answered.
- Compared with command-and-control policies, market-oriented policies allow either more output with the same environmental protection or more environmental protection with the same level of output—or more

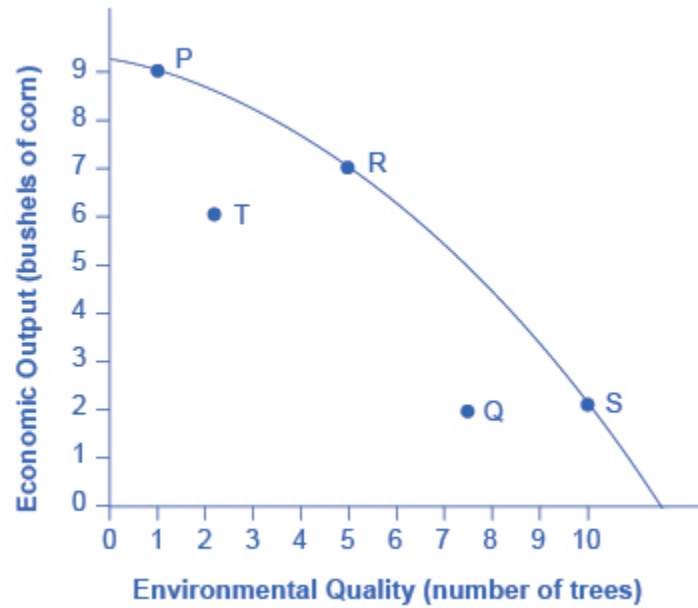


Figure 2.

of both environmental protection and output. Thus, a choice like Q inside the PPF is more likely to represent a command-and-control policy demand than a choice like S on the frontier of the PPF.

CHAPTER 19. POSITIVE EXTERNALITIES AND PUBLIC GOODS

INTRODUCTION TO POSITIVE EXTERNALITIES AND PUBLIC GOODS

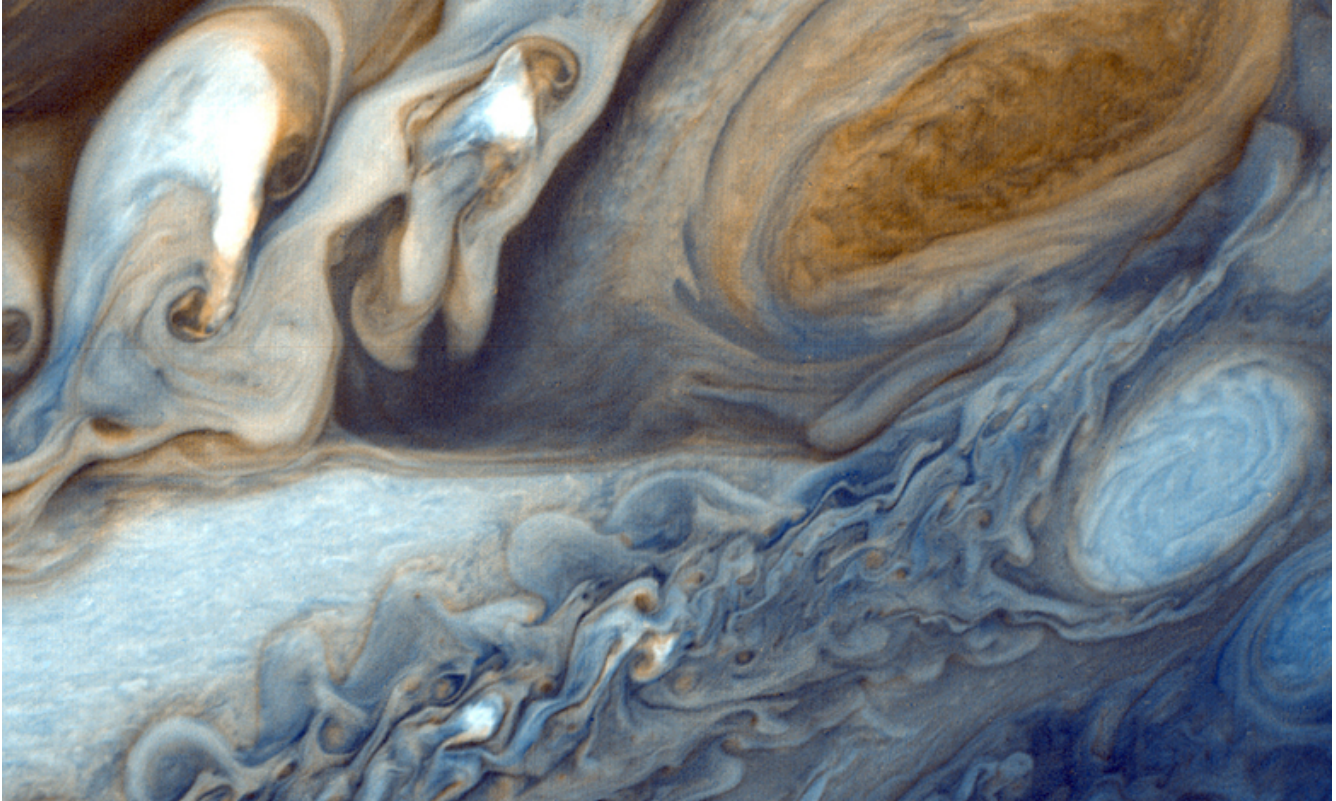


Figure 1. View from Voyager I. Launched by NASA on September 5, 1977, Voyager 1's primary mission was to provide detailed images of Jupiter, Saturn, and their moons. It took this photograph of Jupiter on its journey. In August of 2012, Voyager I entered interstellar space—the first human-made object to do so—and it is expected to send data and images back to earth until 2025. Such a technological feat has a lot to do with economic principles. (Credit: modification of work by NASA/JPL)

THE BENEFITS OF VOYAGER I LIVE ON

The rapid growth of technology has increased our ability to access and process data, to navigate through a busy city, and to communicate with friends on the other side of the globe. The research and development efforts of citizens, scientists, firms, universities, and governments have truly revolutionized the modern economy. To get a sense of how far we have come in a short period of time, let's compare one of humankind's greatest achievements to the smartphone most of us have in our coat pocket.

In 1977 the United States launched Voyager I, a spacecraft originally intended to reach Jupiter and Saturn, to send back

photographs and other cosmic measurements. Voyager I, however, kept going, and going—past Jupiter and Saturn—right out of our solar system. At the time of its launch, Voyager had some of the most sophisticated computing processing power NASA could engineer (8,000 instructions per second), but by the time it left the solar system (in 2012, actually) we Earthlings were using handheld devices that could process 14 billion instructions per second.

Still, the technology of today is a spillover product of the incredible feats accomplished by NASA thirty years ago. NASA research, for instance, is responsible for the kidney dialysis and mammogram machines that we use today. Research in new technologies not only produces private benefits to the investing firm, or in this case to NASA, but it also creates benefits for the broader society. In this way, new knowledge often becomes what economists refer to as a public good. This leads us to the topic of this chapter—technology, positive externalities, public goods, and the role of government in the encouragement of innovation and the social benefits that it provides.

CHAPTER OBJECTIVES

Introduction to Positive Externalities and Public Goods

In this chapter, you will learn about:

- Why the Private Sector Under Invests in Technologies
- How Governments Can Encourage Innovation
- Public Goods

Can you imagine a world in which you did not own a cellular phone or use Wikipedia? New technology changes how people live and work and what they buy. Technology includes the invention of new products, new ways of producing goods and services, and even new ways of managing a company more efficiently. Research and development of technology is the difference between horses and automobiles, between candles and electric lights, between fetching water in buckets and indoor plumbing, and between infection and good health from antibiotics.

In December 2009, ABC News compiled a list of some of the technological breakthroughs that have revolutionized consumer products in the past 10 years:

- GPS tracking devices, originally developed by the defense department and available to consumers in 2000, give users up-to-date information on location and time through satellite technology.
- In 2000, Toyota introduced the Prius hybrid car, which greatly improved fuel efficiency.
- Also in 2000, AT&T offered its customers the ability to text on a mobile phone.
- In 2001, Wikipedia launched a user-generated encyclopedia on the Web.
- Even though Napster died in 2001, the company launched music downloading and file sharing, which revolutionized how consumers get their music and videos.
- Friendster kicked off the social networking business in 2003, and Twitter and Facebook followed.
- In 2003, the Human Genome project was completed. It helps to fight disease and launch new pharmaceutical innovations.

- Also in 2003, the search engine became a way of life for obtaining information quickly. The search engine companies also became innovators in the digital software that dominates mobile devices.
- In 2006, Nintendo launched Wii and changed the way video games are played. Players can now be drawn into the action and use their bodies to respond rather than a handheld device.
- Apple introduced the iPhone in 2007 and launched an entire smartphone industry. In 2015, cell phones now recognize human voices via artificial intelligence.

With all new technologies, however, there are new challenges. This chapter deals with some of these issues: Will private companies be willing to invest in new technology? In what ways does new technology have positive externalities? What motivates inventors? Does government have a role to play in encouraging research and technology? Are there certain types of goods that markets fail to provide efficiently, and that only government can produce? What happens when consumption or production of a product creates positive externalities? Why is it unsurprising when a common resource, like marine fisheries, is overused?

19.1 WHY THE PRIVATE SECTOR UNDER INVESTS IN INNOVATION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify the positive externalities of new technology.
- Explain the difference between private benefits and social benefits and give examples of each.
- Calculate and analyze rates of return

Market competition can provide an incentive for discovering new technology because a **firm** can earn higher profits by finding a way to produce products more cheaply or to create products with characteristics consumers want. As Gregory Lee, CEO of Samsung said, “Relentless pursuit of new innovation is the key principle of our business and enables consumers to discover a world of possibilities with technology.” An innovative firm knows that it will usually have a temporary edge over its competitors and thus an ability to earn above-normal profits before competitors can catch up.

In certain cases, however, **competition** can discourage new **technology**, especially when other firms can quickly copy a new idea. Consider a pharmaceutical firm deciding to develop a new drug. On average, it can cost \$800 million and take more than a decade to discover a new drug, perform the necessary safety tests, and bring the drug to market. If the research and development (R&D) effort fails—and every R&D project has some chance of failure—then the firm will suffer losses and could even be driven out of business. If the project succeeds, then the firm’s competitors may figure out ways of adapting and copying the underlying idea, but without having to pay the costs themselves. As a result, the innovative company will bear the much higher costs of the R&D and will enjoy at best only a small, temporary advantage over the competition.

Many inventors over the years have discovered that their inventions brought them less profit than they might have reasonably expected.

- Eli Whitney (1765–1825) invented the cotton gin, but then southern cotton planters built their own seed-separating devices with a few minor changes in Whitney’s design. When Whitney sued, he found that the courts in southern states would not uphold his patent rights.
- Thomas Edison (1847–1931) still holds the record for most patents granted to an individual.

His first invention was an automatic vote counter, and despite the social benefits, he could not find a government that wanted to buy it.

- Gordon Gould came up with the idea behind the laser in 1957. He put off applying for a patent and, by the time he did apply, other scientists had laser inventions of their own. A lengthy legal battle resulted, in which Gould spent \$100,000 on lawyers, before he eventually received a patent for the laser in 1977. Compared to the enormous social benefits of the laser, Gould received relatively little financial reward.
- In 1936, Turing delivered a paper titled, “On Computable Numbers, with an Application to the Entscheidungsproblem,” in which he presented the notion of a universal machine (later called the “Universal Turing Machine,” and then the “Turing machine”) capable of computing anything that is computable. The central concept of the modern computer was based on Turing’s paper.

A variety of studies by economists have found that the original inventor receives one-third to one-half of the total economic benefits from innovations, while other businesses and new product users receive the rest.

THE POSITIVE EXTERNALITIES OF NEW TECHNOLOGY

Will private firms in a **market economy** under invest in research and technology? If a firm builds a factory or buys a piece of equipment, the firm receives all the economic benefits that result from the investments. However, when a firm invests in new technology, the **private benefits**, or profits, that the firm receives are only a portion of the overall social benefits. The **social benefits** of an innovation take into account the value of all the positive externalities of the new idea or product, whether enjoyed by other companies or society as a whole, as well as the private benefits received by the firm that developed the new technology. As you learned in Environmental Protection and Negative Externalities, **positive externalities** are beneficial spillovers to a third party, or parties.

Consider the example of the Big Drug Company, which is planning its R&D budget for the next year. Economists and scientists working for Big Drug have compiled a list of potential research and development projects and estimated rates of return. (The rate of return is the estimated payoff from the project.) Figure 1 shows how the calculations work. The downward-sloping D_{Private} curve represents the firm’s demand for financial capital and reflects the company’s willingness to borrow to finance research and development projects at various interest rates. Suppose that this firm’s investment in research and development creates a spillover benefit to other firms and households. After all, new innovations often spark other creative endeavors that society also values. If we add the spillover benefits society enjoys to the firm’s private demand for financial capital, we can draw D_{Social} that lies above D_{Private} .

If there was a way for the firm to fully monopolize those social benefits by somehow making them unavailable to the rest of us, the firm’s private demand curve would be the same as society’s demand curve. According to Figure 1 and Table 1, if the going rate of interest on borrowing is 8%, and the company can receive the private benefits of innovation only, then the company would finance \$30 million. Society, at the same rate of 8%, would find it optimal to have \$52 million of borrowing. Unless there is a way for the company to fully enjoy the total benefits, then it will borrow less than the socially optimal level of \$52 million.

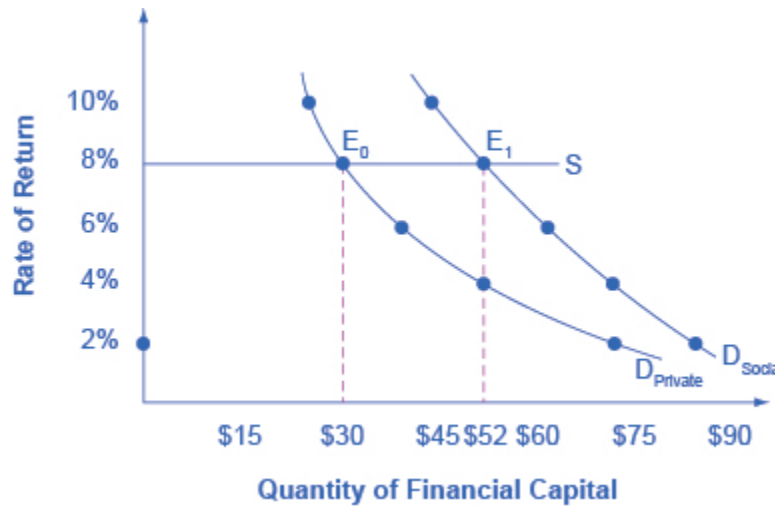


Figure 1. Positive Externalities and Technology. Big Drug faces a cost of borrowing of 8%. If the firm receives only the private benefits of investing in R&D, then its demand curve for financial capital is shown by D_{Private} , and the equilibrium will occur at \$30 million. Because there are spillover benefits, society would find it optimal to have \$52 million of investment. If the firm could keep the social benefits of its investment for itself, its demand curve for financial capital would be D_{Social} and it would be willing to borrow \$52 million.

Rate of Return	D_{Private} (in millions)	D_{Social} (in millions)
2%	\$72	\$84
4%	\$52	\$72
6%	\$38	\$62
8%	\$30	\$52
10%	\$26	\$44

Table 1. Return and Demand for Capital

Big Drug's original demand for financial capital (D_{Private}) is based on the profits received by the firm. However, other pharmaceutical firms and health care companies may learn new lessons about how to treat certain medical conditions and are then able to create their own competing products. The social benefit of the drug takes into account the value of all the positive externalities of the drug. If Big Drug were able to gain this social return instead of other companies, its **demand** for financial capital would shift to the demand curve D_{Social} , and it would be willing to borrow and invest \$52 million. However, if Big Drug is receiving only 50 cents of each dollar of social benefits, the firm will not spend as much on creating new products. The amount it would be willing to spend would fall somewhere in between D_{Private} and D_{Social} .

WHY INVEST IN HUMAN CAPITAL?

The investment in anything, whether it is the construction of a new power plant or research in a new

cancer treatment, usually requires a certain upfront cost with an uncertain future benefit. The investment in education, or human capital, is no different. Over the span of many years, a student and her family invest significant amounts of time and money into that student's education. The idea is that higher levels of educational attainment will eventually serve to increase the student's future productivity and subsequent ability to earn. Once the numbers are crunched, does this investment pay off for the student?

Almost universally, economists have found that the answer to this question is a clear "Yes." For example, several studies of the return to education in the United States estimate that the rate of return to a college education is approximately 10%. Data in Table 2, from the U.S. Bureau of Labor Statistics' *Usual Weekly Earnings of Wage and Salary Workers, Third Quarter 2014*, demonstrate that median weekly earnings are higher for workers who have completed more education. While these rates of return will beat equivalent investments in Treasury bonds or savings accounts, the estimated returns to education go primarily to the individual worker, so these returns are **private rates of return** to education.

	Less than a High School Degree	High School Degree, No College	Bachelor's Degree
Median Weekly Earnings (full-time workers over the age of 25)	\$488	\$668	\$1,101

Table 2. Usual Weekly Earnings of Wage and Salary Workers, Third Quarter 2014. (Source: <http://www.bls.gov/news.release/pdf/wkyeng.pdf>)

What does society gain from investing in the education of another student? After all, if the government is spending taxpayer dollars to subsidize public education, society should expect some kind of return on that spending. Again, economists like **George Psacharopoulos** have found that, across a variety of nations, the **social rate of return** on schooling is also positive. After all, positive externalities exist from investment in education. While not always easy to measure, according to **Walter McMahon**, the positive externalities to education typically include better health outcomes for the population, lower levels of crime, a cleaner environment and a more stable, democratic government. For these reasons, many nations have chosen to use taxpayer dollars to subsidize primary, secondary, and higher education. Education clearly benefits the person who receives it, but a society where most people have a good level of education provides positive externalities for all.

OTHER EXAMPLES OF POSITIVE EXTERNALITIES

Although technology may be the most prominent example of a positive externality, it is not the only one. For example, being vaccinated against disease is not only a protection for the individual, but it has the positive spillover of protecting others who may become infected. When a number of homes in a neighborhood are modernized, updated, and restored, not only does it increase the value of those homes, but the value of other properties in the neighborhood may increase as well.

The appropriate public policy response to a positive externality, like a new technology, is to help the party creating the positive externality receive a greater share of the social benefits. In the case of vaccines, like flu shots, an effective policy might be to provide a subsidy to those who choose to get vaccinated.

Figure 2 shows the market for flu shots. The market demand curve D_{Market} for flu shots reflects only the marginal private benefits (MPB) that the vaccinated individuals receive from the shots. Assuming

that there are no spillover costs in the production of flu shots, the market supply curve is given by the marginal private cost (MPC) of producing the vaccinations.

The equilibrium quantity of flu shots produced in the market, where MPB is equal to MPC, is Q_{Market} and the price of flu shots is P_{Market} . However, spillover benefits exist in this market because others, those who chose not to purchase a flu shot, receive a positive externality in a reduced chance of contracting the flu. When we add the spillover benefits to the marginal private benefit of flu shots, the marginal social benefit (MSB) of flu shots is given by D_{Social} . Because the MPB is greater than MSB, we see that the socially optimal level of flu shots is greater than the market quantity (Q_{Social} exceeds Q_{Market}) and the corresponding price of flu shots, if the market were to produce Q_{Social} , would be at P_{Social} . Unfortunately, the marketplace does not recognize the positive externality and flu shots will go under produced and under consumed.

So how can government try to move the market level of output closer to the socially desirable level of output? One policy would be to provide a subsidy, like a voucher, to any citizen who wishes to get vaccinated. This voucher would act as “income” that could be used to purchase only a flu shot and, if the voucher was exactly equal to the per-unit spillover benefits, would increase market equilibrium to a quantity of Q_{Social} and a price of P_{Social} where MSB equals MSC. Suppliers of the flu shots would receive payment of P_{Social} per vaccination, while consumers of flu shots would redeem the voucher and only pay a price of P_{Subsidy} . When the government uses a subsidy in this way, the socially optimal quantity of vaccinations is produced.

KEY CONCEPTS AND SUMMARY

Competition creates pressure to innovate. However, if new inventions can be easily copied, then the original inventor loses the incentive to invest further in research and development. New technology often has positive externalities; that is, there are often spillovers from the invention of new technology that benefit firms other than the innovator. The social benefit of an invention, once these spillovers are taken into account, typically exceeds the private benefit to the inventor. If inventors could receive a greater share of the broader social benefits for their work, they would have a greater incentive to seek out new inventions.

SELF-CHECK QUESTIONS

1. Are positive externalities reflected in market demand curves? Why or why not?
2. Samsung's R&D investment in digital devices has increased profits by 20%. Is this a private or social benefit?
3. The Gizmo Company is planning to develop new household gadgets. Table 3 shows the company's demand for financial capital for research and development of these gadgets, based on expected rates of return from sales. Now, say that every investment would have an additional 5% social benefit—that is, an investment that pays at least a 6% return to the Gizmo Company will pay at least an 11% return for society as a whole; an investment that pays at least 7% for the Gizmo Company will pay at least 12% for society as a whole, and so on. Answer the questions that follow based on this information.

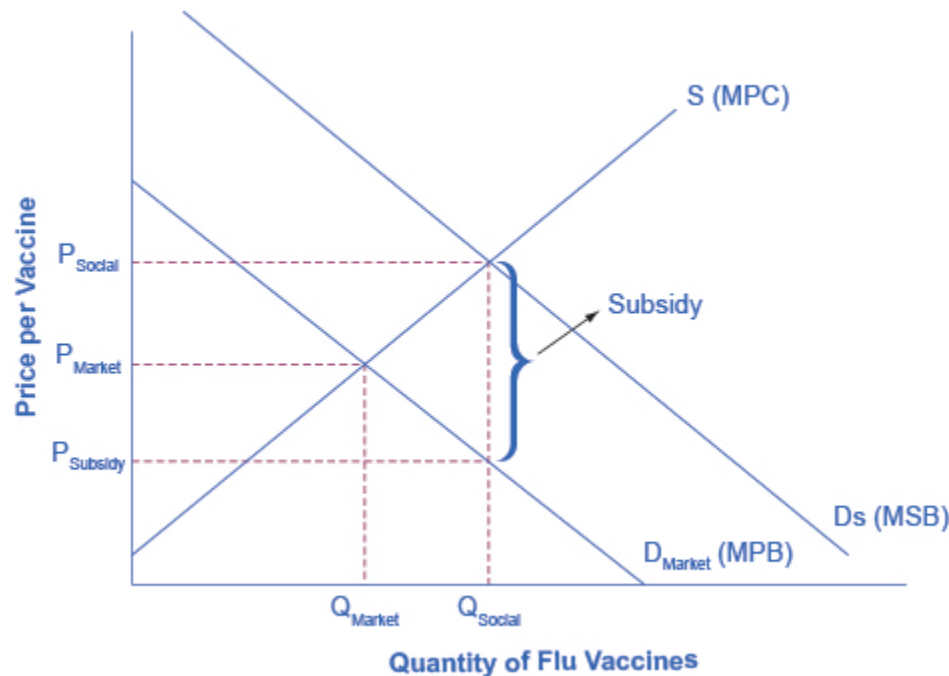


Figure 2. The Market for Flu Shots with Spillover Benefits (A Positive Externality). The market demand curve does not reflect the positive externality of flu vaccinations, so only Q_{Market} will be exchanged. This outcome is inefficient because the marginal social benefit exceeds the marginal social cost. If the government provides a subsidy to consumers of flu shots, equal to the marginal social benefit minus the marginal private benefit, the level of vaccinations can increase to the socially optimal quantity of Q_{Social} .

Estimated Rate of Return	Private profits of the firm from an R&D project (in \$ millions)
10%	\$100
9%	\$102
8%	\$108
7%	\$118
6%	\$133
5%	\$153
4%	\$183
3%	\$223

Table 3.

- a. If the going interest rate is 9%, how much will Gizmo invest in R&D if it receives only the private benefits of this investment?

b. Assume that the interest rate is still 9%. How much will the firm invest if it also receives the social benefits of its investment? (Add an additional 5% return on all levels of investment.)
4. The Junkbuyers Company travels from home to home, looking for opportunities to buy items that would

otherwise be put out with the garbage, but which the company can resell or recycle. Which will be larger, the private or the social benefits?

REVIEW QUESTIONS

1. In what ways do company investments in research and development create positive externalities?
2. Will the demand for borrowing and investing in R&D be higher or lower if there are no external benefits?

CRITICAL THINKING QUESTIONS

Can a company be guaranteed all of the social benefits of a new invention? Why or why not?

PROBLEMS

HighFlyer Airlines wants to build new airplanes with greatly increased cabin space. This will allow HighFlyer Airlines to give passengers more comfort and sell more tickets at a higher price. However, redesigning the cabin means rethinking many other elements of the airplane as well, like the placement of engines and luggage, and the most efficient shape of the plane for moving through the air. HighFlyer Airlines has developed a list of possible methods to increase cabin space, along with estimates of how these approaches would affect costs of operating the plane and sales of airline tickets. Based on these estimates, Table 4 shows the value of R&D projects that provide at least a certain private rate of return. Column 1 = Private Rate of Return. Column 2 = Value of R&D Projects that Return at Least the Private Rate of Return to HighFlyer Airlines. Use the data to answer the following questions.

Private Rate of Return	Value of R&D
12%	\$100
10%	\$200
8%	\$300
6%	\$400
4%	\$500

Table 4.

- a. If the opportunity cost of financial capital for HighFlyer Airlines is 6%, how much should the firm invest in R&D?
- b. Assume that the social rate of return for R&D is an additional 2% on top of the private return; that is, an R&D investment that had a 7% private return to HighFlyer Airlines would have a 9% social return. How much investment is socially optimal at the 6% interest rate?

REFERENCES

- Arias, Omar and Walter W. McMahon. "Dynamic Rates of Return to Education in the U.S." *Economics of Education Review*. 20, 2001. 121–138.
- Biography.com. 2015. "Alan Turing." Accessed April 1, 2015. <http://www.biography.com/people/alan-turing-9512017>.
- Canty Media. 2015. "The World: Life Expectancy (2015) – Top 100+." Accessed April 1, 2015. <http://www.geoba.se/population.php?pc=world&type=15>.
- Hyclak, Thomas, Geraint Johnes, and Robert Thornton. *Fundamentals of Labor Economics*. Boston: Houghton Mifflin Company, 2005.
- McMahon, Walter. *Education and Development: Measuring the Social Benefits*. Oxford: Oxford University Press, 2000.
- National Institute of Health. 2015. "Global Competitiveness—The Importance of U.S. Leadership in Science and Innovation for the Future of Our Economy and Our Health." Accessed April 1, 2015. http://www.nih.gov/about/impact/impact_global.pdf.
- National Science Foundation. 2013. "U.S. R&D Spending Resumes Growth in 2010 and 2011 but Still Lags Behind the Pace of Expansion of the National Economy." Accessed April 1, 2015. <http://www.nsf.gov/statistics/infbrief/nsf13313/>.
- Psacharopoulos, George. "Returns to Investment in Education: A Global Update." *World Development* 22, 1994. 1325–1343.
- Salientes-Narisma, Corrie. "Samsung Shift to Innovative Devices Pay Off." *Inquirer Technology*. Accessed May 15, 2013. <http://technology.inquirer.net/23831/samsungs-shift-to-innovative-devices-pays-off>.

GLOSSARY

positive externalities beneficial spillovers to a third party or parties

private benefits the dollar value of all benefits of a new product or process invented by a company that can be captured by the investing company

private rates of return when the estimated rates of return go primarily to an individual; for example, earning interest on a savings account

social benefits the dollar value of all benefits of a new product or process invented by a company that can be captured by other firms and by society as a whole

social rate of return when the estimated rates of return go primarily to society; for example, providing free education

SOLUTIONS

Answers to Self-Check Questions

1. No. A market demand curve reflects only the private benefits of those who are consuming the product. Positive externalities are benefits that spill over to third parties, so they create social benefits, and are not captured by a market (or private benefit) demand curve.
2. Clearly Samsung is benefiting from the investment, so the 20% increase in profits is a private benefit. If Samsung is unable to capture all of the benefit, perhaps because other companies quickly copy and produce close substitutes, then Samsung's investment will produce social benefits.
3.
 - a. \$102 million.
 - b. If the interest rate is 9%, the cost of financial capital, and the firm can capture the 5% return to society, the firm would invest as if its effective rate of return is 4%, so it will invest \$183 million.
4. When the Junkbuyers Company purchases something for resale, presumably both the buyer and the seller benefit—otherwise, they would not need to make the transaction. However, the company also reduces the amount of garbage produced, which saves money for households and/or for the city that disposes of garbage. So the social benefits are larger than the private benefits.

19.2 HOW GOVERNMENTS CAN ENCOURAGE INNOVATION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the effects of intellectual property rights on social and private rates of return.
- Identify three U.S. Government policies and explain how they encourage innovation

A number of different government policies can increase the incentives to innovate, including: guaranteeing intellectual property rights, government assistance with the costs of research and development, and cooperative research ventures between universities and companies.

INTELLECTUAL PROPERTY RIGHTS

One way to increase new technology is to guarantee the innovator an exclusive right to that new product or process. **Intellectual property** rights include patents, which give the inventor the exclusive legal right to make, use, or sell the invention for a limited time, and copyright laws, which give the author an exclusive legal right over works of literature, music, film/video, and pictures. For example, if a pharmaceutical firm has a **patent** on a new drug, then no other firm can manufacture or sell that drug for twenty-one years, unless the firm with the patent grants permission. Without a patent, the pharmaceutical firm would have to face competition for any successful products, and could earn no more than a normal rate of profit. With a patent, a firm is able to earn monopoly profits on its product for a period of time—which offers an incentive for research and development. In general, how long can “a period of time” be? The Clear it Up discusses patent and copyright protection timeframes for some works you might have heard of.

HOW LONG IS MICKEY MOUSE PROTECTED FROM BEING COPIED?

All patents and copyrights are scheduled to end someday. In 2003, copyright protection for Mickey Mouse was scheduled to run out. Once the copyright had expired, anyone would be able to copy Mickey Mouse cartoons or draw and sell new ones. In 1998, however, Congress passed the Sonny Bono Copyright Term Extension Act. For copyrights owned by companies or other entities, it increased or extended the copyright from 75 years to 95 years after publication. For copyrights owned by individuals, it increased or extended the copyright coverage from 50 years to 70 years after death. Along with protecting Mickey for another 20 years, the copyright extension affected about 400,000 books, movies, and songs.

Figure 1 illustrates how the total number of patent applications filed with the U.S. Patent and Trade-

mark Office, as well as the total number of patents granted, surged in the mid-1990s with the invention of the Internet, and is still going strong today.

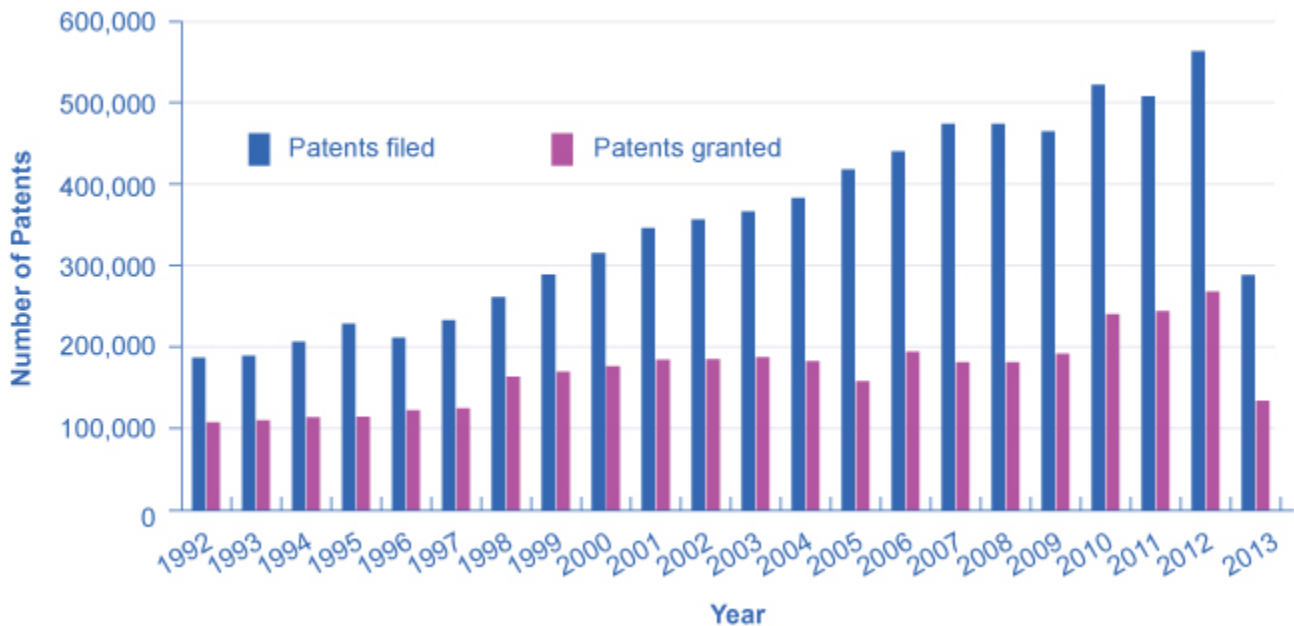


Figure 1. Patents Filed and Granted, 1981–2012. The number of applications filed for patents increased substantially from the mid-1990s into the first years of the 2000s, due in part to the invention of the Internet, which has led to many other inventions and to the 1998 Copyright Term Extension Act. (Source: http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm)

While patents provide an incentive to innovate by protecting the innovator, they are not perfect. For example:

- In countries that already have patents, economic studies show that inventors receive only one-third to one-half of the total economic value of their inventions.
- In a fast-moving high-technology industry like biotechnology or semiconductor design, patents may be almost irrelevant because technology is advancing so quickly.
- Not every new idea can be protected with a patent or a copyright—for example, a new way of organizing a factory or a new way of training employees.
- Patents may sometimes cover too much or be granted too easily. In the early 1970s, Xerox had received over 1,700 patents on various elements of the photocopy machine. Every time Xerox improved the photocopier, it received a patent on the improvement.
- The 21-year time period for a patent is somewhat arbitrary. Ideally, a patent should cover a long enough period of time for the inventor to earn a good return, but not so long that it allows the inventor to charge a monopoly price permanently.

Because patents are imperfect and do not apply well to all situations, alternative methods of improving the rate of return for inventors of new technology are desirable. Some of these possible alternative policies are described in the following sections.

POLICY #1: GOVERNMENT SPENDING ON RESEARCH AND DEVELOPMENT

If the private sector does not have sufficient incentive to carry out research and development, one possibility is for the government to fund such work directly. Government spending can provide direct financial support for research and development (R&D) done at colleges and universities, nonprofit research entities, and sometimes by private firms, as well as at government-run laboratories. While government spending on research and development produces technology that is broadly available for firms to use, it costs taxpayers money and can sometimes be directed more for political than for scientific or economic reasons.

Visit the NASA website and the USDA website to read about government research that would not take place where it left to firms due to the externalities.



The first column of Table 5 shows the sources of total U.S. spending on research and development; the second column shows the total dollars of R&D funding by each source. The third column shows that, relative to the total amount of funding, 26% comes from the federal government, about 67% of R&D is done by industry, and less than 3% is done by universities and colleges. (The percentages below do not add up to exactly 100% due to rounding.)

Sources of R&D Funding	Amount (\$ billions)	Percent of the Total
Federal government	\$133.6	32%
Industry	\$249	60.2%
Universities and colleges	\$12.5	3%
Nonprofits	\$15.1	3.6%
Nonfederal government	\$3.8	0.91%
<i>Total</i>	<i>\$414</i>	

Table 5. U.S. Research and Development Expenditures, 2011. (Source: <http://www.nsf.gov/statistics/infbrief/nsf13313/>)

In the 1960s the federal government paid for about two-thirds of the nation's R&D. Over time, the U.S. economy has come to rely much more heavily on industry-funded R&D. The federal government has tried to focus its direct R&D spending on areas where private firms are not as active. One difficulty with direct government support of R&D is that it inevitably involves political decisions about which projects are worthy. The scientific question of whether research is worthwhile can easily become entangled with considerations like the location of the congressional district in which the research funding is being spent.

POLICY #2: TAX BREAKS FOR RESEARCH AND DEVELOPMENT

A complementary approach to supporting R&D that does not involve the government's close scrutiny of specific projects is to give firms a reduction in taxes depending on how much research and development they do. The federal government refers to this policy as the research and experimentation (R&E) tax credit. According to the Treasury Department: "... the R&E Credit is also a cost-effective policy for stimulating additional private sector investment. Most recent studies find that each dollar of foregone tax revenue through the R&E Tax Credit causes firms to invest at least a dollar in R&D, with some studies finding a benefit to cost ratio of 2 or 2.96."

Visit this website for more information on how the R&E Tax Credit encourages investment.



POLICY #3 COOPERATIVE RESEARCH

State and federal governments support research in a variety of ways. For example, United for Medical Research, a coalition of groups that seek funding for the **National Institutes of Health**, (which is supported by federal grants), states: "NIH-supported research added \$69 billion to our GDP and supported seven million jobs in 2011 alone." The United States remains the leading sponsor of medical-related research spending \$117 billion in 2011. Other institutions, such as the **National Academy of Scientists** and the **National Academy of Engineers**, receive federal grants for innovative projects. The **Agriculture and Food Research Initiative (AFRI)** at the United States Department of Agriculture awards federal grants to projects that apply the best science to the most important agricultural problems, from food safety to childhood obesity. Cooperation between government-funded universities, academies, and the private sector can spur product innovation and create whole new industries.

KEY CONCEPTS AND SUMMARY

Public policy with regard to technology must often strike a balance. For example, patents provide an incentive for inventors, but they should be limited to genuinely new inventions and not extend forever.

Government has a variety of policy tools for increasing the rate of return for new technology and encouraging its development, including: direct government funding of R&D, tax incentives for R&D, protection of intellectual property, and forming cooperative relationships between universities and the private sector.

SELF-CHECK QUESTIONS

1. When a neighborhood is cleaned up and kept neat, there are a number of positive spillovers: higher property values, less crime, happier residents. What types of government policies can encourage neighborhoods to clean up?
2. Education provides both private benefits to those who receive it and broader social benefits for the economy as a whole. Think about the types of policies a government can follow to address the issue of positive spillovers in technology and then suggest a parallel set of policies that governments could follow for addressing positive externalities in education.

REVIEW QUESTIONS

1. Why might private markets tend to provide too few incentives for the development of new technology?
2. What can government do to encourage the development of new technology?

CRITICAL THINKING QUESTIONS

Is it inevitable that government must become involved in supporting investments in new technology?

PROBLEMS

The marginal private costs and the marginal private benefits of a firm producing fuel-efficient cars is represented in the following diagram (show the equilibrium P_{market} , Q_{market}). The government would like to increase the amount of fuel-efficient cars to be produced and sold to Q_{social} . One way that the government can try to increase production of fuel-efficient cars is by making them cheaper to produce, by subsidizing their production. Show, on the same graph, the amount of subsidy needed to increase the equilibrium quantity of fuel-efficient cars to Q_{social} . *Hint: the government is trying to affect production through costs, not benefits.*

REFERENCES

- United States Department of the Treasury. "Research and Experimentation Tax Credit." Accessed November 2013. <http://www.treasury.gov/press-center/news/Pages/investing-in-us-competitive-ness.aspx>.
- U.S. Patent and Trademark Office. 2015. "U.S. Patent Statistics: Calendar Years 1963–2014." Accessed April 10, 2015. http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.pdf.
- United for Medical Research. "Profiles of Prosperity: How NIH-Supported Research Is Fueling Private Sector Growth and Innovation." Introduction. Accessed January 2014. http://www.united-formedicalresearch.com/wp-content/uploads/2013/07/UMR_ProspersityReport_071913a.pdf.

GLOSSARY

intellectual property the body of law including patents, trademarks, copyrights, and trade secret law that protect the right of inventors to produce and sell their inventions

SOLUTIONS

Answers to Self-Check Questions

1. Government programs that either pay for neighborhood clean-up directly or that provide reduced tax payments for those who clean up or fix up their own property could be enacted. It is also easy to imagine how a city might allow its businesses to form a group that would pay for and manage neighborhood cleanup.
2. Government programs that either pay for education directly or that provide loans or reduced tax payments for education could create positive spillovers. A city might allow its businesses to form a group that would coordinate business efforts with schools and local colleges and universities—allowing students to obtain real-world experience in their chosen fields and providing businesses with enthusiastic, trained workers.

19.3 PUBLIC GOODS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify a public good using nonexcludable and nonrivalrous as criteria
- Explain the free rider problem
- Identify several sources of public goods

Even though new technology creates positive externalities so that perhaps one-third or one-half of the social benefit of new inventions spills over to others, the inventor still receives some private return. What about a situation where the positive externalities are so extensive that private firms could not expect to receive any of the social benefit? This kind of good is called a **public good**. Spending on national defense is a good example of a public good. Let's begin by defining the characteristics of a public good and discussing why these characteristics make it difficult for private firms to supply public goods. Then we will see how government may step in to address the issue.

THE DEFINITION OF A PUBLIC GOOD

Economists have a strict definition of a public good, and it does not necessarily include all goods financed through taxes. To understand the defining characteristics of a public good, first consider an ordinary private good, like a piece of pizza. A piece of pizza can be bought and sold fairly easily because it is a separate and identifiable item. However, public goods are not separate and identifiable in this way.

Instead, public goods have two defining characteristics: they are nonexcludable and nonrivalrous. The first characteristic, that a public good is **nonexcludable**, means that it is costly or impossible to exclude someone from using the good. If Larry buys a private good like a piece of pizza, then he can exclude others, like Lorna, from eating that pizza. However, if national defense is being provided, then it includes everyone. Even if you strongly disagree with America's defense policies or with the level of defense spending, the national defense still protects you. You cannot choose to be unprotected, and national defense cannot protect everyone else and exclude you.

The second main characteristic of a public good, that it is **nonrivalrous**, means that when one person uses the public good, another can also use it. With a private good like pizza, if Max is eating the pizza then Michelle cannot also eat it; that is, the two people are rivals in consumption. With a public good

like national defense, Max's consumption of national defense does not reduce the amount left for Michelle, so they are nonrivalrous in this area.

A number of government services are examples of public goods. For instance, it would not be easy to provide fire and police service so that some people in a neighborhood would be protected from the burning and burglary of their property, while others would not be protected at all. Protecting some necessarily means protecting others, too.

Positive externalities and public goods are closely related concepts. Public goods have positive externalities, like police protection or public health funding. Not all goods and services with positive externalities, however, are public goods. Investments in education have huge positive spillovers but can be provided by a private company. Private companies can invest in new inventions such as the Apple iPad and reap profits that may not capture all of the social benefits. Patents can also be described as an attempt to make new inventions into private goods, which are excludable and rivalrous, so that no one but the inventor is allowed to use them during the length of the patent.

THE FREE RIDER PROBLEM OF PUBLIC GOODS

Private companies find it difficult to produce public goods. If a good or service is nonexcludable, like national defense, so that it is impossible or very costly to exclude people from using this good or service, then how can a firm charge people for it?

Visit this website to read about a connection between free riders and “bad music.”



When individuals make decisions about buying a public good, a **free rider** problem can arise, in which people have an incentive to let others pay for the public good and then to “free ride” on the purchases of others. The free rider problem can be expressed in terms of the prisoner’s dilemma game, which is discussed as a representation of oligopoly in Monopolistic Competition and Oligopoly. Say that two people are thinking about contributing to a public good: Rachel and Samuel. When either of them contributes to a public good, such as a local fire department, their personal cost of doing so is \$4 and the social benefit of that person’s contribution is \$6. Because society’s benefit of \$6 is greater than the cost of \$4, the investment is a good idea for society as a whole. The problem is that, while Rachel and Samuel pay for the entire cost of their contribution to the public good, they receive only half of the benefit, because the benefit of the public good is divided equally among the members of society. This sets up the prisoner’s dilemma illustrated in Table 6.

	Samuel (S) Contribute	Samuel (S) Do Not Contribute
Rachel (R) Contribute	R pays \$4, receives \$6, net gain +\$2 S pays \$4, receives \$6, net gain +\$2	R pays \$4, receives \$3, net gain −\$1 S pays \$0, receives \$3, net gain +\$3
Rachel (R) Do Not Contribute	R pays \$0, receives \$3, net gain +\$3 S pays \$4, receives \$3, net gain −\$1	R pays \$0, receives \$0 S pays \$0, receives \$0

Table 6. Contributing to a Public Good as a Prisoner's Dilemma

If neither Rachel nor Samuel contributes to the public good, then there are no costs and no benefits of the public good. Suppose, however, that only Rachel contributes, while Samuel does not. Rachel incurs a cost of \$4, but receives only \$3 of benefit (half of the total \$6 of benefit to society), while Samuel incurs no cost, and yet he also receives \$3 of benefit. In this outcome, Rachel actually loses \$1 while Samuel gains \$3. A similar outcome, albeit with roles reversed, would occur if Samuel had contributed, but Rachel had not. Finally, if both parties contribute, then each incurs a cost of \$4 and each receives \$6 of benefit (half of the total \$12 benefit to society). There is a dilemma with the Prisoner's Dilemma, though. See the Work it Out feature.

THE PROBLEM WITH THE PRISONER'S DILEMMA

The difficulty with the prisoner's dilemma arises as each person thinks through his or her strategic choices.

Step 1. Rachel reasons in this way: If Samuel does not contribute, then I would be a fool to contribute. However, if Samuel does contribute, then I can come out ahead by not contributing.

Step 2. Either way, I should choose not to contribute, and instead hope that I can be a free rider who uses the public good paid for by Samuel.

Step 3. Samuel reasons the same way about Rachel.

Step 4. When both people reason in that way, the public good never gets built, and there is no movement to the option where everyone cooperates—which is actually best for all parties.

THE ROLE OF GOVERNMENT IN PAYING FOR PUBLIC GOODS

The key insight in paying for public goods is to find a way of assuring that everyone will make a contribution and to prevent free riders. For example, if people come together through the political process and agree to pay taxes and make group decisions about the quantity of public goods, they can defeat the free rider problem by requiring, through the law, that everyone contributes.

However, government spending and taxes are not the only way to provide public goods. In some cases, markets can produce public goods. For example, think about radio. It is nonexcludable, since once the radio signal is being broadcast, it would be very difficult to stop someone from receiving it. It is nonrivalrous, since one person listening to the signal does not prevent others from listening as well. Because of these features, it is practically impossible to charge listeners directly for listening to conventional radio broadcasts.

Radio has found a way to collect revenue by selling advertising, which is an indirect way of “charging” listeners by taking up some of their time. Ultimately, consumers who purchase the goods advertised are also paying for the radio service, since the cost of advertising is built into the product cost. In a more recent development, satellite radio companies, such as SiriusXM, charge a regular subscription

fee for streaming music without commercials. In this case, however, the product is excludable—only those who pay for the subscription will receive the broadcast.

Some public goods will also have a mixture of public provision at no charge along with fees for some purposes, like a public city park that is free to use, but the government charges a fee for parking your car, for reserving certain picnic grounds, and for food sold at a refreshment stand.

Read this article to find out what economists say the government should pay for.



In other cases, social pressures and personal appeals can be used, rather than the force of law, to reduce the number of free riders and to collect resources for the public good. For example, neighbors sometimes form an association to carry out beautification projects or to patrol their area after dark to discourage crime. In low-income countries, where social pressure strongly encourages all farmers to participate, farmers in a region may come together to work on a large irrigation project that will benefit all. Many fundraising efforts, including raising money for local charities and for the endowments of colleges and universities, also can be viewed as an attempt to use social pressure to discourage free riding and to generate the outcome that will produce a public benefit.

COMMON RESOURCES AND THE “TRAGEDY OF THE COMMONS”

There are some goods that do not fall neatly into the categories of private good or public good. While it is easy to classify a pizza as a private good and a city park as a public good, what about an item that is nonexcludable and rivalrous, such as the queen conch?

In the Caribbean, the queen conch is a large marine mollusk found in shallow waters of sea grass. These waters are so shallow, and so clear, that a single diver may harvest many conch in a single day. Not only is conch meat a local delicacy and an important part of the local diet, but the large ornate shells are used in art and can be crafted into musical instruments. Because almost anyone with a small boat, snorkel, and mask, can participate in the conch harvest, it is essentially nonexcludable. At the same time, fishing for conch is rivalrous; once a diver catches one conch it cannot be caught by another diver.

Goods that are nonexcludable and rivalrous are called **common resources**. Because the waters of the Caribbean are open to all conch fishermen, and because any conch that *you* catch is conch that *I* cannot catch, common resources like the conch tend to be overharvested.

The problem of overharvesting common resources is not a new one, but ecologist Garret Hardin put the tag “Tragedy of the Commons” to the problem in a 1968 article in the magazine *Science*. Econo-

mists view this as a problem of property rights. Since nobody owns the ocean, or the conch that crawl on the sand beneath it, no one individual has an incentive to protect that resource and responsibly harvest it. To address the issue of overharvesting conch and other marine fisheries, economists typically advocate simple devices like fishing licenses, harvest limits, and shorter fishing seasons. When the population of a species drops to critically low numbers, governments have even banned the harvest until biologists determine that the population has returned to sustainable levels. In fact, such is the case with the conch, the harvesting of which has been effectively banned in the United States since 1986.

Visit this website for more on the queen conch industry.



POSITIVE EXTERNALITIES IN PUBLIC HEALTH PROGRAMS

One of the most remarkable changes in the standard of living in the last several centuries is that people are living longer. Thousands of years ago, human life expectancy is believed to have been in the range of 20 to 30 years. By 1900, average life expectancy in the United States was 47 years. By 2015, life expectancy is 79 years. Most of the gains in life expectancy in the history of the human race happened in the twentieth century.

The rise in life expectancy seems to stem from three primary factors. First, systems for providing clean water and disposing of human waste helped to prevent the transmission of many diseases. Second, changes in public behavior have advanced health. Early in the twentieth century, for example, people learned the importance of boiling bottles before using them for food storage and baby's milk, washing their hands, and protecting food from flies. More recent behavioral changes include reducing the number of people who smoke tobacco and precautions to limit sexually transmitted diseases. Third, medicine has played a large role. Immunizations for diphtheria, cholera, pertussis, tuberculosis, tetanus, and yellow fever were developed between 1890 and 1930. Penicillin, discovered in 1941, led to a series of other antibiotic drugs for bringing infectious diseases under control. In recent decades, drugs that reduce the risks of high blood pressure have had a dramatic effect in extending lives.

These advances in public health have all been closely linked to positive externalities and public goods. Public health officials taught hygienic practices to mothers in the early 1900s and encouraged less smoking in the late 1900s. Many public sanitation systems and storm sewers were funded by government because they have the key traits of public goods. In the twentieth century, many medical discoveries came out of government or university-funded research. Patents and intellectual property rights provided an additional incentive for private inventors. The reason for requiring immunizations, phrased in economic terms, is that it prevents spillovers of illness to others—as well as helping the person immunized.

THE BENEFITS OF VOYAGER I LIVE ON

While we applaud the technology spillovers of NASA's space projects, we should also acknowledge that those benefits are not shared equally. Economists like **Tyler Cowen**, a professor at George Mason University, are seeing more and more evidence of a widening gap between those who have access to rapidly improving technology, and those who do not. According to Cowen author of the recent book, *Average Is Over: Powering America Beyond the Age of the Great Stagnation*, this inequality in access to technology and information is going to deepen the inequality in skills, and ultimately, in wages and global standards of living.

KEY CONCEPTS AND SUMMARY

A public good has two key characteristics: it is nonexcludable and nonrivalrous. Nonexcludable means that it is costly or impossible for one user to exclude others from using the good. Nonrivalrous means that when one person uses the good, it does not prevent others from using it. Markets often have a difficult time producing public goods because free riders will attempt to use the public good without paying for it. The free rider problem can be overcome through measures to assure that users of the public good pay for it. Such measures include government actions, social pressures, and specific situations where markets have discovered a way to collect payments.

SELF-CHECK QUESTIONS

1. Which of the following goods or services are nonexcludable?
 - a. police protection
 - b. streaming music from satellite transmission programs
 - c. roads
 - d. primary education
 - e. cell phone service
2. Are the following goods nonrivalrous in consumption?
 - a. slice of pizza
 - b. laptop computer
 - c. public radio
 - d. ice cream cone

REVIEW QUESTIONS

1. What are the two key characteristics of public goods?
2. Name two public goods and explain why they are public goods.
3. What is the free rider problem?
4. Explain why the federal government funds national defense.

CRITICAL THINKING QUESTIONS

1. How do public television stations, like PBS, try to overcome the free rider problem?
2. Why is a football game on ESPN a quasi-public good but a game on the NBC, CBS, or ABC is a public good?
3. Provide two examples of goods/services that are classified as private goods/services even though they are provided by a federal government.
4. Radio stations, tornado sirens, light houses, and street lights are all public goods in that all are nonrivalrous and nonexclusionary. Therefore why does the government provide tornado sirens, street lights and light houses but not radio stations (other than PBS stations)?

PROBLEMS

Becky and Sarah are sisters who share a room. Their room can easily get messy, and their parents are always telling them to clean it up. Here are the costs and benefits to both Becky and Sarah, of taking the time to clean their room: If both Becky and Sarah clean, they each spends two hours and get a clean room. If Becky decides not to clean and Sarah does all the cleaning, then Sarah spends 10 hours cleaning (Becky spends 0) but Sarah is exhausted. The same would occur for Becky if Sarah decided not to clean—Becky spends 10 hours and becomes exhausted. If both girls decide not to clean, they both have a dirty room.

- a. What is the best outcome for Becky and Sarah? What is the worst outcome? (It would help you to construct a prisoner's dilemma table.)
- b. Unfortunately, we know that the optimal outcome will most likely not happen, and that the worst one will probably be chosen instead. Explain what it is about Becky's and Sarah's reasoning that will lead them both to choose the worst outcome.

REFERENCES

Cowen, Tyler. *Average Is Over: Powering America Beyond the Age of the Great Stagnation*. Dutton Adult, 2013.

Hardin, Garret. "The Tragedy of the Commons." *Science* 162 (3859): 1243–48 (1968).

GLOSSARY

free rider those who want others to pay for the public good and then plan to use the good themselves; if many people act as free riders, the public good may never be provided

nonexcludable when it is costly or impossible to exclude someone from using the good, and thus hard to charge for it

nonrivalrous even when one person uses the good, others can also use it

public good good that is nonexcludable and nonrivalrous, and thus is difficult for market producers to sell to individual consumers

SOLUTIONS

Answers to Self-Check Questions

1.
 - a. Once citizens are protected from crime, it is difficult to exclude someone from this protection, so it is nonexcludable.
 - b. Some satellite radio services, such as SiriusXM, are sold by subscription fee, so it is excludable.
 - c. Once a road is built it is difficult to exclude people, although toll roads can exclude non-payers.
 - d. Primary education can be provided by private companies and so it is excludable.
 - e. Companies sell cell phone service and exclude those who do not pay.
2.
 - a. Two people cannot enjoy the same slice of pizza at the same time, so private goods, such as a slice of pizza, are rivalrous.
 - b. Two people cannot use one laptop at the same time, so they are rivalrous in consumption.
 - c. Public radio can be heard by anyone with a radio, so many people can listen at the same time—the good is nonrivalrous.
 - d. It is difficult for two people to simultaneously eat an ice cream cone, so it is rivalrous in consumption.

CHAPTER 20. POVERTY AND ECONOMIC INEQUALITY

INTRODUCTION TO POVERTY AND ECONOMIC INEQUALITY



Figure 1. Occupying Wall Street. On September 17, 2011, Occupy Wall Street began in New York City's Wall Street financial district. (Credit: modification of work by David Shankbone/Flickr Creative Commons)

OCCUPY WALL STREET

In September 2011, a group of protesters gathered in Zuccotti Park in New York City to decry what they perceived as increasing social and economic inequality in the United States. Calling their protest “Occupy Wall Street,” they argued that the concentration of wealth among the richest 1% in the United States was both economically unsustainable and inequitable, and needed to be changed. The protest then spread to other major cities, and the Occupy movement was born. Why were people so upset? How much wealth is concentrated among the top 1% in our society? How did they acquire so much wealth? These are very real, very important questions in the United States now, and this chapter on poverty and economic inequality will help us address the causes behind this sentiment.

CHAPTER OBJECTIVES

Introduction to Poverty and Economic Inequality

In this chapter, you will learn about:

- Drawing the Poverty Line
- The Poverty Trap
- The Safety Net
- Income Inequality: Measurement and Causes
- Government Policies to Reduce Income Inequality

The labor markets that determine what workers are paid do not take into account how much income a family needs for food, shelter, clothing, and health care. Market forces do not worry about what happens to families when a major local employer goes out of business. Market forces do not take time to contemplate whether those who are earning higher incomes should pay an even higher share of taxes.

However, labor markets do create considerable inequalities of income. In 2014, the median American family income was \$57,939 (the median is the level where half of all families had more than that level and half had less). According to the U.S. Census Bureau, almost nine million U.S. families were classified by the federal government as being below the poverty line in that year. Think about a family of three—perhaps a single mother with two children—attempting to pay for the basics of life on perhaps \$17,916 per year. After paying for rent, healthcare, clothing, and transportation, such a family might have \$6,000 to spend on food. Spread over 365 days, the food budget for the entire family would be about \$17 per day. To put this in perspective, most cities have restaurants where \$17 will buy you an appetizer for one.

This chapter explores how the U.S. government defines poverty, the balance between assisting the poor without discouraging work, and how federal antipoverty programs work. It also discusses income inequality—how economists measure inequality, why inequality has changed in recent decades, the range of possible government policies to reduce inequality, and the danger of a tradeoff that too great a reduction in inequality may reduce incentives for producing output.

20.1 DRAWING THE POVERTY LINE

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain economic inequality and how the poverty line is determined
- Analyze the U.S. poverty rate over time, noting its prevalence among different groups of citizens

Comparisons of high and low incomes raise two different issues: economic inequality and **poverty**. Poverty is measured by the number of people who fall below a certain level of income—called the **poverty line**—that defines the income needed for a basic standard of living. **Income inequality** compares the share of the total income (or wealth) in society that is received by different groups; for example, comparing the share of income received by the top 10% to the share of income received by the bottom 10%.

In the United States, the official definition of the poverty line traces back to a single person: **Mollie Orshansky**. In 1963, Orshansky, who was working for the Social Security Administration, published an article called “Children of the Poor” in a highly useful and dry-as-dust publication called the *Social Security Bulletin*. Orshansky’s idea was to define a poverty line based on the cost of a healthy diet.

Her previous job had been at the U.S. Department of Agriculture, where she had worked in an agency called the Bureau of Home Economics and Human Nutrition. One task of this bureau had been to calculate how much it would cost to feed a nutritionally adequate diet to a family. Orshansky found that the average family spent one-third of its income on food. She then proposed that the poverty line be the amount needed to buy a nutritionally adequate diet, given the size of the family, multiplied by three.

The current U.S. poverty line is essentially the same as the Orshansky poverty line, although the dollar amounts are adjusted each year to represent the same buying power over time. The U.S. poverty line in 2015 ranged from \$11,790 for a single individual to \$25,240 for a household of four people.

Figure 1 shows the U.S. **poverty rate** over time; that is, the percentage of the population below the poverty line in any given year. The poverty rate declined through the 1960s, rose in the early 1980s and early 1990s, but seems to have been slightly lower since the mid-1990s. However, in no year in the last four decades has the poverty rate been less than 11% of the U.S. population—that is, at best about one American in nine is below the poverty line. In recent years, the poverty rate appears to have peaked at 15.9% in 2011 before dropping to 14.5% in 2013. Table 1 compares poverty rates for different groups in 2011. As you will see when we delve further into these numbers, poverty rates are rela-

tively low for whites, for the elderly, for the well-educated, and for male-headed households. Poverty rates for females, Hispanics, and African Americans are much higher than for whites. While Hispanics and African Americans have a higher percentage of individuals living in poverty than others, most people in the United States living below the poverty line are white.

Visit this website for more information on U.S. poverty.

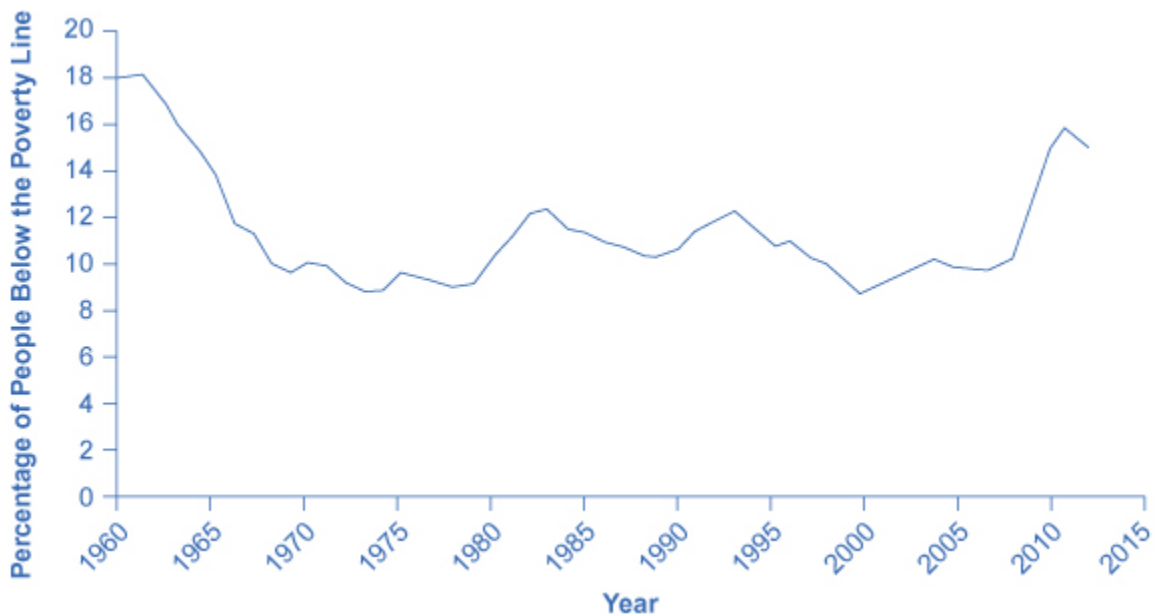


Figure 1. The U.S. Poverty Rate since 1960. The poverty rate fell dramatically during the 1960s, rose in the early 1980s and early 1990s, and, after declining in the 1990s through mid-2000s, rose to 15.9% in 2011, which is close to the 1960 levels. In 2013, the poverty dropped slightly to 14.5%. (Source: U.S. Census Bureau)

Group	Poverty Rate
Females	15.8%
Males	13.1%
White	9.6%
Black	27.1%
Hispanic	23.5%
Under age 18	19.9%
Ages 18–24	20.6%
Ages 25–34	15.9%
Ages 35–44	12.2%
Ages 45–54	10.9%
Ages 55–59	10.7%
Ages 60–64	10.8%
Ages 65 and older	9.5%

Table 1. Poverty Rates by Group, 2013

The concept of a poverty line raises many tricky questions. In a vast country like the United States, should there be a national poverty line? After all, according to the Federal Register, the median household income for a family of four was \$102,552 in New Jersey and \$57,132 in Mississippi in 2013, and prices of some basic goods like housing are quite different between states. The poverty line is based on cash income, which means it does not take into account government programs that provide assistance to the poor in a non-cash form, like **Medicaid** (health care for low-income individuals and families) and food aid. Also, low-income families can qualify for federal housing assistance. (These and other government aid programs will be discussed in detail later in this chapter.)

Should the poverty line be adjusted to take the value of such programs into account? Many economists and policymakers wonder whether the concept of what poverty means in the twenty-first century should be rethought. The following Clear It Up feature explains the poverty lines set by the World Bank for low-income countries around the world.

HOW IS POVERTY MEASURED IN LOW-INCOME COUNTRIES?

The World Bank sets two poverty lines for low-income countries around the world. One poverty line is set at an income of \$1.25/day per person; the other is at \$2/day. By comparison, the U.S. 2015 poverty line of \$20,090 annually for a family of three works out to \$18.35 per person per day.

Clearly, many people around the world are far poorer than Americans, as Table 2 shows. China and India both have more than a billion people; Nigeria is the most populous country in Africa; and Egypt is the most populous country in the Middle East. In all four of those countries, in the mid-2000s, a substantial share of the population subsisted on less than \$2/day. Indeed, about half the world lives on less than \$2.50 a day, and 80 percent of the world lives on less than \$10 per day. (Of course, the cost of food, clothing, and shelter in those countries can be very different from those costs in the United States, so the \$2 and \$2.50 figures may mean greater purchasing power than they would in the United States.)

Country	Share of Population below \$1.25/Day	Share of Population below \$2.00/Day
Brazil (in 2009)	6.1%	10.8%
China (in 2009)	11.8%	27.2%
Egypt (in 2008)	1.7%	15.4%
India (in 2010)	32.7%	68.8%
Mexico (in 2010)	0.7%	4.5%
Nigeria (in 2010)	68.0%	84.5%

Table 2. Poverty Lines for Low-Income Countries, mid-2000s. (Source: <http://data.worldbank.org/indicator/SI.POV.DDAY>)

Any poverty line will be somewhat arbitrary, and it is useful to have a poverty line whose basic definition does not change much over time. If Congress voted every few years to redefine what poverty means, then it would be difficult to compare rates over time. After all, would a lower poverty rate mean that the definition had been changed, or that people were actually better off? Government statisticians at the U.S. Census Bureau have ongoing research programs to address questions like these.

KEY CONCEPTS AND SUMMARY

Wages are influenced by supply and demand in labor markets, which can lead to very low incomes for some people and very high incomes for others. Poverty and income inequality are not the same thing. Poverty applies to the condition of people who cannot afford the necessities of life. Income inequality refers to the disparity between those with higher and lower incomes. The poverty rate is what percentage of the population lives below the poverty line, which is determined by the amount of income that it takes to purchase the necessities of life. Choosing a poverty line will always be somewhat controversial.

SELF-CHECK QUESTIONS

Describe how each of these changes is likely to affect poverty and inequality:

- Incomes rise for low-income and high-income workers, but rise more for the high-income earners.
- Incomes fall for low-income and high-income workers, but fall more for high-income earners.

REVIEW QUESTIONS

- How is the poverty rate calculated?
- What is the poverty line?
- What is the difference between poverty and income inequality?

CRITICAL THINKING QUESTIONS

1. What goods and services would you include in an estimate of the basic necessities for a family of four?
2. If a family of three earned \$20,000, would they be able to make ends meet given the official poverty threshold?

PROBLEMS

1. In country A, the population is 300 million and 50 million people are living below the poverty line. What is the poverty rate?
2. In country B, the population is 900 million and 100 million people are living below the poverty line. What is the poverty rate?

REFERENCES

Ebeling, Ashlea. 2014. "IRS Announces 2015 Estate And Gift Tax Limits." *Forbes*. Accessed March 16, 2015. <http://www.forbes.com/sites/ashleaebeling/2014/10/30/irs-announces-2015-estate-and-gift-tax-limits/>.

Federal Register: The Daily Journal of the United States Government. "State Median Income Estimates for a Four-Person Household: Notice of the Federal Fiscal Year (FFY) 2013 State Median Income Estimates for Use Under the Low Income Home Energy Assistance Program (LIHEAP)." Last modified March 15, 2012. <https://www.federalregister.gov/articles/2012/03/15/2012-6220/state-median-income-estimates-for-a-four-person-household-notice-of-the-federal-fiscal-year-ffy-2013#t-1>.

Luhby, Tami. 2014. "Income is on the rise . . . finally!" Accessed April 10, 2015. <http://money.cnn.com/2014/08/20/news/economy/median-income/>.

Meyer, Ali. 2015. "56,023,000: Record Number of Women Not in Labor Force." *CNSNews.com*. Accessed March 16, 2015. <http://cnsnews.com/news/article/ali-meyer/56023000-record-number-women-not-labor-force>.

Orshansky, Mollie. "Children of the Poor." *Social Security Bulletin*. 26 no. 7 (1963): 3–13. <http://www.ssa.gov/policy/docs/ssb/v26n7/v26n7p3.pdf>.

The World Bank. "Data: Poverty Headcount Ratio at \$1.25 a Day (PPP) (% of Population)." <http://data.worldbank.org/indicator/SI.POV.DDAY>.

U.S. Department of Commerce: United States Census Bureau. "American FactFinder." <http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml>.

U.S. Department of Commerce: United States Census Bureau. "Current Population Survey (CPS): CPS Table Creator." <http://www.census.gov/cps/data/cpstablecreator.html>.

U.S. Department of Commerce: United States Census Bureau. “Income: Table F-6. Regions—Families (All Races) by Median and Mean Income.” <http://www.census.gov/hhes/www/income/data/historical/families/>.

U.S. Department of Commerce: United States Census Bureau. “Poverty: Poverty Thresholds.” Last modified 2012. <http://www.census.gov/hhes/www/poverty/data/threshld/>.

U.S. Department of Health & Human Services. 2015. “2015 Poverty Guidelines.” Accessed April 10, 2015. <http://www.medicaid.gov/medicaid-chip-program-information/by-topics/eligibility/downloads/2015-federal-poverty-level-charts.pdf>.

U.S. Department of Health & Human Services. 2015. “Information on Poverty and Income Statistics: A Summary of 2014 Current Population Survey Data.” Accessed April 13, 2015. http://aspe.hhs.gov/hsp/14/povertyandincomeest/ib_poverty2014.pdf.

GLOSSARY

income inequality when one group receives a disproportionate share of total income or wealth than others

poverty the situation of being below a certain level of income needed for a basic standard of living

poverty line the specific amount of income needed for a basic standard of living

poverty rate percentage of the population living below the poverty line

SOLUTIONS

Answers to Self-Check Questions

- a. Poverty falls, inequality rises.
- b. Poverty rises, inequality falls.

20.2 THE POVERTY TRAP

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the poverty trap, noting how it is impacted by government programs
- Identify potential issues in government programs that seek to reduce poverty
- Calculate a budget constraint line that represents the poverty trap

Can you give people too much help, or the wrong kind of help? When people are provided with food, shelter, healthcare, income, and other necessities, assistance may reduce their incentive to work. Consider a program to fight poverty that works in this reasonable-sounding manner: the government provides assistance to the poor, but as the poor earn income to support themselves, the government reduces the level of assistance it provides. With such a program, every time a poor person earns \$100, the person loses \$100 in government support. As a result, the person experiences no net gain for working. Economists call this problem the **poverty trap**.

Consider the situation faced by a single-parent family. A single mother (earning \$8 an hour) with two children, as illustrated in Figure 1. First, consider the **labor-leisure budget constraint** faced by this family in a situation without government assistance. On the horizontal axis is hours of leisure (or time spent with family responsibilities) increasing in quantity from right to left. Also on the horizontal axis is the number of hours at paid work, going from zero hours on the right to the maximum of 2,500 hours on the left. On the vertical axis is the amount of income per year rising from low to higher amounts of income. The budget constraint line shows that at zero hours of leisure and 2,500 hours of work, the maximum amount of income is \$20,000 ($\$8 \times 2,500$ hours). At the other extreme of the budget constraint line, an individual would work zero hours, earn zero income, but enjoy 2,500 hours of leisure. At point A on the **budget constraint line**, by working 40 hours a week, 50 weeks a year, the utility-maximizing choice is to work a total of 2,000 hours per year and earn \$16,000.

Now suppose that a government antipoverty program guarantees every family with a single mother and two children \$18,000 in income. This is represented on the graph by a horizontal line at \$18,000. With this program, each time the mother earns \$1,000, the government will deduct \$1,000 of its support. Table 3 shows what will happen at each combination of work and government support.

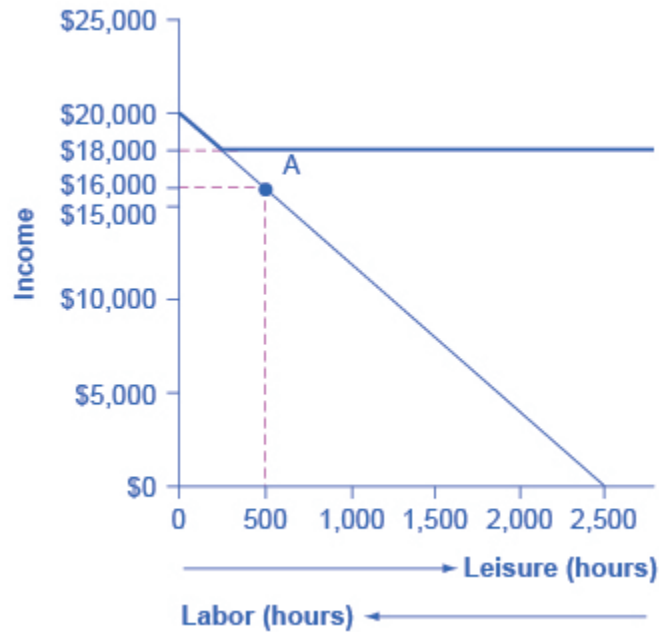


Figure 1. The Poverty Trap in Action. The original choice is 500 hours of leisure, 2,000 hours of work at point A, and income of \$16,000. With a guaranteed income of \$18,000, this family would receive \$18,000 whether it provides zero hours of work or 2,000 hours of work. Only if the family provides, say, 2,300 hours of work does its income rise above the guaranteed level of \$18,000—and even then, the marginal gain to income from working many hours is small.

Amount Worked (hours)	Total Earnings	Government Support	Total Income
0	0	\$18,000	\$18,000
500	\$4,000	\$14,000	\$18,000
1,000	\$8,000	\$10,000	\$18,000
1,500	\$12,000	\$6,000	\$18,000
2,000	\$16,000	\$2,000	\$18,000
2,500	\$20,000	0	\$20,000

Table 3. Total Income at Various Combinations of Work and Support

The new budget line, with the antipoverty program in place, is the horizontal and heavy line that is flat at \$18,000. If the mother does not work at all, she receives \$18,000, all from the government. If she works full time, giving up 40 hours per week with her children, she still ends up with \$18,000 at the end of the year. Only if she works 2,300 hours in the year—which is an average of 44 hours per week for 50 weeks a year—does household income rise to \$18,400. Even in this case, all of her year's work means that household income rises by only \$400 over the income she would receive if she did not work at all. She would need to work 50 hours a week to reach \$20,000.

Indeed, the poverty trap is even stronger than this simplified example shows, because a working

mother will have extra expenses like clothing, transportation, and child care that a nonworking mother will not face, making the economic gains from working even smaller. Moreover, those who do not work fail to build up job experience and contacts, which makes working in the future even less likely.

The bite of the poverty trap can be reduced by designing an antipoverty program so that, instead of reducing government payments by \$1 for every \$1 earned, payments are reduced by some smaller amount instead. The bite of the poverty trap can also be reduced by imposing requirements for work as a condition of receiving benefits and setting a time limit on benefits.

Figure 2 illustrates a government program that guarantees \$18,000 in income, even for those who do not work at all, but then reduces this amount by 50 cents for each \$1 earned. The new, higher budget line in Figure 2 shows that, with this program, additional hours of work will bring some economic gain. Because of the reduction in government income when an individual works, an individual earning \$8.00 will really net only \$4.00 per hour. The vertical intercept of this higher budget constraint line is at \$28,000 ($\$18,000 + 2,500 \text{ hours} \times \$4.00 = \$28,000$). The horizontal intercept is at the point on the graph where \$18,000 and 2500 hours of leisure is set. Table 4 shows the total income differences with various choices of labor and leisure.

However, this type of program raises other issues. First, even if it does not eliminate the incentive to work by reducing government payments by \$1 for every \$1 earned, enacting such a program may still reduce the incentive to work. At least some people who would be working 2,000 hours each year without this program might decide to work fewer hours but still end up with more income—that is, their choice on the new budget line would be like S, above and to the right of the original choice P. Of course, others may choose a point like R, which involves the same amount of work as P, or even a point to the left of R that involves more work.

The second major issue is that when the government phases out its support payments more slowly, the antipoverty program costs more money. Still, it may be preferable in the long run to spend more money on a program that retains a greater incentive to work, rather than spending less money on a program that nearly eliminates any gains from working.

Amount Worked (hours)	Total Earnings	Government Support	Total Income
0	0	\$18,000	\$18,000
500	\$4,000	\$16,000	\$20,000
1,000	\$8,000	\$14,000	\$22,000
1,500	\$12,000	\$12,000	\$24,000
2,000	\$16,000	\$10,000	\$26,000
2,500	\$20,000	\$8,000	\$28,000

Table 4. The Labor-Leisure Tradeoff with Assistance Reduced by 50 Cents for Every Dollar Earned

The next module will consider a variety of government support programs focused specifically on the poor, including welfare, SNAP (food supplement), Medicaid, and the earned income tax credit (EITC). Although these programs vary from state to state, it is generally a true statement that in many states from the 1960s into the 1980s, if poor people worked, their level of income barely rose—or did not rise at all—after the reduction in government support payments was factored in. The following Work It Out feature shows how this happens.

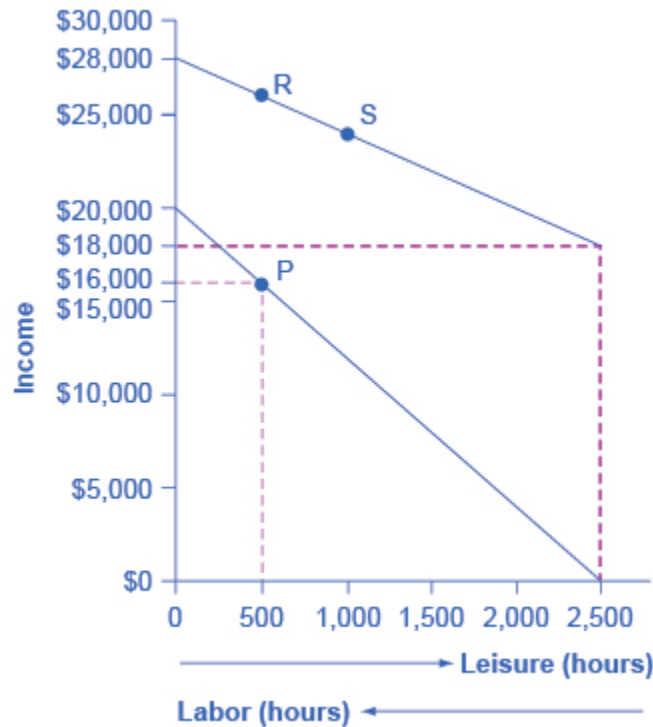


Figure 2. Loosening the Poverty Trap: Reducing Government Assistance by 50 Cents for Every \$1 Earned. On the original labor-leisure opportunity set, the lower budget set shown by the smaller dashed line in the figure, the preferred choice P is 500 hours of leisure and \$16,000 of income. Then, the government created an antipoverty program that guarantees \$18,000 in income even to those who work zero hours, shown by the larger dashed line. In addition, every \$1 earned means phasing out 50 cents of benefits. This program leads to the higher budget set shown in the diagram. The hope is that this program will provide incentives to work the same or more hours, despite receiving income assistance. However, it is possible that the recipients will choose a point on the new budget set like S, with less work, more leisure, and greater income, or a point like R, with the same work and greater income.

CALCULATING A BUDGET CONSTRAINT LINE

Jason earns \$9.00 an hour, and a government antipoverty program provides a floor of \$10,000 guaranteed income. The government reduces government support by \$0.50 for each \$1.00 earned. What are the horizontal and vertical intercepts of the budget constraint line? Assume the maximum hours for work or leisure is 2,500 hours.

Step 1. Determine the amount of the government guaranteed income. In this case, it is \$10,000.

Step 2. Plot that guaranteed income as a horizontal line on the budget constraint line.

Step 3. Determine what Jason earns if he has no income and enjoys 2,500 hours of leisure. In this case, he will receive the guaranteed \$10,000 (the horizontal intercept).

Step 4. Calculate how much Jason's salary will be reduced by due to the reduction in government income. In Jason's case, it will be reduced by one half. He will, in effect, net only \$4.50 an hour.

Step 5. If Jason works 1,000 hours, at a maximum what income will Jason receive? Jason will get the government assistance of \$10,000. He will net only \$4.50 for every hour he chooses to work. If he works 1,000 hours at \$4.50, his earned income is \$4,500 plus the government income of \$10,000. Thus the total maximum income (the vertical intercept) is $\$10,000 + \$4,500 = \$14,500$.

KEY CONCEPTS AND SUMMARY

A poverty trap occurs when government-support payments for the poor decline as the poor earn more income. As a result, the poor do not end up with much more income when they work, because the loss of government support largely or completely offsets any income that is earned by working. The bite of the poverty trap can be reduced by phasing out government benefits more slowly, as well as by imposing requirements for work as a condition of receiving benefits and a time limit on benefits.

SELF-CHECK QUESTIONS

1. Jonathon is a single father with one child. He can work as a server for \$6 per hour for up to 1,500 hours per year. He is eligible for welfare, and so if he does not earn any income, he will receive a total of \$10,000 per year. He can work and still receive government benefits, but for every \$1 of income, his welfare stipend is \$1 less. Create a table similar to Table 4 that shows Jonathan's options. Use four columns, the first showing number of hours to work, the second showing his earnings from work, the third showing the government benefits he will receive, and the fourth column showing his total income (earnings + government support). Sketch a labor-leisure diagram of Jonathan's opportunity set with and without government support.
2. Imagine that the government reworks the welfare policy that was affecting Jonathan in question 1, so that for each dollar someone like Jonathan earns at work, his government benefits diminish by only 30 cents. Reconstruct the table from question 1 to account for this change in policy. Draw Jonathan's labor-leisure opportunity sets, both for before this welfare program is enacted and after it is enacted.

REVIEW QUESTIONS

1. How does the poverty trap discourage people from working?
2. How can the effect of the poverty trap be reduced?

CRITICAL THINKING QUESTIONS

1. Self-Check Question 1 and Self-Check Question 2 asked you to describe the labor-leisure tradeoff for

- Jonathon. Since, in the first example, there is no monetary incentive for Jonathon to work, explain why he may choose to work anyway. Explain what the opportunity costs of working and not working might be for Jonathon in each example. Using your tables and graphs from Self-Check Question 1 and Self-Check Question 2, analyze how the government welfare system affects Jonathan's incentive to work.
2. Explain how you would create a government program that would give an incentive for labor to increase hours and keep labor from falling into the poverty trap.

PROBLEMS

Susan is a single mother with three children. She can earn \$8 per hour and works up to 2,000 hours per year. However, if she does not earn any income at all, she will receive government benefits totaling \$16,000 per year. For every \$1 of income she earns, her level of government support will be reduced by \$1. Create a table, patterned after Table 6. The first column should show Susan's choices of how many hours to work per year, up to 2,000 hours. The second column should show her earnings from work. The third column should show her level of government support, given her earnings. The final column should show her total income, combining earnings and government support.

GLOSSARY

poverty trap antipoverty programs set up so that government benefits decline substantially as people earn more income—as a result, working provides little financial gain

EXERCISES

Answers to Self-Check Questions

1. Jonathon's options for working and total income are shown in the following table. His labor-leisure diagram is shown in the figure following the table.

Number of Work Hours	Earnings from Work	Government Benefits	Total Income
1,500	\$9,000	\$1,000	\$10,000
1,200	\$7,200	\$2,800	\$10,000
900	\$5,400	\$4,600	\$10,000
600	\$3,600	\$6,400	\$10,000
300	\$1,800	\$8,200	\$10,000
0	\$0	\$10,000	\$10,000

Table 5.

2. The following table shows a policy where only 30 cents in government support is pulled right back for every \$1 of income earned. Jonathon's labor-leisure diagram is shown in the figure following the table. "Opportunity set after program" extends from (0, \$16,300) to (1,500, \$10,000). "Opportunity set before program" slopes downward from (0, \$9,000) to (1,500, \$0).

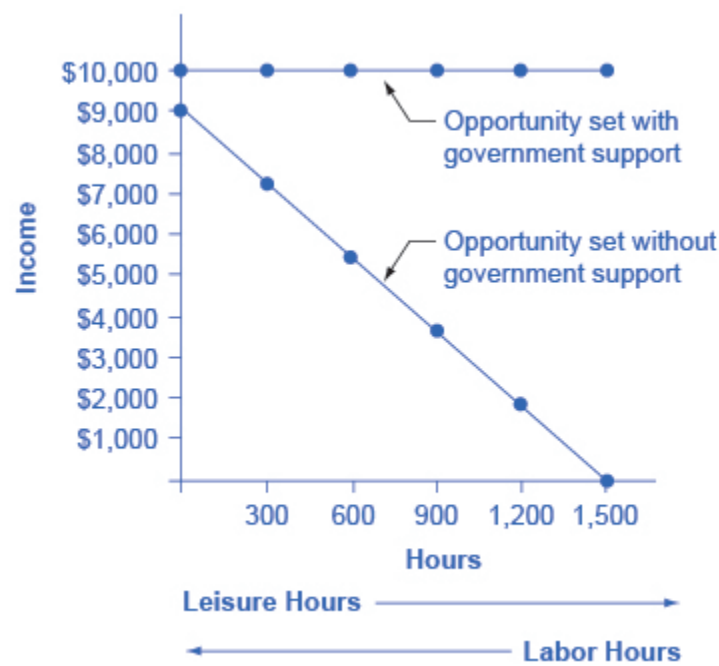


Figure 3.

Number of Work Hours	Earnings from Work	Government Benefits	Total Income
1,500	\$9,000	\$7,300	\$16,300
1,200	\$7,200	\$7,840	\$15,040
900	\$5,400	\$8,380	\$13,780
600	\$3,600	\$8,920	\$12,520
300	\$1,800	\$9,460	\$22,260
0	\$0	\$10,000	\$10,000

Table 6.

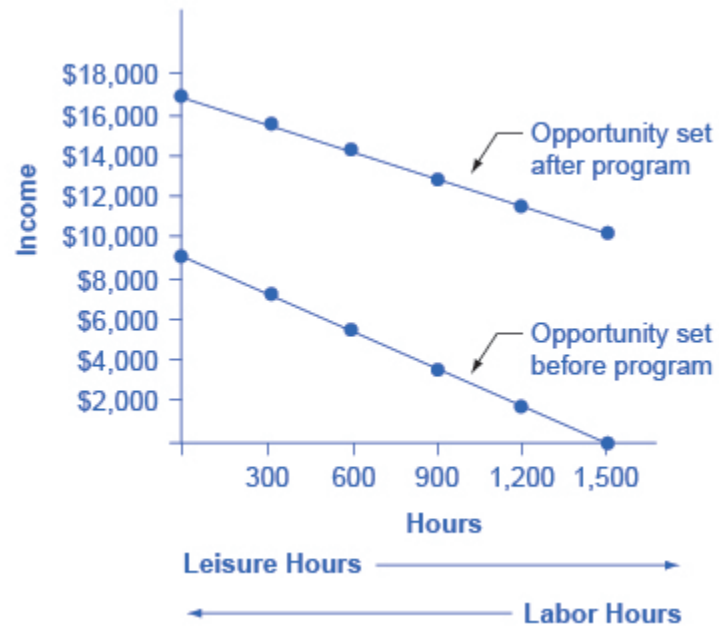


Figure 4.

20.3 THE SAFETY NET

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify the antipoverty government programs that compose the safety net
- Explain the primary goals of the safety net programs and how these programs have changed over time
- Discuss the complexities of these safety net programs and why they can be controversial

The U.S. government has implemented a number of programs to assist those below the poverty line and those who have incomes just above the poverty line, who are referred to as the **near-poor**. Such programs are called the **safety net**, in recognition of the fact that they offer some protection for those who find themselves without jobs or income.

TEMPORARY ASSISTANCE FOR NEEDY FAMILIES

From the Great Depression of the 1930s until 1996, the United States' most visible antipoverty program was **Aid to Families with Dependent Children (AFDC)**, which provided cash payments to mothers with children who were below the poverty line. This program was often just called “welfare.” In 1996, Congress passed and President Bill Clinton signed into law the Personal Responsibility and Work Opportunity Reconciliation Act, more commonly called the “welfare reform act.” The new law replaced AFDC with **Temporary Assistance for Needy Families (TANF)**.

Visit this website to watch a video of President Bill Clinton’s Welfare Reform speech.



TANF brought several dramatic changes in how welfare operated. Under the old AFDC program, states set the level of welfare benefits that they would pay to the poor, and the federal government

guaranteed it would chip in some of the money as well. The federal government's welfare spending would rise or fall depending on the number of poor people, and on how each state set its own welfare contribution.

Under TANF, however, the federal government gives a fixed amount of money to each state. The state can then use the money for almost any program with an antipoverty component: for example, the state might use the money to give cash to poor families, or to reduce teenage pregnancy, or even to raise the high school graduation rate. However, the federal government imposed two key requirements. First, if states are to keep receiving the TANF grants, they must impose work requirements so that most of those receiving TANF benefits are working (or attending school). Second, no one can receive TANF benefits with federal money for more than a total of five years over his or her lifetime. The old AFDC program had no such work requirements or time limits.

TANF attempts to avoid the poverty trap by requiring that welfare recipients work and by limiting the length of time they can receive benefits. In its first few years, the program was quite successful. The number of families receiving payments in 1995, the last year of AFDC, was 4.8 million. By 2012, according to the Congressional Research Service, the average number of families receiving payments under TANF was 1.8 million—a decline of more than half.

TANF benefits to poor families vary considerably across states. For example, again according to the Congressional Research Service, in 2011 the highest monthly payment in Alaska to a single mother with two children was \$923, while in Mississippi the highest monthly payment to that family was \$170. These payments reflect differences in states' cost of living. Total spending on TANF was approximately \$16.6 billion in 1997. As of 2012, spending was at \$12 billion, an almost 28% decrease, split about evenly between the federal and state governments. When you take into account the effects of inflation, the decline is even greater. Moreover, there seemed little evidence that poor families were suffering a reduced standard of living as a result of TANF—although, on the other side, there was not much evidence that poor families had greatly improved their total levels of income, either.

THE EARNED INCOME TAX CREDIT (EITC)

The **earned income tax credit (EITC)**, first passed in 1975, is a method of assisting the working poor through the tax system. The EITC is one of the largest assistance program for low-income groups, and projections for 2013 expected 26 million households to take advantage of it at an estimated cost of \$50 billion. In 2013, for example, a single parent with two children would have received a tax credit of \$5,372 up to an income level of \$17,530. The amount of the tax break increases with the amount of income earned, up to a point. The earned income tax credit has often been popular with both economists and the general public because of the way it effectively increases the payment received for work.

What about the danger of the poverty trap that every additional \$1 earned will reduce government support payments by close to \$1? To minimize this problem, the earned income tax credit is phased out slowly. According to the Tax Policy Center, for a single-parent family with two children in 2013, the credit is not reduced at all (but neither is it increased) as earnings rise from \$13,430 to \$17,530. Then, for every \$1 earned above \$17,530, the amount received from the credit is reduced by 21.06 cents, until the credit phases out completely at an income level of \$46,227.

Figure 1 illustrates that the earned income tax credits, child tax credits, and the TANF program all

cost the federal government money—either in direct outlays or in loss of tax revenues. CTC stands for the government tax cuts for the child tax credit.

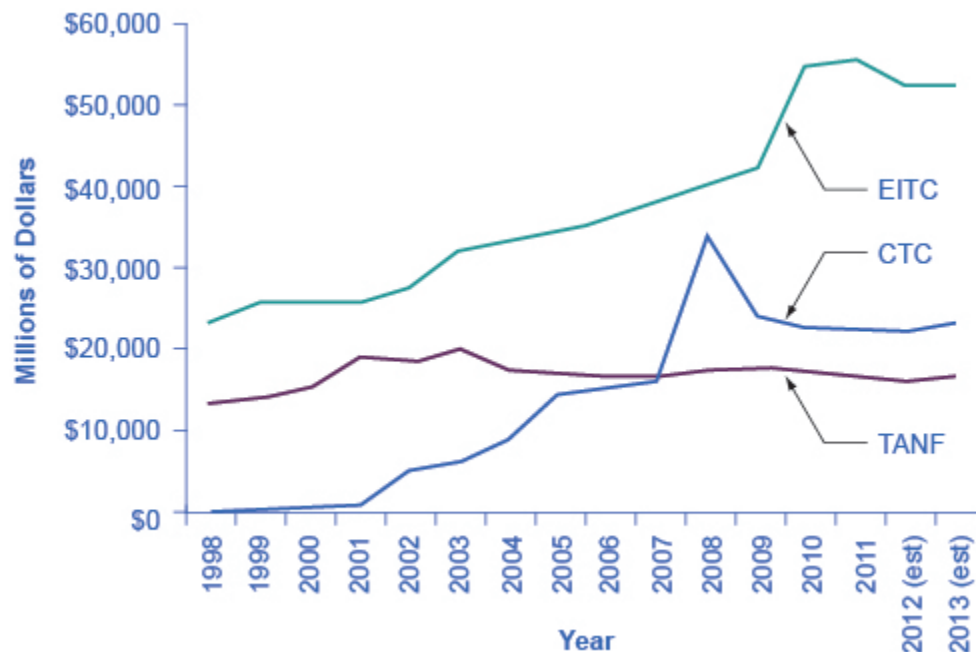


Figure 1. Real Federal Spending on CTC, EITC, and TANF, 1975-2013. EITC increased from more than \$20 billion in 2000 to over an estimated \$50 billion by 2013, far exceeding estimated 2013 outlays in the CTC (Child Tax Credits) and TANF of over \$20 billion and \$10 billion, respectively. (Source: Office of Management and Budget)

In recent years, the EITC has become a hugely expensive government program for providing income assistance to the poor and near-poor, costing about \$60 billion in 2012. In that year, the EITC provided benefits to about 27 million families and individuals and, on average, is worth about \$2,296 per family (with children), according to the Tax Policy Center. One reason that the TANF law worked as well as it did is that the EITC was greatly expanded in the late 1980s and again in the early 1990s, which increased the returns to work for low-income Americans.

SUPPLEMENTAL NUTRITION ASSISTANCE PROGRAM (SNAP)

Often called “food stamps,” **Supplemental Nutrition Assistance Program (SNAP)** is a federally funded program, started in 1964, in which each month poor people receive a card like a debit card that they can use to buy food. The amount of food aid for which a household is eligible varies by income, number of children, and other factors but, in general, households are expected to spend about 30% of their own net income on food, and if 30% of their net income is not enough to purchase a nutritionally adequate diet, then those households are eligible for SNAP.

SNAP can contribute to the poverty trap. For every \$100 earned, the government assumes that a family can spend \$30 more for food, and thus reduces its eligibility for food aid by \$30. This decreased benefit is not a complete disincentive to work—but combined with how other programs reduce benefits as income increases, it adds to the problem. SNAP, however, does try to address the poverty trap with its own set of work requirements and time limits.

Why give debit cards and not just cash? Part of the political support for SNAP comes from a belief that since the cards must be spent on food, they cannot be “wasted” on other forms of consumption. From an economic point of view, however, the belief that cards must increase spending on food seems wrong-headed. After all, say that a poor family is spending \$2,500 per year on food, and then it starts receiving \$1,000 per year in SNAP aid. The family might react by spending \$3,500 per year on food (income plus aid), or it might react by continuing to spend \$2,500 per year on food, but use the \$1,000 in food aid to free up \$1,000 that can now be spent on other goods. So it is reasonable to think of SNAP cards as an alternative method, along with TANF and the earned income tax credit, of transferring income to the working poor.

Indeed, anyone eligible for TANF is also eligible for SNAP, although states can expand eligibility for food aid if they wish to do so. In some states, where TANF welfare spending is relatively low, a poor family may receive more in support from SNAP than from TANF. In 2014, about 40 million people received food aid at an annual cost of about \$76 billion, with an average monthly benefit of about \$287 per person per month. SNAP participation increased by 70% between 2007 and 2011, from 26.6 million participants to 45 million. According to the Congressional Budget Office, this dramatic rise in participation was caused by the **Great Recession** of 2008–2009 and rising food prices.

The federal government deploys a range of income security programs that are funded through departments such as Health and Human Services, Agriculture, and Housing and Urban Development (HUD) (see Figure 2). According to the Office of Management and Budget, collectively, these three departments provided an estimated \$62 billion of aid through programs such as supplemental feeding programs for women and children, subsidized housing, and energy assistance. The federal government also transfers funds to individual states through special grant programs.

The safety net includes a number of other programs: government-subsidized school lunches and breakfasts for children from low-income families; the **Special Supplemental Food Program for Women, Infants and Children (WIC)**, which provides food assistance for pregnant women and newborns; the Low Income Home Energy Assistance Program, which provides help with home heating bills; housing assistance, which helps pay the rent; and Supplemental Security Income, which provides cash support for the disabled and the elderly poor.

MEDICAID

Medicaid was created by Congress in 1965 and is a joint health insurance program entered into by both the states and the federal government. The federal government helps fund Medicaid, but each state is responsible for administering the program, determining the level of benefits, and determining eligibility. It provides medical insurance for certain low-income people, including those below the poverty line, with a focus on families with children, the elderly, and the disabled. About one-third of Medicaid spending is for low-income mothers with children. While an increasing share of the program funding in recent years has gone to pay for nursing home costs for the elderly poor. The program ensures that a basic level of benefits is provided to Medicaid participants, but because each state sets eligibility requirements and provides varying levels of service, the program differs from state to state.

In the past, a common problem has been that many low-paying jobs pay enough to a breadwinner so that a family could lose its eligibility for Medicaid, yet the job does not offer health insurance benefits. A poor parent considering such a job might choose not to work rather than lose health insurance for

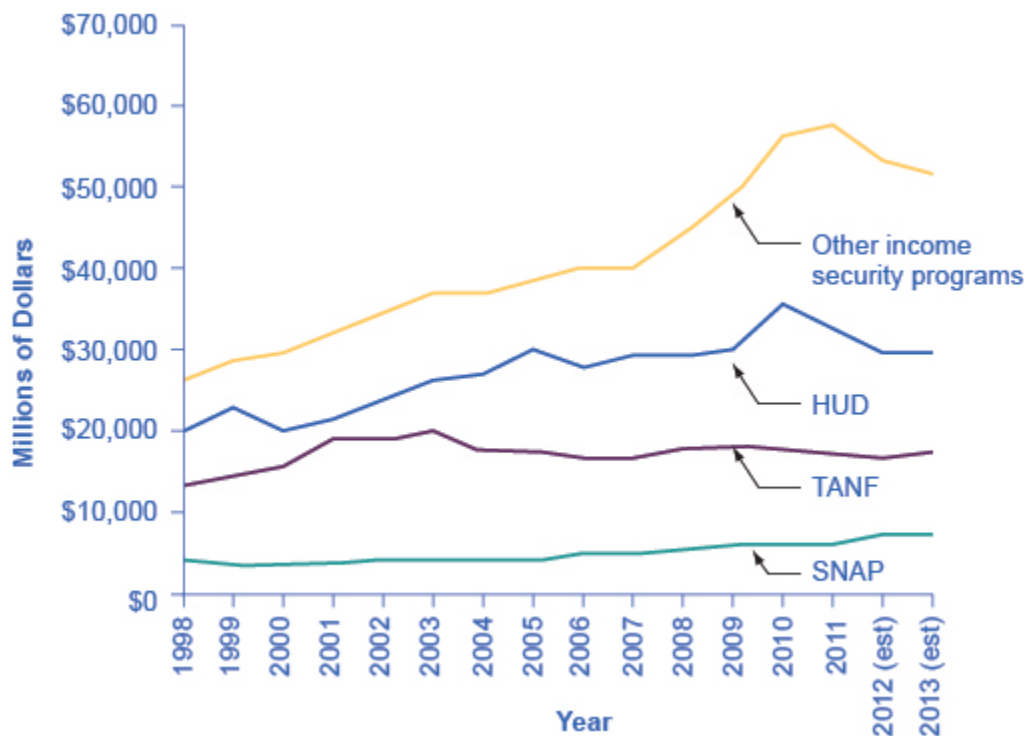


Figure 2. Expenditure Comparison of TANF, SNAP, HUD, and Other Income Security Programs, 1988–2013 (est.). Total expenditures on income security continued to rise between 1988 and 2010, while payments for TANF have increased from \$13 billion in 1998 to an estimated \$17.3 billion in 2013. SNAP has seen relatively small increments. These two programs comprise a relatively small portion of the estimated \$106 billion dedicated to income security in 2013. Note that other programs and housing programs increased dramatically during the 2008 and 2010 time periods. (Source: Table 12.3 Section 600 Income Security, <https://www.whitehouse.gov/sites/default/files/omb/budget/fy2013/assets/hist.pdf>)

his or her children. In this way, health insurance can become a part of the poverty trap. Many states recognized this problem in the 1980s and 1990s and expanded their Medicaid coverage to include not just the poor, but the near-poor earning up to 135% or even 185% of the poverty line. Some states also guaranteed that children would not lose coverage if their parents worked.

These expanded guarantees cost the government money, of course, but they also helped to encourage those on welfare to enter the labor force. As of 2014, approximately 69.7 million people participated in Medicaid. Of those enrolled, almost half are children. Healthcare expenditures, however, are highest for the elderly population, which comprises approximately 25% of participants. As Figure 3 (a) indicates, the largest number of households that enroll in Medicaid are those with children. Lower-income adults are the next largest group enrolled in Medicaid at 28%. The blind and disabled are 16% of those enrolled, and seniors are 9% of those enrolled. Figure 3 (b) shows how much actual Medicaid dollars are spent for each group. Out of total Medicaid spending, more is spent on seniors (20%) and the blind and disabled (44%). So, 64% of all Medicaid spending goes to seniors, the blind, and disabled. Children receive 21% of all Medicaid spending, followed by adults at 15%.

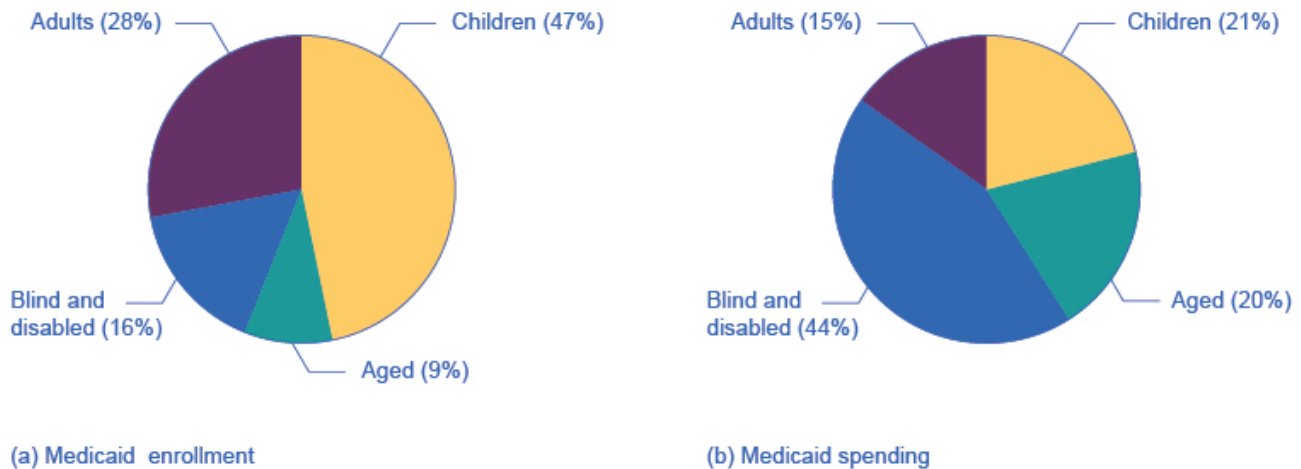


Figure 3. Medicaid Enrollment and Spending. Part (a) shows the Medicaid enrollment by different populations, with children comprising the largest percentage at 47%, followed by adults at 28%, and the blind and disabled at 16%. Part (b) shows that Medicaid spending is principally for the blind and disabled, followed by the elderly. Although children are the largest population covered by Medicaid, expenditures on children are only at 21%.

KEY CONCEPTS AND SUMMARY

The group of government programs that assist the poor are called the safety net. In the United States, prominent safety net programs include Temporary Assistance to Needy Families (TANF), the Supplemental Nutrition Assistance Program (SNAP), the earned income tax credit (EITC), Medicaid, and the Special Supplemental Food Program for Women, Infants, and Children (WIC).

SELF-CHECK QUESTIONS

1. We have discovered that the welfare system discourages recipients from working because the more income they earn, the less welfare benefits they receive. How does the earned income tax credit attempt to loosen the poverty trap?
2. How does the TANF attempt to loosen the poverty trap?

REVIEW QUESTIONS

1. Who are the near-poor?
2. What is the safety net?
3. Briefly explain the differences between TANF, the earned income tax credit, SNAP, and Medicaid.

CRITICAL THINKING QUESTIONS

1. Many critics of government programs to help low-income individuals argue that these programs create a poverty trap. Explain how programs such as TANF, EITC, SNAP, and Medicaid will affect low-income individuals and whether or not you think these programs will benefit families and children.
2. Think about the business cycle: during a recession, unemployment increases; it decreases in an expansionary phase. Explain what happens to TANF, SNAP, and Medicaid programs at each phase of the business cycle (recession, trough, expansion, and peak).

REFERENCES

Congressional Budget Office. 2015. "The Effects of Potential Cuts in SNAP Spending on Households With Different Amounts of Income." Accessed April 13, 2015. <https://www.cbo.gov/publication/49978>.

Falk, Gene. Congressional Research Service. "The Temporary Assistance for Needy Families (TANF) Block Grant: Responses to Frequently Asked Questions." Last modified October 17, 2013. <http://www.fas.org/sgp/crs/misc/RL32760.pdf>.

Library of Congress. "Congressional Research Service." <http://www.loc.gov/crsinfo/about/>.

Office of Management and Budget. "Fiscal Year 2013 Historical Tables: Budget of the U.S. Government." <http://www.whitehouse.gov/sites/default/files/omb/budget/fy2013/assets/hist.pdf>.

Tax Policy Center: Urban Institute and Brookings Institution. "The Tax Policy Briefing Book: Taxation and the Family: What is the Earned Income Tax Credit?" <http://www.taxpolicycenter.org/briefing-book/key-elements/family/eitc.cfm>.

GLOSSARY

earned income tax credit (EITC) a method of assisting the working poor through the tax system

Medicaid a federal–state joint program enacted in 1965 that provides medical insurance for certain (not all) low-income people, including the near-poor as well as those below the poverty line, and focusing on low-income families with children, the low-income elderly, and the disabled

near-poor those who have incomes just above the poverty line

safety net the group of government programs that provide assistance to the poor and the near-poor

Supplemental Nutrition Assistance Program (SNAP) a federally funded program, started in 1964, in which each month poor people receive SNAP cards they can use to buy food

SOLUTIONS

Answers to Self-Check Questions

1. The earned income tax credit works like this: a poor family receives a tax break that increases according to how much they work. Families that work more get more. In that sense it loosens the poverty trap by

encouraging work. As families earn above the poverty level, the earned income tax credit is gradually reduced. For those near-poor families, the earned income tax credit is a partial disincentive to work.

2. TANF attempts to loosen the poverty trap by providing incentives to work in other ways. Specifically, it requires that people work (or complete their education) as a condition of receiving TANF benefits, and it places a time limit on benefits.

20.4 INCOME INEQUALITY: MEASUREMENT AND CAUSES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the distribution of income, and analyze the sources of income inequality in a market economy
- Measure income distribution in quintiles
- Calculate and graph a Lorenz curve
- Show income inequality through demand and supply diagrams

Poverty levels can be subjective based on the overall income levels of a country; typically poverty is measured based on a percentage of the median income. Income inequality, however, has to do with the distribution of that income, in terms of which group receives the most or the least income. Income inequality involves comparing those with high incomes, middle incomes, and low incomes—not just looking at those below or near the poverty line. In turn, measuring income inequality means dividing up the population into various groups and then comparing the groups, a task that can be carried out in several ways, as the next Clear It Up feature shows.

HOW DO YOU SEPARATE POVERTY AND INCOME INEQUALITY?

Poverty can change even when inequality does not move at all. Imagine a situation in which income for everyone in the population declines by 10%. Poverty would rise, since a greater share of the population would now fall below the poverty line. However, inequality would be the same, because everyone suffered the same proportional loss. Conversely, a general rise in income levels over time would keep inequality the same, but reduce poverty.

It is also possible for income inequality to change without affecting the poverty rate. Imagine a situation in which a large number of people who already have high incomes increase their incomes by even more. Inequality would rise as a result—but the number of people below the poverty line would remain unchanged.

Why did inequality of household income increase in the United States in recent decades? Indeed, a trend toward greater income inequality has occurred in many countries around the world, although the effect has been more powerful in the U.S. economy. Economists have focused their explanations for the increasing inequality on two factors that changed more or less continually from the 1970s into the 2000s. One set of explanations focuses on the changing shape of American households; the other focuses on greater inequality of wages, what some economists call “winner take all” **labor markets**. We will begin with how we measure inequality, and then consider the explanations for growing inequality in the United States.

MEASURING INCOME DISTRIBUTION BY QUINTILES

One common way of measuring income inequality is to rank all households by income, from lowest to highest, and then to divide all households into five groups with equal numbers of people, known as **quintiles**. This calculation allows for measuring the distribution of income among the five groups compared to the total. The first quintile is the lowest fifth or 20%, the second quintile is the next lowest, and so on. Income inequality can be measured by comparing what share of the total income is earned by each quintile.

U.S. income distribution by quintile appears in Table 7. In 2011, for example, the bottom quintile of the income distribution received 3.2% of income; the second quintile received 8.4%; the third quintile, 14.3%; the fourth quintile, 23.0%; and the top quintile, 51.14%. The final column of Table 7 shows what share of income went to households in the top 5% of the income distribution: 22.3% in 2011. Over time, from the late 1960s to the early 1980s, the top fifth of the income distribution typically received between about 43% to 44% of all income. The share of income that the top fifth received then begins to rise. According to the Census Bureau, much of this increase in the share of income going to the top fifth can be traced to an increase in the share of income going to the top 5%. The quintile measure shows how income inequality has increased in recent decades.

Year	Lowest Quintile	Second Quintile	Third Quintile	Fourth Quintile	Highest Quintile	Top 5%
1967	4.0	10.8	17.3	24.2	43.6	17.2
1970	4.1	10.8	17.4	24.5	43.3	16.6
1975	4.3	10.4	17.0	24.7	43.6	16.5
1980	4.2	10.2	16.8	24.7	44.1	16.5
1985	3.9	9.8	16.2	24.4	45.6	17.6
1990	3.8	9.6	15.9	24.0	46.6	18.5
1995	3.7	9.1	15.2	23.3	48.7	21.0
2000	3.6	8.9	14.8	23.0	49.8	22.1
2005	3.4	8.6	14.6	23.0	50.4	22.2
2010	3.3	8.5	14.6	23.4	50.3	21.3
2013	3.2	8.4	14.4	23.0	51	22.2

Table 7. Share of Aggregate Income Received by Each Fifth and Top 5% of Households, 1967–2013. (Source: U.S. Census Bureau, Table 2)

It can also be useful to divide the income distribution in ways other than quintiles; for example, into tenths or even into percentiles (that is, hundredths). A more detailed breakdown can provide additional insights. For example, the last column of Table 7 shows the income received by the top 5% percent of the income distribution. Between 1980 and 2013, the share of income going to the top 5% increased by 5.7 percentage points (from 16.5% in 1980 to 22.2% in 2013). From 1980 to 2013 the share of income going to the top quintile increased by 7.0 percentage points (from 44.1% in 1980 to 51% in 2013). Thus, the top 20% of householders (the fifth quintile) received over half (51%) of all the income in the United States in 2013.

LORENZ CURVE

The data on income inequality can be presented in various ways. For example, you could draw a bar

graph that showed the share of income going to each fifth of the income distribution. Figure 1 presents an alternative way of showing inequality data in what is called a **Lorenz curve**. The Lorenz curve shows the cumulative share of population on the horizontal axis and the cumulative percentage of total income received on the vertical axis.

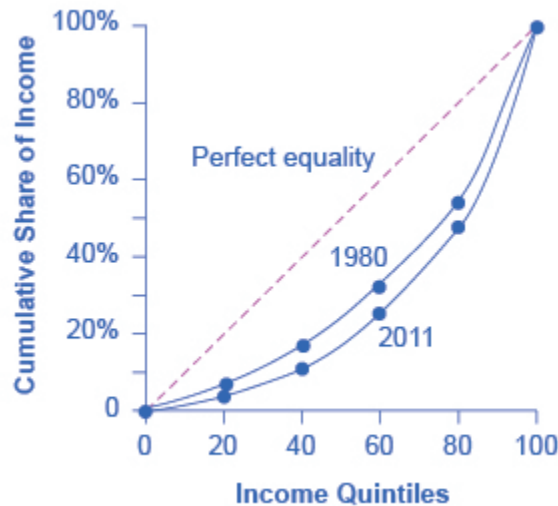


Figure 1. The Lorenz Curve. A Lorenz curve graphs the cumulative shares of income received by everyone up to a certain quintile. The income distribution in 1980 was closer to the perfect equality line than the income distribution in 2011—that is, the U.S. income distribution became more unequal over time.

Every Lorenz curve diagram begins with a line sloping up at a 45-degree angle, shown as a dashed line in Figure 1. The points along this line show what perfect equality of the income distribution looks like. It would mean, for example, that the bottom 20% of the income distribution receives 20% of the total income, the bottom 40% gets 40% of total income, and so on. The other lines reflect actual U.S. data on inequality for 1980 and 2011.

The trick in graphing a Lorenz curve is that you must change the shares of income for each specific quintile, which are shown in the first column of numbers in Table 8, into cumulative income, shown in the second column of numbers. For example, the bottom 40% of the cumulative income distribution will be the sum of the first and second quintiles; the bottom 60% of the cumulative income distribution will be the sum of the first, second, and third quintiles, and so on. The final entry in the cumulative income column needs to be 100%, because by definition, 100% of the population receives 100% of the income.

Income Category	Share of Income in 1980 (%)	Cumulative Share of Income in 1980 (%)	Share of Income in 2013 (%)	Cumulative Share of Income in 2013 (%)
First quintile	4.2	4.2	3.2	3.2
Second quintile	10.2	14.4	8.4	11.6
Third quintile	16.8	31.2	14.4	26.0
Fourth quintile	24.7	55.9	23.0	49.0
Fifth quintile	44.1	100.0	51.0	100.0

Table 8. Calculating the Lorenz Curve

In a Lorenz curve diagram, a more unequal distribution of income will loop farther down and away from the 45-degree line, while a more equal distribution of income will move the line closer to the 45-degree line. The greater inequality of the U.S. income distribution between 1980 and 2013 is illustrated in Figure 1 because the Lorenz curve for 2013 is farther from the 45-degree line than the Lorenz curve for 1980. The Lorenz curve is a useful way of presenting the quintile data that provides an image of all the quintile data at once. The next Clear It Up feature shows how income inequality differs in various countries compared to the United States.

HOW DOES ECONOMIC INEQUALITY VARY AROUND THE WORLD?

The U.S. economy has a relatively high degree of income inequality by global standards. As Table 9 shows, based on a variety of national surveys done for a selection of years in the last five years of the 2000s (with the exception of Germany, and adjusted to make the measures more comparable), the U.S. economy has greater inequality than Germany (along with most Western European countries). The region of the world with the highest level of income inequality is Latin America, illustrated in the numbers for Brazil and Mexico. The level of inequality in the United States is lower than in some of the low-income countries of the world, like China and Nigeria, or some middle-income countries like the Russian Federation. However, not all poor countries have highly unequal income distributions; India provides a counterexample.

Country	Survey Year	First Quintile	Second Quintile	Third Quintile	Fourth Quintile	Fifth Quintile
United States	2013	3.2%	8.4%	14.4%	23.0%	51.0%
Germany	2000	8.5%	13.7%	17.8%	23.1%	36.9%
Brazil	2009	2.9%	7.1%	12.4%	19.0%	58.6%
Mexico	2010	4.9%	8.8%	13.3%	20.2%	52.8%
China	2009	4.7%	9.7%	15.3%	23.2%	47.1%
India	2010	8.5%	12.1%	15.7%	20.8%	42.8%
Russia	2009	6.1%	10.4%	14.8%	21.3%	47.1%
Nigeria	2010	4.4%	8.3%	13.0%	20.3%	54.0%

Table 9. Income Distribution in Select Countries. (Source: U.S. data from U.S. Census Bureau Table 2. Other data from The World Bank Poverty and Inequality Data Base, <http://databank.worldbank.org/data/views/reports/tableview.aspx#>)

Visit this website to watch a video of wealth inequality across the world.



CAUSES OF GROWING INEQUALITY: THE CHANGING COMPOSITION OF AMERICAN HOUSEHOLDS

In 1970, 41% of married women were in the labor force, but by 2015, according to the Bureau of Labor Statistics, 56.7% of married women were in the labor force. One result of this trend is that more households have two earners. Moreover, it has become more common for one high earner to marry another high earner. A few decades ago, the common pattern featured a man with relatively high earnings, such as an executive or a doctor, marrying a woman who did not earn as much, like a secretary or a nurse. Often, the woman would leave paid employment, at least for a few years, to raise a family. However, now doctors are marrying doctors and executives are marrying executives, and mothers with high-powered careers are often returning to work while their children are quite young. This pattern of households with two high earners tends to increase the proportion of high-earning households.

According to data in the National Journal, even as two-earner couples have increased, so have single-parent households. Of all U.S. families, 13.1% were headed by single mothers; the poverty rate among single-parent households tends to be relatively high.

These changes in family structure, including the growth of single-parent families who tend to be at the lower end of the income distribution, and the growth of two-career high-earner couples near the top end of the income distribution, account for roughly half of the rise in income inequality across households in recent decades.

Visit this website to watch a video that illustrates the distribution of wealth in the United States.



CAUSES OF GROWING INEQUALITY: A SHIFT IN THE DISTRIBUTION OF WAGES

Another factor behind the rise in U.S. income inequality is that earnings have become less equal since the late 1970s. In particular, the earnings of high-skilled labor relative to low-skilled labor have

increased. Winner-take-all labor markets result from changes in technology, which have increased global demand for “stars,”—whether the best CEO, doctor, basketball player, or actor. This global demand pushes salaries far above productivity differences versus educational differences. One way to measure this change is to take the earnings of workers with at least a four-year college bachelor’s degree (including those who went on and completed an advanced degree) and divide them by the earnings of workers with only a high school degree. The result is that those in the 25–34 age bracket with college degrees earned about 1.67 times as much as high school graduates in 2010, up from 1.59 times in 1995, according to U.S. Census data. Winner-take-all labor market theory argues that the salary gap between the median and the top 1 percent is not due to educational differences.

Economists use the demand and supply model to reason through the most likely causes of this shift. According to the National Center for Education Statistics, in recent decades, the supply of U.S. workers with college degrees has increased substantially; for example, 840,000 four-year bachelor’s degrees were conferred on Americans in 1970; in 2009–2010, 1,602,480 such degrees were conferred—an increase of about 90%. In Figure 2, this shift in supply to the right, from S_0 to S_1 , should result in a lower equilibrium wage for high-skilled labor. Thus, the increase in the price of high-skilled labor must be explained by a greater demand, like the movement from D_0 to D_1 . Evidently, combining both the increase in supply and in demand has resulted in a shift from E_0 to E_1 , and a resulting higher wage.

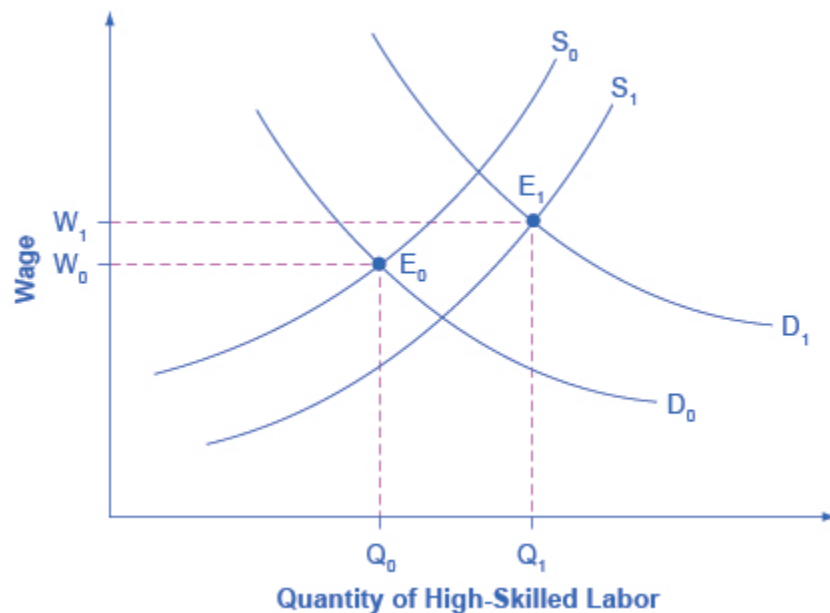


Figure 2. Why Would Wages Rise for High-Skilled Labor? The proportion of workers attending college has increased in recent decades, so the supply curve for high-skilled labor has shifted to the right, from S_0 to S_1 . If the demand for high-skilled labor had remained at D_0 , then this shift in supply would have led to lower wages for high-skilled labor. However, the wages for high-skilled labor, especially if there is a large global demand, have increased even with the shift in supply to the right. The explanation must lie in a shift to the right in demand for high-skilled labor, from D_0 to D_1 . The figure shows how a combination of the shift in supply, from S_0 to S_1 , and the shift in demand, from D_0 to D_1 , led to both an increase in the quantity of high-skilled labor hired and also to a rise in the wage for such labor, from W_0 to W_1 .

What factors would cause the demand for high-skilled labor to rise? The most plausible explanation is that while the explosion in new information and communications technologies over the last several decades has helped many workers to become more productive, the benefits have been especially great for high-skilled workers like top business managers, consultants, and design professionals. The new technologies have also helped to encourage **globalization**, the remarkable increase in international trade over the last few decades, by making it more possible to learn about and coordinate economic interactions all around the world. In turn, the rising impact of foreign trade in the U.S. economy has opened up greater opportunities for high-skilled workers to sell their services around the world. And lower-skilled workers have to compete with a larger supply of similarly skilled workers around the globe.

The market for high-skilled labor can be viewed as a race between forces of supply and demand. Additional education and on-the-job training will tend to increase the supply of high-skilled labor and to hold down its relative wage. Conversely, new technology and other economic trends like globalization tend to increase the demand for high-skilled labor and push up its relative wage. The greater inequality of wages can be viewed as a sign that demand for skilled labor is increasing faster than supply. On the other hand, if the supply of lower skilled workers exceeds the demand, then average wages in the lower quintiles of the income distribution will decrease. The combination of forces in the high-skilled and low-skilled labor markets leads to increased income disparity.

KEY CONCEPTS AND SUMMARY

Measuring inequality involves making comparisons across the entire distribution of income, not just the poor. One way of doing this is to divide the population into groups, like quintiles, and then calculate what share of income is received by each group. An alternative approach is to draw Lorenz curves, which compare the cumulative income actually received to a perfectly equal distribution of income. Income inequality in the United States increased substantially from the late 1970s and early 1980s into the 2000s. The two most common explanations cited by economists are changes in the structure of households that have led to more two-earner couples and single-parent families, and the effect of new information and communications technology on wages.

SELF-CHECK QUESTIONS

1. A group of 10 people have the following annual incomes: \$24,000, \$18,000, \$50,000, \$100,000, \$12,000, \$36,000, \$80,000, \$10,000, \$24,000, \$16,000. Calculate the share of total income received by each quintile of this income distribution. Do the top and bottom quintiles in this distribution have a greater or larger share of total income than the top and bottom quintiles of the U.S. income distribution?
2. Table 10 shows the share of income going to each quintile of the income distribution for the United Kingdom in 1979 and 1991. Use this data to calculate what the points on a Lorenz curve would be, and sketch the Lorenz curve. How did inequality in the United Kingdom shift over this time period? How can you see the patterns in the quintiles in the Lorenz curves?

Share of Income	1979	1991
Top quintile	39.7%	42.9%
Fourth quintile	24.8%	22.7%
Middle quintile	17.0%	16.3%
Second quintile	11.5%	11.5%
Bottom quintile	7.0%	6.6%

Table 10. Income Distribution in the United Kingdom, 1979 and 1991

- Using two demand and supply diagrams, one for the low-wage labor market and one for the high-wage labor market, explain how information technology can increase income inequality if it is a complement to high-income workers like salespeople and managers, but a substitute for low-income workers like file clerks and telephone receptionists.
- Using two demand and supply diagrams, one for the low-wage labor market and one for the high-wage labor market, explain how a program that increased educational levels for a substantial number of low-skill workers could reduce income inequality.

REVIEW QUESTIONS

- Who is included in the top income quintile?
- What is measured on the two axes of a Lorenz curve?
- If a country had perfect income equality what would the Lorenz curve look like?
- How has the inequality of income changed in the U.S. economy since the late 1970s?
- What are some reasons why a certain degree of inequality of income would be expected in a market economy?
- What are the main reasons economists give for the increase in inequality of incomes?

CRITICAL THINKING QUESTIONS

- Explain how a country may experience greater equality in the distribution of income, yet still experience high rates of poverty *Hint*: Look at the Clear It Up “How is poverty measured in low-income countries?” and compare to Table 7.
- The demand for skilled workers in the United States has been increasing. To increase the supply of skilled workers, many argue that immigration reform to allow more skilled labor into the United States is needed. Explain whether you agree or disagree.
- Explain a situation using the supply and demand for skilled labor in which the increased number of college graduates leads to depressed wages. Given the rising cost of going to college, explain why a college education will or will not increase income inequality.

PROBLEMS

A group of 10 people have the following annual incomes: \$55,000, \$30,000, \$15,000, \$20,000, \$35,000, \$80,000, \$40,000, \$45,000, \$30,000, \$50,000. Calculate the share of total income received by each quintile of this income distribution. Do the top and bottom quintiles in this distribution have a greater or larger share of total income than the top and bottom quintiles of the U.S. income distribution for 2005?

REFERENCES

Frank, Robert H., and Philip J. Cook. *The Winner-Take-All Society*. New York: Martin Kessler Books at The Free Press, 1995.

Institute of Education Sciences: National Center for Education Statistics. "Fast Facts: Degrees Conferred by Sex and Race." <http://nces.ed.gov/fastfacts/display.asp?id=72>.

Nhan, Doris. "Census: More in U.S. Report Nontraditional Households." *National Journal*. Last modified May 1, 2012. <http://www.nationaljournal.com/thenextamerica/demographics/census-more-in-u-s-report-nontraditional-households-20120430>.

U.S. Bureau of Labor Statistics: BLS Reports. "Report 1040: Women in the Labor Force: A Databook." Last modified March 26, 2013. <http://www.bls.gov/cps/wlf-databook-2012.pdf>.

U.S. Department of Commerce: United States Census Bureau. "Income: Table H-2. Share of Aggregate Income Received by Each Fifth and Top 5 Percent of Households." <http://www.census.gov/hhes/www/income/data/historical/household/>.

United States Census Bureau. 2014. "2013 Highlights." Accessed April 13, 2015. <http://www.census.gov/hhes/www/poverty/about/overview/>.

United States Census Bureau. 2014. "Historical Income Tables: Households: Table H-2 Share of Aggregate Income Received by Each Fifth and Top 5% of Income. All Races." Accessed April 13, 2015. <http://www.census.gov/hhes/www/income/data/historical/household/>.

GLOSSARY

Lorenz curve a graph that compares the cumulative income actually received to a perfectly equal distribution of income; it shows the share of population on the horizontal axis and the cumulative percentage of total income received on the vertical axis

quintile dividing a group into fifths, a method often used to look at distribution of income

SOLUTIONS

Answers to Self-Check Questions

1. A useful first step is to rank the households by income, from lowest to highest. Then, since there are 10 households total, the bottom quintile will be the bottom two households, the second quintile will be the

third and fourth households, and so on up to the top quintile. The quintiles and percentage of total income for the data provided are shown in the following table. Comparing this distribution to the U.S. income distribution for 2005, the top quintile in the example has a smaller share of total income than in the U.S. distribution and the bottom quintile has a larger share. This pattern usually means that the income distribution in the example is more equal than the U.S. distribution.

Income	Quintile	% of Total Income
\$10,000	Total first quintile income: \$22,000	6.0%
\$12,000		
\$16,000	Total second quintile income: \$34,000	9.2%
\$18,000		
\$24,000	Total third quintile income: \$48,000	13.0%
\$24,000		
\$36,000	Total fourth quintile income: \$86,000	23.2%
\$50,000		
\$80,000	Total top quintile income: \$180,000	48.6%
\$100,000		
\$370,000	Total Income	

Table 11.

- Just from glancing at the quintile information, it is fairly obvious that income inequality increased in the United Kingdom over this time: The top quintile is getting a lot more, and the lowest quintile is getting a bit less. Converting this information into a Lorenz curve, however, is a little trickier, because the Lorenz curve graphs the cumulative distribution, not the amount received by individual quintiles. Thus, as explained in the text, you have to add up the individual quintile data to convert the data to this form. The following table shows the actual calculations for the share of income in 1979 versus 1991. The figure following the table shows the perfect equality line and the Lorenz curves for 1979 and 1991. As shown, the income distribution in 1979 was closer to the perfect equality line than the income distribution in 1991—that is, the United Kingdom income distribution became more unequal over time.

Share of income received	1979	1991
Bottom 20%	7.0%	6.6%
Bottom 40%	18.5%	18.1%
Bottom 60%	35.5%	34.4%
Bottom 80%	60.3%	57.1%
All 100%	100.0%	100.0%

Table 12.

- In the market for low-wage labor, information technology shifts the demand for low-wage labor to the left. One reason is that technology can often substitute for low-wage labor in certain kinds of telephone or bookkeeping jobs. In addition, information technology makes it easier for companies to manage connections with low-wage workers in other countries, thus reducing the demand for low-wage workers in the United States. In the market for high-wage labor, information technology shifts the demand for high-wage labor to the right. By using the new information and communications technologies, high-wage labor can become more productive and can oversee more tasks than before. The following figure illustrates these

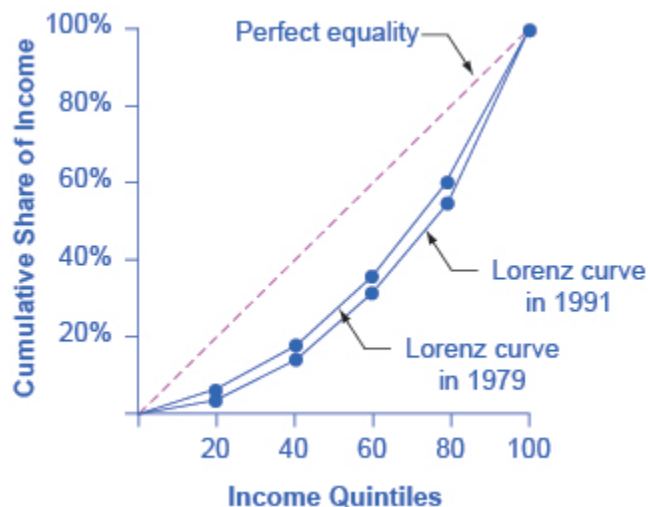


Figure 3.

two labor markets. The combination of lower wages for low-wage labor and higher wages for high-wage labor means greater inequality.

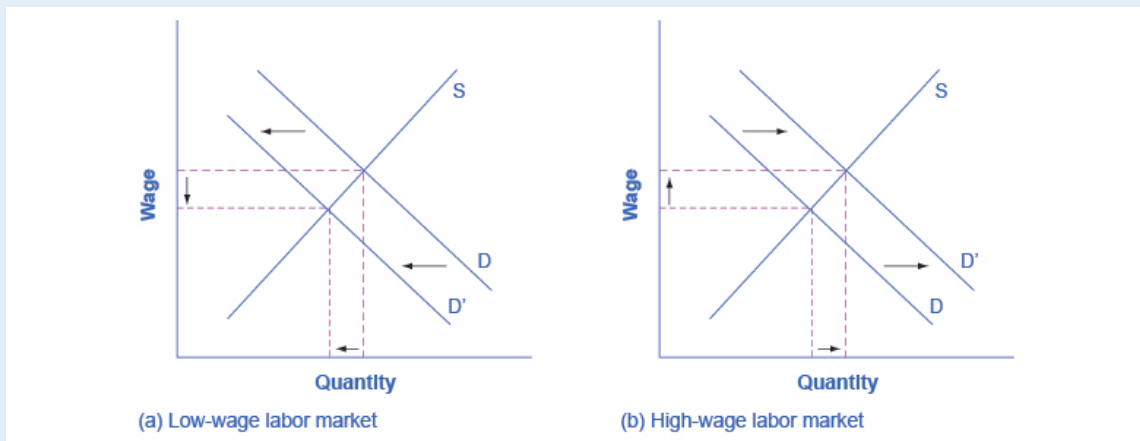


Figure 4.

4. In the market for low-wage labor, a skills program will shift supply to the left, which will tend to drive up wages for the remaining low-skill workers. In the market for high-wage labor, a skills program will shift supply to the right (because after the training program there are now more high-skilled workers at every wage), which will tend to drive down wages for high-skill workers. The combination of these two programs will result in a lesser degree of inequality. The following figure illustrates these two labor markets. In the market for high-wage labor, a skills program will shift supply to the right, which will tend to drive down wages for high-skill workers.

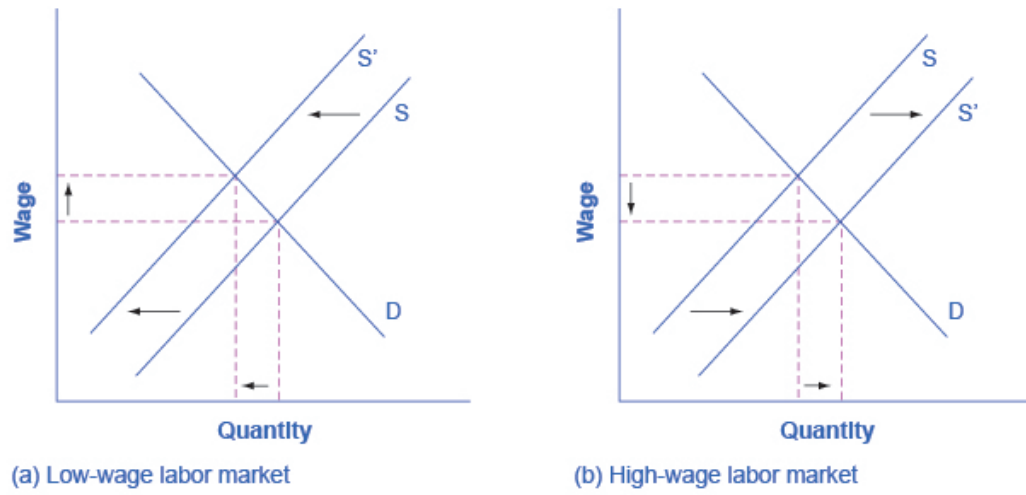


Figure 5.

20.5 GOVERNMENT POLICIES TO REDUCE INCOME INEQUALITY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the arguments for and against government intervention in a market economy
- Identify beneficial ways to reduce the economic inequality in a society
- Show the tradeoff between incentives and income equality

No society should expect or desire complete equality of income at a given point in time, for a number of reasons. First, most workers receive relatively low earnings in their first few jobs, higher earnings as they reach middle age, and then lower earnings after retirement. Thus, a society with people of varying ages will have a certain amount of income inequality. Second, people's preferences and desires differ. Some are willing to work long hours to have income for large houses, fast cars and computers, luxury vacations, and the ability to support children and grandchildren.

These factors all imply that a snapshot of inequality in a given year does not provide an accurate picture of how people's incomes rise and fall over time. Even if some degree of economic inequality is expected at any point in time, how much inequality should there be? There is also the difference between income and wealth, as the following Clear It Up feature explains.

HOW DO YOU MEASURE WEALTH VERSUS INCOME INEQUALITY?

Income is a flow of money received, often measured on a monthly or an annual basis; **wealth** is the sum of the value of all assets, including money in bank accounts, financial investments, a pension fund, and the value of a home. In calculating wealth all debts must be subtracted, such as debt owed on a home mortgage and on credit cards. A retired person, for example, may have relatively little income in a given year, other than a pension or Social Security. However, if that person has saved and invested over time, the person's accumulated wealth can be quite substantial.

In the United States, the wealth distribution is more unequal than the income distribution, because differences in income can accumulate over time to make even larger differences in wealth. However, the degree of inequality in the wealth distribution can be measured with the same tools we use to measure the inequality in the income distribution, like quintile measurements. Data on wealth are collected once every three years in the Survey of Consumer Finance.

Even if they cannot answer the question of how much inequality is too much, economists can still play an important role in spelling out policy options and tradeoffs. If a society decides to reduce the level of economic inequality, it has three main sets of tools: redistribution from those with high incomes to

those with low incomes; trying to assure that a ladder of opportunity is widely available; and a tax on inheritance.

REDISTRIBUTION

Redistribution means taking income from those with higher incomes and providing income to those with lower incomes. Earlier in this chapter, we considered some of the key government policies that provide support for the poor: the welfare program TANF, the earned income tax credit, SNAP, and Medicaid. If a reduction in inequality is desired, these programs could receive additional funding.

The programs are paid for through the federal income tax, which is a **progressive tax system** designed in such a way that the rich pay a higher percent in income taxes than the poor. Data from household income tax returns in 2009 shows that the top 1% of households had an average income of \$1,219,700 per year in pre-tax income and paid an average federal tax rate of 28.9%. The **effective income tax**, which is total taxes paid divided by total income (all sources of income such as wages, profits, interest, rental income, and government transfers such as veterans' benefits), was much lower. The effective tax paid by the top 1% of householders was 20.4%, while the bottom two quintiles actually paid negative effective income taxes, because of provisions like the earned income tax credit. News stories occasionally report on a high-income person who has managed to pay very little in taxes, but while such individual cases exist, according to the Congressional Budget Office, the typical pattern is that people with higher incomes pay a higher average share of their income in federal income taxes.

Of course, the fact that some degree of redistribution occurs now through the federal income tax and government antipoverty programs does not settle the questions of how much redistribution is appropriate, and whether more redistribution should occur.

THE LADDER OF OPPORTUNITY

Economic inequality is perhaps most troubling when it is not the result of effort or talent, but instead is determined by the circumstances under which a child grows up. One child attends a well-run grade school and high school and heads on to college, while parents help out by supporting education and other interests, paying for college, a first car, and a first house, and offering work connections that lead to internships and jobs. Another child attends a poorly run grade school, barely makes it through a low-quality high school, does not go to college, and lacks family and peer support. These two children may be similar in their underlying talents and in the effort they put forth, but their economic outcomes are likely to be quite different.

Public policy can attempt to build a ladder of opportunities so that, even though all children will never come from identical families and attend identical schools, each child has a reasonable opportunity to attain an economic niche in society based on their interests, desires, talents, and efforts. Some of those initiatives include those shown in Table 13.

Children	College Level	Adults
<ul style="list-style-type: none"> • Improved day care • Enrichment programs for preschoolers • Improved public schools • After school and community activities • Internships and apprenticeships 	<ul style="list-style-type: none"> • Widespread loans and grants for those in financial need • Public support for a range of institutions from two-year community colleges to large research universities 	<ul style="list-style-type: none"> • Opportunities for retraining and acquiring new skills • Prohibiting discrimination in job markets and housing on the basis of race, gender, age, and disability
	–	–
	–	–
	–	–

Table 13. Public Policy Initiatives

The United States has often been called a land of opportunity. Although the general idea of a ladder of opportunity for all citizens continues to exert a powerful attraction, specifics are often quite controversial. Society can experiment with a wide variety of proposals for building a ladder of opportunity, especially for those who otherwise seem likely to start their lives in a disadvantaged position. Such policy experiments need to be carried out in a spirit of open-mindedness, because some will succeed while others will not show positive results or will cost too much to enact on a widespread basis.

INHERITANCE TAXES

There is always a debate about inheritance taxes. It goes like this: On the one hand, why should people who have worked hard all their lives and saved up a substantial nest egg not be able to give their money and possessions to their children and grandchildren? In particular, it would seem un-American if children were unable to inherit a family business or a family home. On the other hand, many Americans are far more comfortable with inequality resulting from high-income people who earned their money by starting innovative new companies than they are with inequality resulting from high-income people who have inherited money from rich parents.

The United States does have an **estate tax**—that is, a tax imposed on the value of an inheritance—which suggests a willingness to limit how much wealth can be passed on as an inheritance. However, according to the Center on Budget and Policy Priorities, in 2015 the estate tax applied only to those leaving inheritances of more than \$5.43 million and thus applies to only a tiny percentage of those with high levels of wealth.

THE TRADEOFF BETWEEN INCENTIVES AND INCOME EQUALITY

Government policies to reduce poverty or to encourage economic equality, if carried to extremes, can injure incentives for economic output. The poverty trap, for example, defines a situation where guaranteeing a certain level of income can eliminate or reduce the incentive to work. An extremely high degree of redistribution, with very high taxes on the rich, would be likely to discourage work and entrepreneurship. Thus, it is common to draw the tradeoff between economic output and equality, as shown in Figure 1 (a). In this formulation, if society wishes a high level of economic output, like point A, it must also accept a high degree of inequality. Conversely, if society wants a high level of equality,

like point B, it must accept a lower level of economic output because of reduced incentives for production.

This view of the tradeoff between economic output and equality may be too pessimistic, and Figure 1 (b) presents an alternate vision. Here, the tradeoff between economic output and equality first slopes up, in the vicinity of choice C, suggesting that certain programs might increase both output and economic equality. For example, the policy of providing free public education has an element of redistribution, since the value of the public schooling received by children of low-income families is clearly higher than what low-income families pay in taxes. A well-educated population, however, is also an enormously powerful factor in providing the skilled workers of tomorrow and helping the economy to grow and expand. In this case, equality and economic growth may complement each other.

Moreover, policies to diminish inequality and soften the hardship of poverty may sustain political support for a market economy. After all, if society does not make some effort toward reducing inequality and poverty, the alternative might be that people would rebel against market forces. Citizens might seek economic security by demanding that their legislators pass laws forbidding employers from ever laying off workers or reducing wages, or laws that would impose price floors and price ceilings and shut off international trade. From this viewpoint, policies to reduce inequality may help economic output by building social support for allowing markets to operate.

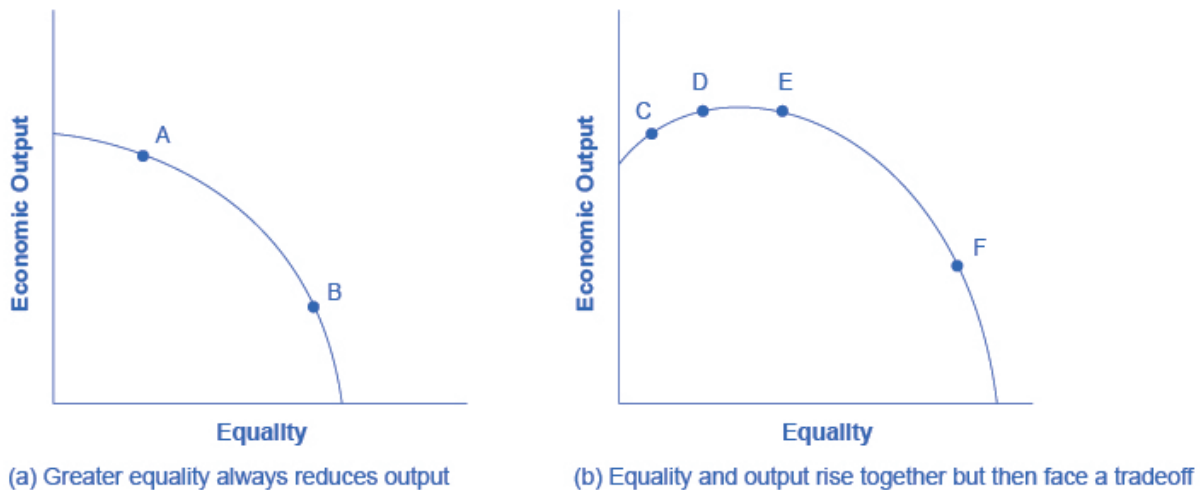


Figure 1. The Tradeoff between Incentives and Economic Equality. (a) Society faces a trade-off where any attempt to move toward greater equality, like moving from choice A to B, involves a reduction in economic output. (b) Situations can arise like point C, where it is possible both to increase equality and also to increase economic output, to a choice like D. It may also be possible to increase equality with little impact on economic output, like the movement from choice D to E. However, at some point, too aggressive a push for equality will tend to reduce economic output, as in the shift from E to F.

The tradeoff in Figure 1 (b) then flattens out in the area between points D and E, which reflects the pattern that a number of countries that provide similar levels of income to their citizens—the United States, Canada, the nations of the European Union, Japan, Australia—have different levels of inequality. The pattern suggests that countries in this range could choose a greater or a lesser degree of inequality without much impact on economic output. Only if these countries push for a much higher level of equality, like at point F, will they experience the diminished incentives that lead to lower levels of economic output. In this view, while a danger always exists that an agenda to reduce poverty or

inequality can be poorly designed or pushed too far, it is also possible to discover and design policies that improve equality and do not injure incentives for economic output by very much—or even improve such incentives.

OCCUPY WALL STREET

The Occupy movement took on a life of its own over the last few months of 2011, bringing to light issues faced by many people on the lower end of the income distribution. The contents of this chapter indicate that there is a significant amount of income inequality in the United States. The question is: What should be done about it?

The Great Recession of 2008–2009 caused unemployment to rise and incomes to fall. Many people attribute the recession to mismanagement of the financial system by bankers and financial managers—those in the 1% of the income distribution—but those in lower quintiles bore the greater burden of the recession through unemployment. This seemed to present the picture of inequality in a different light: the group that seemed responsible for the recession was not the group that seemed to bear the burden of the decline in output. A burden shared can bring a society closer together; a burden pushed off onto others can polarize it.

On one level, the problem with trying to reduce income inequality comes down to whether you still believe in the American Dream. If you believe that one day you will have your American Dream—a large income, large house, happy family, or whatever else you would like to have in life—then you do not necessarily want to prevent anyone else from living out their dream. You certainly would not want to run the risk that someone would want to take part of your dream away from you. So there is some reluctance to engage in a redistributive policy to reduce inequality.

However, when those for whom the likelihood of living the American Dream is very small are considered, there are sound arguments in favor of trying to create greater balance. As the text indicated, a little more income equality, gained through long-term programs like increased education and job training, can increase overall economic output. Then everyone is made better off. And the 1% will not seem like such a small group any more.

KEY CONCEPTS AND SUMMARY

Policies that can affect the level of economic inequality include redistribution between rich and poor, making it easier for people to climb the ladder of opportunity; and estate taxes, which are taxes on inheritances. Pushing too aggressively for economic equality can run the risk of decreasing economic incentives. However, a moderate push for economic equality can increase economic output, both through methods like improved education and by building a base of political support for market forces.

SELF-CHECK QUESTIONS

1. Here is one hypothesis: A well-funded social safety net can increase economic equality but will reduce economic output. Explain why this might be so, and sketch a production possibility curve that shows this tradeoff.
2. Here is a second hypothesis: A well-funded social safety net may lead to less regulation of the market economy. Explain why this might be so, and sketch a production possibility curve that shows this tradeoff.
3. Which set of policies is more likely to cause a tradeoff between economic output and equality: policies of redistribution or policies aimed at the ladder of opportunity? Explain how the production possibility frontier tradeoff between economic equality and output might look in each case.

4. Why is there reluctance on the part of some in the United States to redistribute income so that greater equality can be achieved?

REVIEW QUESTIONS

1. Identify some public policies that can reduce the level of economic inequality.
2. Describe how a push for economic equality might reduce incentives to work and produce output. Then describe how a push for economic inequality might not have such effects.

CRITICAL THINKING QUESTIONS

1. What do you think is more important to focus on when considering inequality: income inequality or wealth inequality?
2. To reduce income inequality, should the marginal tax rates on the top 1% be increased?
3. Redistribution of income occurs through the federal income tax and government antipoverty programs. Explain whether or not this level of redistribution is appropriate and whether more redistribution should occur.
4. How does a society or a country make the decision about the tradeoff between equality and economic output? *Hint:* Think about the political system.
5. Explain what the long- and short-term consequences are of not promoting equality or working to reduce poverty.

REFERENCES

Board of Governors of the Federal Reserve System. "Research Resources: Survey of Consumer Finances." Last modified December 13, 2013. <http://www.federalreserve.gov/econresdata/scf/scfindex.htm>.

Congressional Budget Office. "The Distribution of Household Income and Federal Taxes, 2008 and 2009." Last modified July 10, 2012. <http://www.cbo.gov/publication/43373>.

Huang, Chye-Ching, and Nathaniel Frentz. "Myths and Realities About the Estate Tax." *Center on Budget and Policy Priorities*. Last modified August 29, 2013. <http://www.cbpp.org/files/estatetaxmyths.pdf>.

GLOSSARY

effective income tax percentage of total taxes paid divided by total income

estate tax a tax imposed on the value of an inheritance

income a flow of money received, often measured on a monthly or an annual basis

progressive tax system a tax system in which the rich pay a higher percentage of their income in taxes, rather than a higher absolute amount

redistribution taking income from those with higher incomes and providing income to those with lower incomes

wealth the sum of the value of all assets, including money in bank accounts, financial investments, a pension fund, and the value of a home

SOLUTIONS

Answers to Self-Check Questions

1. A very strong push for economic equality might include extremely high taxes on high-wage earners to pay for extremely large government social payments for the poor. Such a policy could limit incentives for the high-wage workers, lock the poor into a poverty trap, and thus reduce output. The PPF in this case will have the standard appearance: it will be downward sloping.
2. For the second hypothesis, a well-funded social safety net might make people feel that even if their company goes bankrupt or they need to change jobs or industries, they will have some degree of protection. As a result, people may be more willing to allow markets to work without interference, and not to lobby as hard for rules that would prevent layoffs, set price controls, or block foreign trade. In this case, safety net programs that increase equality could also allow the market to work more freely in a way that could increase output. In this case, at least some portion of the PPF between equality and economic output would slope up.
3. Pure redistribution is more likely to cause a sharp tradeoff between economic output and equality than policies aimed at the ladder of opportunity. A production possibility frontier showing a strict tradeoff between economic output and equality will be downward sloping. A PPF showing that it is possible to increase equality, at least to some extent, while either increasing output or at least not diminishing it would have a PPF that first rises, perhaps has a flat area, and then falls.
4. Many view the redistribution of income to achieve greater equality as taking away from the rich to pay the poor, or as a “zero sum” game. By taking taxes from one group of people and redistributing them to another, the tax system is robbing some of the American Dream.

CHAPTER 21. ISSUES IN LABOR MARKETS: UNIONS, DISCRIMINATION, IMMIGRATION

INTRODUCTION TO ISSUES IN LABOR MARKETS: UNIONS, DISCRIMINATION, IMMIGRATION



Figure 1. Arguing for Collective Bargaining. In 2011, thousands of people in Wisconsin protested against a bill that would eliminate the right to collective bargaining over everything except wages. (Credit: modification of work by Fibonacci Blue/Flickr Creative Commons)

COLLECTIVE BARGAINING IN WISCONSIN

In 2011, thousands of people crowded into the Wisconsin State Capitol rotunda carrying placards reading “Kill the Bill.” What were they protesting? The newly elected Wisconsin governor, Scott Walker, supported a bill proposed by Republican state legislators that would have effectively eliminated most collective bargaining rights of public sector union employees. Collective bargaining laws require employers to sit down and negotiate with the representative union of their employees. The governor argued that the state needed to close a multi-billion-dollar deficit, so legislators proposed a Budget Repair Act that would eliminate collective bargaining over everything but wages. The bill passed and was signed into law after a significant level of drama that saw Democratic legislators leaving the state so that there would not be enough legislators in house to continue the debate or bring the bill to a vote. The law proved so unpopular that Governor Walker faced a recall vote in 2012. The recall attempt was defeated, but the law has been subjected to numerous court reviews. The discussion about the role of collective bargaining is not over.

Why was a bill like this proposed? Are collective bargaining rights necessary for public sector employees? How would an economist respond to such a bill? This chapter lays out the changing role of unions in U.S. labor markets.

CHAPTER OBJECTIVES

Introduction to Issues in Labor Markets: Unions, Discrimination, Immigration

In this chapter, you will learn about:

- Labor Unions
- Employment Discrimination
- Immigration

When a job applicant is bargaining with an employer for a position, the applicant is often at a disadvantage—needing the job more than the employer needs that particular applicant. John Bates Clark (1847–1938), often named as the first great American economist, wrote in 1907: “In the making of the wages contract the individual laborer is always at a disadvantage. He has something which he is obliged to sell and which his employer is not obliged to take, since he [that is, the employer] can reject single men with impunity.”

To give workers more power, the U.S. government has passed, in response to years of labor protests, a number of laws to create a more equal balance of power between workers and employers. These laws include some of the following:

- Setting minimum hourly wages
- Setting maximum hours of work (at least before employers pay overtime rates)
- Prohibiting child labor
- Regulating health and safety conditions in the workplace
- Preventing discrimination on the basis of race, ethnicity, gender, sexual orientation, and age
- Requiring employers to provide family leave
- Requiring employers to give advance notice of layoffs
- Covering workers with unemployment insurance
- Setting a limit on the number of immigrant workers from other countries

Table 1 lists some prominent U.S. workplace protection laws. Many of the laws listed in the table were only the start of labor market regulations in these areas and have been followed, over time, by other related laws, regulations, and court rulings.

Law	Protection
National Labor-Management Relations Act of 1935 (the “Wagner Act”)	Establishes procedures for establishing a union that firms are obligated to follow; sets up the National Labor Relations Board for deciding disputes
Social Security Act of 1935	Under Title III, establishes a state-run system of unemployment insurance, in which workers pay into a state fund when they are employed and received benefits for a time when they are unemployed
Fair Labor Standards Act of 1938	Establishes the minimum wage, limits on child labor, and rules requiring payment of overtime pay for those in jobs that are paid by the hour and exceed 40 hours per week
Taft-Hartley Act of 1947	Allows states to decide whether all workers at a firm can be required to join a union as a condition of employment; in the case of a disruptive union strike, permits the president to declare a “cooling-off period” during which workers have to return to work
Civil Rights Act of 1964	Title VII of the Act prohibits discrimination in employment on the basis of race, gender, national origin, religion, or sexual orientation
Occupational Health and Safety Act of 1970	Creates the Occupational Safety and Health Administration (OSHA), which protects workers from physical harm in the workplace
Employee Retirement and Income Security Act of 1974	Regulates employee pension rules and benefits
Pregnancy Discrimination Act of 1978	Prohibits discrimination against women in the workplace who are planning to get pregnant or who are returning to work after pregnancy
Immigration Reform and Control Act of 1986	Prohibits hiring of illegal immigrants; requires employers to ask for proof of citizenship; protects rights of legal immigrants
Worker Adjustment and Retraining Notification Act of 1988	Requires employers with more than 100 employees to provide written notice 60 days before plant closings or large layoffs
Americans with Disabilities Act of 1990	Prohibits discrimination against those with disabilities and requires reasonable accommodations for them on the job
Family and Medical Leave Act of 1993	Allows employees to take up to 12 weeks of unpaid leave per year for family reasons, including birth or family illness
Pension Protection Act of 2006	Penalizes firms for underfunding their pension plans and gives employees more information about their pension accounts
Lilly Ledbetter Fair Pay Act of 2009	Restores protection for pay discrimination claims on the basis of sex, race, national origin, age, religion, or disability

Table 1. Prominent U.S. Workplace Protection Laws

This chapter covers three issues in the labor markets: labor unions, discrimination against women or minority groups, and immigration and U.S. labor market issues.

21.1 UNIONS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the concept of labor unions, including membership levels and wages
- Evaluate arguments for and against labor unions
- Analyze reasons for the decline in U.S. union membership

A **labor union** is an organization of workers that negotiates with employers over wages and working conditions. A labor union seeks to change the balance of power between employers and workers by requiring employers to deal with workers collectively, rather than as individuals. Thus, negotiations between unions and firms are sometimes called **collective bargaining**.

The subject of labor unions can be controversial. Supporters of labor unions view them as the workers' primary line of defense against efforts by profit-seeking firms to hold down wages and benefits. Critics of labor unions view them as having a tendency to grab as much as they can in the short term, even if it means injuring workers in the long run by driving firms into bankruptcy or by blocking the new technologies and production methods that lead to economic growth. We will start with some facts about union membership in the United States.

FACTS ABOUT UNION MEMBERSHIP AND PAY

According to the U.S. Bureau of Labor and Statistics, about 11.1% of all U.S. workers belong to unions. Following are some of the facts provided by the bureau for 2014:

- 12.0% of U.S. male workers belong to unions; 10.5% of female workers do
- 11.1% of white workers, 13.4 % of black workers, and 9.8 % of Hispanic workers belong to unions
- 12.5% of full-time workers and 6.0% of part-time workers are union members
- 4.2% of workers ages 16–24 belong to unions, as do 14% of workers ages 45-54
- Occupations in which relatively high percentages of workers belong to unions are the federal government (26.9% belong to a union), state government (31.3%), local government (41.7%); transportation and utilities (20.6%); natural resources, construction, and maintenance (16.3%); and production, transportation, and material moving (14.7%)

- Occupations that have relatively low percentages of unionized workers are agricultural workers (1.4%), financial services (1.1%), professional and business services (2.4%), leisure and hospitality (2.7%), and wholesale and retail trade (4.7%)

In summary, the percentage of workers belonging to a union is higher for men than women; higher for blacks than for whites or Hispanics; higher for the 45–64 age range; and higher among workers in government and manufacturing than workers in agriculture or service-oriented jobs. Table 2 lists the largest U.S. labor unions and their membership.

Union	Membership
National Education Association (NEA)	3.2 million
Service Employees International Union (SEIU)	2.1 million
American Federation of Teachers (AFT)	1.5 million
International Brotherhood of Teamsters (IBT)	1.4 million
The American Federation of State, County, and Municipal Workers (AFSCME)	1.3 million
United Food and Commercial Workers International Union	1.3 million
United Steelworkers	1.2 million
International Union, United Automobile, Aerospace and Agricultural Implement Workers of America (UAW)	990,000
International Association of Machinists and Aerospace Workers	720,000
International Brotherhood of Electrical Workers (IBEW)	675,000

Table 2. The Largest American Unions in 2013. (Source: U.S. Department of Labor, Bureau of Labor Statistics)

In terms of pay, benefits, and hiring, U.S. unions offer a good news/bad news story. The good news for unions and their members is that their members earn about 20% more than nonunion workers, even after adjusting for factors such as years of work experience and education level. The bad news for unions is that the share of U.S. workers who belong to a labor union has been steadily declining for 50 years, as shown in Figure 1. About one-quarter of all U.S. workers belonged to a union in the mid-1950s, but only 11.1% of U.S. workers are union members today. If you leave out workers employed by the government (which includes teachers in public schools), only 6.6% of the workers employed by private firms now work for a union.

The following section analyzes the higher pay union workers receive compared the pay rates for nonunion workers. The following section analyzes declining union membership levels. An overview of these two issues will allow us to discuss many aspects of how unions work.

HIGHER WAGES FOR UNION WORKERS

Why might union workers receive higher pay? What are the limits on how much higher pay they can receive? To analyze these questions, let's consider a situation where all firms in an industry must negotiate with a single union, and no **firm** is allowed to hire nonunion labor. If no labor union existed in this market, then equilibrium (E) in the labor market would occur at the intersection of the demand for labor (D) and the supply of labor (S) in Figure 2. The union can, however, threaten that, unless firms agree to the wages they demand, the workers will strike. As a result, the labor union manages to achieve, through negotiations with the firms, a union wage of W_u for its members, above what the equilibrium wage would otherwise have been.

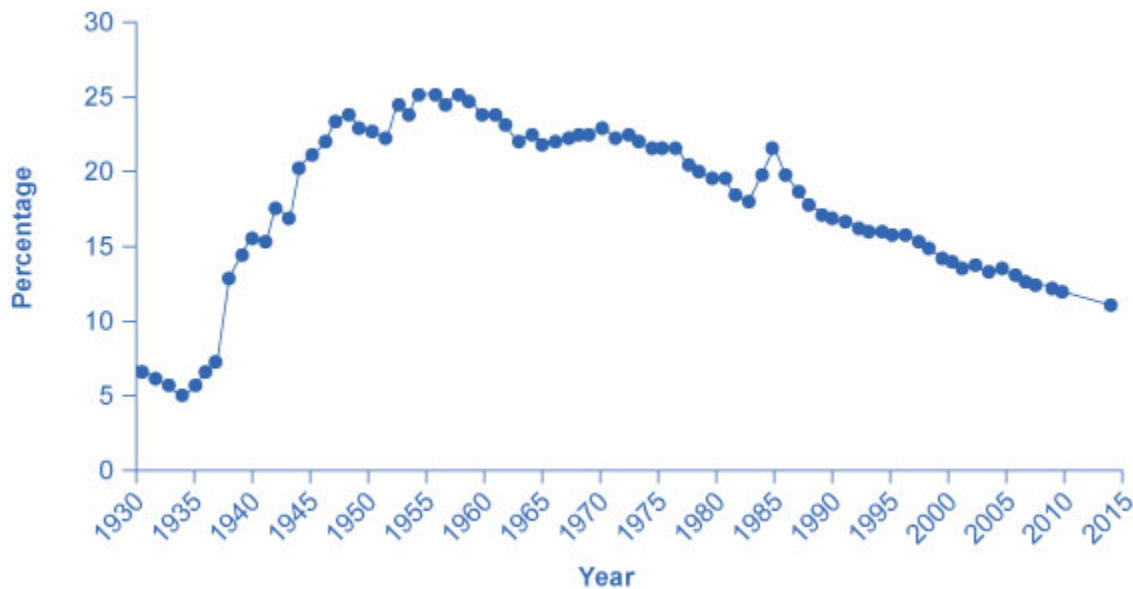


Figure 1. Percentage of Wage and Salary Workers Who Are Union Members. The share of wage and salary workers who belong to unions rose sharply in the 1930s and 1940s, but has tailed off since then to 11.1% of all workers in 2014.

This labor market situation resembles what a **monopoly firm** does in selling a product, but in this case a union is a monopoly selling labor to firms. At the higher union wage W_u , the firms in this industry will hire less labor than they would have hired in equilibrium. Moreover, an excess supply of workers want union jobs, but firms will not be hiring for such jobs.

From the union point of view, workers who receive higher wages are better off. However, notice that the quantity of workers (Q_d) hired at the union wage W_u is smaller than the quantity Q_e that would have been hired at the original equilibrium wage. A sensible union must recognize that when it pushes up the wage, it also reduces the incentive of firms to hire. This situation does not necessarily mean that union workers are fired. Instead, it may be that when union workers move on to other jobs or retire, they are not always replaced. Or perhaps when a firm expands production, it expands employment somewhat less with a higher union wage than it would have done with the lower equilibrium wage. Or perhaps a firm decides to purchase inputs from nonunion producers, rather than producing them with its own highly paid unionized workers. Or perhaps the firm moves or opens a new facility in a state or country where unions are less powerful.

From the firm's point of view, the key question is whether the higher wage of union workers is matched by higher productivity. If so, then the firm can afford to pay the higher union wages and, indeed, the demand curve for "unionized" labor could actually shift to the right. This could reduce the job losses as the equilibrium employment level shifts to the right and the difference between the equilibrium and the union wages will have been reduced. If worker unionization does not increase productivity, then the higher union wage will cause lower profits or losses for the firm.

Union workers might have higher productivity than nonunion workers for a number of reasons. First, higher wages may elicit higher productivity. Second, union workers tend to stay longer at a given job, a trend that reduces the employer's costs for training and hiring and results in workers with more years of experience. Many unions also offer job training and apprenticeship programs.

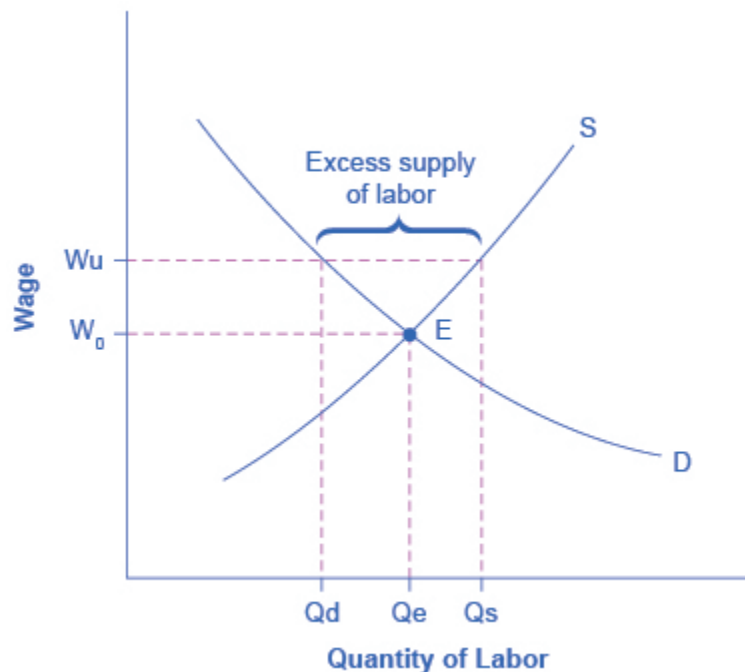


Figure 2. Union Wage Negotiations. Without a union, the equilibrium at E would have involved the wage W_e and the quantity of labor Q_e . However, the union is able to use its bargaining power to raise the wage to W_u . The result is an excess supply of labor for union jobs. That is, a quantity of labor supplied, Q_s is greater than firms' quantity demanded for labor, Q_d .

In addition, firms that are confronted with union demands for higher wages may choose production methods that involve more physical capital and less labor, resulting in increased labor productivity. Table 3 provides an example. Assume that a firm can produce a home exercise cycle with three different combinations of labor and manufacturing equipment. Say that labor is paid \$16 an hour (including benefits) and the machines for manufacturing cost \$200 each. Under these circumstances, the total cost of producing a home exercise cycle will be lowest if the firm adopts the plan of 50 hours of labor and one machine, as the table shows. Now, suppose that a union negotiates a wage of \$20 an hour including benefits. In this case, it makes no difference to the firm whether it uses more hours of labor and fewer machines or less labor and more machines, though it might prefer to use more machines and to hire fewer union workers. (After all, machines never threaten to strike—but they do not buy the final product or service either.) In the final column of the table, the wage has risen to \$24 an hour. In this case, the firm clearly has an incentive for using the plan that involves paying for fewer hours of labor and using three machines. If management responds to union demands for higher wages by investing more in machinery, then union workers can be more productive because they are working with more or better physical capital equipment than the typical nonunion worker. However, the firm will need to hire fewer workers.

Hours of Labor	Number of Machines	Cost of Labor + Cost of Machine \$16/hour	Cost of Labor + Cost of Machine \$20/hour	Cost of Labor + Cost of Machine \$24/hr
30	3	$\$480 + \$600 = \$1,080$	$\$600 + \$600 = \$1,200$	$\$720 + \$600 = \$1,320$
40	2	$\$640 + \$400 = \$1,040$	$\$800 + \$400 = \$1,200$	$\$960 + \$400 = \$1,360$
50	1	$\$800 + \$200 = \$1,000$	$\$1,000 + \$200 = \$1,200$	$\$1,200 + \$200 = \$1,400$

Table 3. Three Production Choices to Manufacture a Home Exercise Cycle

In some cases, unions have discouraged the use of labor-saving physical capital equipment—out of the reasonable fear that new machinery will reduce the number of union jobs. For example, in 2002, the union representing longshoremen who unload ships and the firms that operate shipping companies and port facilities staged a work stoppage that shut down the ports on the western coast of the United States. Two key issues in the dispute were the desire of the shipping companies and port operators to use handheld scanners for record-keeping and computer-operated cranes for loading and unloading ships—changes which the union opposed, along with overtime pay. President Obama threatened to use the Labor Management Relations Act of 1947—commonly known as the **Taft-Hartley Act**—where a court can impose an 80-day “cooling-off period” in order to allow time for negotiations to proceed without the threat of a work stoppage. Federal mediators were called in, and the two sides agreed to a deal in February 2015. The ultimate agreement allowed the new technologies, but also kept wages, health, and pension benefits high for workers. In the past, presidential use of the Taft-Hartley Act sometimes has made labor negotiations more bitter and argumentative but, in this case, it seems to have smoothed the road to an agreement.

In other instances, unions have proved quite willing to adopt new technologies. In one prominent example, during the 1950s and 1960s, the **United Mineworkers union** demanded that mining companies install labor-saving machinery in the mines. The mineworkers’ union realized that over time, the new machines would reduce the number of jobs in the mines, but the union leaders also knew that the mine owners would have to pay higher wages if the workers became more productive, and mechanization was a necessary step toward greater productivity.

In fact, in some cases union workers may be more willing to accept new technology than nonunion workers, because the union workers believe that the union will negotiate to protect their jobs and wages, whereas nonunion workers may be more concerned that the new technology will replace their jobs. In addition, union workers, who typically have higher job market experience and training, are likely to suffer less and benefit more than non-union workers from the introduction of new technology. Overall, it is hard to make a definitive case that union workers as a group are always either more or less welcoming to new technology than are nonunion workers.

THE DECLINE IN U.S. UNION MEMBERSHIP

The proportion of U.S. workers belonging to unions has declined dramatically since the early 1950s. Economists have offered a number of possible explanations:

- The shift from manufacturing to service industries
- The force of globalization and increased competition from foreign producers
- A reduced desire for unions because of the workplace protection laws now in place

- U.S. legal environment that makes it relatively more difficult for unions to organize workers and expand their membership

Let's discuss each of these four explanations in more detail.

A first possible explanation for the decline in the share of U.S. workers belonging to unions involves the patterns of job growth in the manufacturing and service sectors of the economy shown in Figure 3. The U.S. economy had about 15 million manufacturing jobs in 1960. This total rose to 19 million by the late 1970s and then declined to 17 million in 2013. Meanwhile, the number of jobs in service industries and in government combined rose from 35 million in 1960 to over 118 million by 2013, according to the Bureau of Labor Statistics. Because over time unions were stronger in manufacturing than in service industries, the growth in jobs was not happening where the unions were. It is interesting to note that several of the biggest unions in the country are made up of government workers, including the **American Federation of State, County and Municipal Employees (AFSCME)**; the **Service Employees International Union**; and the **National Education Association**. The membership of each of these unions is listed in Table 2. Outside of government employees, however, unions have not had great success in organizing the service sector.

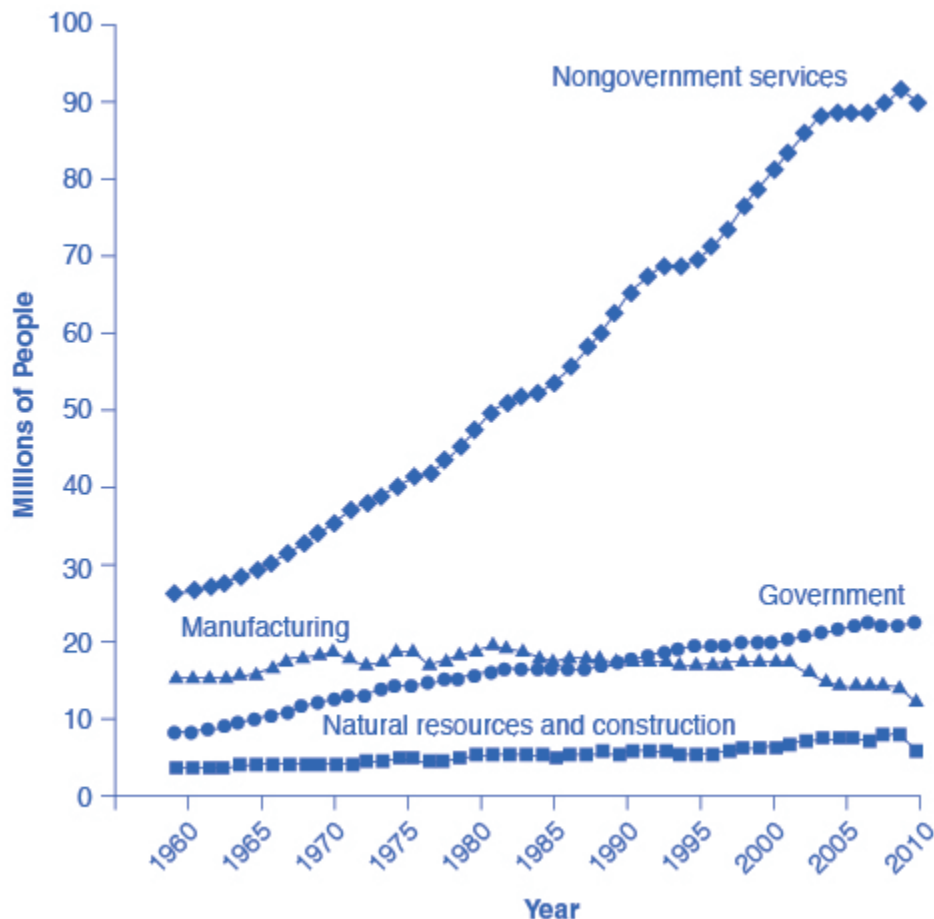


Figure 3. The Growth of Service Jobs. Jobs in services have increased dramatically in the last few decades. Jobs in government have increased modestly. Jobs in manufacturing have not changed much, although they have trended down in recent years. Source: U.S. Department of Labor, Bureau of Labor Statistics.

A second explanation for the decline in the share of unionized workers looks at import competition. Starting in the 1960s, U.S. carmakers and steelmakers faced increasing competition from Japanese and European manufacturers. As sales of imported cars and steel rose, the number of jobs in U.S. auto manufacturing fell. This industry is heavily unionized. Not surprisingly, membership in the United Auto Workers, which was 975,000 in 1985, had fallen to roughly 390,000 by 2015. Import competition not only decreases the employment in sectors where unions were once strong, but also decreases the bargaining power of unions in those sectors. However, as we have seen, unions that organize public-sector workers, who are not threatened by import competition, have continued to see growth.

A third possible reason for the decline in the number of union workers is that citizens often call on their elected representatives to pass laws concerning work conditions, overtime, parental leave, regulation of pensions, and other issues. Unions offered strong political support for these laws aimed at protecting workers but, in an ironic twist, the passage of those laws then made many workers feel less need for unions.

These first three possible reasons for the decline of unions are all somewhat plausible, but they have a common problem. Most other developed economies have experienced similar economic and political trends, such as the shift from manufacturing to services, globalization, and increasing government social benefits and regulation of the workplace. Clearly there are cultural differences between countries as to their acceptance of unions in the workplace. The share of the population belonging to unions in other countries is very high compared with the share in the United States. Table 4 shows the proportion of workers in a number of the world's high-income economies who belong to unions. The United States is near the bottom, along with France and Spain. The last column shows union coverage, defined as including those workers whose wages are determined by a union negotiation even if the workers do not officially belong to the union. In the United States, union membership is almost identical to union coverage. However, in many countries, the wages of many workers who do not officially belong to a union are still determined by collective bargaining between unions and firms.

Country	Union Density: Percentage of Workers Belonging to a Union	Union Coverage: Percentage of Workers Whose Wages Are Determined by Union Bargaining
Austria	37%	99%
France	9%	95%
Germany	26%	63%
Japan	22%	23%
Netherlands	25%	82%
Spain	11.3%	81%
Sweden	82%	92%
United Kingdom	29%	35%
United States	11.1%	12.5%

Table 4. International Comparisons of Union Membership and Coverage in 2012. (Source, CIA World Factbook, retrieved from www.cia.gov)

These international differences in union membership suggest a fourth reason for the decline of union membership in the United States: perhaps U.S. laws are less friendly to the formation of unions than such laws in other countries. The close connection between union membership and a friendly legal

environment is apparent in the history of U.S. unions. The great rise in union membership in the 1930s followed the passage of the **National Labor-Management Relations Act** of 1935, which specified that workers had a right to organize unions and that management had to give them a fair chance to do so. The U.S. government strongly encouraged the formation of unions during the early 1940s in the belief that unions would help to coordinate the all-out production efforts needed during World War II. However, after World War II came the passage of the Taft-Hartley Act of 1947, which gave states the power to allow workers to opt out of the union in their workplace if they so desired. This law made the legal climate less encouraging to those seeking to form unions, and union membership levels soon started declining.

The procedures for forming a union differ substantially from country to country. For example, the procedures in the United States and those in Canada are strikingly different. When a group of workers wish to form a union in the United States, they announce this fact and an election date is set when the employees at a firm will vote in a secret ballot on whether to form a union. Supporters of the union lobby for a “yes” vote, and the management of the firm lobbies for a “no” vote—often even hiring outside consultants for assistance in swaying workers to vote “no.” In Canada, by contrast, a union is formed when a sufficient proportion of workers (usually about 60%) sign an official card saying that they want a union. There is no separate “election date.” The management of Canadian firms is limited by law in its ability to lobby against the union. In addition, though it is illegal to discriminate and fire workers based on their union activity in the United States, the penalties are slight, making this a not so costly way of deterring union activity. In short, forming unions is easier in Canada—and in many other countries—than in the United States.

In summary, union membership in the United States is lower than in many other high-income countries, a difference that may be due to different legal environments and cultural attitudes toward unions.

Visit this website to read about recent protests regarding minimum wage for fast food employees.



KEY CONCEPTS AND SUMMARY

A labor union is an organization of workers that negotiates as a group with employers over compensation and work conditions. Union workers in the United States are paid more on average than other workers with comparable education and experience. Thus, either union workers must be more productive to match this higher pay or the higher pay will lead employers to find ways of hiring fewer union workers than they otherwise would. American union membership has been falling for decades. Some possible reasons include the shift of jobs to service industries; greater competition from global-

ization; the passage of worker-friendly legislation; and U.S. laws that are less favorable to organizing unions.

SELF-CHECK QUESTIONS

- Table 5 shows the quantity demanded and supplied in the labor market for driving city buses in the town of Unionville, where all the bus drivers belong to a union.

Wage Per Hour	Quantity of Workers Demanded	Quantity of Workers Supplied
\$14	12,000	6,000
\$16	10,000	7,000
\$18	8,000	8,000
\$20	6,000	9,000
\$22	4,000	10,000
\$24	2,000	11,000

Table 5.

- What would the equilibrium wage and quantity be in this market if no union existed?
 - Assume that the union has enough negotiating power to raise the wage to \$4 per hour higher than it would otherwise be. Is there now excess demand or excess supply of labor?
- Do unions typically oppose new technology out of a fear that it will reduce the number of union jobs? Why or why not?
 - Compared with the share of workers in most other high-income countries, is the share of U.S. workers whose wages are determined by union bargaining higher or lower? Why or why not?
 - Are firms with a high percentage of union employees more likely to go bankrupt because of the higher wages that they pay? Why or why not?
 - Do countries with a higher percentage of unionized workers usually have less growth in productivity because of strikes and other disruptions caused by the unions? Why or why not?

REVIEW QUESTIONS

- What is a labor union?
- Why do employers have a natural advantage in bargaining with employees?
- What are some of the most important laws that protect employee rights?
- How does the presence of a labor union change negotiations between employers and workers?
- What is the long-term trend in American union membership?
- Would you expect the presence of labor unions to lead to higher or lower pay for worker-members? Would you expect a higher or lower quantity of workers hired by those employers? Explain briefly.
- What are the main causes for the recent trends in union membership rates in the United States? Why are union rates lower in the United States than in many other developed countries?

CRITICAL THINKING QUESTIONS

1. Are unions and technological improvements complementary? Why or why not?
2. Will union membership continue to decline? Why or why not?

REFERENCES

AFL-CIO. "Training and Apprenticeships." <http://www.aflcio.org/Learn-About-Unions/Training-and-Apprenticeships>.

Central Intelligence Agency. "The World Factbook." <https://www.cia.gov/library/publications/the-world-factbook/index.html>.

Clark, John Bates. *Essentials of Economic Theory: As Applied to Modern Problems of Industry and Public Policy*. New York: A. M. Kelley, 1907, 501.

United Auto Workers (UAW). "About: Who We Are." <http://www.uaw.org/page/who-we-are>.

United States Department of Labor: Bureau of Labor Statistics. "Economic News Release: Union Members Summary." Last modified January 23, 2013. <http://www.bls.gov/news.release/union2.nr0.htm>.

United States Department of Labor, Bureau of Labor Statistics. 2015. "Economic News; Union Members Summary." Accessed April 13, 2015. <http://www.bls.gov/news.release/union2.nr0.htm>.

GLOSSARY

collective bargaining negotiations between unions and a firm or firms

labor union an organization of workers that negotiates with employers over wages and working conditions

SOLUTIONS

Answers to Self-Check Questions

1.
 - a. With no union, the equilibrium wage rate would be \$18 per hour and there would be 8,000 bus drivers.
 - b. If the union has enough negotiating power to raise the wage to \$4 per hour higher than under the original equilibrium, the new wage would be \$22 per hour. At this wage, 4,000 workers would be demanded while 10,000 would be supplied, leading to an excess supply of 6,000 workers.
2. Unions have sometimes opposed new technology out of a fear of losing jobs, but in other cases unions have helped to facilitate the introduction of new technology because unionized workers felt that the union was looking after their interests or that their higher skills meant that their jobs were essentially protected. And the new technologies meant increased productivity.

3. In a few other countries (such as France and Spain), the percentage of workers belonging to a union is similar to that in the United States. Union membership rates, however, are generally lower in the United States. When the share of workers whose wages are determined by union negotiations is considered, the United States ranks by far the lowest (because in countries like France and Spain, union negotiations often determine pay even for nonunion employees).
4. No. While some unions may cause firms to go bankrupt, other unions help firms to become more competitive. No overall pattern exists.
5. From a social point of view, the benefits of unions and the costs seem to counterbalance. There is no evidence that in countries with a higher percentage of unionized workers, the economies grow more or less slowly.

21.2 EMPLOYMENT DISCRIMINATION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Analyze earnings gaps based on race and gender
- Explain the impact of discrimination in a competitive market
- Identify U.S. public policies designed to reduce discrimination

Discrimination involves acting on the belief that members of a certain group are inferior solely because of a factor such as race, gender, or religion. There are many types of discrimination but the focus here will be on discrimination in labor markets, which arises if workers with the same skill levels—as measured by education, experience, and expertise—receive different pay or have different job opportunities because of their race or gender.

EARNINGS GAPS BY RACE AND GENDER

A possible signal of labor market discrimination is when one group is paid less than another. Figure 1 shows the average wage of black workers as a ratio of the average wage of white workers and the average wage of female workers as a ratio of the average wage of male workers. Research by the economists **Francine Blau** and **Laurence Kahn** shows that the gap between the earnings of women and men did not move much in the 1970s, but has declined since the 1980s. According to the U.S. Census, the gap between the earnings of blacks and whites diminished in the 1970s, but has not changed in 50 years. In both gender and race, an earnings gap remains.

An earnings gap between average wages, in and of itself, does not prove that discrimination is occurring in the labor market. We need to apply the same productivity characteristics to all parties (employees) involved. Gender discrimination in the labor market occurs when women are paid less than men despite having comparable levels of education, experience, and expertise. (Read the Clear It Up about the sex-discrimination suit brought against Wal-Mart.) Similarly, racial discrimination in the labor market exists when racially diverse employees are paid less than their coworkers of the majority race despite having comparable levels of education, experience, and expertise. To bring a successful gender discrimination lawsuit, a female employee must prove that she is paid less than a male employee who holds a similar job, with similar educational attainment, and with similar expertise. Likewise, someone who wants to sue on the grounds of racial discrimination must prove that he or she is paid less than an employee of another race who holds a similar job, with similar educational attainment, and with similar expertise.

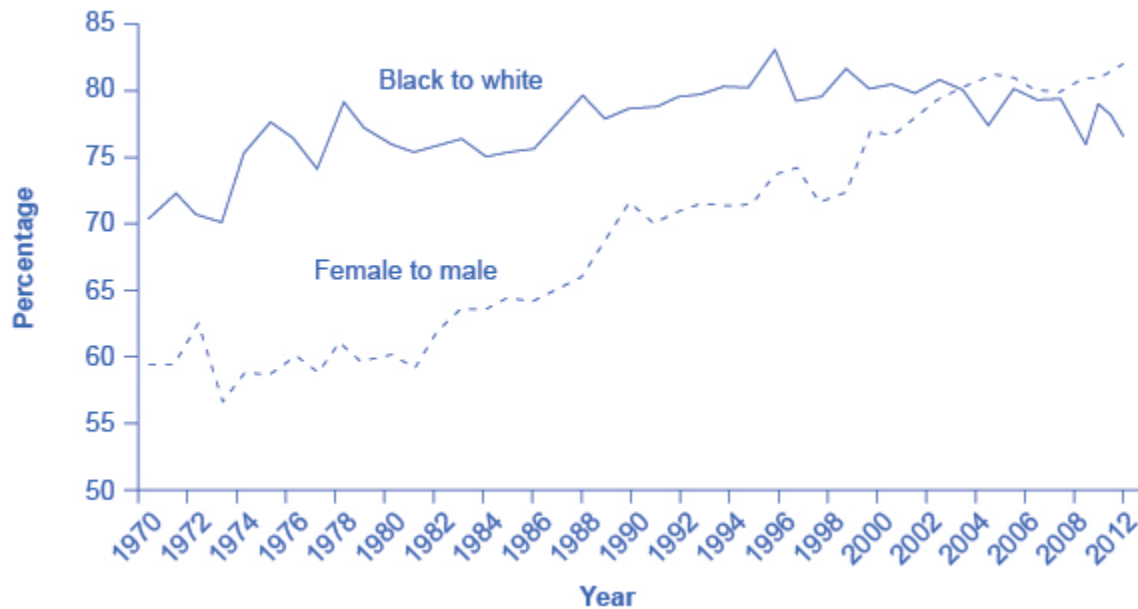


Figure 1. Wage Ratios by Sex and Race. The ratio of wages for black workers to white workers rose substantially in the late 1960s and through the 1970s, but has not changed much since then. The ratio of wages for female to male workers changed little through the 1970s, but has risen substantially since the 1980s. In both cases, a gap remains between the average wages of black and white workers and between the average wages of female and male workers. Source: U.S.

Department of Labor, Bureau of Labor Statistics.

WHAT WAS THE SEX-DISCRIMINATION CASE AGAINST WAL-MART?

In one of the largest class-action sex-discrimination cases in U.S. history, 1.2 million female employees of Wal-Mart claimed that the company engaged in wage and promotion discrimination. In 2011, the Supreme Court threw out the case on the grounds that the group was too large and too diverse for the case to be considered a class action suit. Lawyers for the women regrouped and are now suing in smaller groups. Part of the difficulty for the female employees is that the court said that pay and promotion decisions were made by local managers and were not necessarily policies of the company as a whole. Consequently, female Wal-Mart employees in Texas are arguing that their new suit will challenge the management of a “discrete group of regional district and store managers.” They claim these managers made biased pay and promotion decisions. However, in 2013, a smaller California class action suit against the company was again rejected by a federal district court.

On other issues, Wal-Mart made the news again in 2013 when the National Labor Relations Board found Wal-Mart guilty of illegally penalizing and firing workers who took part in labor protests and strikes. Wal-Mart has already paid \$11.7 million in back wages and compensation damages to women in Kentucky who were denied jobs due to their sex.

INVESTIGATING THE FEMALE/MALE EARNINGS GAP

As a result of changes in law and culture, women began to enter the paid workforce in substantial numbers in the mid- to late-twentieth century. By 2014, 58.1% of adult women held jobs while 72.0% of adult men did. Moreover, along with entering the workforce, women began to ratchet up their education levels. In 1971, 44% of undergraduate college degrees went to women; by 2014, women

received 56% of bachelor's degrees. In 1970, women received 5.4% of the degrees from law schools and 8.4% of the degrees from medical schools. By 2014, women were receiving 47% of the law degrees and 48.0% of the medical degrees. These gains in education and experience have reduced the female/male wage gap over time. However, concerns remain about the extent to which women have not yet assumed a substantial share of the positions at the top of the largest companies or in the U.S. Congress.

There are factors that can lower women's average wages. Women are likely to bear a disproportionately large share of household responsibilities. A mother of young children is more likely to drop out of the labor force for several years or work on a reduced schedule than is the father. As a result, women in their 30s and 40s are likely, on average, to have less job experience than men. In the United States, childless women with the same education and experience levels as men are typically paid comparably. However, women with families and children are typically paid about 7% to 14% less than other women of similar education and work experience. (Meanwhile, married men earn about 10% to 15% more than single men with comparable education and work experience.)

The different patterns of family responsibilities possibly could be called discrimination, but it is primarily rooted in America's social patterns of discrimination, which involve the roles that fathers and mothers play in child-rearing, rather than discrimination by employers in hiring and salary decisions.

Visit this website to read more about the persistently low numbers of women in executive roles in business and in the U.S. Congress.



INVESTIGATING THE BLACK/WHITE EARNINGS GAP

Blacks experienced blatant labor market discrimination during much of the twentieth century. Until the passage of the **Civil Rights Act of 1964**, it was legal in many states to refuse to hire a black worker, regardless of the credentials or experience of that worker. Moreover, blacks were often denied access to educational opportunities, which in turn meant that they had lower levels of qualifications for many jobs. At least one economic study has shown that the 1964 law is partially responsible for the narrowing of the gap in black-white earnings in the late 1960s and into the 1970s; for example, the ratio of total earnings of black male workers to white male workers rose from 62% in 1964 to 75.3% in 2013, according to the Bureau of Labor Statistics.

However, the earnings gap between black and white workers has not changed as much as the earnings gap between men and women has in the last half century. The remaining racial gap seems related both to continuing differences in education levels and to the presence of discrimination. Table 6 shows that the percentage of blacks who complete a four-year college degree remains substantially lower than

the percentage of whites who complete college. According to the U.S. Census, both whites and blacks have higher levels of educational attainment than Hispanics and lower levels than Asians. The lower average levels of education for black workers surely explain part of the earnings gap. In fact, black women who have the same levels of education and experience as white women receive, on average, about the same level of pay. One study shows that white and black college graduates have identical salaries immediately after college; however, the racial wage gap widens over time, an outcome that suggests the possibility of continuing discrimination. Another study conducted a field experiment by responding to job advertisements with fictitious resumes with either very African American sounding names or very white sounding names and found out that white names received 50 percent more callbacks for interviews. This is suggestive of discrimination in job opportunities. Further, as the following Clear It Up feature explains, there is evidence to support that discrimination in the housing market is connected to employment discrimination.

	White	Hispanic	Black	Asian
Completed four years of high school or more	89.0%	65.0%	84.2%	88.9%
Completed four years of college or more	29.0%	14.0%	17.0%	52.0%

Table 6. Educational Attainment by Race and Ethnicity in 2012. (Source: <http://www.census.gov/hhes/socdemo/education/data/cps/2014/tables.html>)

HOW IS DISCRIMINATION IN THE HOUSING MARKET CONNECTED TO EMPLOYMENT DISCRIMINATION?

In a recent study by the Housing and Urban Development (HUD) department, black homebuyers who ask to look at homes for sale are shown 18 percent fewer homes compared to white homebuyers. Asians are shown 19 percent fewer properties. Additionally, Hispanics experience more discrimination in renting apartments and undergo stiffer credit checks than white renters. In a 2012 study conducted by the U.S. Department of Housing and Urban Development and the nonprofit Urban Institute, Hispanic testers who contacted agents about advertised rental units were given information about 12 percent fewer units available and were shown seven percent fewer units than white renters. The \$9 million study, based on research in 28 metropolitan areas, concluded that blatant “door slamming” forms of discrimination are on the decline but that the discrimination that does exist is harder to detect, and as a result, more difficult to remedy. According to the *Chicago Tribune*, HUD Secretary Shaun Donovan told reporters, “Just because it’s taken on a hidden form doesn’t make it any less harmful. You might not be able to move into that community with the good schools.”

The lower levels of education for black workers can also be a result of discrimination—although it may be pre-labor market discrimination, rather than direct discrimination by employers in the labor market. For example, if discrimination in housing markets causes black families to live clustered together in certain poorer neighborhoods, then the black children will continue to have lower educational attainment than their white counterparts and, consequently, not be able to obtain the higher paying jobs that require higher levels of education. Another element to consider is that in the past, when blacks were effectively barred from many high-paying jobs, getting additional education could have seemed somewhat pointless, because the educational degrees would not pay off. Even though labor market discrimination has been legally abolished, it can take some time to establish a culture and a tradition of valuing education highly. Additionally, a legacy of past discrimination may contribute to an attitude that blacks will have a difficult time succeeding in academic subjects. In any case, the impact of social discrimination in labor markets is more complicated than seeking to punish a few bigoted employers.

COMPETITIVE MARKETS AND DISCRIMINATION

Gary Becker (b. 1930), who won the Nobel Prize in economics in 1992, was one of the first to analyze discrimination in economic terms. Becker pointed out that while competitive markets can allow some

employers to practice discrimination, it can also provide profit-seeking firms with incentives not to discriminate. Given these incentives, Becker explored the question of why discrimination persists.

If a business is located in an area with a large minority population and refuses to sell to minorities, it will cut into its own profits. If some businesses run by bigoted employers refuse to pay women and/or minorities a wage based on their productivity, then other profit-seeking employers can hire these workers. In a **competitive market**, if the owners of a business care more about the color of money than about the color of skin, they will have an incentive to make buying, selling, hiring, and promotion decisions strictly based on economic factors.

The power of markets to offer at least a degree of freedom to oppressed groups should not be underestimated. In many countries, cohesive minority groups like Jews and emigrant Chinese have managed to carve out a space for themselves through their economic activities, despite legal and social discrimination against them. Many immigrants, including those who come to the United States, have taken advantage of economic freedom to make new lives for themselves. However, history teaches that market forces alone are unlikely to eliminate discrimination. After all, discrimination against African Americans persisted in the market-oriented U.S. economy during the century between President Abraham Lincoln's Emancipation Proclamation, which freed the slaves in 1863, and the passage of the Civil Rights Act of 1964—and has continued since then, too.

So why does discrimination persist in competitive markets? Gary Becker sought to explain this persistence. Discriminatory impulses can emerge at a number of levels: among managers, among workers, and among customers. Consider the situation of a manager who is not personally prejudiced, but who has many workers or customers who are prejudiced. If that manager treats minority groups or women fairly, the manager may find it hurts the morale of prejudiced co-workers or drives away prejudiced customers. In such a situation, a policy of nondiscrimination could reduce the firm's profits. After all, a business firm is part of society, and a firm that does not follow the societal norms is likely to suffer. Market forces alone are unlikely to overwhelm strong social attitudes about discrimination.

Visit this website to read more about wage discrimination.



PUBLIC POLICIES TO REDUCE DISCRIMINATION

A first public policy step against discrimination in the labor market is to make it illegal. For example, the **Equal Pay Act of 1963** said that men and women who do equal work at a company must be paid the same. The Civil Rights Act of 1964 prohibits employment discrimination based on race, color, religion, sex, or national origin. The **Age Discrimination in Employment Act** of 1967 prohibited discrimination on the basis of age against individuals who are 40 years of age or older. The **Civil**

Rights Act of 1991 provides monetary damages in cases of intentional employment discrimination. The Pregnancy Discrimination Act of 1978 was aimed at prohibiting discrimination against women in the workplace who are planning to get pregnant, are pregnant, or are returning after pregnancy. Passing a law, however, is only part of the answer, since discrimination by prejudiced employers may be less important than broader social patterns.

These laws against discrimination have reduced the gender wage gap. A study by the Department of Labor in 2007 compared salaries of men and women who have similar educational achievement, work experience, and occupation and found that the gender wage gap is only 5%.

In the case of the earnings gap between blacks and whites (and also between Hispanics and whites), probably the single largest step that could be taken at this point in U.S. history to close the earnings gap would be to reduce the gap in educational achievement. Part of the answer to this issue involves finding ways to improve the performance of schools, which is a highly controversial topic in itself. In addition, the education gap is unlikely to close unless black and Hispanic families and peer groups strengthen their culture of support for educational achievement.

Affirmative action is the name given to active efforts by government or businesses that give special rights to minorities in hiring and promotion to make up for past discrimination. Affirmative action, in its limited and not especially controversial form, means making an effort to reach out to a broader range of minority candidates for jobs. In its more aggressive and controversial form, affirmative action required government and companies to hire a specific number or percentage of minority employees. However, the U.S. Supreme Court has ruled against state affirmative action laws. Today, affirmative action policies are applied only to federal contractors who have lost a discrimination lawsuit. This type of redress is enforced by the federal **Equal Employment Opportunity Commission (EEOC)**.

AN INCREASINGLY DIVERSE WORKFORCE

Racial and ethnic diversity is on the rise in the U.S. population and work force. As Figure 2 shows, while the white Americans composed 78% of the population in 2012, the U.S. Bureau of the Census projects that whites will be 69% of the U.S. population by 2060. The proportion of U.S. citizens who are of Hispanic background is predicted to rise substantially. Moreover, in addition to expected changes in the population, diversity is being increased in the workforce as the women who entered the workforce in the 1970s and 1980s are now moving up the promotion ladders within their organizations.

Fortune-telling is not economics, but it still can be clarifying to speculate about the future. Optimists argue that the growing proportions of minority workers will knock over remaining discriminatory barriers. The economy will benefit as an increasing proportion of workers from traditionally disadvantaged groups have a greater opportunity to fulfill their potential. Pessimists worry that the social tensions between men and women and between ethnic groups will rise and that workers will be less productive as a result. Anti-discrimination policy, at its best, seeks to help society move toward the more optimistic outcome.

KEY CONCEPTS AND SUMMARY

Discrimination occurs in a labor market when workers with the same economic characteristics, such

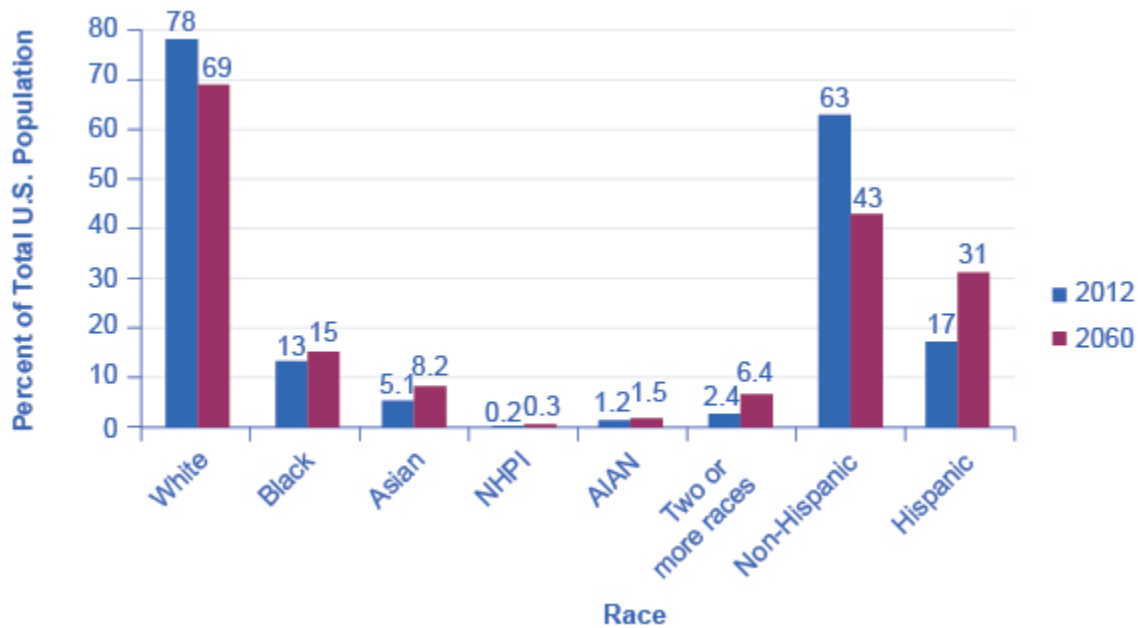


Figure 2. Projected Changes in America's Racial and Ethnic Diversity. This figure shows projected changes in the ethnic makeup of the U.S. population by 2060. Note that "NHPI" stands for Native Hawaiian and Other Pacific Islander. "AIAN" stands for American Indian and Alaska Native.

Source: US Department of Commerce

as education, experience, and skill, are paid different amounts because of race, gender, religion, age, or disability status. In the United States, female workers on average earn less than male workers, and black workers on average earn less than white workers. There is controversy over the extent to which these earnings gaps can be explained by discrimination or by differences in factors like education and job experience. Free markets can allow discrimination to occur; but the threat of a loss of sales or a loss of productive workers can also create incentives for a firm not to discriminate. A range of public policies can be used to reduce earnings gaps between men and women or between white and other racial/ethnic groups: requiring equal pay for equal work, and attaining more equal educational outcomes.

SELF-CHECK QUESTIONS

1. Explain in each of the following situations how market forces might give a business an incentive to act in a less discriminatory fashion.
 - a. A local flower delivery business run by a bigoted white owner notices that many of its local customers are black.
 - b. An assembly line has traditionally only hired men, but it is having a hard time hiring sufficiently qualified workers.
 - c. A biased owner of a firm that provides home health care services would like to pay lower wages to Hispanic workers than to other employees.

2. Does the earnings gap between the average wages of females and the average wages of males prove labor market discrimination? Why or why not?

REVIEW QUESTIONS

1. Describe how the earnings gap between men and women has evolved in recent decades.
2. Describe how the earnings gap between blacks and whites has evolved in recent decades.
3. Does a gap between the average earnings of men and women, or between whites and blacks, prove that employers are discriminating in the labor market? Explain briefly.
4. Will a free market tend to encourage or discourage discrimination? Explain briefly.
5. What policies, when used together with antidiscrimination laws, might help to reduce the earnings gap between men and women or between white and black workers?
6. Describe how affirmative action is applied in the labor market.

CRITICAL THINKING QUESTIONS

1. If it is not profitable to discriminate, why does discrimination persist?
2. If a company has discriminated against minorities in the past, should it be required to give priority to minority applicants today? Why or why not?

REFERENCES

Anderson, Deborah J., Melissa Binder, and Kate Krause. "The Motherhood Wage Penalty Revisited: Experience, Heterogeneity, Work Effort, and Work-Schedule Flexibility." *Industrial and Labor Relations Review*. no. 2 (2003): 273–294. http://www2.econ.iastate.edu/classes/econ321/orazem/anderson_motherhood-penalty.pdf.

Turner, Margery Austin, Rob Santos, Diane K. Levy, Doug Wissoker, Claudia Aranda, Rob Pitingolo, and The Urban Institute. "Housing Discrimination Against Racial and Ethnic Minorities 2012." *U.S. Department of Housing and Urban Development*. Last modified June 2013. http://www.huduser.org/Publications/pdf/HUD-514_HDS2012.pdf.

Austin, Algernon. "The Unfinished March: An Overview." *Economic Policy Institute*. Last modified June 18, 2013. <http://www.epi.org/publication/unfinished-march-overview/>.

Bertrand, Marianne, and Sendhil Mullainathan. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review*. no. 4 (2004): 991–1013. <https://www.aeaweb.org/articles.php?doi=10.1257/0002828042002561&fnd=s>.

Blau, Francine D., and Laurence M. Kahn. "The Gender Pay Gap: Have Women Gone as Far as They Can?" *Academy of Management Perspectives*. no. 1 (2007): 7–23. 10.5465/AMP.2007.24286161.

Card, David, and Alan B. Kruger. "Trends in Relative Black–White Earnings Revisited (Working Paper #310)." *Princeton University and the National Bureau of Economic Research*. December 1992. <http://harris.princeton.edu/pubs/pdfs/310.pdf>.

Donovan, Theresa. Jurist. "Federal Judge Rejects Class Status in Wal-Mart Discrimination Suit." Last modified August 5, 2013. <http://jurist.org/paperchase/2013/08/federal-judge-rejects-class-status-in-wal-mart-discrimination-suit.php>.

Harris, Elizabeth A. "Labor Panel Finds Illegal Punishments at Walmart." *The New York Times*, November 18, 2013. http://www.nytimes.com/2013/11/19/business/labor-panel-finds-illegal-punishments-at-walmart.html?_r=1&.

Kolesnikova, Natalia A., and Yang Liu. Federal Reserve Bank of St. Louis: The Regional Economist. "Gender Wage Gap May Be Much Smaller Than Most Think." Last modified October 2011. <http://www.stlouisfed.org/publications/re/articles/?id=2160>.

Podmolik, Mary Ellen. "HUD Finds Housing Discrimination 'Hidden' But Prevalent." *Chicago Tribune: Business*, June 12, 2013. http://articles.chicagotribune.com/2013-06-12/business/ct-biz-0612-housing-discrimination-20130612_1_renters-testers-chicago-area.

Spreen, Thomas Luke. United States Department of Labor: Bureau of Labor Statistics. "Recent College Graduates in the U.S. Labor Force: Data from the Current Population Survey." *Monthly Labor Review* (February, 2013). <http://www.bls.gov/opub/mlr/2013/02/art1full.pdf>.

U.S. Bureau of Labor Statistics: BLS Reports. 2014. "Women in the Labor Force: A Databook (Report 1052)." Accessed April 13, 2015. <http://www.bls.gov/opub/reports/cps/women-in-the-labor-force-a-databook-2014.pdf>.

U.S. Equal Employment Opportunity Commission, "Walmart to Pay More than \$11.7 Million to Settle EEOC Sex Discrimination Suit." Last modified March 1, 2010. <http://www.eeoc.gov/eeoc/newsroom/release/3-1-10.cfm>.

United States Census Bureau. 2014. "Table 1. Educational Attainment of the Population 18 Years and Over, by Age, Sex, Race, and Hispanic Origin: 2014." Accessed April 13, 2015. <http://www.census.gov/hhes/socdemo/education/data/cps/2014/tables.html>.

United States Department of Labor: Bureau of Labor Statistics. "News Release: Usual Weekly Earnings of Wage and Salary Workers (Third Quarter 2013)." November 1, 2013. <http://www.bls.gov/news.release/pdf/wkyeng.pdf>.

Warner, Judith. 2014. "Fact Sheet: The Women's Leadership Gap: Women's Leadership by the Numbers." Center for American Progress. Accessed March 16, 2015. <https://www.americanprogress.org/issues/women/report/2014/03/07/85457/fact-sheet-the-womens-leadership-gap/>.

Weinberger, Catherine J., and Lois Joy. "The Relative Earnings of Black College Graduates, 1980–2001." In *Race and Economic Opportunity in the 21st Century*, edited by Marlene Kim. New York: Routledge, 2007. <http://www.econ.ucsb.edu/~weinberg/grads.pdf>.

GLOSSARY

affirmative action active efforts by government or businesses that give special rights to minorities in hiring, promotion, or access to education to make up for past discrimination

discrimination actions based on the belief that members of a certain group or groups are in some way inferior solely because of a factor such as race, gender, or religion

SOLUTIONS

Answers to Self-Check Questions

1.
 - a. Firms have a profit incentive to sell to everyone, regardless of race, ethnicity, religion, or gender.
 - b. A business that needs to hire workers to expand may also find that if it draws only from its accustomed pool of workers—say, white men—it lacks the workers it needs to expand production. Such a business would have an incentive to hire more women and minorities.
 - c. A discriminatory business that is underpaying its workers may find those workers leaving for jobs with another employer who offers better pay. This market pressure could cause the discriminatory business to behave better.
2. No. The earnings gap does not prove discrimination because it does not compare the wages of men and women in the same job who have the same amounts of education, experience, and productivity.

21.3 IMMIGRATION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Analyze historical patterns of immigration
- Explain economic effects of immigration
- Evaluate notable proposals for immigration reform

Most Americans would be outraged if a law prevented them from moving to another city or another state. However, when the conversation turns to crossing national borders and are about other people arriving in the United States, laws preventing such movement often seem more reasonable. Some of the tensions over immigration stem from worries over how it might affect a country's culture, including differences in language, and patterns of family, authority, or gender relationships. Economics does not have much to say about such cultural issues. Some of the worries about immigration do, however, have to do with its effects on wages and income levels, and how it affects government taxes and spending. On those topics, economists have insights and research to offer.

HISTORICAL PATTERNS OF IMMIGRATION

Supporters and opponents of immigration look at the same data and see different patterns. Those who express concern about immigration levels to the United States point to graphics like Figure 1 which shows total inflows of **immigrants** decade by decade through the twentieth century. Clearly, the level of immigration has been high and rising in recent years, reaching and exceeding the towering levels of the early twentieth century. However, those who are less worried about immigration point out that the high immigration levels of the early twentieth century happened when total population was much lower. Since the U.S. population roughly tripled during the twentieth century, the seemingly high levels in immigration in the 1990s and 2000s look relatively smaller when they are divided by the population.

Where have the immigrants come from? Immigrants from Europe were more than 90% of the total in the first decade of the twentieth century, but less than 20% of the total by the end of the century. By the 2000s, about half of U.S. immigration came from the rest of the Americas, especially Mexico, and about a quarter came from various countries in Asia.

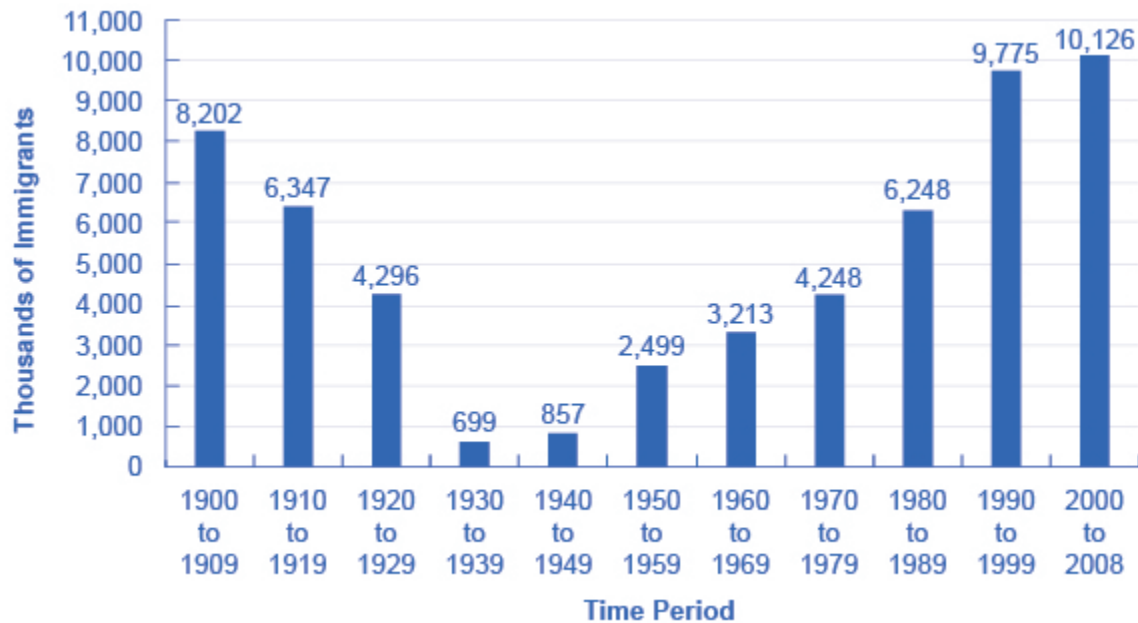


Figure 1. Immigration Since 1900. The number of immigrants in each decade declined between 1900 and the 1940s, but has risen sharply in recent decades. (Source: U.S. Department of Homeland Security, Yearbook of Immigration Statistics: 2011, Table 1)

ECONOMIC EFFECTS OF IMMIGRATION

A surge of immigration can affect the economy in a number of different ways. In this section, we will consider how immigrants might benefit the rest of the economy, how they might affect wage levels, and how they might affect government spending at the federal and local level.

To understand the economic consequences of immigration, consider the following scenario. Imagine that the immigrants entering the United States matched the existing U.S. population in age range, education, skill levels, family size, occupations, and so on. How would immigration of this type affect the rest of the U.S. economy? Immigrants themselves would be much better off, because their standard of living would be higher in the United States. Immigrants would contribute to both increased production and increased consumption. Given enough time for adjustment, the range of jobs performed, income earned, taxes paid, and public services needed would not be much affected by this kind of immigration. It would be as if the population simply increased a little.

Now, consider the reality of recent immigration to the United States. Immigrants are not identical to the rest of the U.S. population. About one-third of immigrants over the age of 25 lack a high school diploma. As a result, many of the recent immigrants end up in jobs like restaurant and hotel work, lawn care, and janitorial work. This kind of immigration represents a shift to the right in the supply of unskilled labor for a number of jobs, which will lead to lower wages for these jobs. The middle- and upper-income households that purchase the services of these unskilled workers will benefit from these lower wages. However, low-skilled U.S. workers who must compete with low-skilled immigrants for jobs will tend to suffer from immigration.

The difficult policy questions about immigration are not so much about the overall gains to the rest of the economy, which seem to be real but small in the context of the U.S. economy, as they are about the disruptive effects of immigration in specific labor markets. One disruptive effect, as just noted, is that

immigration weighted toward low-skill workers tends to reduce wages for domestic low-skill workers. A study by **Michael S. Clune** found that for each 10% rise in the number of employed immigrants with no more than a high school diploma in the labor market, high school students reduced their annual number of hours worked by 3%. The effects on wages of low-skill workers are not large—perhaps in the range of decline of about 1%. These effects are likely kept low, in part, because of the legal floor of federal and state minimum wage laws. In addition, immigrants are also thought to contribute to increased demand for local goods and services which can stimulate the local low skilled labor market. It is also possible that employers, in face of abundant low-skill workers may choose production processes which are more labor intensive than otherwise would have been. These various factors would explain the small negative wage effect observed among the native low-skill workers as a result of immigration.

Another potential disruptive effect is the impact on the budgets of state and local government. Many of the costs imposed by immigrants are costs that arise in state-run programs, like the cost of public schooling and of welfare benefits. However, many of the taxes that immigrants pay are federal taxes like income taxes and Social Security taxes. Many immigrants do not own property (such as homes and cars), so they do not pay property taxes, which are one of the main sources of state and local tax revenue. Though they do pay sales taxes, which are state and local, and the landlords of property they rent pay property taxes. According to the nonprofit Rand Corporation, the effects of immigration on taxes are generally positive at the federal level, but they are negative at the state and local levels in places where there are many low-skilled immigrants.

Visit this website to obtain more context regarding immigration.



PROPOSALS FOR IMMIGRATION REFORM

The Congressional Jordan Commission of the 1990s proposed reducing overall levels of immigration and refocusing U.S. immigration policy to give priority to immigrants with a higher level of skills. In the labor market, focusing on high-skilled immigrants would help prevent any negative effects on the wages of low-skilled workers. For government budgets, higher-skilled workers find jobs more quickly, earn higher wages, and pay more in taxes. Several other immigration-friendly countries, notably Canada and Australia, have immigration systems where those with high levels of education or job skills have a much better chance of obtaining permission to immigrate. For the United States, high tech companies regularly ask for a more lenient immigration policy to admit a greater quantity of highly skilled workers. In addition, a current immigration issue deals with the so-called “DREAM Act” legislation not yet passed by Congress, which would offer a path to citizenship for illegal immigrants brought to the United States before the age of 16. However, some state legislatures, such as California, have passed their own Dream Acts.

If the United States decided to reduce immigration substantially, the economic losses would likely be small relative to the overall economy. If the United States decided to increase immigration substantially, the U.S. economy certainly is large enough to afford some additional assistance to low-wage workers or to local governments that might be adversely affected by immigration. Whether immigration levels are increased, decreased, or left the same, the quality of the debate over immigration policy would be improved by an explicit recognition of who receives economic benefits from immigration and who bears its costs.

COLLECTIVE BARGAINING IN WISCONSIN

Should we end collective bargaining rights for government employees? In an effort to reduce the budget deficit, a contentious Wisconsin law prohibited most public employees from collectively bargaining on anything except wages. Legislators in Wisconsin argued that public safety is so important that public safety workers should be exempted from this. They could not risk firefighters and police going on strike. All firms and employees know that pensions and benefits are expensive; and there was a \$3.6 billion budget deficit in Wisconsin that Governor Walker and legislators wanted to decrease. A lingering question is: should the unions have been more willing to shoulder a greater burden of the cost of those benefits? That question suggests that it is the cost, not necessarily the role of the union itself, which is the problem. After all, unions were founded to reduce the disadvantage that single employees face when bargaining with employers. Because so many government employees are union members, collective bargaining is even more important for them.

Ultimately, the benefit of unions is in the impact they have on economic productivity and output. The more productive the union workers become as a result of collective bargaining, the better off the economy will be.

The long-term repercussions of the Wisconsin law have yet to be realized. As a result of this bill, wage increases higher than the rate of inflation for Wisconsin public sector employees must be voted upon. Imagine if you are working for the Wisconsin government, and are able to find a higher-paying job in the private sector. What will you do? If you decide to leave because your options are better elsewhere, then the government must replace you. How will the government find workers to replace you? For some sectors of the government, reduced numbers of workers may mean greater efficiency. For other sectors, though, reduced numbers of government workers may mean reduced services.

KEY CONCEPTS AND SUMMARY

The recent level of U.S. immigration is at a historically high level if measured in absolute numbers, but not if measured as a share of population. The overall gains to the U.S. economy from immigration are real but relatively small. However, immigration also causes effects like slightly lower wages for low-skill workers and budget problems for certain state and local governments.

SELF-CHECK QUESTIONS

If immigration is reduced, what is the impact on the wage for low-skilled labor? Explain.

REVIEW QUESTIONS

1. Have levels of immigration to the United States been relatively high or low in recent years? Explain.
2. How would you expect immigration by primarily low-skill workers to affect American low-skilled workers?

3. What factors can explain the relatively small effect of low-skilled immigration on the wages of low-skilled workers?

CRITICAL THINKING QUESTIONS

1. If the United States allows a greater quantity of highly skilled workers, what will be the impact on the average wages of highly skilled employees?
2. If all countries eliminated all barriers to immigration, would global economic growth increase? Why or why not?

REFERENCES

Clune, Michael S. "The Fiscal Impacts of Immigrants: A California Case Study." In *The Immigration Debate: Studies on the Economic, Demographic, and Fiscal Effects of Immigration*, edited by James P. Smith and Barry Edmonston. Washington, DC: National Academy Press, 1998, 120–182. http://www.nap.edu/openbook.php?record_id=5985&page=120.

Smith, James P. "Immigration Reform." *Rand Corporation: Rand Review*. <http://www.rand.org/pubs/periodicals/rand-review/issues/2012/fall/leadership/immigration-reform.html>.

U.S. Department of Homeland Security: Office of Immigration Statistics. "2011 Yearbook of Immigration Statistics." September 2012. http://www.dhs.gov/sites/default/files/publications/immigration-statistics/yearbook/2011/ois_yb_2011.pdf.

SOLUTIONS

Answers to Self-Check Questions

If a large share of immigrants have relatively low skills, then reducing the number of immigrants would shift the supply curve of low-skill labor back to the left, which would tend to raise the equilibrium wage for low-skill labor.

CHAPTER 22. INFORMATION, RISK, AND INSURANCE

INTRODUCTION TO INFORMATION, RISK, AND INSURANCE



Figure 1. President Obama’s Health Care Reform. The Patient Protection and Affordable Care Act has become a controversial topic—one which relates strongly to the topic of this chapter. (Credit: modification of work by Daniel Borman/Flickr Creative Commons)

WHAT’S THE BIG DEAL WITH OBAMACARE?

In August 2009, many members of the U.S. Congress used their summer recess to return to their home districts and hold town hall-style meetings to discuss President Obama’s proposed changes to the U.S. healthcare system. This was officially known as the **Patient Protection and Affordable Care Act (PPACA)** or as the **Affordable Care Act (ACA)**, but was more popularly known as Obamacare. The bill’s opponents’ claims ranged from the charge that the changes were unconstitutional and would add \$750 billion to the deficit, to extreme claims about the inclusion of things like the implantation of microchips and so-called “death panels” that decide which critically-ill patients receive care and which do not.

Why did people react so strongly? After all, the intent of the law is to make healthcare insurance more affordable, to allow more people to get insurance, and to reduce the costs of healthcare. For each year from 2000 to 2011, these costs grew at least double the rate of inflation. In 2014, healthcare spending accounted for around 24% of all federal government spend-

ing. In the United States, we spend more for our healthcare than any other high-income nation. Yet in 2015, over 32 million people in the United States, about 13.2%, were without insurance. Even today, however, several years after the Act was signed into law and after it was mostly upheld by the Supreme Court, a 2015 Kaiser Foundation poll found that 43% of likely voters viewed it unfavorably. Why is this?

The debate over the ACA and healthcare reform could take an entire textbook, but what this chapter will do is introduce the basics of insurance and the problems insurance companies face. It is these problems, and how insurance companies respond to them that, in part, explain the ACA.

CHAPTER OBJECTIVES

Introduction to Information, Risk, and Insurance

In this chapter, you will learn about:

- The Problem of Imperfect Information and Asymmetric Information
- Insurance and Imperfect Information

Every purchase is based on a belief about the satisfaction that the good or service will provide. In turn, these beliefs are based on the information that the buyer has available. For many products, the information available to the buyer or the seller is imperfect or unclear, which can either make buyers regret past purchases or avoid making future ones.

This chapter discusses how imperfect and asymmetric information affect markets. The first module of the chapter discusses how asymmetric information affects markets for goods, labor, and financial capital. When buyers have less information about the quality of the good (for example, a gemstone) than sellers do, sellers may be tempted to mislead buyers. If a buyer cannot have at least some confidence in the quality of what is being purchased, then he will be reluctant or unwilling to purchase the products. Thus, mechanisms are needed to bridge this information gap, so buyers and sellers can engage in a transaction.

The second module of the chapter discusses insurance markets, which also face similar problems of imperfect information. For example, a car insurance company would prefer to sell insurance only to those who are unlikely to have auto accidents—but it is hard for the firm to identify those perfectly safe drivers. Conversely, buyers of car insurance would like to persuade the auto insurance company that they are safe drivers and should pay only a low price for insurance. If insurance markets cannot find ways to grapple with these problems of imperfect information, then even people who have low or average risks of making claims may not be able to purchase insurance. The chapter on financial markets (markets for stocks and bonds) will show that the problems of imperfect information can be especially poignant. Imperfect information cannot be eliminated, but it can often be managed.

22.1 THE PROBLEM OF IMPERFECT INFORMATION AND ASYMMETRIC INFORMATION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Analyze the impact of both imperfect information and asymmetric information
- Evaluate the role of advertisements in creating imperfect information
- Identify ways to reduce the risk of imperfect information
- Explain how imperfect information can affect price, quantity, and quality

Consider a purchase that many people make at important times in their lives: buying expensive jewelry. In May 1994, Doree Lynn bought an expensive ring from a jeweler in Washington, D.C., which included an emerald that cost \$14,500. Several years later, the emerald fractured. Lynn took it to another jeweler who found that cracks in the emerald had been filled with an epoxy resin. Lynn sued the original jeweler in 1997 for selling her a treated emerald without telling her, and won. The case publicized a number of little-known facts about precious stones. Most emeralds have internal flaws, and so they are soaked in clear oil or an epoxy resin to hide the flaws and make the color more deep and clear. Clear oil can leak out over time, and epoxy resin can discolor with age or heat. However, using clear oil or epoxy to “fill” emeralds is completely legal, as long as it is disclosed.

After Doree Lynn’s lawsuit, the NBC news show “Dateline” bought emeralds at four prominent jewelry stores in New York City in 1997. All the sales clerks at these stores, unaware that they were being recorded on a hidden camera, said the stones were untreated. When the emeralds were tested at a laboratory, however, it was discovered they had all been treated with oil or epoxy. Emeralds are not the only gemstones that are treated. Diamonds, topaz, and tourmaline are also often irradiated to enhance colors. The general rule is that all treatments to gemstones should be revealed, but often disclosure is not made. As such, many buyers face a situation of **asymmetric information**, where the both parties involved in an economic transaction have an unequal amount of information (one party knows much more than the other).

Many economic transactions are made in a situation of **imperfect information**, where either the buyer, the seller, or both, are less than 100% certain about the qualities of what is being bought and sold. Also, the transaction may be characterized by asymmetric information, in which one party has more information than the other regarding the economic transaction. Let’s begin with some examples of how imperfect information complicates transactions in goods, labor, and **financial capital markets**. The presence of imperfect information can easily cause a decline in prices or quantities of prod-

ucts sold. However, buyers and sellers also have incentives to create mechanisms that will allow them to make mutually beneficial transactions even in the face of imperfect information.

If you are unclear about the difference between asymmetric information and imperfect information, read the following Clear It Up feature.

WHAT IS THE DIFFERENCE BETWEEN IMPERFECT AND ASYMMETRIC INFORMATION?

For a market to reach **equilibrium** sellers and buyers must have full information about the product's price and quality. If there is limited information, then buyers and sellers may not be able to transact or will possibly make poor decisions. Imperfect information refers to the situation where buyers and/or sellers do not have all of the necessary information to make an informed decision about the price or quality of a product. The term imperfect information simply means that not all the information necessary to make an informed decision is known to the buyers and/or sellers. Asymmetric information is the condition where one party, either the buyer or the seller, has more information about the quality or price of the product than the other party. In either case (imperfect or asymmetric information) buyers or sellers need remedies to make more informed decisions.

"LEMONS" AND OTHER EXAMPLES OF IMPERFECT INFORMATION

Consider Marvin, who is trying to decide whether to buy a used car. Let's assume that Marvin is truly clueless about what happens inside a car's engine. He is willing to do some background research, like reading *Consumer Reports* or checking websites that offer information about makes and models of used cars and what they should cost. He might pay a mechanic to inspect the car. Even after devoting some money and time collecting information, however, Marvin still cannot be absolutely sure that he is buying a high-quality used car. He knows that he might buy the car, drive it home, and use it for a few weeks before discovering that car is a "lemon," which is slang for a defective product (especially a car).

Imagine that Marvin shops for a used car and finds two that look very similar in terms of mileage, exterior appearances, and age. One car costs \$4,000, while the other car costs \$4,600. Which car should Marvin buy?

If Marvin was choosing in a world of perfect information, the answer would be simple: he should buy the cheaper car. But Marvin is operating in a world of imperfect information, where the sellers likely know more about the car's problems than he does, and have an incentive to hide the information. After all, the more problems that are disclosed, the lower the car's selling price.

What should Marvin do? First, he needs to understand that even with imperfect information, prices still reflect information. Typically, used cars are more expensive on some dealer lots because the dealers have a trustworthy reputation to uphold. Those dealers try to fix problems that may not be obvious to their customers, in order to create good word of mouth about their vehicles' long term reliability. The short term benefits of selling their customers a "lemon" could cause a quick collapse in the dealer's reputation and a loss of long term profits. On other lots that are less well-established, one can find cheaper used cars, but the buyer takes on more risk when a dealer's reputation has little at stake. The cheapest cars of all often appear on Craigslist, where the individual seller has no reputation to defend. In sum, cheaper prices do carry more risk, so Marvin should balance his appetite for risk versus the potential headaches of many more unanticipated trips to the repair shop.

Similar problems with imperfect information arise in labor and financial capital markets. Consider

Greta, who is applying for a job. Her potential employer, like the used car buyer, is concerned about ending up with a “lemon”—in this case a poor quality employee. The employer will collect information about Greta’s academic and work history. In the end, however, a degree of uncertainty will inevitably remain regarding Greta’s abilities, which are hard to demonstrate without actually observing her on the job. How can a potential employer screen for certain attributes, such as motivation, timeliness, ability to get along with others, and so on? Employers often look to trade schools and colleges to pre-screen candidates. Employers may not even interview a candidate unless he has a degree and, sometimes, a degree from a particular school. Employers may also view awards, a high grade point average, and other accolades as a signal of hard work, perseverance, and ability. Employers may also seek references for insights into key attributes such as energy level, work ethic, and so on.

HOW IMPERFECT INFORMATION CAN AFFECT EQUILIBRIUM PRICE AND QUANTITY

The presence of imperfect information can discourage both buyers and sellers from participating in the market. Buyers may become reluctant to participate because they cannot determine the quality of a product. Sellers of high-quality or medium-quality goods may be reluctant to participate, because it is difficult to demonstrate the quality of their goods to buyers—and since buyers cannot determine which goods have higher quality, they are likely to be unwilling to pay a higher price for such goods.

A market with few buyers and few sellers is sometimes referred to as a **thin market**. By contrast, a market with many buyers and sellers is called a **thick market**. When imperfect information is severe and buyers and sellers are discouraged from participating, markets may become extremely thin as a relatively small number of buyer and sellers attempt to communicate enough information that they can agree on a price.

WHEN PRICE MIXES WITH IMPERFECT INFORMATION ABOUT QUALITY

A buyer confronted with imperfect information will often believe that the price being charged reveals something about the quality of the product. For example, a buyer may assume that a gemstone or a used car that costs more must be of higher quality, even though the buyer is not an expert on gemstones. Think of the expensive restaurant where the food must be good because it is so expensive or the shop where the clothes must be stylish because they cost so much, or the gallery where the art must be great, because it costs so much. If you are hiring a lawyer, you might assume that a lawyer who charges \$400 per hour must be better than a lawyer who charges \$150 per hour. In these cases, price can act as a signal of quality.

When buyers use the market price to draw inferences about the quality of products, then markets may have trouble reaching an **equilibrium price** and quantity. Imagine a situation where a used car dealer has a lot full of used cars that do not seem to be selling, and so the dealer decides to cut the prices of the cars to sell a greater quantity. In a market with imperfect information, many buyers may assume that the lower price implies low-quality cars. As a result, the lower price may not attract more customers. Conversely, a dealer who raises prices may find that customers assume that the higher price means that cars are of higher quality; as a result of raising prices, the dealer might sell more cars. (Whether or not consumers always behave rationally, as an economist would see it, is the subject of the following Clear It Up feature.)

The idea that higher prices might cause a greater quantity demanded and that lower prices might cause a lower quantity demanded runs exactly counter to the basic model of demand and supply

(as outlined in the Demand and Supply chapter). These contrary effects, however, will reach natural limits. At some point, if the price is high enough, the quantity demanded will decline. Conversely, when the price declines far enough, buyers will increasingly find value even if the quality is lower. In addition, information eventually becomes more widely known. An overpriced restaurant that charges more than the quality of its food is worth to many buyers will not last forever.

IS CONSUMER BEHAVIOR RATIONAL?

There is a lot of human behavior out there that mainstream economists have tended to call “irrational” since it is consistently at odds with economists’ **utility maximizing** models. The typical response is for economists to brush these behaviors aside and call them “anomalies” or unexplained quirks.

“If only you knew more economics, you would not be so irrational,” is what many mainstream economists seem to be saying. A group known as **behavioral economists** has challenged this notion, because so much of this so-called “quirky” behavior is extremely common among us. For example, a conventional economist would say that if you lost a \$10 bill today, and also got an extra \$10 in your paycheck, you should feel perfectly neutral. After all, $-\$10 + \$10 = \$0$. You are the same financially as you were before. However, behavioral economists have done research that shows many people will feel some negative emotion—anger, frustration, and so forth—after those two things happen. We tend to focus more on the loss than the gain. This is known as “loss aversion,” where a \$1 loss pains us 2.25 times more than a \$1 gain helps us, according to the economists Daniel Kahneman and Amos Tversky in a famous 1979 *Econometrica* paper. This has implications for investing, as people tend to “overplay” the stock market by reacting more to losses than to gains.

Behavioral economics also tries to explain why people make seemingly irrational decisions in the presence of different situations, or how the decision is “framed.” A popular example is outlined here: Imagine you have the opportunity to buy an alarm clock for \$20 in Store A. Across the street, you learn, is the exact same clock at Store B for \$10. You might say it is worth your time—a five minute walk—to save \$10. Now, take a different example: You are in Store A buying a \$300 phone. Five minutes away, at Store B, the same phone is \$290. You again save \$10 by taking a five minute walk. Do you do it? Surprisingly, it is likely that you would not. Mainstream economists would say “\$10 is \$10” and that it would be irrational to make a five minute walk for \$10 in one case and not the other. However, behavioral economists have pointed out that most of us evaluate outcomes relative to a reference point—here the cost of the product—and think of gains and losses as percentages rather than using actual savings.

Which view is right? Both have their advantages, but behavioral economists have at least shed a light on trying to describe and explain systematic behavior which previously has been dismissed as irrational. If most of us are engaged in some “irrational behavior,” perhaps there are deeper underlying reasons for this behavior in the first place.

MECHANISMS TO REDUCE THE RISK OF IMPERFECT INFORMATION

If you were selling a good like emeralds or used cars where imperfect information is likely to be a problem, how could you reassure possible buyers? If you were buying a good where imperfect information is a problem, what would it take to reassure you? Buyers and sellers in the goods market rely on reputation as well as guarantees, warranties, and service contracts to assure product quality; in the labor market, occupational licenses and certifications are used to assure competency, while in financial capital market cosigners and collateral are used as insurance against unforeseen, detrimental events.

In the goods market, the seller of a good might offer a **money-back guarantee**, an agreement that functions as a promise of quality. This strategy may be especially important for a company that sells goods through mail-order catalogs or over the web, whose customers cannot see the actual products, because it encourages people to buy something even if they are not certain they want to keep it.

L.L. Bean started using money-back-guarantees in 1911, when the founder stitched waterproof shoe rubbers together with leather shoe tops, and sold them as hunting shoes. He guaranteed satisfaction. However, the stitching came apart and, out of the first batch of 100 pairs that were sold, 90 pairs were returned. L.L. Bean took out a bank loan, repaired all of the shoes, and replaced them. The L.L. Bean reputation for customer satisfaction began to spread. Many firms today offer money-back-guarantees for a few weeks or months, but L.L. Bean offers a complete money-back guarantee. Anything you have bought from L.L. Bean can always be returned, no matter how many years later or what condition the product is in, for a full money-back guarantee.

L.L. Bean has very few stores. Instead, most of its sales are made by mail, telephone, or, now, through their website. For this kind of firm, imperfect information may be an especially difficult problem, because customers cannot see and touch what they are buying. A combination of a money-back guarantee and a reputation for quality can help for a mail-order firm to flourish.

Visit this website to read about the origin of Eddie Bauer's 100% customer satisfaction guarantee.



Sellers may offer a **warranty**, which is a promise to fix or replace the good, at least for a certain period of time. The seller may also offer a buyer a chance to buy a **service contract**, where the buyer pays an extra amount and the seller agrees to fix anything that goes wrong for a set time period. Service contracts are often used with large purchases such as cars, appliances and even houses.

Guarantees, warranties, and service contracts are examples of explicit reassurance that sellers provide. In many cases, firms also offer unstated guarantees. For example, some movie theaters might refund the cost of a ticket to a customer who walks out complaining about the show. Likewise, while restaurants do not generally advertise a money-back guarantee or exchange policies, many restaurants allow customers to exchange one dish for another or reduce the price of the bill if the customer is not satisfied.

The rationale for these policies is that firms want repeat customers, who in turn will recommend the business to others; as such, establishing a good reputation is of paramount importance. When buyers know that a firm is concerned about its reputation, they are less likely to worry about receiving a poor-quality product. For example, a well-established grocery store with a good reputation can often charge a higher price than a temporary stand at a local farmer's market, where the buyer may never see the seller again.

Sellers of labor provide information through resumes, recommendations, school transcripts, and examples of their work. **Occupational licenses** are also used to establish quality in the **labor market**. Occupational licenses, which are typically issued by government agencies, show that a worker has

completed a certain type of education or passed a certain test. Some of the professionals who must hold a license are doctors, teachers, nurses, engineers, accountants, and lawyers. In addition, most states require a license to work as a barber, an embalmer, a dietitian, a massage therapist, a hearing aid dealer, a counselor, an insurance agent, and a real estate broker. Some other jobs require a license in only one state. Minnesota requires a state license to be a field archeologist. North Dakota has a state license for bait retailers. In Louisiana, a state license is needed to be a “stress analyst” and California requires a state license to be a furniture upholsterer. According to a 2013 study from the University of Chicago, about 29% of U.S. workers have jobs that require occupational licenses.

Occupational licenses have their downside as well, as they represent a barrier to entry to certain industries. This makes it more difficult for new entrants to compete with incumbents, which can lead to higher prices and less consumer choice. In industries that require licenses, the government has decided that the additional information provided by licenses outweighs the negative effect on competition.

ARE ADVERTISERS ALLOWED TO BENEFIT FROM IMPERFECT INFORMATION?

Many advertisements seem full of imperfect information—at least by what they imply. Driving a certain car, drinking a particular soda, or wearing a certain shoe are all unlikely to bring fashionable friends and fun automatically, if at all. The government rules on advertising, enforced by the **Federal Trade Commission (FTC)**, allow advertising to contain a certain amount of exaggeration about the general delight of using a product. They, however, also demand that if a claim is presented as a fact, it must be true.

Legally, deceptive advertising dates back to the 1950s when Colgate-Palmolive created a television advertisement that seemed to show Rapid Shave shaving cream being spread on sandpaper and then the sand was shaved off the sandpaper. What the television advertisement actually showed was sand sprinkled on Plexiglas—without glue—and then scraped aside by the razor.

In the 1960s, in magazine advertisements for Campbell’s vegetable soup, the company was having problems getting an appetizing picture of the soup, because the vegetables kept sinking. So they filled a bowl with marbles and poured the soup over the top, so that the bowl appeared to be crammed with vegetables.

In the late 1980s, the Volvo Company filmed a television advertisement that showed a monster truck driving over cars, crunching their roofs—all except for the Volvo, which did not crush. However, the FTC found in 1991 that the roof of the Volvo used in the filming had been reinforced with an extra steel framework, while the roof supports on the other car brands had been cut.

The Wonder Bread Company ran television advertisements featuring “Professor Wonder,” who said that because Wonder Bread contained extra calcium, it would help children’s minds work better and improve their memory. The FTC objected, and in 2002 the company agreed to stop running the advertisements.

As can be seen in each of these cases, factual claims about the product’s performance are often checked, at least to some extent, by the Federal Trade Commission. Language and images that are exaggerated or ambiguous, but not actually false, are allowed in advertising. Untrue “facts” are not allowed. In any case, an old Latin saying applies when watching advertisements: *Caveat emptor*—that is, “let the buyer beware.”

On the buyer’s side of the labor market, a standard precaution against hiring a “lemon” of an employee is to specify that the first few months of employment are officially a trial or probationary period, and that the worker can be let go for any reason or no reason after that time. Sometimes workers also receive lower pay during this trial period.

In the **financial capital market**, before a bank makes a loan, it requires a prospective borrower fill

out forms regarding the sources of income; in addition, the bank conducts a credit check on the individual's past borrowing. Another approach is to require a **cosigner** on a loan; that is, another person or firm who legally pledges to repay some or all of the money if the original borrower does not do so. Yet another approach is to require **collateral**, often property or equipment that the bank would have a right to seize and sell if the loan is not repaid.

Buyers of goods and services cannot possibly become experts in evaluating the quality of gemstones, used cars, lawyers, and everything else they buy. Employers and lenders cannot be perfectly omniscient about whether possible workers will turn out well or potential borrowers will repay loans on time. But the mechanisms mentioned above can reduce the risks associated with imperfect information so that the buyer and seller are willing to proceed.

KEY CONCEPTS AND SUMMARY

Many economic transactions are made in a situation of imperfect information, where either the buyer, the seller, or both are less than 100% certain about the qualities of what is being bought and sold. When information about the quality of products is highly imperfect, it may be difficult for a market to exist.

A “lemon” is the name given to a product that turns out, after the purchase, to have low quality. When the seller has more accurate information about the quality of the product than the buyer, the buyer will be hesitant to buy, out of fear of purchasing a “lemon.”

Markets have many ways to deal with imperfect information. In goods markets, buyers facing imperfect information about products may depend upon money-back guarantees, warranties, service contracts, and reputation. In labor markets, employers facing imperfect information about potential employees may turn to resumes, recommendations, occupational licenses for certain jobs, and employment for trial periods. In capital markets, lenders facing imperfect information about borrowers may require detailed loan applications and credit checks, cosigners, and collateral.

SELF-CHECK QUESTIONS

1. For each of the following purchases, say whether you would expect the degree of imperfect information to be relatively high or relatively low:
 - a. Buying apples at a roadside stand
 - b. Buying dinner at the neighborhood restaurant around the corner
 - c. Buying a used laptop computer at a garage sale
 - d. Ordering flowers over the Internet for your friend in a different city
2. Why is there asymmetric information in the labor market? What signals can an employer look for that might indicate the traits they are seeking in a new employee?

REVIEW QUESTIONS

1. Why might it be difficult for a buyer and seller to agree on a price when imperfect information exists?
2. What do economists (and used-car dealers) mean by a “lemon”?
3. What are some of the ways a seller of goods might reassure a possible buyer who is faced with imperfect information?
4. What are some of the ways a seller of labor (that is, someone looking for a job) might reassure a possible employer who is faced with imperfect information?
5. What are some of the ways that someone looking for a loan might reassure a bank that is faced with imperfect information about whether the loan will be repaid?

CRITICAL THINKING QUESTIONS

1. You are on the board of directors of a private high school, which is hiring new tenth-grade science teachers. As you think about hiring someone for a job, what are some mechanisms you might use to overcome the problem of imperfect information?
2. A website offers a place for people to buy and sell emeralds, but information about emeralds can be quite imperfect. The website then enacts a rule that all sellers in the market must pay for two independent examinations of their emerald, which are available to the customer for inspection.
 - a. How would you expect this improved information to affect demand for emeralds on this website?
 - b. How would you expect this improved information to affect the quantity of high-quality emeralds sold on the website?

PROBLEMS

Using Critical Thinking Question 2, sketch the effects in parts (a) and (b) on a single supply and demand diagram. What prediction would you make about how the improved information alters the equilibrium quantity and price?

REFERENCES

Center on Budget and Policy Priorities. 2015. “Policy Basics: Where Do Our Federal Tax Dollars Go?” Accessed April 1, 2015. <http://www.cbpp.org/cms/?fa=view&id=1258>.

Consumer Reports. “Consumer Reports.org.” <http://www.consumerreports.org/cro/index.htm>.

Federal Trade Commission. “About the Federal Trade Commission.” Last modified October 17, 2013. <http://www.ftc.gov/ftc/about.shtm>.

Huffington Post. 2015. “HUFFPOLLSTER: Poll Shows Uptick In Obamacare Favorable Rating.” Accessed April 1, 2015. http://www.huffingtonpost.com/2015/03/19/affordable-care-act-fav_n_6900938.html.

Kahneman, Daniel, and Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*. 47, no. 2 (1979): 263-291. http://www.princeton.edu/~kahneman/docs/Publications/prospect_theory.pdf.

Rasmussen Reports, LLC. "Rasmussen Reports." <http://www.rasmussenreports.com/>.

GLOSSARY

asymmetric information a situation where the seller or the buyer has more information than the other regarding the quality of the item being sold

collateral something valuable—often property or equipment—that a lender would have a right to seize and sell if the loan is not repaid

cosigner another person or firm who legally pledges to repay some or all of the money on a loan if the original borrower does not do so

imperfect information a situation where either the buyer or the seller, or both, are uncertain about the qualities of what is being bought and sold

money-back guarantee a promise that the buyer's money will be refunded under certain conditions

occupational license licenses issued by government agencies, which indicate that a worker has completed a certain type of education or passed a certain test

service contract the buyer pays an extra amount and the seller agrees to fix anything specified in the contract that goes wrong for a set time period

warranty a promise to fix or replace the good for a certain period of time

SOLUTIONS

Answers to Self-Check Questions

1.
 - a. Imperfect information is relatively low; after all, you can see the apples.
 - b. Imperfect information is relatively low. The neighborhood restaurant probably has a certain local reputation.
 - c. Imperfect information is relatively high. How can you tell whether the computer is really in good working order? Why are they selling it?
 - d. Imperfect information is relatively high. What do those flowers really look like?
2. Asymmetric information often exists in the labor market because employers cannot observe many key employee attributes until after the person is hired. Employees, however, know whether they are energetic or detailed-oriented. Employers, therefore, often seek schools to pre-screen candidates. Employers may not even interview a candidate unless he has a degree and often a degree from a particular school. Employers may also view awards, a high grade point average, and other accolades as a signal of hard work, perseverance, and ability. Finally, employers seek references for insights into key attributes such as energy level, work ethic, and so on.

22.2 INSURANCE AND IMPERFECT INFORMATION

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how insurance works
- Identify and evaluate various forms of government and social insurance
- Discuss the problems caused by moral hazard and adverse selection
- Analyze the impact of government regulation of insurance

Insurance is a method that households and firms use to prevent any single event from having a significant detrimental financial effect. Generally, households or firms with insurance make regular payments, called **premiums**. The insurance company prices these premiums based on the probability of certain events occurring among a pool of people. Members of the group who then suffer a specified bad experience receive payments from this pool of money.

Many people have several kinds of insurance: health insurance that pays when they receive medical care; car insurance that pays if they are the driver in an automobile accident; house or renter's insurance that pays if possessions are stolen or damaged by fire; and life insurance, which pays for the family if the principal dies. Table 1 lists a set of insurance markets.

Type of Insurance	Who Pays for It?	It Pays Out When . . .
Health insurance	Employers and individuals	Medical expenses are incurred
Life insurance	Employers and individuals	Policyholder dies
Automobile insurance	Individuals	Car is damaged, stolen, or causes damage to others
Property and homeowner's insurance	Homeowners and renters	Dwelling is damaged or burglarized
Liability insurance	Firms and individuals	An injury occurs for which you are partly responsible
Malpractice insurance	Doctors, lawyers, and other professionals	A poor quality of service is provided that causes harm to others

Table 1. Some Insurance Markets

All insurance involves imperfect information in both an obvious way and in a deeper way. At an obvious level, future events cannot be predicted with certainty. For example, it cannot be known with certainty who will have a car accident, become ill, die, or have his home robbed in the next year. Imperfect information also applies to estimating the risk that something will happen to any individ-

ual. It is difficult for an insurance company to estimate the risk that, say, a particular 20-year-old male driver from New York City will have an accident, because even within that group, some drivers will drive more safely than others. Thus, adverse events occur out of a combination of people's characteristics and choices that make the risks higher or lower and then the good or bad luck of what actually happens.

HOW INSURANCE WORKS

A simplified example of automobile insurance might work this way. Suppose that a group of 100 drivers can be divided into three groups. In a given year, 60 of those people have only a few door dings or chipped paint, which costs \$100 each. Another 30 of the drivers have medium-sized accidents that cost an average of \$1,000 in damages, and 10 of the drivers have large accidents that cost \$15,000 in damages. For the moment, let's imagine that at the beginning of any year, there is no way of identifying the drivers who are low-risk, medium-risk, or high-risk. The total damage incurred by car accidents in this group of 100 drivers will be \$186,000, that is:

$$\begin{aligned} \text{Total damage} & (60 \times \$100) + (30 \times \$1,000) + (10 \times \$15,000) \\ & \$600 + \$30,000 + \$150,000 \\ & \$186,000 \end{aligned}$$

If each of the 100 drivers pays a premium of \$1,860 each year, the insurance company will collect the \$186,000 that is needed to cover the costs of the accidents that occur.

Since insurance companies have such a large number of clients, they are able to negotiate with providers of health care and other services for lower rates than the individual would be able to get, thus increasing the benefit to consumers of becoming insured and saving the insurance company itself money when it pays out claims.

Insurance companies receive income, as shown in Figure 1, from insurance premiums and investment income. Investment income is derived from investing the funds that insurance companies received in the past but did not pay out as insurance claims in prior years. The insurance company receives a rate of return from investing these funds or reserves. The investments are typically made in fairly safe, liquid (easy to convert into cash) investments, as the insurance companies needs to be able to readily access these funds when a major disaster strikes.



Figure 1. An Insurance Company: What Comes In, What Goes Out. Money flows into an insurance company through premiums and investments and out through the payment of claims and operating expenses.

GOVERNMENT AND SOCIAL INSURANCE

Federal and state governments run a number of insurance programs. Some of the programs look much like **private insurance**, in the sense that the members of a group makes steady payments into a fund, and those in the group who suffer an adverse experience receive payments. Other programs protect against risk, but without an explicit fund being set up. Following are some examples.

- **Unemployment insurance:** Employers in every state pay a small amount for unemployment insurance, which goes into a fund that is used to pay benefits to workers for a period of time, usually six months, after they lose their jobs.
- **Pension insurance:** Employers that offer pensions to their retired employees are required by law to pay a small fraction of what they are setting aside for pensions to the Pension Benefit Guarantee Corporation, which is used to pay at least some pension benefits to workers if a company goes bankrupt and cannot pay the pensions it has promised.
- **Deposit insurance:** Banks are required by law to pay a small fraction of their deposits to the Federal Deposit Insurance Corporation, which goes into a fund that is used to pay depositors the value of their bank deposits up to \$250,000 (the amount was raised from \$100,000 to \$250,000 in 2008) if the bank should go bankrupt.
- **Workman's compensation insurance:** Employers are required by law to pay a small percentage of the salaries that they pay into funds, typically run at the state level, that are used to pay benefits to workers who suffer an injury on the job.
- **Retirement insurance:** All workers pay a percentage of their income into Social Security and into Medicare, which then provides income and health care benefits to the elderly. Social Security and Medicare are not literally "insurance" in the sense that those currently contributing to the fund are not eligible for benefits. They function like insurance, however, in the sense that regular payments are made into the programs today in exchange for benefits to be received in the case of a later event—either becoming old or becoming sick when old. Such programs are sometimes called "social insurance."

The major additional costs to insurance companies, other than the payment of claims, are the costs of running a business: the administrative costs of hiring workers, administering accounts, and processing insurance claims. For most insurance companies, the insurance premiums coming in and the claims payments going out are much larger than the amounts earned by investing money or the administrative costs.

Thus, while factors like investment income earned on reserves, administrative costs, and groups with different risks complicate the overall picture, a fundamental law of insurance must hold true: The average person's payments into insurance over time must cover 1) the average person's claims, 2) the costs of running the company, and 3) leave room for the firm's profits. This law can be boiled down to the idea that average premiums and average insurance payouts must be approximately equal.

RISK GROUPS AND ACTUARIAL FAIRNESS

Not all of those who purchase insurance face the same risks. Some people may be more likely, because of genetics or personal habits, to fall sick with certain diseases. Some people may live in an area where car theft or home robbery is more likely than others. Some drivers are safer than others. A **risk group** can be defined as a group that shares roughly the same risks of an adverse event occurring.

Insurance companies often classify people into risk groups, and charge lower premiums to those with lower risks. If people are not separated into risk groups, then those with low-risk must pay for those with high risks. In the simple example of how car insurance works, given earlier, 60 drivers had very low damage of \$100 each, 30 drivers had medium-sized accidents that cost \$1,000 each, and 10 of the drivers had large accidents that cost \$15,000. If all 100 of these drivers pay the same \$1,860, then those with low damages are in effect paying for those with high damages.

If it is possible to classify drivers according to risk group, then each group can be charged according to its expected losses. For example, the insurance company might charge the 60 drivers who seem safest of all \$100 apiece, which is the average value of the damages they cause. Then the intermediate group could pay \$1,000 apiece and the high-cost group \$15,000 each. When the level of insurance premiums that someone pays is equal to the amount that an average person in that risk group would collect in insurance payments, the level of insurance is said to be “actuarially fair.”

Classifying people into risk groups can be controversial. For example, if someone had a major automobile accident last year, should that person be classified as a high-risk driver who is likely to have similar accidents in the future, or as a low-risk driver who was just extremely unlucky? The driver is likely to claim to be low-risk, and thus someone who should be in a risk group with those who pay low insurance premiums in the future. The insurance company is likely to believe that, on average, having a major accident is a signal of being a high-risk driver, and thus try to charge this driver higher insurance premiums. The next two sections discuss the two major problems of imperfect information in insurance markets—called moral hazard and adverse selection. Both problems arise from attempts to categorize those purchasing insurance into risk groups.

THE MORAL HAZARD PROBLEM

Moral hazard refers to the case when people engage in riskier behavior with insurance than they would if they did not have insurance. For example, if you have health insurance that covers the cost of visiting the doctor, you may be less likely to take precautions against catching an illness that might require a doctor’s visit. If you have car insurance, you will worry less about driving or parking your car in ways that make it more likely to get dented. In another example, a business without insurance might install absolute top-level security and fire sprinkler systems to guard against theft and fire. If it is insured, that same business might only install a minimum level of security and fire sprinkler systems.

Moral hazard cannot be eliminated, but insurance companies have some ways of reducing its effect. Investigations to prevent insurance fraud are one way of reducing the extreme cases of moral hazard. Insurance companies can also monitor certain kinds of behavior; to return to the example from above, they might offer a business a lower rate on property insurance if the business installs a top-level security and fire sprinkler system and has those systems inspected once a year.

Another method to reduce moral hazard is to require the injured party to pay a share of the costs. For example, insurance policies often have **deductibles**, which is an amount that the insurance policyholder must pay out of their own pocket before the insurance coverage starts paying. For example, auto insurance might pay for all losses greater than \$500. Health insurance policies often have a **copayment**, in which the policyholder must pay a small amount; for example, a person might have to pay \$20 for each doctor visit, and the insurance company would cover the rest. Another method of cost-sharing is **coinsurance**, which means that the insurance company covers a certain percentage of

the cost. For example, insurance might pay for 80% of the costs of repairing a home after a fire, but the homeowner would pay the other 20%.

All of these forms of cost-sharing discourage moral hazard, because people know that they will have to pay something out of their own pocket when they make an insurance claim. The effect can be powerful. One prominent study found that when people face moderate deductibles and copayments for their health insurance, they consume about one-third less in medical care than people who have complete insurance and do not pay anything out of pocket, presumably because deductibles and copayments reduce the level of moral hazard. However, those who consumed less health care did not seem to have any difference in health status.

A final way of reducing moral hazard, which is especially applicable to health care, is to focus on the incentives of providers of health care, rather than consumers. Traditionally, most health care in the United States has been provided on a **fee-for-service** basis, which means that medical care providers are paid for the services they provide and are paid more if they provide additional services. However, in the last decade or so, the structure of healthcare provision has shifted to an emphasis on health maintenance organizations (HMOs). A **health maintenance organization (HMO)** provides health-care that receives a fixed amount per person enrolled in the plan—regardless of how many services are provided. In this case, a patient with insurance has an incentive to demand more care, but the health-care provider, which is receiving only a fixed payment, has an incentive to reduce the moral hazard problem by limiting the quantity of care provided—as long as it will not lead to worse health problems and higher costs later. Today, many doctors are paid with some combination of managed care and fee-for-service; that is, a flat amount per patient, but with additional payments for the treatment of certain health conditions.

Imperfect information is the cause of the moral hazard problem. If an insurance company had perfect information on risk, it could simply raise its premiums every time an insured party engages in riskier behavior. However, an insurance company cannot monitor all the risks that people take all the time and so, even with various checks and cost-sharing, moral hazard will remain a problem.

Visit this website to read about the relationship between health care and behavioral economics.



THE ADVERSE SELECTION PROBLEM

Adverse selection refers to the problem in which the buyers of insurance have more information about whether they are high-risk or low-risk than the insurance company does. This creates an asymmetric information problem for the insurance company because buyers who are high-risk tend to want to buy more insurance, without letting the insurance company know about their higher risk. For

example, someone purchasing health insurance or life insurance probably knows more about their family's health history than an insurer can reasonably find out even with a costly investigation; someone purchasing car insurance may know that they are a high-risk driver who has not yet had a major accident—but it is hard for the insurance company to collect information about how people actually drive.

To understand how adverse selection can strangle an insurance market, recall the situation of 100 drivers who are buying automobile insurance, where 60 drivers had very low damages of \$100 each, 30 drivers had medium-sized accidents that cost \$1,000 each, and 10 of the drivers had large accidents that cost \$15,000. That would equal \$186,000 in total payouts by the insurance company. Imagine that, while the insurance company knows the overall size of the losses, it cannot identify the high-risk, medium-risk, and low-risk drivers. However, the drivers themselves know their risk groups. Since there is asymmetric information between the insurance company and the drivers, the insurance company would likely set the price of insurance at \$1,860 per year, to cover the average loss (not including the cost of overhead and profit). The result is that those with low risks of only \$100 will likely decide not to buy insurance; after all, it makes no sense for them to pay \$1,860 per year when they are likely only to experience losses of \$100. Those with medium risks of a \$1,000 accident will not buy insurance either. So the insurance company ends up only selling insurance for \$1,860 to high-risk drivers who will average \$15,000 in claims apiece. So the insurance company ends up losing a lot of money. If the insurance company tries to raise its premiums to cover the losses of those with high risks, then those with low or medium risks will be even more discouraged from buying insurance.

Rather than face such a situation of adverse selection, the insurance company may decide not to sell insurance in this **market** at all. If an insurance market is to exist, then one of two things must happen. First, the insurance company might find some way of separating insurance buyers into risk groups with some degree of accuracy and charging them accordingly, which in practice often means that the insurance company tries not to sell insurance to those who may pose high risks. Or second, those with low risks must be required to buy insurance, even if they have to pay more than the actuarially fair amount for their risk group. The notion that people can be required to purchase insurance raises the issue of government laws and regulations that influence the insurance industry.

U.S. HEALTH CARE IN AN INTERNATIONAL CONTEXT

The United States is the only high-income country in the world where most health insurance is paid for and provided by private firms. Greater government involvement in the provision of health insurance is one possible way of addressing moral hazard and adverse selection problems.

The moral hazard problem with health insurance is that when people have insurance, they will demand higher quantities of health care. In the United States, private healthcare insurance tends to encourage an ever-greater demand for healthcare services, which healthcare providers are happy to fulfill. Table 2 shows that on a per-person basis, U.S. healthcare spending towers above other countries. It should be noted that while healthcare expenditures in the United States are far higher than healthcare expenditures in other countries, the health outcomes in the United States, as measured by life expectancy and lower rates of childhood mortality, tend to be lower. Health outcomes, however, may not be significantly affected by healthcare expenditures. Many studies have shown that a country's health is more closely related to diet, exercise, and genetic factors than to healthcare expenditure. This fact further emphasizes that the United States is spending very large amounts on medical care with little obvious health gain.

In the U.S. health insurance market, the main way of solving this adverse selection problem is that health insurance is often sold through groups based on place of employment, or, under The Affordable Care Act, from a state government sponsored health exchange market. From an insurance company's point of view, selling insurance through an employer mixes together a group of people—some with high risks of future health problems and some with lower risks—and thus reduces the insurance firm's fear of attracting only those who have high risks. However, many small companies do not provide health insurance to their employees, and many lower-paying jobs do not include health insurance. Even after all U.S. government programs that provide health insurance for the elderly and the poor are taken into account, approximately 32 million Americans were without health insurance in 2015. While a government-controlled system can avoid the adverse selection problem entirely by providing at least basic health insurance for all, another option is to mandate that all Americans buy health insurance from some provider by preventing providers from denying individuals based on pre-existing conditions. Indeed, this approach was adopted in the Patient Protection and Affordable Care Act, which is discussed later on in this chapter.

Country	Health Care Spending per Person (in 2008)	Male Life Expectancy at Birth, in Years (in 2012)	Female Life Expectancy at Birth, in Years (in 2012)	Male Chance of Dying before Age 5, per 1,000 (in 2012)	Female Chance of Dying before Age 5, per 1,000 (in 2012)
United States	\$7,538	76	81	8	7
Germany	\$3,737	78	83	4	4
France	\$3,696	78	85	4	4
Canada	\$4,079	79	84	6	5
United Kingdom	\$3,129	78	83	5	4

Table 2. A Comparison of Healthcare Spending Across Select Countries. (Source: 2010 OECD study and World Fact Book)

At its best, the largely private U.S. system of health insurance and healthcare delivery provides an extraordinarily high quality of care, along with generating a seemingly endless parade of life-saving innovations. But the system also struggles to control its high costs and to provide basic medical care to all. Other countries have lower costs and more equal access, but they often struggle to provide rapid access to health care and to offer the near-miracles of the most up-to-date medical care. The challenge is a healthcare system that strikes the right balance between quality, access, and cost.

GOVERNMENT REGULATION OF INSURANCE

The U.S. insurance industry is primarily regulated at the state level; indeed, since 1871 there has been a **National Association of Insurance Commissioners** that brings together these state regulators to exchange information and strategies. The state insurance regulators typically attempt to accomplish two things: to keep the price of insurance low and to make sure that everyone has insurance. These goals, however, can conflict with each other and also become easily entangled in politics.

If insurance premiums are set at actuarially fair levels, so that people end up paying an amount that accurately reflects their risk group, certain people will end up paying a lot. For example, if health insurance companies were trying to cover people who already have a chronic disease like AIDS, or who were elderly, they would charge these groups very high premiums for health insurance, because their expected health care costs are quite high. Women in the age bracket 18–44 consume, on aver-

age, about 65% more in health care spending than men. Young male drivers have more car accidents than young female drivers. Thus, actuarially fair insurance would tend to charge young men much more for car insurance than young women. Because people in high-risk groups would find themselves charged so heavily for insurance, they might choose not to buy insurance at all.

State insurance regulators have sometimes reacted by passing rules that attempt to set low premiums for insurance. Over time, however, the fundamental law of insurance must hold: the average amount received by individuals must equal the average amount paid in premiums. When rules are passed to keep premiums low, insurance companies try to avoid insuring any high-risk or even medium-risk parties. If a state legislature passes strict rules requiring insurance companies to sell to everyone at low prices, the insurance companies always have the option of withdrawing from doing business in that state. For example, the insurance regulators in New Jersey are well-known for attempting to keep auto insurance premiums low, and more than 20 different insurance companies stopped doing business in the state in the late 1990s and early 2000s. Similarly, in 2009, State Farm announced that it was withdrawing from selling property insurance in Florida.

In short, government regulators cannot force companies to charge low prices and provide high levels of insurance coverage—and thus take losses—for a sustained period of time. If insurance premiums are going to be set below the actuarially fair level for a certain group, some other group will have to make up the difference. There are two other groups who can make up the difference: taxpayers or other buyers of insurance.

In some industries, the U.S. government has decided free markets will not provide insurance at an affordable price, and so the government pays for it directly. For example, private health insurance is too expensive for many people whose incomes are too low. To combat this, the U.S. government, together with the states, runs the Medicaid program, which provides health care to those with low incomes. Private health insurance also does not work well for the elderly, because their average health care costs can be very high. Thus, the U.S. government started the Medicare program, which provides health insurance to all those over age 65. Other government-funded health-care programs are aimed at military veterans, as an added benefit, and children in families with relatively low incomes.

Another common government intervention in insurance markets is to require that everyone buy certain kinds of insurance. For example, most states legally require car owners to buy auto insurance. Likewise, when a bank loans someone money to buy a home, the person is typically required to have homeowner's insurance, which protects against fire and other physical damage (like hailstorms) to the home. A legal requirement that everyone must buy insurance means that insurance companies do not need to worry that those with low risks will avoid buying insurance. Since insurance companies do not need to fear adverse selection, they can set their prices based on an average for the market, and those with lower risks will, to some extent, end up subsidizing those with higher risks. However, even when laws are passed requiring people to purchase insurance, insurance companies cannot be compelled to sell insurance to everyone who asks—at least not at low cost. Thus, insurance companies will still try to avoid selling insurance to those with high risks whenever possible.

The government cannot pass laws that make the problems of moral hazard and adverse selection disappear, but the government can make political decisions that certain groups should have insurance, even though the private market would not otherwise provide that insurance. Also, the government can impose the costs of that decision on taxpayers or on other buyers of insurance.

THE PATIENT PROTECTION AND AFFORDABLE CARE ACT

In March of 2010, President Obama signed into law the **Patient Protection and Affordable Care Act (PPACA)**. This highly contentious law began to be phased in over time starting in October of 2013. The goal of the act is to bring the United States closer to universal coverage. Some of the key features of the plan include:

- **Individual mandate:** All individuals, who do not receive health care through their employer or through a government program (for example, Medicare), are required to have health insurance or pay a fine. The individual mandate's goal was to reduce the adverse selection problem and keep prices down by requiring all consumers—even the healthiest ones—to have health insurance. Without the need to guard against adverse selection (whereby only the riskiest consumers buy insurance) by raising prices, health insurance companies could provide more reasonable plans to their customers.
- Each state is required to have health insurance exchanges whereby insurance companies compete for business. The goal of the exchanges is to improve competition in the market for health insurance.
- **Employer mandate:** All employers with more than 50 employees must offer health insurance to their employees.

The **Affordable Care Act (ACA)** will be funded through additional taxes to include:

- Increase the Medicare tax by 0.9 percent and add a 3.8 percent tax on unearned income for high income taxpayers.
- Charge an annual fee on health insurance providers.
- Impose other taxes such as a 2.3% tax on manufacturers and importers of certain medical devices.

Many people and politicians have sought to overturn the bill. Those that oppose the bill believe it violates an individual's right to choose whether to have insurance or not. In 2012, a number of states challenged the law on the basis that the individual mandate provision is unconstitutional. In June of 2012, the U.S. Supreme Court ruled in a 5–4 decision that the individual mandate is actually a tax, so it is constitutional as the federal government has the right to tax the populace.

WHAT'S THE BIG DEAL WITH OBAMACARE?

What is it that the Affordable Care Act (ACA) will actually do? To begin with, we should note that it is a massively complex law, with a large number of parts, some of which were implemented immediately, and others that will start every year from 2013 through 2020.

As noted in the chapter, people face ever-increasing healthcare costs in the United States. Those with health insurance demand more health care, pushing up the cost. This is one of the problems the ACA is attempting to fix, in part by introducing regulations designed to control increases in healthcare costs. One example is the regulation that caps the amount healthcare providers can spend on administrative costs. Another is a requirement that healthcare providers switch to electronic medical records (EMRs), which will reduce administrative costs.

Another component of the ACA is the requirement that states establish health insurance exchanges, or markets, where people without health insurance, and businesses that do not provide it for their employees, can shop for different insurance

plans. Setting up these exchanges reduces the imperfections in the market for insurance and, by adding to the supply of insurance plans, may lead to lower prices if the supply increases more than demand. Also, people who are uninsured tend to use emergency rooms for treatment—the most expensive form of healthcare. Given that there are over 40 million uninsured citizens in the United States, this has contributed significantly to rising costs. Capping administrative costs, requiring the use of EMRs, and establishing health insurance markets for those currently uninsured, are all components of the ACA that are intended to help control increases in healthcare costs.

Over the years, the ranks of the uninsured in the United States have grown as rising prices, designed to offset the problem of distinguishing the high-risk from the low-risk person, have pushed employers and individuals out of the market. Also, insurance companies have increasingly used pre-existing conditions to determine if someone is high risk, and thus they either charge prices based on average costs, or they choose not to insure these groups. This has also contributed to the over 32 million uninsured. The ACA addresses this problem by providing that people with preexisting conditions cannot be denied health insurance.

This presents another selection problem because those with pre-existing conditions are a high-risk group. Taken as a separate group, the law of insurance says they should pay higher prices for insurance. Since they cannot be singled out, prices go up for everyone, and low-risk people leave the group. As the high-risk group gets sicker and more risky, prices go up again, and still more people leave the group, creating an upward spiral in prices. To offset this selection problem, the ACA includes an employer and individual mandate requirement. All businesses and individuals must purchase health insurance. At the time of this writing, the actual impact of the Patient Protection and Affordable Care Act is still unknown. Due to political opposition and some difficulties with meeting deadlines, several parts of the law have been delayed, and it will be some time before economists are able to collect enough data to determine whether the law has, in fact, increased coverage and lowered costs as was its intent.

KEY CONCEPTS AND SUMMARY

Insurance is a way of sharing risk. A group of people pay premiums for insurance against some unpleasant event, and those in the group who actually experience the unpleasant event then receive some compensation. The fundamental law of insurance is that what the average person pays in over time must be very similar to what the average person gets out. In an actuarially fair insurance policy, the premiums that a person pays to the insurance company are the same as the average amount of benefits for a person in that risk group. Moral hazard arises in insurance markets because those who are insured against a risk will have less reason to take steps to avoid the costs from that risk.

Many insurance policies have deductibles, copayments, or coinsurance. A deductible is the maximum amount that the policyholder must pay out-of-pocket before the insurance company pays the rest of the bill. A copayment is a flat fee that an insurance policy-holder must pay before receiving services. Coinsurance requires the policyholder to pay a certain percentage of costs. Deductibles, copayments, and coinsurance reduce moral hazard by requiring the insured party to bear some of the costs before collecting insurance benefits.

In a fee-for-service health financing system, medical care providers are reimbursed according to the cost of services they provide. An alternative method of organizing health care is through health maintenance organizations (HMOs), where medical care providers are reimbursed according to the number of patients they handle, and it is up to the providers to allocate resources between patients who receive more or fewer health care services. Adverse selection arises in insurance markets when insurance buyers know more about the risks they face than does the insurance company. As a result, the

insurance company runs the risk that low-risk parties will avoid its insurance because it is too costly for them, while high-risk parties will embrace it because it looks like a good deal to them.

SELF-CHECK QUESTIONS

Why is it difficult to measure health outcomes?

REVIEW QUESTIONS

1. What is an insurance premium?
2. In an insurance system, would you expect each person to receive in benefits pretty much what they pay in premiums? Or is it just that the average benefits paid will equal the average premiums paid?
3. What is an actuarially fair insurance policy?
4. What is the problem of moral hazard?
5. How can moral hazard lead to insurance being more costly than was expected?
6. Define deductibles, copayments, and coinsurance.
7. How can deductibles, copayments, and coinsurance reduce moral hazard?
8. What is the key difference between a fee-for-service healthcare system and a system based on health maintenance organizations?
9. How might adverse selection make it difficult for an insurance market to operate?
10. What are some of the metrics used to measure health outcomes?

CRITICAL THINKING QUESTIONS

1. How do you think the problem of moral hazard might have affected the safety of sports such as football and boxing when safety regulations started requiring that players wear more padding?
2. To what sorts of customers would an insurance company offer a policy with a high copay? What about a high premium with a lower copay?

PROBLEMS

Imagine that 50-year-old men can be divided into two groups: those who have a family history of cancer and those who do not. For the purposes of this example, say that 20% of a group of 1,000 men have a family history of cancer, and these men have one chance in 50 of dying in the next year, while the other 80% of men have one chance in 200 of dying in the next year. The insurance company is selling a policy that will pay \$100,000 to the estate of anyone who dies in the next year.

- a. If the insurance company were selling life insurance separately to each group, what would be the actuarially fair premium for each group?

- b. If an insurance company were offering life insurance to the entire group, but could not find out about family cancer histories, what would be the actuarially fair premium for the group as a whole?
- c. What will happen to the insurance company if it tries to charge the actuarially fair premium to the group as a whole rather than to each group separately?

REFERENCES

Central Intelligence Agency. "The World Factbook." <https://www.cia.gov/library/publications/the-world-factbook/index.html>.

National Association of Insurance Commissioners. "National Association of Insurance Commissioners & The Center for Insurance Policy and Research." <http://www.naic.org/>.

OECD. "The Organisation for Economic Co-operation and Development (OECD)." <http://www.oecd.org/about/>.

USA Today. 2015. "Uninsured Rates Drop Dramatically under Obamacare." Accessed April 1, 2015. <http://www.usatoday.com/story/news/nation/2015/03/16/uninsured-rates-drop-sharply-under-obamacare/24852325/>.

Thaler, Richard H., and Sendhil Mullainathan. "The Concise Encyclopedia of Economics: Behavioral Economics." *Library of Economics and Liberty*. <http://www.econlib.org/library/Enc/BehavioralEconomics.html>.

Henry J. Kaiser Family Foundation, The. "Health Reform: Summary of the Affordable care Act." Last modified April 25, 2013. <http://kff.org/health-reform/fact-sheet/summary-of-new-health-reform-law/>.

GLOSSARY

adverse selection when groups with inherently higher risks than the average person seek out insurance, thus straining the insurance system

coinsurance when an insurance policyholder pays a percentage of a loss, and the insurance company pays the remaining cost

copayment when an insurance policyholder must pay a small amount for each service, before insurance covers the rest

deductible an amount that the insurance policyholders must pay out of their own pocket before the insurance coverage pays anything

fee-for-service when medical care providers are paid according to the services they provide

health maintenance organization (HMO) an organization that provides health care and is paid a fixed amount per person enrolled in the plan—regardless of how many services are provided

insurance method of protecting a person from financial loss, whereby policy holders make regular payments to an insurance entity; the insurance firm then remunerates a group member who suffers significant financial damage from an event covered by the policy

moral hazard when people have insurance against a certain event, they are less likely to guard against that event occurring

premium payment made to an insurance company

risk group a group that shares roughly the same risks of an adverse event occurring

SOLUTIONS

Answers to Self-Check Questions

It is almost impossible to distinguish whether a health outcome such as life expectancy was the result of personal preferences that might affect health and longevity, such as diet, exercise, certain risky behavior, and consumption of certain items like tobacco, or the result of expenditures on health care (for example, annual check-ups).

CHAPTER 23. FINANCIAL MARKETS

INTRODUCTION TO FINANCIAL MARKETS



Figure 1. Building Home Equity. Many people choose to purchase their home rather than rent. This chapter explores how the global financial crisis has influenced home ownership. (Credit: modification of work by Diana Parkhouse/Flickr Creative Commons)

THE HOUSING BUBBLE AND THE FINANCIAL CRISIS OF 2007

In 2006, housing equity in the United States peaked at \$13 trillion. That means that the market prices of homes, less what was still owed on the loans used to buy these houses, equaled \$13 trillion. This was a very good number, since the equity represented the value of the financial asset most U.S. citizens owned.

However, by 2008 this number had gone down to \$8.8 trillion, and it declined further still in 2009. Combined with the decline in value of other financial assets held by U.S. citizens, by 2010, U.S. homeowners' wealth had declined by \$14 trillion! This is a staggering result, and it affected millions of lives: people had to alter their retirement decisions, housing decisions, and other important consumption decisions. Just about every other large economy in the world suffered a decline in the market value of financial assets, as a result of the global financial crisis of 2008–2009.

This chapter will explain why people buy houses (other than as a place to live), why they buy other types of financial assets, and why businesses sell those financial assets in the first place. The chapter will also give us insight into why financial markets and assets go through boom and bust cycles like the one described here.

CHAPTER OBJECTIVES

Introduction to Financial Markets

In this chapter, you will learn about:

- How Businesses Raise Financial Capital
- How Households Supply Financial Capital
- How to Accumulate Personal Wealth

When a firm needs to buy new equipment or build a new facility, it often must go to the financial market to raise funds. Usually firms will add capacity during an economic expansion when profits are on the rise and consumer demand is high. Business investment is one of the critical ingredients needed to sustain economic growth. Even in the sluggish economy of 2009, U.S. firms invested \$1.4 trillion in new equipment and structures, in the hope that these investments would generate profits in the years ahead.

Between the end of the recession in 2009 through the second quarter 2013, profits for the S&P 500 companies grew to 9.7 % despite the weak economy, with much of that amount driven by cost cutting and reductions in input costs, according to the *Wall Street Journal*. Figure 2 shows corporate profits after taxes (adjusted for inventory and capital consumption). Despite the steep decline in quarterly net profit in 2008, profits have recovered and surpassed pre-Recession levels.



Figure 2. Corporate Profits After Tax (Adjusted for Inventory and Capital Consumption). Until 2008, corporate profits after tax have generally continued to increase each year. There was a significant drop in profits during 2008 and into 2009. The profit trend has since continued to increase each year, though at a less steady or consistent rate. (Source: **Federal Reserve Economic Data (FRED)** <https://research.stlouisfed.org/fred2/series/CPATAX>)

Many firms, from huge companies like General Motors to startup firms writing computer software, do not have the financial resources within the firm to make all the desired investments. These firms need financial capital from outside investors, and they are willing to pay interest for the opportunity to get a rate of return on the investment for that financial capital.

On the other side of the financial capital market, suppliers of financial capital, like households, wish to use their savings in a way that will provide a return. Individuals cannot, however, take the few thousand dollars that they save in any given year, write a letter to General Motors or some other firm, and negotiate to invest their money with that firm. **Financial capital markets** bridge this gap: that is, they find ways to take the inflow of funds from many separate suppliers of financial capital and transform it into the funds desired by demanders of financial capital. Such financial markets include stocks, bonds, bank loans, and other financial investments.

Visit this website to read more about financial markets.



Our perspective then shifts to consider how these financial investments appear to suppliers of capital such as the households that are saving funds. Households have a range of investment options: bank accounts, certificates of deposit, money market mutual funds, bonds, stocks, stock and bond mutual funds, housing, and even tangible assets like gold. Finally, the chapter investigates two methods for becoming rich: a quick and easy method that does not work very well at all, and a slow, reliable method that can work very well indeed over a lifetime.

23.1 HOW BUSINESSES RAISE FINANCIAL CAPITAL

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe financial capital and how it relates to profits
- Discuss the purpose and process of borrowing, bonds, and corporate stock
- Explain how firms choose between sources of financial capital

Firms often make decisions that involve spending money in the present and expecting to earn profits in the future. Examples include when a **firm** buys a machine that will last 10 years, or builds a new plant that will last for 30 years, or starts a research and development project. Firms can raise the financial capital they need to pay for such projects in four main ways: (1) from early-stage investors; (2) by reinvesting profits; (3) by borrowing through banks or bonds; and (4) by selling stock. When owners of a business choose sources of financial capital, they also choose how to pay for them.

EARLY STAGE FINANCIAL CAPITAL

Firms that are just beginning often have an idea or a prototype for a product or service to sell, but few customers, or even no customers at all, and thus are not earning profits. Such firms face a difficult problem when it comes to raising financial capital: How can a firm that has not yet demonstrated any ability to earn profits pay a rate of return to financial investors?

For many small businesses, the original source of money is the owner of the business. Someone who decides to start a restaurant or a gas station, for instance, might cover the startup costs by dipping into his or her own bank account, or by borrowing money (perhaps using a home as collateral). Alternatively, many cities have a network of well-to-do individuals, known as “angel investors,” who will put their own money into small new companies at an early stage of development, in exchange for owning some portion of the firm.

Venture capital firms make financial investments in new companies that are still relatively small in size, but that have potential to grow substantially. These firms gather money from a variety of individual or institutional investors, including banks, institutions like college endowments, insurance companies that hold financial reserves, and corporate pension funds. Venture capital firms do more than just supply money to small startups. They also provide advice on potential products, customers, and key employees. Typically, a venture capital fund invests in a number of firms, and then investors in that fund receive returns according to how the fund as a whole performs.

The amount of money invested in venture capital fluctuates substantially from year to year: as one example, venture capital firms invested more than \$48.3 billion in 2014, according to the **National Venture Capital Association**. All early-stage investors realize that the majority of small startup businesses will never hit it big; indeed, many of them will go out of business within a few months or years. They also know that getting in on the ground floor of a few huge successes like a Netflix or an Amazon.com can make up for a lot of failures. Early-stage investors are therefore willing to take large risks in order to be in a position to gain substantial returns on their investment.

PROFITS AS A SOURCE OF FINANCIAL CAPITAL

If firms are earning profits (their revenues are greater than costs), they can choose to reinvest some of these profits in equipment, structures, and research and development. For many established companies, reinvesting their own profits is one primary source of financial capital. Companies and firms just getting started may have numerous attractive investment opportunities, but few current profits to invest. Even large firms can experience a year or two of earning low profits or even suffering losses, but unless the firm can find a steady and reliable source of financial capital so that it can continue making real investments in tough times, the firm may not survive until better times arrive. Firms often need to find sources of financial capital other than profits.

BORROWING: BANKS AND BONDS

When a firm has a record of at least earning significant revenues, and better still of earning profits, the firm can make a credible promise to pay interest, and so it becomes possible for the firm to borrow money. Firms have two main methods of borrowing: banks and bonds.

A bank loan for a firm works in much the same way as a loan for an individual who is buying a car or a house. The firm borrows an amount of money and then promises to repay it, including some rate of interest, over a predetermined period of time. If the firm fails to make its loan payments, the bank (or banks) can often take the firm to court and require it to sell its buildings or equipment to make the loan payments.

Another source of financial capital is a bond. A **bond** is a financial contract: a borrower agrees to repay the amount that was borrowed and also a rate of interest over a period of time in the future. A **corporate bond** is issued by firms, but bonds are also issued by various levels of government. For example, a **municipal bond** is issued by cities, a state bond by U.S. states, and a **Treasury bond** by the federal government through the **U.S. Department of the Treasury**. A bond specifies an amount that will be borrowed, the interest rate that will be paid, and the time until repayment.

A large company, for example, might issue bonds for \$10 million; the firm promises to make interest payments at an annual rate of 8%, or \$800,000 per year and then, after 10 years, will repay the \$10 million it originally borrowed. When a firm issues bonds, the total amount that is borrowed is divided up. A firm seeks to borrow \$50 million by issuing bonds, might actually issue 10,000 bonds of \$5,000 each. In this way, an individual investor could, in effect, loan the firm \$5,000, or any multiple of that amount. Anyone who owns a bond and receives the interest payments is called a **bondholder**. If a firm issues bonds and fails to make the promised interest payments, the bondholders can take the firm to court and require it to pay, even if the firm needs to raise the money by selling buildings or equipment. However, there is no guarantee the firm will have sufficient assets to pay off the bonds. The bondholders may get back only a portion of what they loaned the firm.

Bank borrowing is more customized than issuing bonds, so it often works better for relatively small firms. The bank can get to know the firm extremely well—often because the bank can monitor sales and expenses quite accurately by looking at deposits and withdrawals. Relatively large and well-known firms often issue bonds instead. They use bonds to raise new financial capital that pays for investments, or to raise capital to pay off old bonds, or to buy other firms. However, the idea that banks are usually used for relatively smaller loans and bonds for larger loans is not an ironclad rule: sometimes groups of banks make large loans and sometimes relatively small and lesser-known firms issue bonds.

CORPORATE STOCK AND PUBLIC FIRMS

A **corporation** is a business that “incorporates”—that is owned by shareholders that have limited liability for the debt of the company but share in its profits (and losses). Corporations may be private or public, and may or may not have stock that is publicly traded. They may raise funds to finance their operations or new investments by raising capital through the sale of stock or the issuance of bonds.

Those who buy the stock become the owners, or **shareholders**, of the firm. **Stock** represents ownership of a firm; that is, a person who owns 100% of a company’s stock, by definition, owns the entire company. The stock of a company is divided into **shares**. Corporate giants like IBM, AT&T, Ford, General Electric, Microsoft, Merck, and Exxon all have millions of shares of stock. In most large and well-known firms, no individual owns a majority of the shares of the stock. Instead, large numbers of shareholders—even those who hold thousands of shares—each have only a small slice of the overall ownership of the firm.

When a company is owned by a large number of shareholders, there are three questions to ask:

1. How and when does the company get money from the sale of its stock?
2. What rate of return does the company promise to pay when it sells stock?
3. Who makes decisions in a company owned by a large number of shareholders?

First, a firm receives money from the sale of its stock only when the company sells its own stock to the public (the public includes individuals, mutual funds, insurance companies, and pension funds). A firm’s first sale of stock to the public is called an **initial public offering (IPO)**. The IPO is important for two reasons. For one, the IPO, and any stock issued thereafter, such as stock held as treasury stock (shares that a company keeps in their own treasury) or new stock issued later as a secondary offering, provides the funds to repay the early-stage investors, like the angel investors and the venture capital firms. A venture capital firm may have a 40% ownership in the firm. When the firm sells stock, the venture capital firm sells its part ownership of the firm to the public. A second reason for the importance of the IPO is that it provides the established company with financial capital for a substantial expansion of its operations.

Most of the time when corporate stock is bought and sold, however, the firm receives no financial return at all. If you buy shares of stock in General Motors, you almost certainly buy them from the current owner of those shares, and General Motors does not receive any of your money. This pattern should not seem particularly odd. After all, if you buy a house, the current owner gets your money, not the original builder of the house. Similarly, when you buy shares of stock, you are buying a small slice of ownership of the firm from the existing owner—and the firm that originally issued the stock is not a part of this transaction.

Second, when a firm decides to issue stock, it must recognize that investors will expect to receive a rate of return. That rate of return can come in two forms. A firm can make a direct payment to its shareholders, called a **dividend**. Alternatively, a financial investor might buy a share of stock in Wal-Mart for \$45 and then later sell that share of stock to someone else for \$60, for a gain of \$15. The increase in the value of the stock (or of any asset) between when it is bought and when it is sold is called a **capital gain**.

Third: Who makes the decisions about when a firm will issue stock, or pay dividends, or re-invest profits? To understand the answers to these questions, it is useful to separate firms into two groups: private and public.

A **private company** is owned by the people who run it on a day-to-day basis. A private company can be run by individuals, in which case it is called a **sole proprietorship**, or it can be run by a group, in which case it is a **partnership**. A private company can also be a corporation, but with no publicly issued stock. A small law firm run by one person, even if it employs some other lawyers, would be a sole proprietorship. A larger law firm may be owned jointly by its partners. Most private companies are relatively small, but there are some large private corporations, with tens of billions of dollars in annual sales, that do not have publicly issued stock, such as farm products dealer Cargill, the Mars candy company, and the Bechtel engineering and construction firm.

When a firm decides to sell stock, which in turn can be bought and sold by financial investors, it is called a **public company**. Shareholders own a public company. Since the shareholders are a very broad group, often consisting of thousands or even millions of investors, the shareholders vote for a board of directors, who in turn hire top executives to run the firm on a day-to-day basis. The more shares of stock a shareholder owns, the more votes that shareholder is entitled to cast for the company's board of directors.

In theory, the board of directors helps to ensure that the firm is run in the interests of the true owners—the shareholders. However, the top executives who run the firm have a strong voice in choosing the candidates who will be on their board of directors. After all, few shareholders are knowledgeable enough or have enough of a personal incentive to spend energy and money nominating alternative members of the board.

HOW FIRMS CHOOSE BETWEEN SOURCES OF FINANCIAL CAPITAL

There are clear patterns in how businesses raise financial capital. These patterns can be explained in terms of imperfect information, which as discussed in Information, Risk, and Insurance, is a situation where buyers and sellers in a market do not both have full and equal information. Those who are actually running a firm will almost always have more information about whether the firm is likely to earn profits in the future than outside investors who provide financial capital.

Any young startup firm is a risk; indeed, some startup firms are only a little more than an idea on paper. The firm's founders inevitably have better information about how hard they are willing to work, and whether the firm is likely to succeed, than anyone else. When the founders put their own money into the firm, they demonstrate a belief in its prospects. At this early stage, angel investors and venture capitalists try to overcome the imperfect information, at least in part, by knowing the managers and their business plan personally and by giving them advice.

Accurate information is sometimes not available because **corporate governance**, the name econo-

mists give to the institutions that are supposed to watch over top executives, fails, as the following Clear It Up feature on Lehman Brothers shows.

HOW DID LACK OF CORPORATE GOVERNANCE LEAD TO THE LEHMAN BROTHERS FAILURE?

In 2008, Lehman Brothers was the fourth largest U.S. investment bank, with 25,000 employees. The firm had been in business for 164 years. On September 15, 2008, Lehman Brothers filed for Chapter 11 bankruptcy protection. There are many causes of the Lehman Brothers failure. One area of apparent failure was the lack of oversight by the Board of Directors to keep managers from undertaking excessive risk. Part of the oversight failure, according to Tim Geithner's April 10, 2010, testimony to Congress, can be attributed to the Executive Compensation Committee's emphasis on short-term gains without enough consideration of the risks. In addition, according to the court examiner's report, the Lehman Brother's Board of Directors paid too little attention to the details of the operations of Lehman Brothers and also had limited financial service experience.

The board of directors, elected by the shareholders, is supposed to be the first line of corporate governance and oversight for top executives. A second institution of corporate governance is the auditing firm hired to go over the financial records of the company and certify that everything looks reasonable. A third institution of corporate governance is outside investors, especially large shareholders like those who invest large mutual funds or pension funds. In the case of Lehman Brothers, corporate governance failed to provide investors with accurate financial information about the firm's operations.

As a firm becomes at least somewhat established and its strategy appears likely to lead to profits in the near future, knowing the individual managers and their business plans on a personal basis becomes less important, because information has become more widely available regarding the company's products, revenues, costs, and profits. As a result, other outside investors who do not know the managers personally, like bondholders and shareholders, are more willing to provide financial capital to the firm.

At this point, a firm must often choose how to access financial capital. It may choose to borrow from a bank, issue bonds, or issue stock. The great disadvantage of borrowing money from a bank or issuing bonds is that the firm commits to scheduled interest payments, whether or not it has sufficient income. The great advantage of borrowing money is that the firm maintains control of its operations and is not subject to shareholders. Issuing stock involves selling off ownership of the company to the public and becoming responsible to a board of directors and the shareholders.

The benefit of issuing stock is that a small and growing firm increases its visibility in the financial markets and can access large amounts of financial capital for expansion, without worrying about paying this money back. If the firm is successful and profitable, the board of directors will need to decide upon a dividend payout or how to reinvest profits to further grow the company. Issuing and placing stock is expensive, requires the expertise of investment bankers and attorneys, and entails compliance with reporting requirements to shareholders and government agencies, such as the federal Securities and Exchange Commission.

KEY CONCEPTS AND SUMMARY

Companies can raise early-stage financial capital in several ways: from their owners' or managers' personal savings, or credit cards and from private investors like angel investors and venture capital firms.

A bond is a financial contract through which a borrower agrees to repay the amount that was borrowed. A bond specifies an amount that will be borrowed, the amounts that will be repaid over time based on the interest rate when the bond is issued, and the time until repayment. Corporate bonds are issued by firms; municipal bonds are issued by cities, state bonds by U.S. states, and Treasury bonds by the federal government through the U.S. Department of the Treasury.

Stock represents ownership of a firm. The stock of a company is divided into shares. A firm receives financial capital when it sells stock to the public. A company's first sale of stock to the public is called the initial public offering (IPO). However, a firm does not receive any funds when one shareholder sells stock in the firm to another investor. The rate of return on stock is received in two forms: dividends and capital gains.

A private company is usually owned by the people who run it on a day-to-day basis, although it can be run by hired managers. A private company owned and run by an individual is called a sole proprietorship, while a firm owned run by a group is called a partnership. When a firm decides to sell stock that can be bought and sold by financial investors, then the firm is owned by its shareholders—who in turn elect a board of directors to hire top day-to-day management—and is called a public company. Corporate governance is the name economists give to the institutions that are supposed to watch over top executives, though it does not always work.

SELF-CHECK QUESTIONS

1. Answer these three questions about early-stage corporate finance:
 - a. Why do very small companies tend to raise money from private investors instead of through an IPO?
 - b. Why do small, young companies often prefer an IPO to borrowing from a bank or issuing bonds?
 - c. Who has better information about whether a small firm is likely to earn profits, a venture capitalist or a potential bondholder, and why?
2. From a firm's point of view, how is a bond similar to a bank loan? How are they different?

REVIEW QUESTIONS

1. What are the most common ways for start-up firms to raise financial capital?
2. Why can firms not just use their own profits for financial capital, with no need for outside investors?
3. Why are banks more willing to lend to well-established firms?
4. What is a bond?
5. What does a share of stock represent?
6. When do firms receive money from the sale of stock in their firm and when do they not receive money?
7. What is a dividend?
8. What is a capital gain?
9. What is the difference between a private company and a public company?
10. How do the shareholders who own a company choose the actual managers of the company?

CRITICAL THINKING QUESTIONS

1. If you owned a small firm that had become somewhat established, but you needed a surge of financial capital to carry out a major expansion, would you prefer to raise the funds through borrowing or by issuing stock? Explain your choice.
2. Explain how a company can fail when the safeguards that should be in place fail.

PROBLEMS

The Darkroom Windowshade Company has 100,000 shares of stock outstanding. The investors in the firm own the following numbers of shares: investor 1 has 20,000 shares; investor 2 has 18,000 shares; investor 3 has 15,000 shares; investor 4 has 10,000 shares; investor 5 has 7,000 shares; and investors 6 through 11 have 5,000 shares each. What is the minimum number of investors it would take to vote to change the top management of the company? If investors 1 and 2 agree to vote together, can they be certain of always getting their way in how the company will be run?

REFERENCES

National Venture Capital Association. "Recent Stats & Studies." http://www.nvca.org/index.php?option=com_content&view=article&id=344&Itemid=103Update.

Freddie Mac. 2015. "Freddie Mac Update: March 2015." Accessed April 13, 2015. <http://www.freddiemac.com/investors/pdf/investor-presentation.pdf>.

Former, Jamie D. "Should Your Small Business Go Public? Consider the Benefits and Risks of Becoming a Publicly Traded Company." *U.S. Small Business Administration: Community Blog (blog)*. Publication date March 23, 2010. <http://www.sba.gov/community/blogs/community-blogs/business-law-advisor/should-your-small-business-go-public-consider-0>.

GLOSSARY

bond a financial contract through which a borrower like a corporation, a city or state, or the federal government agrees to repay the amount that was borrowed and also a rate of interest over a period of time in the future

bondholder someone who owns bonds and receives the interest payments

capital gain a financial gain from buying an asset, like a share of stock or a house, and later selling it at a higher price

corporate bond a bond issued by firms that wish to borrow

corporate governance the name economists give to the institutions that are supposed to watch over top executives in companies owned by shareholders

corporation a business owned by shareholders who have limited liability for the company's debt yet a share of the company's profits; may be private or public and may or may not have publicly-traded stock

dividend a direct payment from a firm to its shareholders

initial public offering (IPO) the first sale of shares of stock by a firm to outside investors

municipal bonds a bond issued by cities that wish to borrow

partnership a company run by a group as opposed to an individual

private company a firm owned by the people who run it on a day-to-day basis

public company a firm that has sold stock to the public, which in turn can be bought and sold by investors

shareholders people who own at least some shares of stock in a firm

shares the stock of a firm, divided into individual portions

sole proprietorship a company run by an individual as opposed to a group

stock a claim on partial ownership of a specific firm

Treasury bond a bond issued by the federal government through the U.S. Department of the Treasury

venture capital financial investments in new companies that are still relatively small in size, but that have potential to grow substantially

SOLUTIONS

Answers to Self-Check Questions

1.
 - a. The management of small companies might rather do an IPO right away, but until they get the company up and running, most people would pay very much for the stock because of the risks involved.
 - b. A small company may be earning few or zero profits, and its owners want to reinvest their earnings in the future growth of the company. If this company issues bonds or borrows money, it is obligated to make interest payments, which can eat up the company's cash. If the company issues stock, it is not obligated to make payments to anyone (although it may choose to pay dividends).
 - c. Venture capitalists are private investors who can keep close tabs on the management and strategy of the company—and thus reduce the problems of imperfect information about whether the firm is being well run. Venture capitalists often own a substantial portion of the firm and have much better information than a typical shareholder would.
2. From a firm's point of view, a bond is very similar to a bank loan. Both are ways of borrowing money. Both require paying interest. The major difference is who must be persuaded to lend money: a bank loan requires persuading the bank, while issuing bonds requires persuading a number of separate bondholders. Since a bank often knows a great deal about a firm (especially if the firm has its accounts with that bank), bank loans are more common where imperfect information would otherwise be a problem.

23.2 HOW HOUSEHOLDS SUPPLY FINANCIAL CAPITAL

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Show the relationship between savers, banks, and borrowers
- Calculate bond yield
- Contrast bonds, stocks, mutual funds, and assets
- Explain the tradeoffs between return and risk

The ways in which firms would prefer to raise funds are only half the story of financial markets. The other half is what those households and individuals who supply funds desire, and how they perceive the available choices. The focus of our discussion now shifts from firms on the demand side of financial capital markets to households on the supply side of those markets. The mechanisms for saving available to households can be divided into several categories: deposits in bank accounts; bonds; stocks; money market mutual funds; stock and bond mutual funds; and housing and other tangible assets like owning gold. Each of these investments needs to be analyzed in terms of three factors: (1) the expected rate of return it will pay; (2) the risk that the return will be much lower or higher than expected; and (3) the **liquidity** of the investment, which refers to how easily money or financial assets can be exchanged for a good or service. We will do this analysis as we discuss each of these investments in the sections below. First, however, we need to understand the difference between expected rate of return, risk, and actual rate of return.

EXPECTED RATE OF RETURN, RISK, AND ACTUAL RATE OF RETURN

The **expected rate of return** refers to how much a project or an investment is expected to return to the investor, either in future interest payments, capital gains, or increased profitability. It is usually the average return over a period of time, usually in years or even decades. **Risk** measures the uncertainty of that project's profitability. There are several types of risk, including default risk and interest rate risk. Default risk, as its name suggests, is the risk that the borrower fails to pay back the bond. Interest rate risk is the danger that you might buy a long term bond at a 6% interest rate right before market rates suddenly raise, so had you waited, you could have gotten a similar bond that paid 9%. A high-risk investment is one for which a wide range of potential payoffs is reasonably probable. A low-risk investment will have actual returns that are fairly close to its expected rate of return year after year. A high-risk investment will have actual returns that are much higher than the expected rate of return in some months or years and much lower in other months or years. The **actual rate of return**

refers to the total rate of return, including capital gains and interest paid on an investment at the end of a period of time.

BANK ACCOUNTS

An intermediary is one who stands between two other parties; for example, a person who arranges a blind date between two other people is one kind of intermediary. In financial capital markets, banks are an example of a **financial intermediary**—that is, an institution that operates between a saver who deposits funds in a bank and a borrower who receives a loan from that bank. When a bank serves as a financial intermediary, unlike the situation with a couple on a blind date, the saver and the borrower never meet. In fact, it is not even possible to make direct connections between those who deposit funds in banks and those who borrow from banks, because all funds deposited end up in one big pool, which is then loaned out.

Figure 1 illustrates the position of banks as a financial intermediary, with a pattern of deposits flowing into a bank and loans flowing out, and then repayment of the loans flowing back to the bank, with interest payments for the original savers.

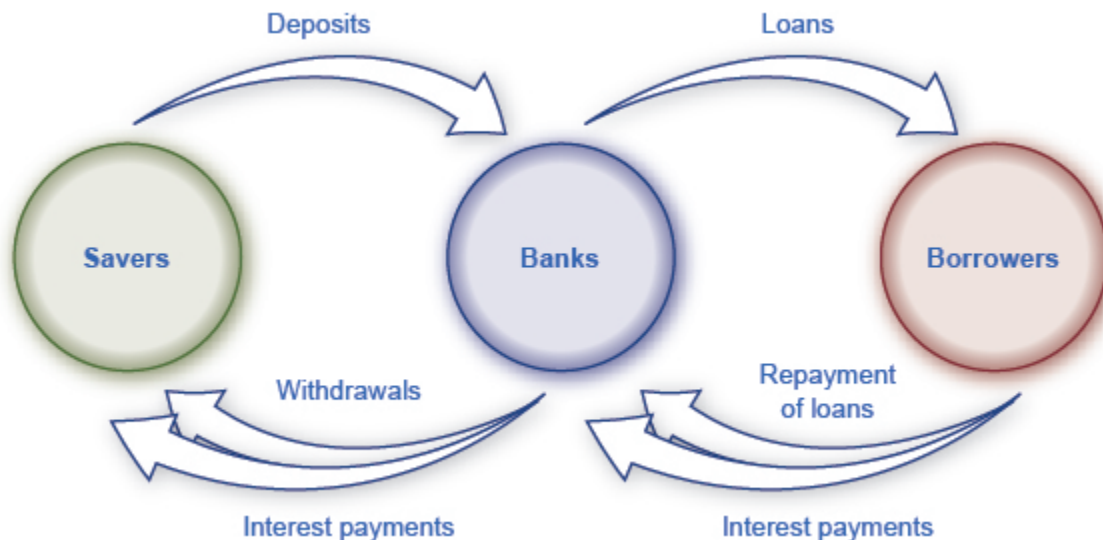


Figure 1. Banks as Financial Intermediaries. Banks are a financial intermediary because they stand between savers and borrowers. Savers place deposits with banks, and then receive interest payments and withdraw money. Borrowers receive loans from banks, and repay the loans with interest.

Banks offer a range of accounts to serve different needs. A **checking account** typically pays little or no interest, but it facilitates transactions by giving you easy access to your money, either by writing a check or by using a **debit card** (that is, a card which works like a credit card, except that purchases are immediately deducted from your checking account rather than being billed separately through a credit card company). A **savings account** typically pays some interest rate, but getting the money typically requires you to make a trip to the bank or an automatic teller machine (or you can access the funds electronically). The lines between checking and savings accounts have blurred in the last couple of decades, as many banks offer checking accounts that will pay an interest rate similar to a savings

account if you keep a certain minimum amount in the account, or conversely, offer savings accounts that allow you to write at least a few checks per month.

Another way to deposit savings at a bank is to use a **certificate of deposit (CD)**. With a CD, as it is commonly called, you agree to deposit a certain amount of money, often measured in thousands of dollars, in the account for a stated period of time, typically ranging from a few months to several years. In exchange, the bank agrees to pay a higher interest rate than for a regular savings account. While you can withdraw the money before the allotted time, as the advertisements for CDs always warn, there is “a substantial penalty for early withdrawal.”

Figure 2 shows the annual rate of interest paid on a six-month, one-year, and five-year CD since 1984, as reported by Bankrate.com. The interest rates paid by savings accounts are typically a little lower than the CD rate, because financial investors need to receive a slightly higher rate of interest as compensation for promising to leave deposits untouched for a period of time in a CD, and thus giving up some liquidity.



Figure 2. Interest Rates on Six-Month, One-Year, and Five-Year Certificates of Deposit. The interest rates on certificates of deposit have fluctuated over time. The high interest rates of the early 1980s are indicative of the relatively high inflation rate in the United States at that time.

Interest rates fluctuate with the business cycle, typically increasing during expansions and decreasing during a recession. Note the steep decline in CD rates since 2008, the beginning of the Great Recession.

The great advantages of bank accounts are that financial investors have very easy access to their money, and also money in bank accounts is extremely safe. In part, this safety arises because a bank account offers more security than keeping a few thousand dollars in the toe of a sock in your underwear drawer. In addition, the **Federal Deposit Insurance Corporation (FDIC)** protects the savings of the average person. Every bank is required by law to pay a fee to the FDIC, based on the size of its deposits. Then, if a bank should happen to go bankrupt and not be able to repay depositors, the FDIC guarantees that all customers will receive their deposits back up to \$250,000.

The bottom line on bank accounts looks like this: low risk means low rate of return but high liquidity.

BONDS

An investor who buys a bond expects to receive a rate of return. However, bonds vary in the rates of return that they offer, according to the riskiness of the borrower. An interest rate can always be divided up into three components (as explained in Choice in a World of Scarcity): compensation for delaying consumption, an adjustment for an inflationary rise in the overall level of prices, and a risk premium that takes the borrower's riskiness into account.

The U.S. government is considered to be an extremely safe borrower, so when the U.S. government issues Treasury bonds, it can pay a relatively low rate of interest. Firms that appear to be safe borrowers, perhaps because of their sheer size or because they have consistently earned profits over time, will still pay a higher interest rate than the U.S. government. Firms that appear to be riskier borrowers, perhaps because they are still growing or their businesses appear shaky, will pay the highest interest rates when they issue bonds. Bonds that offer high interest rates to compensate for their relatively high chance of default are called **high yield bonds** or **junk bonds**. A number of today's well-known firms issued junk bonds in the 1980s when they were starting to grow, including Turner Broadcasting and Microsoft.

Visit this website to read about Treasury bonds.



A bond issued by the U.S. government or a large corporation may seem to be relatively low risk: after all, the issuer of the bond has promised to make certain payments over time, and except for rare cases of bankruptcy, these payments will be made. If the issuer of a corporate bond fails to make the payments that it owes to its bondholders, the bondholders can require that the company declare bankruptcy, sell off its assets, and pay them as much as it can. Even in the case of junk bonds, a wise investor can reduce the risk by purchasing bonds from a wide range of different companies since, even if a few firms go broke and do not pay, they are not all likely to go bankrupt.

As we noted before, bonds carry an interest rate risk. For example, imagine you decide to buy a 10-year bond that would pay an annual interest rate of 8%. Soon after you buy the bond, interest rates on bonds rise, so that now similar companies are paying an annual rate of 12%. Anyone who buys a bond now can receive annual payments of \$120 per year, but since your bond was issued at an interest rate of 8%, you have tied up \$1,000 and receive payments of only \$80 per year. In the meaningful sense of opportunity cost, you are missing out on the higher payments that you could have received. Furthermore, the amount you should be willing to pay now for future payments can be calculated. To place a present discounted value on a future payment, decide what you would need in the present to

equal a certain amount in the future. This calculation will require an interest rate. For example, if the interest rate is 25%, then a payment of \$125 a year from now will have a present discounted value of \$100—that is, you could take \$100 in the present and have \$125 in the future. (This is discussed further in the appendix on Present Discounted Value.)

In financial terms, a bond has several parts. A bond is basically an “I owe you” note that is given to an investor in exchange for capital (money). The bond has a **face value**. This is the amount the borrower agrees to pay the investor at maturity. The bond has a **coupon rate** or interest rate, which is usually semi-annual, but can be paid at different times throughout the year. (Bonds used to be paper documents with coupons that were clipped and turned in to the bank to receive interest.) The bond has a **maturity date** when the borrower will pay back its face value as well as its last interest payment. Combining the bond’s face value, interest rate, and maturity date, and market interest rates, allows a buyer to compute a bond’s **present value**, which is the most that a buyer would be willing to pay for a given bond. This may or may not be the same as the face value.

The **bond yield** measures the rate of return a bond is expected to pay over time. Bonds are bought not only when they are issued; they are also bought and sold during their lifetimes. When buying a bond that has been around for a few years, investors should know that the interest rate printed on a bond is often not the same as the bond yield, even on new bonds. Read the next Work It Out feature to see how this happens.

CALCULATING THE BOND YIELD

You have bought a \$1,000 bond whose coupon rate is 8%. To calculate your return or yield, follow these steps:

1. Assume the following:

Face value of a bond : \$1,000

Coupon rate : 8%

Annual payment : \$80 per year

2. Consider the risk of the bond. If this bond carries no risk, then it would be safe to assume that the bond will sell for \$1,000 when it is issued and pay the purchaser \$80 per year until its maturity, at which time the final interest payment will be made and the original \$1,000 will be repaid. Now, assume that over time the interest rates prevailing in the economy rise to 12% and that there is now only one year left to this bond’s maturity. This makes the bond an unattractive investment, since an investor can find another bond that perhaps pays 12%. To induce the investor to buy the 8% bond, the bond seller will lower its price below its face value of \$1,000.
3. Calculate the price of the bond when its interest rate is less than the market interest rate. The expected payments from the bond one year from now are \$1,080, because in the bond’s last year the issuer of the bond will make the final interest payment and then also repay the original \$1,000. Given that interest rates are now 12%, you know that you could invest \$964 in an alternative investment and receive \$1,080 a year from now; that is:

$$\$964(1 + 0.12) = \$1080$$

Therefore, you will not pay more than \$964 for the original \$1,000 bond.

4. Consider that the investor will receive the \$1,000 face value, plus \$80 for the last year’s interest payment. The yield on the bond will be:

$$(\$1080 - \$964)/\$964 = 12\%$$

The yield, or total return, means interest payments, plus capital gains. Note that the interest or coupon rate of 8% did not change. When interest rates rise, bonds previously issued at lower interest rates will sell for less than face value. Conversely, when interest rates fall, bonds previously issued at higher interest rates will sell for more than face value.

Figure 3 shows bond yield for two kinds of bonds: 10-year Treasury bonds (which are officially called “notes”) and corporate bonds issued by firms that have been given an AAA rating as relatively safe borrowers by **Moody’s**, an independent firm that publishes such ratings. Even though corporate bonds pay a higher interest rate, because firms are riskier borrowers than the federal government, the rates tend to rise and fall together. Treasury bonds typically pay more than bank accounts, and corporate bonds typically pay a higher interest rate than Treasury bonds.

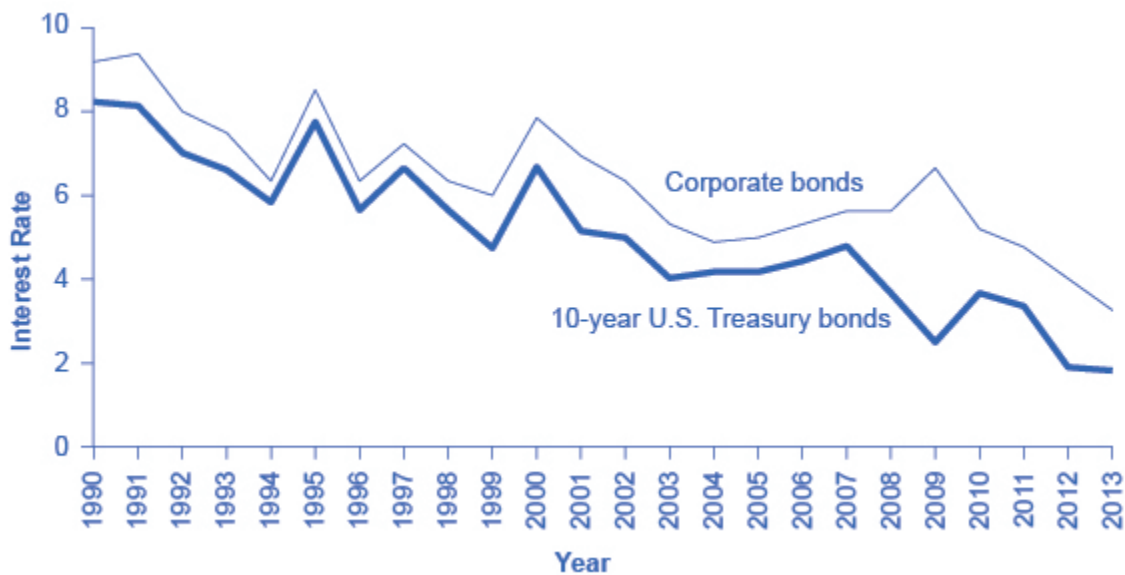


Figure 3. Interest Rates for Corporate Bonds and Ten-Year U.S. Treasury Bonds. The interest rates for corporate bonds and U.S. Treasury bonds (officially “notes”) rise and fall together, depending on conditions for borrowers and lenders in financial markets for borrowing. The corporate bonds always pay a higher interest rate, to make up for the higher risk they have of defaulting compared with the U.S. government.

The bottom line for bonds: rate of return—low to moderate, depending on the risk of the borrower; risk—low to moderate, depending on whether interest rates in the economy change substantially after the bond is issued; liquidity—moderate, because the bond needs to be sold before the investor regains the cash.

STOCKS

As stated earlier, the rate of return on a financial investment in a share of stock can come in two forms: as dividends paid by the firm and as a capital gain achieved by selling the stock for more than you paid. The range of possible returns from buying stock is mind-bending. Firms can decide to pay

dividends or not. A stock price can rise to a multiple of its original price or sink all the way to zero. Even in short periods of time, well-established companies can see large movements in the price of their stock. For example, in July 1, 2011, Netflix stock peaked at \$295 per share; one year later, on July 30, 2012, it was at \$53.91 per share; in 2015, it had recovered to \$414. When Facebook went public, its shares of stock sold for around \$40 per share, but in 2015, they were selling for slightly over \$83.

The reasons why stock prices fall and rise so abruptly will be discussed below, but first you need to know how we measure stock market performance. There are a number of different ways of measuring the overall performance of the stock market, based on averaging the stock prices of different subsets of companies. Perhaps the best-known measure of the stock markets is the Dow Jones Industrial Average, which is based on the stock prices of 30 large U.S. companies. Another gauge of stock market performance, the Standard & Poor's 500, follows the stock prices of the 500 largest U.S. companies. The Wilshire 5000 tracks the stock prices of essentially all U.S. companies that have stock the public can buy and sell.

Other measures of stock markets focus on where stocks are traded. For example, the New York Stock Exchange monitors the performance of stocks that are traded on that exchange in New York City. The Nasdaq stock market includes about 3,600 stocks, with a concentration of technology stocks. Table 1 lists some of the most commonly cited measures of U.S. and international stock markets.

Measure of the Stock Market	Comments
Dow Jones Industrial Average (DJIA): http://indexes.dowjones.com	Based on 30 large companies from a diverse set of representative industries, chosen by analysts at Dow Jones and Company. The index was started in 1896.
Standard & Poor's 500: http://www.standardandpoors.com	Based on 500 large U.S. firms, chosen by analysts at Standard & Poor's to represent the economy as a whole.
Wilshire 5000: http://www.wilshire.com	Includes essentially all U.S. companies with stock ownership. Despite the name, this index includes about 7,000 firms.
New York Stock Exchange: http://www.nyse.com	The oldest and largest U.S. stock market, dating back to 1792. It trades stocks for 2,800 companies of all sizes. It is located at 18 Broad St. in New York City.
NASDAQ: http://www.nasdaq.com	Founded in 1971 as an electronic stock market, allowing people to buy or sell from many physical locations. It has about 3,600 companies.
FTSE: http://www.ftse.com	Includes the 100 largest companies on the London Stock Exchange. Pronounced "footsie." Originally stood for Financial Times Stock Exchange.
Nikkei: http://www.nikkei.co.jp/nikkeiinfo/en/	Nikkei stands for <i>Nihon Keizai Shimbun</i> , which translates as the Japan Economic Journal, a major business newspaper in Japan. Index includes the 225 largest and most actively traded stocks on the Tokyo Stock Exchange.
DAX: http://www.exchange.de	Tracks 30 of the largest companies on the Frankfurt, Germany, stock exchange. DAX is an abbreviation for <i>Deutscher Aktien Index</i> .

Table 1. Some Measures of Stock Markets

The trend in the stock market is generally up over time, but with some large dips along the way. Figure 4 shows the path of the **Standard & Poor's 500 index** (which is measured on the left-hand vertical axis) and the **Dow Jones Index** (which is measured on the right-hand vertical axis). Broad measures of the stock market, like the ones listed here, tend to move together. The S&P 500 Index is the weighted average market capitalization of the firms selected to be in the index. The Dow Jones Industrial Average is the price weighted average of 30 industrial stocks tracked on the New York Stock Exchange.

When the Dow Jones average rises from 5,000 to 10,000, you know that the average price of the stocks

in that index has roughly doubled. Figure 4 shows that stock prices did not rise much in the 1970s, but then started a steady climb in the 1980s. From 2000 to 2013, stock prices bounced up and down, but ended up at about the same level.

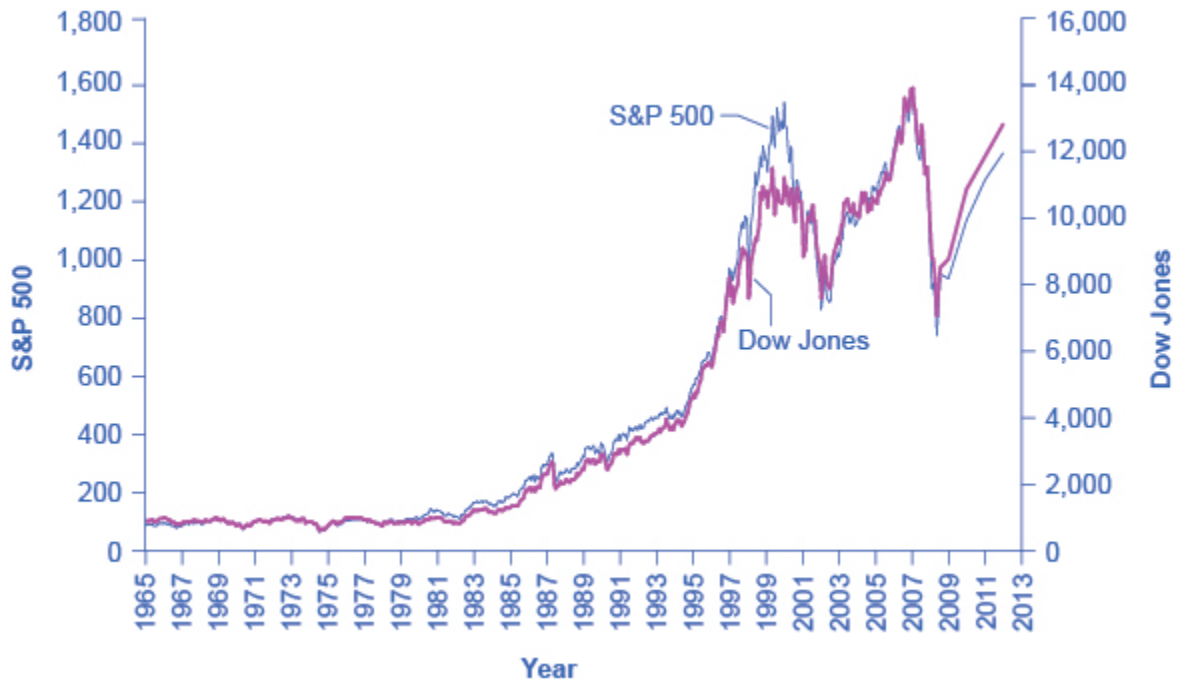


Figure 4. The Dow Jones Industrial Index and the Standard & Poor's 500, 1965–2013. Stock prices rose dramatically from the 1980s up to about 2000. From 2000 to 2013, stock prices bounced up and down, but ended up at about the same level.

Table 2 shows the total annual rate of return an investor would have received from buying the stocks in the S&P 500 index over recent decades. The total return here includes both dividends paid by these companies and also capital gains arising from increases in the value of the stock. (For technical reasons related to how the numbers are calculated, the dividends and capital gains do not add exactly to the total return.) From the 1950s to the 1980s, the average firm paid annual dividends equal to about 4% of the value of its stock. Since the 1990s, dividends have dropped and now often provide a return closer to 1% to 2%. In the 1960s and 1970s, the gap between percent earned on capital gains and dividends was much closer than it has been since the 1980s. In the 1980s and 1990s, however, capital gains were far higher than dividends. In the 2000s, dividends remained low and, while stock prices fluctuated, they ended the decade roughly where they had started.

Period	Total Annual Return	Capital Gains	Dividends
1950–1959	19.25%	13.58%	4.99%
1960–1969	7.78%	4.39%	3.25%
1970–1979	5.88%	1.60%	4.20%
1980–1989	17.55%	12.59%	4.40%
1990–1999	18.21%	15.31%	2.51%
2000–2009	–1.00%	–2.70%	1.70%
2010	15.06%	13.22%	1.84%
2011	2.11%	0.04%	2.07%
2012	16.00%	13.87%	2.13%

Table 2. Annual Returns on S&P 500 Stocks, 1950–2012

The overall pattern is that stocks as a group have provided a high rate of return over extended periods of time, but this return comes with risks. The market value of individual companies can rise and fall substantially, both over short time periods and over the long run. During extended periods of time like the 1970s or the first decade of the 2000s, the overall return on the stock market can be quite modest. The stock market can sometimes fall sharply, as it did in 2008.

The bottom line on investing in stocks is that the rate of return over time will be high, but the risks are also high, especially in the short run; liquidity is also high since stock in publicly held companies can be readily sold for spendable money.

MUTUAL FUNDS

Buying stocks or bonds issued by a single company is always somewhat risky. An individual firm may find itself buffeted by unfavorable supply and demand conditions or hurt by unlucky or unwise managerial decisions. Thus, a standard recommendation from financial investors is **diversification**, which means buying stocks or bonds from a wide range of companies. A saver who diversifies is following the old proverb: “Don’t put all your eggs in one basket.” In any broad group of companies, some firms will do better than expected and some will do worse—but the extremes have a tendency to cancel out extreme increases and decreases in value.

Purchasing a diversified group of the stocks or bonds has gotten easier in the Internet age, but it remains something of a task. To simplify the process, companies offer **mutual funds**, which are organizations that buy a range of stocks or bonds from different companies. The financial investor buys shares of the mutual fund, and then receives a return based on how the fund as a whole performs. In 2012, according to the Investment Company Factbook, about 44% of U.S. households had a financial investment in a mutual fund—including many people who have their retirement savings or pension money invested in this way.

Mutual funds can be focused in certain areas: one mutual fund might invest only in stocks of companies based in Indonesia, or only in bonds issued by large manufacturing companies, or only in stock of biotechnology companies. At the other end of the spectrum, a mutual fund might be quite broad; at the extreme, some mutual funds own a tiny share of every firm in the stock market, and thus the value of the mutual fund will fluctuate with the average of the overall stock market. A mutual fund that seeks only to mimic the overall performance of the market is called an **index fund**.

Diversification can offset some of the risks of individual stocks rising or falling. Even investors who buy an indexed mutual fund designed to mimic some measure of the broad stock market, like the Standard & Poor's 500, had better buckle their seatbelts against some ups and downs, like those the stock market experienced in the first decade of the 2000s. In 2008 average U.S. stock funds declined 38%, reducing the wealth of individuals and households. This steep drop in value hit hardest those who were close to retirement and were counting on their stock funds to supplement retirement income.

The bottom line on investing in mutual funds is that the rate of return over time will be high; the risks are also high, but the risks and returns for an individual mutual fund will be lower than those for an individual stock. As with stocks, liquidity is also high provided the mutual fund or stock index fund is readily traded.

HOUSING AND OTHER TANGIBLE ASSETS

Households can also seek a rate of return by purchasing tangible assets, especially housing. About two-thirds of U.S. households own their own home. An owner's **equity** in a house is the monetary value the owner would have after selling the house and repaying any outstanding bank loans used to buy the house. For example, imagine that you buy a house for \$200,000, paying 10% of the price as a down payment and taking out a bank loan for the remaining \$180,000. Over time, you pay off some of your bank loan, so that only \$100,000 remains, and the value of the house on the market rises to \$250,000. At that point, your equity in the home is the value of the home minus the value of the loan outstanding, which is \$150,000. For many middle-class Americans, home equity is their single greatest financial asset. The total value of all home equity held by U.S. households was \$11.3 trillion at the end of 2015, according to Federal Reserve Data.

Investment in a house is tangibly different from bank accounts, stocks, and bonds because a house offers both a financial and a nonfinancial return. If you buy a house to live in, part of the return on your investment occurs from your consumption of "housing services"—that is, having a place to live. (Of course, if you buy a home and rent it out, you receive rental payments for the housing services you provide, which would offer a financial return.) Buying a house to live in also offers the possibility of a capital gain from selling the house in the future for more than you paid for it. There can, however, be different outcomes, as the Clear It Up on the housing market shows.

Housing prices have usually risen steadily over time; for example, the median sales price for an existing one-family home was \$122,900 in 1990, but \$294,000 in 2015. Over these 23 years, home prices increased an average of 3.1% per year, which is an average financial return over this time. Figure 5 shows U.S. Census data for the median average sales price of a house in the United States over this time period.

Go to this website to experiment with a compound annual growth rate calculator.



However, the possible capital gains from rising housing prices are riskier than these national price averages. Certain regions of the country or metropolitan areas have seen drops in housing prices over time. The median housing price for the United States as a whole fell almost 7% in 2008 and again in 2009, dropping the median price from \$247,900 to \$216,700. As of 2015, home values had almost recovered to their pre-recession levels.

Visit this website to watch the trailer for *Inside Job*, a movie that explores the modern financial crisis.



Investors can also put money into other tangible assets such as gold, silver, and other precious metals, or in duller commodities like sugar, cocoa, coffee, orange juice, oil, and natural gas. The return on these investments derives from the saver's hope of buying low, selling high, and receiving a capital gain. Investing in, say, gold or coffee offers relatively little in the way of nonfinancial benefits to the user (unless the investor likes to caress gold or gaze upon a warehouse full of coffee). Indeed, typically investors in these commodities never even see the physical goods; instead, they sign a contract that takes ownership of a certain quantity of these commodities, which are stored in a warehouse, and later they sell the ownership to someone else. As one example, from 1981 to 2005, the price of gold generally fluctuated between about \$300 and \$500 per ounce, but then rose sharply to over \$1,100 per ounce by early 2010.

A final area of tangible assets are “collectibles” like paintings, fine wine, jewelry, antiques, or even baseball cards. Most collectibles provide returns both in the form of services or of a potentially higher selling price in the future. You can use paintings by hanging them on the wall; jewelry by wearing it; baseball cards by displaying them. You can also hope to sell them someday for more than you paid for them. However, the evidence on prices of collectibles, while scanty, is that while they may go through periods where prices skyrocket for a time, you should not expect to make a higher-than-average rate of return over a sustained period of time from investing in this way.

The bottom line on investing in tangible assets: rate of return—moderate, especially if you can receive

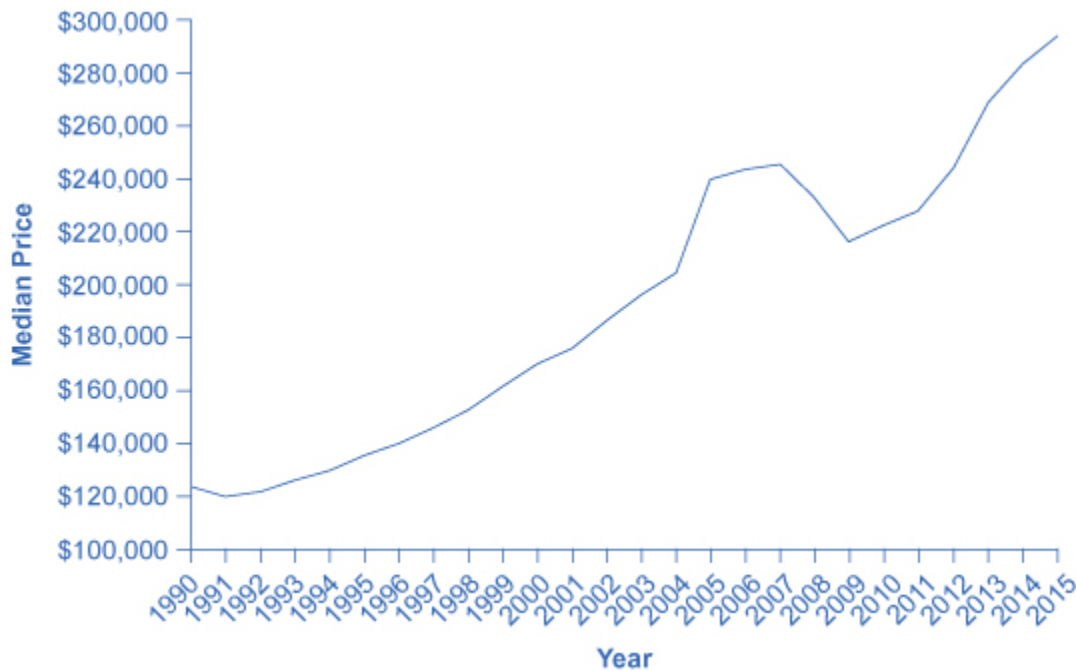


Figure 5. The Median Average Sales Price for New Single-Family Homes, 1990–2015. The median price is the price where half of sales prices are higher and half are lower. The median sales price for an new one-family home was \$122,900 in 1990. It rose as high as \$248,000 in 2007, before falling to \$232,000 in 2008. In 2015, the median sales price was \$294,000. Of course, this national figure conceals many local differences, like the areas where housing prices are higher or lower, or how housing prices have risen or fallen at certain times. (Source: U.S. Census)

nonfinancial benefits from, for example, living in the house; risk—moderate for housing or high if you buy gold or baseball cards; liquidity—low, because it often takes considerable time and energy to sell a house or a piece of fine art and turn your capital gain into cash. The next Clear It Up feature explains the issues in the recent U.S. housing market crisis.

WHAT WAS ALL THE COMMOTION IN THE RECENT U.S. HOUSING MARKET?

The cumulative average growth rate in housing prices from 1981 to 2000 was 5.1%. The price of an average U.S. home then took off from 2003 to 2005, rising at more than 10% per year. No serious analyst believed this rate of growth was sustainable; after all, if housing prices grew at, say, 11% per year over time, the average price of a home would more than double every seven years. However, at the time many serious analysts saw no reason for deep concern. After all, housing prices often change in fits and starts, like all prices, and a price surge for a few years is often followed by prices that are flat or even declining a bit as local markets adjust.

The sharp rise in housing prices was driven by a high level of demand for housing. Interest rates were low, so people were encouraged to borrow money to buy a house. Banks became much more flexible in their lending, making what were called “subprime” loans. Banks loaned money with low, or sometimes no, down payment. They offered loans with very low payments for the first two years, but then much higher payments after that; the idea was that housing prices would keep rising, so the borrower would just refinance the mortgage two years in the future, and thus would not ever have to make the higher payments. Some banks even offered so-called NINJA loans, which meant a loan given even though the borrower had No Income, No Job or Assets.

In retrospect, these loans seem nearly crazy. Many borrowers figured, however, that as long as housing prices kept rising, it made sense to buy. Many lenders used a process called “securitizing,” in which they sold their mortgages to financial companies, which put all the mortgages into a big pool, creating large financial securities, and then re-sold these mortgage-backed securities to investors. In this way, the lenders off-loaded the risks of the mortgages to investors. Investors were interested in mortgage-backed securities as they appeared to offer a steady stream of income, provided the mortgages were repaid. Investors relied on the ratings agencies to assess the credit risk associated with the mortgage backed securities. In hindsight, it appears that the credit agencies were far too lenient in their ratings of many of the securitized loans. Bank and financial regulators watched the steady rise in the market for mortgage-backed securities, but saw no reason at the time to intervene.

When housing prices turned down, many households that had borrowed when prices were high found that what they owed the bank was more than their home was worth. Many banks believed that they had diversified by selling their individual loans and instead buying securities based on mortgage loans from all over the country. After all, banks thought back in 2005, the average price of a house had not declined at any time since the Great Depression of the 1930s. These securities based on mortgage loans, however, turned out to be far riskier than expected. The bust in housing prices weakened the finances of both banks and households, and thus helped bring on the Great Recession of 2008–2009.

THE TRADEOFFS BETWEEN RETURN AND RISK

The discussion of financial investments has emphasized the expected rate of return, the risk, and the liquidity of each investment. Table 3 summarizes these characteristics.

Financial Investment	Return	Risk	Liquidity
Checking account	Very low	Very little	Very high
Savings account	Low	Very little	High
Certificate of deposit	Low to medium	Very little	Medium
Stocks	High	Medium to high	Medium
Bonds	Medium	Low to medium	Medium
Mutual funds	Medium to high	Medium to high	Medium to high
Housing	Medium	Medium	Low
Gold	Medium	High	Low
Collectibles	Low to medium	High	Low

Table 3. Key Characteristics for Financial Investments

The household investment choices listed here display a tradeoff between the expected return and the degree of risk involved. Bank accounts have very low risk and very low returns; bonds have higher risk but higher returns; and stocks are riskiest of all but have the potential for still higher returns. In effect, the higher average return compensates for the higher degree of risk. If risky assets like stocks did not also offer a higher average return, then few investors would want them.

This tradeoff between return and risk complicates the task of any financial investor: Is it better to invest safely or to take a risk and go for the high return? Ultimately, choices about risk and return will be based on personal preferences. However, it is often useful to examine risk and return in the context of different time frames.

The high returns of stock market investments refer to a high average return that can be expected over

a period of several years or decades. The high risk of such investments refers to the fact that in shorter time frames, from months to a few years, the rate of return may fluctuate a great deal. Thus, a person near retirement age, who already owns a house, may prefer reduced risk and certainty about retirement income. For young workers, just starting to make a reasonably profitable living, it may make sense to put most of their savings for retirement in stocks. Stocks are risky in the short term, to be sure, but when the worker can look forward to several decades during which stock market ups and downs can even out, stocks will typically pay a much higher return over that extended period than will bonds or bank accounts. Thus, tradeoffs between risk and return must be considered in the context of where the investor is in life.

KEY CONCEPTS AND SUMMARY

All investments can be categorized according to three key characteristics: average expected return, degree of risk, and liquidity. To get a higher rate of return, an investor must typically accept either more risk or less liquidity. Banks are an example of a financial intermediary, an institution that operates to coordinate supply and demand in the financial capital market. Banks offer a range of accounts, including checking accounts, savings accounts, and certificates of deposit. Under the federal deposit insurance program, banks purchase insurance against the risk of a bank failure.

A typical bond promises the financial investor a series of payments over time, based on the interest rate at the time the bond is issued, and then repayment of what was borrowed. Bonds that offer a high rate of return but also a relatively high chance of defaulting on the payments are called high yield or junk bonds. The bond yield is the rate of return that a bond promises to pay at the time of purchase. Even when bonds make payments based on a fixed rate of interest, they are somewhat risky, because if interest rates rise for the economy as a whole, an investor who owns bonds issued at lower interest rates is now locked into the low rate and suffers a loss.

Changes in the price of a stock depend on changes in expectations about future profits. Investing in any individual firm is somewhat risky, so investors are wise to practice diversification, which means investing in a range of companies. A mutual fund purchases an array of stocks and/or bonds. An investor in the mutual fund then receives a return depending on the overall performance of the investments made by the fund as a whole. A mutual fund that seeks to imitate the overall behavior of the stock market is called an index fund.

Housing and other tangible assets can also be regarded as forms of financial investment, which pay a rate of return in the form of capital gains. Housing can also offer a nonfinancial return—specifically, you can live in it.

SELF-CHECK QUESTIONS

1. Calculate the equity each of these people has in his or her home:
 - a. Fred just bought a house for \$200,000 by putting 10% as a down payment and borrowing the rest from the bank.
 - b. Freda bought a house for \$150,000 in cash, but if she were to sell it now, it would sell for \$250,000.

- c. Frank bought a house for \$100,000. He put 20% down and borrowed the rest from the bank. However, the value of the house has now increased to \$160,000 and he has paid off \$20,000 of the bank loan.
2. Which has a higher average return over time: stocks, bonds, or a savings account? Explain your answer.
3. Investors sometimes fear that a high-risk investment is especially likely to have low returns. Is this fear true? Does a high risk mean the return must be low?

REVIEW QUESTIONS

1. Why are banks called “financial intermediaries”?
2. Name several different kinds of bank account. How are they different?
3. Why are bonds somewhat risky to buy, even though they make predetermined payments based on a fixed rate of interest?
4. Why should a financial investor care about diversification?
5. What is a mutual fund?
6. What is an index fund?
7. How is buying a house to live in a type of financial investment?
8. Why is it hard to forecast future movements in stock prices?

CRITICAL THINKING QUESTIONS

1. What are some reasons why the investment strategy of a 30-year-old might differ from the investment strategy of a 65-year-old?
2. Explain why a financial investor in stocks cannot earn high capital gains simply by buying companies with a demonstrated record of high profits.

PROBLEMS

1. Imagine that a \$10,000 ten-year bond was issued at an interest rate of 6%. You are thinking about buying this bond one year before the end of the ten years, but interest rates are now 9%.
 - a. Given the change in interest rates, would you expect to pay more or less than \$10,000 for the bond?
 - b. Calculate what you would actually be willing to pay for this bond.
2. Suppose Ford Motor Company issues a five year bond with a face value of \$5,000 that pays an annual coupon payment of \$150.

- a. What is the interest rate Ford is paying on the borrowed funds?
- b. Suppose the market interest rate rises from 3% to 4% a year after Ford issues the bonds. Will the value of the bond increase or decrease?

REFERENCES

Howley, Kathleen M. Bloomberg. "Home Value Highest Since '07 as U.S. Houses Make Cash." Last modified March 26, 2013. <http://www.bloomberg.com/news/2013-03-26/home-value-highest-since-07-as-u-s-houses-make-cash-mortgages.html>.

NASDAQ.com. 2015. "Facebook, Inc. Historical Stock Prices." Accessed March 28, 2015. <http://www.nasdaq.com/symbol/fb/historical>.

National Association of Realtors. 2015. "Existing-Home Sales: Latest News." Accessed April 1, 2015. <http://www.realtor.org/topics/existing-home-sales/data>.

PricewaterhouseCoopers LLP. 2015. "Annual Venture Capital Investment Tops \$48 Billion in 2014, Reaching Highest Level in Over a Decade, According to the Moneytree Report." Accessed April 1, 2015. <http://www.pwc.com/us/en/press-releases/2015/annual-venture-capital-investment-tops-48-billion.jhtml>.

Rooney, Ben. "Trading Program Sparked May 'Flash Crash.'" *CNN Money*. Last modified October 1, 2010. http://money.cnn.com/2010/10/01/markets/SEC_CFTC_flash_crash/index.htm.

Investment Company Institute. "2013 Investment Company Fact Book, Chapter 6: Characteristics of Mutual Fund Owners." http://icifactbook.org/fb_ch6.html.

GLOSSARY

actual rate of return the total rate of return, including capital gains and interest paid on an investment at the end of a period of time

bond yield the rate of return a bond is expected to pay at the time of purchase

certificate of deposit (CD) a mechanism for a saver to deposit funds at a bank and promise to leave them at the bank for a time, in exchange for a higher rate of interest

checking account a bank account that typically pays little or no interest, but that gives easy access to money, either by writing a check or by using a "debit card"

coupon rate the interest rate paid on a bond; can be annual or semi-annual

debit card a card that lets the person make purchases, and the cost is immediately deducted from that person's checking account

diversification investing in a wide range of companies to reduce the level of risk

equity the monetary value a homeowner would have after selling the house and repaying any outstanding bank loans used to buy the house

expected rate of return how much a project or an investment is expected to return to the investor, either in future interest payments, capital gains, or increased profitability

face value the amount that the bond issuer or borrower agrees to pay the investor

financial intermediary an institution, like a bank, that receives money from savers and provides funds to borrowers

high yield bonds bonds that offer relatively high interest rates to compensate for their relatively high chance of default

index fund a mutual fund that seeks only to mimic the overall performance of the market

junk bonds see high yield bonds

liquidity refers to how easily money or financial assets can be exchanged for a good or service

maturity date the date that a bond must be repaid

mutual funds funds that buy a range of stocks or bonds from different companies, thus allowing an investor an easy way to diversify

present value a bond's current price at a given time

risk a measure of the uncertainty of that project's profitability

savings account a bank account that pays an interest rate, but withdrawing money typically requires a trip to the bank or an automatic teller machine

SOLUTIONS

Answers to Self-Check Questions

1.
 - a. Remember, equity is the market value of the house minus what is still owed to the bank. Thus: the value of the house is \$200,000, Fred owes \$180,000 to the bank, and his equity is \$20,000.
 - b. The value of Freda's house is \$250,000. It does not matter what price she bought it for. She owes zero to the bank, so her equity is the whole \$250,000.
 - c. The value of Frank's house is \$160,000. He owes \$60,000 to the bank (the original \$80,000 minus the \$20,000 he has paid off the loan). His equity is \$100,000.
2. Over a sustained period of time, stocks have an average return higher than bonds, and bonds have an average return higher than a savings account. This is because in any given year the value of a savings account changes very little. In contrast, stock values can grow or decline by a very large amount (for example, the S&P 500 increased 26% in 2009 after declining 37% in 2008. The value of a bond, which depends largely on interest rate fluctuations, varies far less than a stock, but more than a savings account.
3. When people believe that a high-risk investment must have a low return, they are getting confused between what risk and return mean. Yes, a high-risk investment might have a low return, but it might also have a high return. Risk refers to the fact that a wide range of outcomes is possible. However, a high-risk investment must, on average, expect a relatively high return or else no one would be willing to take the risk. Thus, it is quite possible—even likely—for an investment to have high risk and high return. Indeed, the reason that an investment has a high expected return is that it also has a high risk.

23.3 HOW TO ACCUMULATE PERSONAL WEALTH

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the random walk theory
- Calculate simple and compound interest
- Evaluate how capital markets transform financial capital

Getting rich may seem straightforward enough. Figure out what companies are going to grow and earn high profits in the future, or figure out what companies are going to become popular for everyone else to buy. Those companies are the ones that will pay high dividends or whose stock price will climb in the future. Then, buy stock in those companies. Presto! Multiply your money!

Why is this path to riches not as easy as it sounds? This module first discusses the problems with picking stocks, and then discusses a more reliable but undeniably duller method of accumulating personal wealth.

WHY IT IS HARD TO GET RICH QUICK: THE RANDOM WALK THEORY

The chief problem with attempting to buy **stock** in companies that will have higher prices in the future is that many other financial investors are trying to do the same thing. Thus, in attempting to get rich in the stock market, it is no help to identify a company that is going to earn high profits if many other investors have already reached the same conclusion, because the stock price will already be high, based on the expected high level of future profits.

The idea that stock prices are based on expectations about the future has a powerful and unexpected implication. If expectations determine stock price, then shifts in expectations will determine shifts in the stock price. Thus, what matters for predicting whether the stock price of a company will do well is not whether the company will actually earn profits in the future. Instead, you must find a company that is widely believed at present to have poor prospects, but that will actually turn out to be a shining star. Brigades of stock market analysts and individual investors are carrying out such research 24 hours a day.

The fundamental problem with predicting future stock winners is that, by definition, no one can predict the future news that alters expectations about profits. Because stock prices will shift in response to unpredictable future news, these prices will tend to follow what mathematicians call a “random

walk with a trend.” The “random walk” part means that, on any given day, stock prices are just as likely to rise as to fall. “With a trend” means that over time, the upward steps tend to be larger than the downward steps, so stocks do gradually climb.

If stocks follow a random walk, then not even financial professionals will be able to choose those that will beat the average consistently. While some investment advisers are better than average in any given year, and some even succeed for a number of years in a row, the majority of financial investors do not outguess the market. If we look back over time, it is typically true that half or two-thirds of the mutual funds that attempted to pick stocks which would rise more than the market average actually ended up doing worse than the market average. For the average investor who reads the business pages of the newspaper over a cup of coffee in the morning, the odds of doing better than full-time professionals is not very good at all. Trying to pick the stocks that will gain a great deal in the future is a risky and unlikely way to become rich.

GETTING RICH THE SLOW, BORING WAY

Many U.S. citizens can accumulate a large amount of wealth during their lifetimes, if they make two key choices. The first is to complete additional education and training. In 2014, the U.S. Census Bureau reported median earnings for households where the main earner had only a high school degree of \$33,124; for those with a two-year associate degree, median earnings were \$40,560 and for those with a four-year bachelor’s degree, median income was \$54,340. Learning is not only good for you, but it pays off financially, too.

The second key choice is to start saving money early in life, and to give the power of compound interest a chance. Imagine that at age 25, you save \$3,000 and place that money into an account that you do not touch. In the long run, it is not unreasonable to assume a 7% real annual rate of return (that is, 7% above the rate of inflation) on money invested in a well-diversified stock portfolio. After 40 years, using the formula for compound interest, the original \$3,000 investment will have multiplied nearly fifteen fold:

$$3,000(1 + .07)^{40} = \$44,923$$

Having \$45,000 does not make you a millionaire. Notice, however, that this tidy sum is the result of saving \$3,000 exactly once. Saving that amount every year for several decades—or saving more as income rises—will multiply the total considerably. This type of wealth will not rival the riches of Microsoft CEO Bill Gates, but remember that only half of Americans have any money in mutual funds at all. Accumulating hundreds of thousands of dollars by retirement is a perfectly achievable goal for a well-educated person who starts saving early in life—and that amount of accumulated wealth will put you at or near the top 10% of all American households. The following Work It Out feature shows the difference between simple and compound interest, and the power of compound interest.

SIMPLE AND COMPOUND INTEREST

Simple interest is an interest rate calculation only on the principal amount.

Step 1. Learn the formula for simple interest:

$$\textit{Principal} \times \textit{Rate} \times \textit{Time} = \textit{Interest}$$

Step 2. Practice using the simple interest formula.

Example 1: \$100 Deposit at a simple interest rate of 5% held for one year is:

$$\$100 \times 0.05 \times 1 = \$5$$

Simple interest in this example is \$5.

Example 2: \$100 Deposit at a simple interest rate of 5% held for three years is:

$$\$100 \times 0.05 \times 3 = \$15$$

Simple interest in this example is \$15.

Step 3. Calculate the total future amount using this formula:

$$Total\ future\ amount = principal + interest$$

Step 4. Put the two simple interest formulas together.

$$Total\ future\ amount\ (with\ simple\ interest) = Principal + (Principal \times Rate \times Time)$$

Step 5. Apply the simple interest formula to our three year example.

$$Total\ future\ amount\ (with\ simple\ interest) = \$100 + (\$100 \times 0.05 \times 3) = \$115$$

Compound interest is an interest rate calculation on the principal plus the accumulated interest.

Step 6. To find the compound interest, we determine the difference between the future value and the present value of the principal. This is accomplished as follows:

$$Future\ Value - Principal \times (1 + interest\ rate)^{time}$$

$$Compound\ interest = Future\ Value - Present\ Value$$

Step 7. Apply this formula to our three-year scenario. Follow the calculations in Table 4.

Year 1	
Amount in Bank	\$100
Bank Interest Rate	5%
Total	\$105
	$\$100 + (\$100 \times 0.05)$
Year 2	
Amount in Bank	\$105
Bank Interest Rate	5%
Total	\$110.25
	$\$105 + (\$105 \times .05)$
Year 3	
Amount in Bank	\$110.25
Bank Interest Rate	5%
Total	\$115.75
	$\$110.25 + (\$110.25 \times .05)$
Compound interest	$\$115.75 - \$100 = \$15.75$

Table 4.

Step 8. Note that, after three years, the total is \$115.75. Therefore the total compound interest is \$15.75. This is \$0.75 more than was obtained with simple interest. While this may not seem like much, keep in mind that we were only working with \$100 and over a relatively short time period. Compound interest can make a huge difference with larger sums of money and over longer periods of time.

Getting additional education and saving money early in life obviously will not make you rich overnight. Additional education typically means putting off earning income and living as a student for more years. Saving money often requires choices like driving an older or less expensive car, living in a smaller apartment or buying a smaller house, and making other day-to-day sacrifices. For most people, the tradeoffs for achieving substantial personal wealth will require effort, patience, and sacrifice.

HOW CAPITAL MARKETS TRANSFORM FINANCIAL FLOWS

Financial capital markets have the power to repackage money as it moves from those who supply financial capital to those who demand it. Banks accept checking account deposits and turn them into long-term loans to companies. Individual firms sell shares of stock and issue bonds to raise capital. Firms make and sell an astonishing array of goods and services, but an investor can receive a return on the company's decisions by buying stock in that company. Stocks and bonds are sold and resold by financial investors to one another. Venture capitalists and angel investors search for promising small companies. Mutual funds combine the stocks and bonds—and thus, indirectly, the products and investments—of many different companies.

Visit this website to read an article about how austerity can work.



In this chapter, we discussed the basic mechanisms of financial markets. (A more advanced course in economics or finance will consider more sophisticated tools.) The fundamentals of those financial capital markets remain the same: Firms are trying to raise financial capital and households are looking for a desirable combination of rate of return, risk, and liquidity. Financial markets are society's mechanisms for bringing together these forces of demand and supply.

THE HOUSING BUBBLE AND THE FINANCIAL CRISIS OF 2007

The housing boom and bust in the United States, and the resulting multi-trillion-dollar decline in home equity, started with the fall of home prices starting in 2007. As home values fell, many home prices fell below the amount owed on the mortgage and owners stopped paying and defaulted on their loan. Banks found that their assets (loans) became worthless. Many financial institutions around the world had invested in mortgage-backed securities, or had purchased insurance on mortgage-backed securities. When housing prices collapsed, the value of those financial assets collapsed as well. The asset side of the banks' balance sheets dropped, causing bank failures and bank runs. Around the globe, financial institutions were bankrupted or nearly so. The result was a large decrease in lending and borrowing, referred to as a freezing up of available credit. When credit dries up, the economy is on its knees. The crisis was not limited to the United States. Iceland, Ireland, the United Kingdom, Spain, Portugal, and Greece all had similar housing boom and bust cycles, and similar credit freezes.

If businesses cannot access financial capital, they cannot make physical capital investments. Those investments ultimately lead to job creation. So when credit dried up, businesses invested less, and they ultimately laid off millions of workers. This caused incomes to drop, which caused demand to drop. In turn businesses sold less, so they laid off more workers. Compounding these events, as economic conditions worsened, financial institutions were even less likely to make loans.

To make matters even worse, as businesses sold less, their expected future profit decreased, and this led to a drop in stock prices. Combining all these effects led to major decreases in incomes, demand, consumption, and employment, and to the Great Recession, which in the United States officially lasted from December 2007 to June 2009. During this time, the unemployment rate rose from 5% to a peak of 10.1%. Four years after the recession officially ended, unemployment was still stubbornly high, at 7.6%, and 11.8 million people were still unemployed.

As the world's leading consumer, if the United States goes into recession, it usually drags other countries down with it. The Great Recession was no exception. With few exceptions, U.S. trading partners also entered into recessions of their own, of varying lengths, or suffered slower economic growth. Like the United States, many European countries also gave direct financial assistance, so-called bailouts, to the institutions that make up their financial markets. There was good reason to do this. Financial markets bridge the gap between demanders and suppliers of financial capital. These institutions and markets need to function in order for an economy to invest in new financial capital.

However, much of this bailout money was borrowed, and this borrowed money contributed to another crisis in Europe. Because of the impact on their budgets of the financial crisis and the resulting bailouts, many countries found themselves with unsustainably high deficits. They chose to undertake austerity measures, large decreases in government spending and large tax increases, in order to reduce their deficits. Greece, Ireland, Spain, and Portugal have all had to undertake relatively severe austerity measures. The ramifications of this crisis have spread; the viability of the euro has even been called into question.

KEY CONCEPTS AND SUMMARY

It is extremely difficult, even for financial professionals, to predict changes in future expectations and thus to choose the stocks whose price is going to rise in the future. Most Americans can accumulate considerable financial wealth if they follow two rules: complete significant additional education and training after graduating from high school and start saving money early in life.

SELF-CHECK QUESTIONS

1. What is the total amount of interest collected from a \$5,000 loan after three years with a simple interest rate of 6%?
2. If you receive \$500 in simple interest on a loan that you made for \$10,000 for 5 years, what was the interest rate you charged?
3. You open a 5-year CD for \$1,000 that pays 2% interest, compounded annually. What is the value of that CD at the end of the 5 years?

REVIEW QUESTIONS

1. What are the two key choices U.S. citizens need to make that determines their relative wealth?
2. Is investing in housing always a very safe investment?

CRITICAL THINKING QUESTIONS

1. Explain what happens in an economy when the financial markets limit access to capital. How does this affect economic growth and employment?
2. You and your friend have opened an account on E-Trade and have each decided to select five similar companies in which to invest. You are diligent in monitoring your selections, tracking prices, current events, and actions taken by the company. Your friend chooses his companies randomly, pays no attention to the financial news, and spends his leisure time focused on everything besides his investments. Explain what might be the performance for each of your portfolios at the end of the year.
3. How do bank failures cause the economy to go into recession?

PROBLEMS

1. How much money do you have to put into a bank account that pays 10% interest compounded annually to have \$10,000 in ten years?
2. Many retirement funds charge an administrative fee each year equal to 0.25% on managed assets. Suppose that Alexx and Spenser each invest \$5,000 in the same stock this year. Alexx invests directly and earns 5% a year. Spenser uses a retirement fund and earns 4.75%. After 30 years, how much more will Alexx have than Spenser?

REFERENCES

U.S. Department of Commerce: United States Census Bureau. "Income: Table H-13. Educational Attainment of Householder—Households with Householder 25 Years Old and Over by Median and Mean Income." <http://www.census.gov/hhes/www/income/data/historical/household/>.

United States Department of Labor. Bureau of Labor Statistics. 2015. "Table 9. Quartiles and Selected Deciles of Usual Weekly Earnings of Full-Time Wage and Salary Workers by Selected Characteristics, 2014 Annual Averages." Accessed April 1, 2015. <http://www.bls.gov/news.release/wkyeng.t09.htm>.

GLOSSARY

compound interest an interest rate calculation on the principal plus the accumulated interest

simple interest an interest rate calculation only on the principal amount

SOLUTIONS

Answers to Self-Check Questions

$$\begin{aligned}
 1. \quad & \text{Principal} + (\text{principal} \times \text{rate} \times \text{time}) \\
 & \$5,000 + (\$5,000 \times 0.06 \times 3) = \$5,900
 \end{aligned}$$

$$2. \quad \begin{array}{l} \text{Principal} + (\text{principal} \times \text{rate} \times \text{time}) \\ \text{Interest} \quad \text{Principal} \times \text{rate} \times \text{time} \end{array}$$

$$\$500 \quad \$10,000 \times \text{rate} \times 5 \text{ years}$$

$$\$500 \quad \$50,000 \times \text{rate}$$

$$\$500/\$50,000 \quad \text{rate}$$

$$\text{Rate} \quad 1\%$$

$$\text{Principal}(1 + \text{interest rate})^{\text{time}} \quad \$1,000(1 + 0.02)^5$$

$$3. \quad \$1,104.08$$

CHAPTER 24. PUBLIC ECONOMY

INTRODUCTION TO PUBLIC ECONOMY

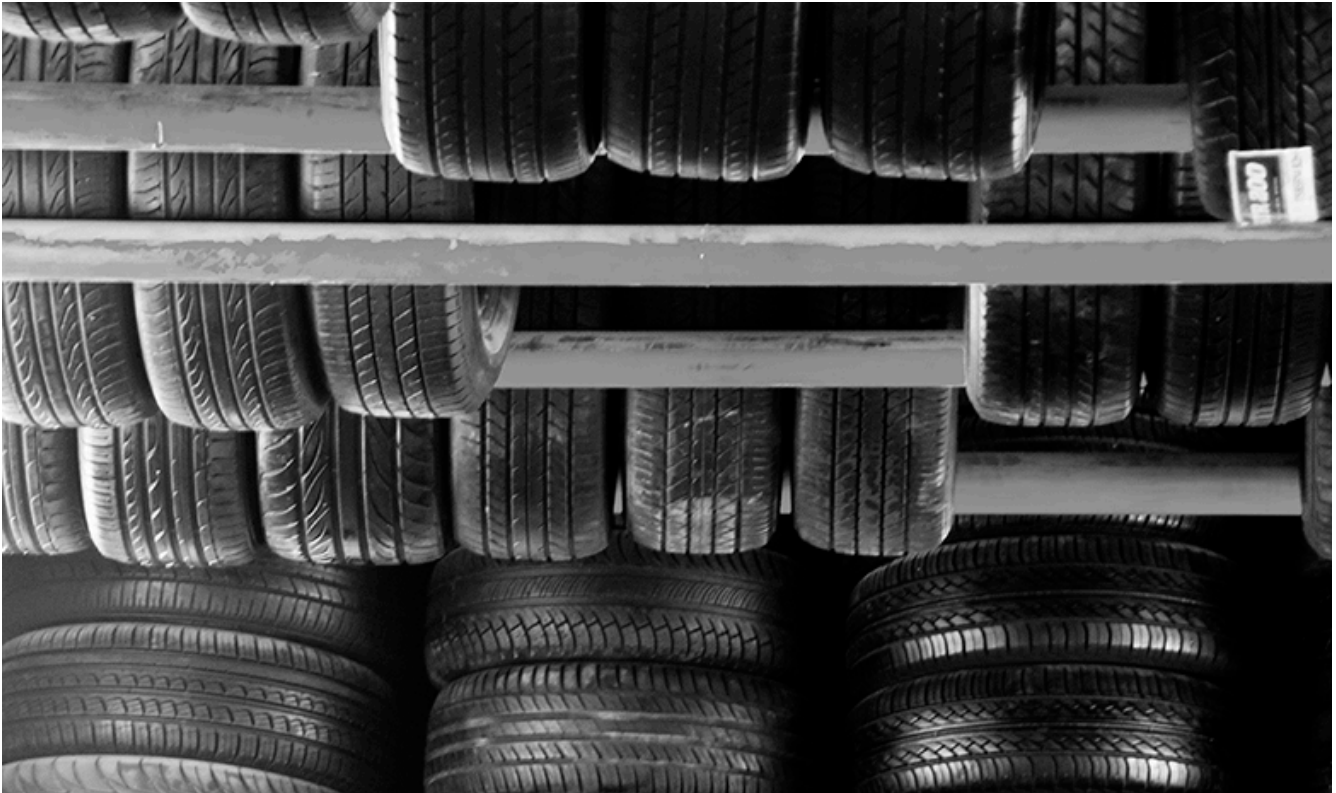


Figure 1. Domestic Tires? While these tires may all appear similar, some are made in the United States and others are not. Those that are not could be subject to a tariff that could cause the cost of all tires to be higher.
(Credit: modification of work by Jayme del Rosario/Flickr Creative Commons)

CHINESE TIRE TARIFFS

Do you know where the tires on your car are made? If they were imported, they may be subject to a tariff (a tax on imported goods) that could raise the price of your car. What do you think about that tariff? Would you write to your representative or your senator about it? Would you start a Facebook or Twitter campaign?

Most people are unlikely to fight this kind of tax or even inform themselves about the issue in the first place. In *The Logic of Collective Action* (1965), economist Mancur Olson challenged the popular idea that, in a democracy, the majority view will prevail, and in doing so launched the modern study of public economy, sometimes referred to as public choice, a subtopic of microeconomics. In this chapter, we will look at the economics of government policy, why smaller, more organized groups have an incentive to work hard to get certain policies enacted, and why lawmakers ultimately make decisions that may result in bad economic policy.

CHAPTER OBJECTIVES

Introduction to Public Economy

In this chapter, you will learn about:

- Voter Participation and Costs of Elections
- Special Interest Politics
- Flaws in the Democratic System of Government

As President Abraham Lincoln famously said in his 1863 *Gettysburg Address*, democratic governments are supposed to be “of the people, by the people, and for the people.” Can we rely on democratic governments to enact sensible economic policies? After all, they react to voters, not to analyses of demand and supply curves. The main focus of an economics course is, naturally enough, to analyze the characteristics of markets and purely economic institutions. But political institutions also play a role in allocating society’s scarce resources, and economists have played an active role, along with other social scientists, in analyzing how such political institutions work.

Other chapters of this book discuss situations in which market forces can sometimes lead to undesirable results: monopoly, imperfect competition, and antitrust policy; negative and positive externalities; poverty and inequality of incomes; failures to provide insurance; and financial markets that may go from boom to bust. Many of these chapters suggest that government economic policies could be aimed at addressing these issues.

However, just as markets can face issues and problems that lead to undesirable outcomes, a democratic system of government can also make mistakes, either by enacting policies that do not benefit society as a whole or by failing to enact policies that would have benefited society as a whole. This chapter discusses some practical difficulties of democracy from an economic point of view: the actors in the political system are presumed to follow their own self-interest, which is not necessarily the same as the public good. For example, many of those who are eligible to vote do not, which obviously raises questions about whether a democratic system will reflect everyone’s interests. Benefits or costs of government action are sometimes concentrated on small groups, which in some cases may organize and have a disproportionately large impact on politics and in other cases may fail to organize and end up neglected. A legislator who worries about support from voters in his or her district may focus on spending projects specific to the district without sufficient concern for whether this spending is in the interest of the nation.

When more than two choices exist, the principle that the majority of voters should decide may not always make logical sense, because situations can arise where it becomes literally impossible to decide what the “majority” prefers. Government may also be slower than private firms to correct its mistakes, because government agencies do not face competition or the threat of new entry.

24.1 VOTER PARTICIPATION AND COSTS OF ELECTIONS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the significance of rational ignorance
- Evaluate the impact of election expenses

In U.S. presidential elections over the last few decades, about 55% to 65% of voting-age citizens actually voted, according to the U.S. Census. In congressional elections when there is no presidential race, or in local elections, the turnout is typically lower, often less than half the eligible voters. In other countries, the share of adults who vote is often higher. For example, in national elections since the 1980s in Germany, Spain, and France, about 75% to 80% of those of voting age cast ballots. Even this total falls well short of 100%. Some countries have laws that require voting, among them Australia, Belgium, Italy, Greece, Turkey, Singapore, and most Latin American nations. At the time the United States was founded, voting was mandatory in Virginia, Maryland, Delaware, and Georgia. Even if the law can require people to vote, however, no law can require that each voter cast an informed or a thoughtful vote. Moreover, in the United States and in most countries around the world, the freedom to vote has also typically meant the freedom *not* to vote.

Why do people not vote? Perhaps they do not care too much about who wins, or they are uninformed about who is running, or they do not believe their vote will matter or change their lives in any way. Indeed, these reasons are probably tied together, since people who do not believe their vote matters will not bother to become informed or care who wins. Economists have suggested why a **utility-maximizing** person might rationally decide not to vote or not to become informed about the election. While a few elections in very small towns may be decided by a single vote, in most elections of any size, the margin of victory is measured in hundreds, thousands, or even millions of votes. A rational voter will recognize that one vote is extremely unlikely to make a difference. This theory of **rational ignorance** holds that people will not vote if the costs of becoming informed and voting are too high, or they feel their vote will not be decisive in the election.

In a 1957 work, *An Economic Theory of Democracy*, the economist **Anthony Downs** stated the problem this way: “It seems probable that for a great many citizens in a democracy, rational behavior excludes any investment whatever in political information per se. No matter how significant a difference between parties is revealed to the rational citizen by his free information, or how uncertain he is about which party to support, he realizes that his vote has almost no chance of influencing the outcome... He will not even utilize all the free information available, since assimilating it takes time.” In his classic

1948 novel *Walden Two*, the psychologist B. F. Skinner puts the issue even more succinctly via one of his characters, who states: “The chance that one man’s vote will decide the issue in a national election...is less than the chance that he will be killed on his way to the polls.” The following Clear It Up feature explores another aspect of the election process: spending.

HOW MUCH IS TOO MUCH TO SPEND ON AN ELECTION?

According to a report from *The New York Times*, the 2012 elections for president, Congress, and state and local offices, saw a total of about \$5.8 billion spent. The money raised went to the campaigns, including advertising, fundraising, travel, and staff. Many people worry that politicians spend too much time raising money and end up entangled with special interest groups that make major donations. Critics would prefer a system that restricts what candidates can spend, perhaps in exchange for limited public campaign financing or free television advertising time.

How much spending on campaigns is too much? Five billion dollars will buy a lot of potato chips, but in the U.S. economy, which exceeded \$16 trillion in 2012, the \$5.8 billion spent on political campaigns was about 1/25 of 1% of the overall economy. Here is another way to think about campaign spending. Total government spending programs in 2009, including federal and state governments, was about \$5.1 trillion, so the cost of choosing the people who would determine how this money would be spent was about 1/10 of 1% of that. In the context of the enormous U.S. economy, \$5.8 billion is not as much money as it sounds. U.S. consumers spend about \$2 billion per year on toothpaste and \$7 billion on hair care products. In 2008, Proctor and Gamble spent \$4.8 billion on advertising. It may not be sensible to believe the United States is going to decide its presidential elections for less than we spend on toothpaste or than Proctor and Gamble spends on advertisements.

Whatever we believe about whether candidates and their parties spend too much or too little on elections, the U.S. Supreme Court has placed limits on how government can limit campaign spending. In a 1976 decision, *Buckley v. Valeo*, the Supreme Court emphasized that the First Amendment to the U.S. Constitution specifies freedom of speech. The federal government and states can offer candidates a voluntary deal in which government makes some public financing available to candidates, but only if the candidates agree to abide by certain spending limits. Of course, candidates can also voluntarily agree to set certain spending limits if they wish. But government cannot forbid people or organizations to raise and spend money above these limits if they choose.

In 2002, Congress passed and President George W. Bush signed into law the **Bipartisan Campaign Reform Act (BCRA)**. The relatively noncontroversial portions of the act strengthen the rules requiring full and speedy disclosure of who contributes money to campaigns. However, some controversial portions of the Act limit the ability of individuals and groups to make certain kinds of political donations and they ban certain kinds of advertising in the months leading up to an election. These bans were called into question after the release of two films: Michael Moore’s *Fahrenheit 9/11* and Citizens United’s *Hillary: The Movie*. At question was whether each film sought to discredit political candidates for office too close to an election, in violation of the BCRA. Moore’s film was found by lower courts not to violate the Act, while Citizens United’s was. The fight reached the Supreme Court, as *Citizens United v. Federal Election Commission*, saying that the First Amendment protects the rights of corporations as well as individuals to donate to political campaigns. The Court ruled, in a 5–4 decision, that the spending limits were unconstitutional. This controversial decision, which essentially allows unlimited contributions by corporations to political action committees, overruled several previous decisions and will likely be revisited in the future, due to the strength of the public reaction. For now, it has resulted in a sharp increase in election spending.

While many U.S. adults do not bother to vote in presidential elections, more than half do. What motivates them? Research on voting behavior has shown that people who are more settled or more “connected” to society tend to vote more frequently. According to the *Washington Post*, more married people vote than single people. Those with a job vote more than the unemployed. Those who have lived longer in a neighborhood are more likely to vote than newcomers. Those who report that they

know their neighbors and talk to them are more likely to vote than socially isolated people. Those with a higher income and level of education are also more likely to vote. These factors suggest that politicians are likely to focus more on the interests of married, employed, well-educated people with at least a middle-class level of income than on the interests of other groups. For example, those who vote may tend to be more supportive of financial assistance for the two-year and four-year colleges they expect their children to attend than they are of medical care or public school education aimed at families of the poor and unemployed.

Visit this website to see a breakdown of how different groups voted in 2012.



A number of proposals have been offered to encourage greater voter turnout: making it easier to register to vote, keeping the polls open for more hours, or even moving Election Day to the weekend, when fewer people need to worry about jobs or school commitments. However, the changes that have been made do not seem to have caused a long-term upward trend in the number of people voting. After all, casting an informed vote will always impose some costs of time and energy. It is not clear how to strengthen people's feeling of connectedness to society in a way that will lead to a substantial increase in voter turnout. Without greater voter turnout, however, politicians elected by the votes of 60% or fewer of the population may not enact economic policy in the best interests of 100% of the population. Meanwhile, countering a long trend toward making voting easier, many states have recently erected new voting laws that critics say are actually barriers to voting. States have passed laws reducing early voting, restricting groups who are organizing get-out-the-vote efforts, enacted strict photo ID laws, as well as laws that require showing proof of U.S. citizenship. The ACLU argues that while these laws profess to prevent voter fraud, they are in effect making it harder for individuals to cast their vote.

KEY CONCEPTS AND SUMMARY

The theory of rational ignorance says voters will recognize that their single vote is extremely unlikely to influence the outcome of an election. As a consequence, they will choose to remain uninformed about issues and not vote. This theory helps explain why voter turnout is so low in the United States.

SELF-CHECK QUESTIONS

1. Based on the theory of rational ignorance, what should we expect to happen to voter turnout as the Internet makes information easier to obtain?
2. What is the cost of voting in an election?

REVIEW QUESTIONS

How does rational ignorance discourage voting?

CRITICAL THINKING QUESTIONS

1. What are some reasons people might find acquiring information about politics and voting rational, in contrast to rational ignorance theory?
2. What are some possible ways to encourage voter participation and overcome rational ignorance?
3. Given that rational ignorance discourages some people from becoming informed about elections, is it necessarily a good idea to encourage greater voter turnout? Why or why not?

REFERENCES

Olson, Mancur. *The Logic of Collective Action: Public Goods and the Theory of Groups*, (Boston: Harvard University Press, 1965).

Downs, Anthony. *An Economic Theory of Democracy*. New York: Harper, 1957.

Skinner, B.F. *Walden Two*. Indianapolis: Hackett Publishing Company, Inc., 1948.

American Civil Liberties Union (ACLU). "Voting Rights." <https://www.aclu.org/voting-rights>.

The New York Times. "The 2012 Money Race: Compare the Candidates." <http://elections.nytimes.com/2012/campaign-finance>.

The Washington Post. "Exit Polls 2012: Where Americans Stood this Election." Last modified November 7, 2012. <http://www.washingtonpost.com/wp-srv/special/politics/2012-exit-polls/national-breakdown/>.

Cornell University Law School: Legal Information Institute. "Citizens United v. Federal Election Committee (No. 08-205)." Last modified January 21, 2010. <http://www.law.cornell.edu/supct/html/08-205.ZS.html>.

U.S. Department of Commerce: United States Census Bureau. "Voting and Registration: Historical Time Series Tables; Table A-9: Reported Voting Rates in Presidential Election Years, by Selected Characteristics: November 1964 to 2012." <http://www.census.gov/hhes/www/socdemo/voting/publications/historical/index.html>

GLOSSARY

rational ignorance the theory that rational people will not vote if the costs of becoming informed and voting are too high or because they know their vote will not be decisive in the election

SOLUTIONS

Answers to Self-Check Questions

1. All other things being equal, voter turnout should increase as the cost of casting an informed vote decreases.
2. The cost in time of voting, transportation costs to and from the polling place, and any additional time and effort spent becoming informed about the candidates.

24.2 SPECIAL INTEREST POLITICS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how special interest groups and lobbyists can influence campaigns and elections
- Describe pork-barrel spending and logrolling

Many political issues are of intense interest to a relatively small group, as noted above. For example, many U.S. drivers do not much care where the tires for their car were made—they just want good quality as inexpensively as they can get it. In September 2009, President Obama and Congress enacted a **tariff** (taxes added on imported goods) on tires imported from China that would increase the import price of Chinese tires by 35 percent in its first year, 30 percent in its second year, and 25 percent in its third year. Interestingly, the U.S. companies that make tires did not favor this step, because most of them also import tires from China and other countries. (See Globalization and Protectionism for more on tariffs.) However, the United Steelworkers union, which had seen jobs in the tire industry fall by 5,000 over the previous five years, lobbied fiercely for the tariff to be enacted. With this tariff, the cost of all tires increased significantly. (See the closing Bring It Home feature at the end of this chapter for more information on the tire tariff.)

Special interest groups are groups that are small in number relative to the nation, but quite well organized and focused on a specific issue. A special interest group can pressure legislators to enact public policies that do not benefit society as a whole. Imagine an environmental rule to reduce air pollution that will cost 10 large companies \$8 million each, for a total cost of \$80 million. The social benefits from enacting this rule provide an average benefit of \$10 for every person in the United States, for a total of about \$3 trillion. Even though the benefits are far higher than the costs for society as a whole, the 10 companies are likely to lobby much more fiercely to avoid \$8 million in costs than the average person is to argue for \$10 worth of benefits.

As this example suggests, we can relate the problem of special interests in politics to an issue raised in Environmental Protection and Negative Externalities about economic policy with respect to **negative externalities** and pollution—the problem called **regulatory capture** (which we defined in Monopoly and Antitrust Policy). In legislative bodies and agencies that write laws and regulations about how much corporations will pay in taxes, or rules for safety in the workplace, or instructions on how to satisfy environmental regulations, you can be sure the specific industry affected has **lobbyists** who study every word and every comma. They talk with the legislators who are writing the legislation and suggest alternative wording. They contribute to the campaigns of legislators on the key commit-

tees—and may even offer those legislators high-paying jobs after they have left office. As a result, it often turns out that those who are being regulated can exercise considerable influence over the regulators.

Visit this website to read about lobbying.



In the early 2000s, about 40 million people in the United States were eligible for **Medicare**, a government program that provides health insurance for those 65 and older. On some issues, the elderly are a powerful interest group. They donate money and time to political campaigns, and in the 2012 presidential election, 70% of those over age 65 voted, while just 49% of those aged 18 to 24 cast a ballot, according to the U.S. Census.

In 2003, Congress passed and President George Bush signed into law a substantial expansion of Medicare that helped the elderly to pay for prescription drugs. The prescription drug benefit cost the federal government about \$40 billion in 2006, and the Medicare system projects that the annual cost will rise to \$121 billion by 2016. The political pressure to pass a prescription drug benefit for Medicare was apparently quite high, while the political pressure to assist the 40 million with no health insurance at all was considerably lower. One reason might be that senior citizens are represented by **AARP**, a well-funded and well-organized lobbying group, while there is no umbrella organization to lobby for those without health insurance.

In the battle over passage of the 2010 **Affordable Care Act (ACA)**, which became known as “Obamacare,” there was heavy lobbying on all sides by insurance companies and pharmaceutical companies. However, a lobby group, **Health Care for America Now (HCAN)**, was financed by labor unions and community groups and was organized to offset corporate lobbying. HCAN, spending \$60 million dollars, was successful in helping to get the legislation passed which added new regulations on insurance companies and a mandate that all individuals will obtain health insurance by 2014. The following Work It Out feature further explains voter incentives and lobbyist influence.

PAYING TO GET YOUR WAY

Suppose Congress proposes a tax on carbon emissions for certain factories in a small town of 10,000 people. The tax is estimated to reduce pollution to such an extent that it will benefit each resident by an equivalent of \$300. The tax will also reduce profits to the town’s two large factories by \$1 million each. How much should the factory owners be willing to spend to fight the passage of the tax, and how much should the townspeople be willing to pay to support it? Why is society unlikely to achieve the optimal outcome?

Step 1. The two factory owners each stand to lose \$1 million if the tax is passed, so each should be willing to spend up to

that amount to prevent the passage of the tax, a combined sum of \$2 million. Of course, in the real world, there is no guarantee that lobbying efforts will be successful, so the factory owners may choose to invest an amount that is substantially lower.

Step 2. There are 10,000 townspeople, each standing to benefit by \$300 if the tax passes. Theoretically, then, they should be willing to spend up to \$3 million ($10,000 \times \300) to ensure passage. (Again, in the real world with no guarantees of success, they may choose to spend less.)

Step 3. It is costly and difficult for 10,000 people to coordinate in such a way as to influence public policy. Since each person stands to gain only \$300, many may feel lobbying is not worth the effort.

Step 4. The two factory owners, however, find it very easy and profitable to coordinate their activities, so they have a greater incentive to do so.

Special interests may develop a close relationship with one political party, so their ability to influence legislation rises and falls as that party moves in or out of power. A special interest may even hurt a political party if it appears to a number of voters that the relationship is too cozy. In a close election, a small group that has been under-represented in the past may find that it can tip the election one way or another—so that group will suddenly receive considerable attention. Democratic institutions produce an ebb and flow of political parties and interests and thus offer both opportunities for special interests and ways of counterbalancing those interests over time.

IDENTIFIABLE WINNERS, ANONYMOUS LOSERS

A number of economic policies produce gains whose beneficiaries are easily identifiable, but costs that are partly or entirely shared by a large number who remain anonymous. A democratic political system probably has a bias toward those who are identifiable.

For example, policies that impose **price controls**—like rent control—may look as if they benefit renters and impose costs only on landlords. However, when landlords then decide to reduce the number of rental units available in the area, a number of people who would have liked to rent an apartment end up living somewhere else because no units were available. These would-be renters have experienced a cost of rent control, but it is hard to identify who they are.

Similarly, policies that block imports will benefit the firms that would have competed with those imports—and workers at those firms—who are likely to be quite visible. Consumers who would have preferred to purchase the imported products, and who thus bear some costs of the protectionist policy, are much less visible.

Specific tax breaks and spending programs also have identifiable winners and impose costs on others who are hard to identify. Special interests are more likely to arise from a group that is easily identifiable, rather than from a group where some of those who suffer may not even recognize they are bearing costs.

PORK BARRELS AND LOGROLLING

Politicians have an incentive to ensure that government money is spent in their home state or district, where it will benefit their constituents in a direct and obvious way. Thus, when legislators are negotiating over whether to support a piece of legislation, they commonly ask each other to include **pork-barrel spending**, legislation that benefits mainly a single political district. Pork-barrel spending is

another case in which democracy is challenged by concentrated benefits and widely dispersed costs: the benefits of pork-barrel spending are obvious and direct to local voters, while the costs are spread over the entire country. Read the following Clear It Up feature for more information on pork-barrel spending.

HOW MUCH IMPACT CAN PORK-BARREL SPENDING HAVE?

U.S. Senator Robert C. Byrd of West Virginia, who was originally elected to the Senate in 1958 and served until 2010, is widely regarded as one of the masters of pork-barrel politics, directing a steady stream of federal funds to his home state. A journalist once compiled a list of structures in West Virginia at least partly funded by the government and named after Byrd: “the Robert C. Byrd Highway; the Robert C. Byrd Locks and Dam; the Robert C. Byrd Institute; the Robert C. Byrd Life Long Learning Center; the Robert C. Byrd Honors Scholarship Program; the Robert C. Byrd Green Bank Telescope; the Robert C. Byrd Institute for Advanced Flexible Manufacturing; the Robert C. Byrd Federal Courthouse; the Robert C. Byrd Health Sciences Center; the Robert C. Byrd Academic and Technology Center; the Robert C. Byrd United Technical Center; the Robert C. Byrd Federal Building; the Robert C. Byrd Drive; the Robert C. Byrd Hilltop Office Complex; the Robert C. Byrd Library; and the Robert C. Byrd Learning Resource Center; the Robert C. Byrd Rural Health Center.” This list does not include government-funded projects in West Virginia that were not named after Byrd. Of course, we would have to analyze each of these expenditures in detail to figure out whether they should be treated as pork-barrel spending or whether they provide widespread benefits that reach beyond West Virginia. At least some of them, or a portion of them, certainly would fall into that category. Because there are currently no term limits for Congressional representatives, those who have been in office longer generally have more power to enact pork-barrel projects.

The amount spent on individual pork-barrel projects is small, but many small projects can add up to a substantial total. A nonprofit watchdog organization, called Citizens against Government Waste, produces an annual report, the *Pig Book* that attempts to quantify the amount of pork-barrel spending, focusing on items that were requested by only one member of Congress, that were passed into law without any public hearings, or that serve only a local purpose. Whether any specific item qualifies as pork can be controversial, of course, but at least by this measure, pork-barrel spending totaled \$2.7 billion in 2014.

Pork-barrel spending can be encouraged by **logrolling**, an action in which all members of a group of legislators agree to vote for a package of otherwise unrelated laws that they individually favor. For example, if one member of the U.S. Congress suggests building a new bridge or hospital in his or her own congressional district, the other members might oppose it. However, if 51% of the legislators come together, they can pass a bill that includes a bridge or hospital for every one of their districts.

As a reflection of this interest of legislators in their own districts, the U.S. government has typically spread out its spending on military bases and weapons programs to congressional districts all across the country. In part, this is done to help create a situation that encourages members of Congress to vote in support of defense spending.

KEY CONCEPTS AND SUMMARY

Special interest politics arises when a relatively small group, called a special interest group, each of whose members has a large interest in a political outcome, devotes a lot of time and energy to lobbying for the group’s preferred choice. Meanwhile, the large majority, each of whose members has only a small interest in this issue, pays no attention.

Pork-barrel spending is defined as legislation whose benefits are concentrated on a single district while the costs are spread widely over the country. Logrolling refers to a situation in which two or more legislators agree to vote for each other's legislation, which can then encourage pork-barrel spending in many districts.

SELF-CHECK QUESTIONS

1. What is the main factor preventing a large community from influencing policy in the same way as a special interest group?
2. Why might legislators vote to impose a tariff on Egyptian cotton, when consumers in their districts would benefit from its availability?

REVIEW QUESTIONS

1. How can a small special interest group win in a situation of majority voting when the benefits it seeks flow only to a small group?
2. How can pork-barrel spending occur in a situation of majority voting when it benefits only a small group?
3. Why do legislators vote for spending projects in districts that are not their own?

CRITICAL THINKING QUESTIONS

1. When Microsoft was founded, the company devoted very few resources to lobbying activities. After a high-profile antitrust case against it, however, the company began to lobby heavily. Why does it make financial sense for companies to invest in lobbyists?
2. Special interest groups are often made up of representatives of competing firms. Why are competitors sometimes willing to cooperate in order to form lobbying associations?
3. Special interests do not oppose regulations in all cases. The Marketplace Fairness Act of 2013 would require online merchants to collect sales taxes from their customers in other states. Why might a large online retailer like Amazon.com support such a measure?
4. To ensure safety and efficacy, the Food and Drug Administration regulates the medicines that are allowed to be sold in the United States. Sometimes this means a drug must be tested for years before it is allowed to reach the market. The winners in this system are easily identifiable as those who are protected from unsafe drugs that might otherwise harm them. Who are the more anonymous losers who suffer from strict medical regulations?
5. How is it possible to bear a cost without realizing it? What are some examples of policies that affect people in ways they may not even be aware of?
6. Is pork-barrel spending always a bad thing? Can you think of some examples of pork-barrel projects, perhaps from your own district, that have had positive results?

PROBLEMS

Say that the government is considering a ban on smoking in restaurants in Tobaccoville. There are 1 million people living there, and each would benefit by \$200 from this smoking ban. However, there are two large tobacco companies in Tobaccoville and the ban would cost them \$5 million each. What are the total costs and benefits of this proposed policy? Do you think it will be passed?

REFERENCES

Roper Center Public Opinion Archives. "A Brief History of Public Opinion on the Government's Role in Providing Health Care." Last modified September 23, 2013.

U.S. Department of Commerce: United States Census Bureau. "Voting and Registration: Historical Time Series Tables; Table A-10: Reported Registration Rates in Presidential Election Years, by Selected Characteristics: November 1968 to 2012." <http://www.census.gov/hhes/www/socdemo/voting/publications/historical/index.html>

GLOSSARY

logrolling the situation in which groups of legislators all agree to vote for a package of otherwise unrelated laws that they individually favor

pork-barrel spending spending that benefits mainly a single political district

special interest groups groups that are small in number relative to the nation, but well organized and thus exert a disproportionate effect on political outcomes

SOLUTIONS

Answers to Self-Check Questions

1. The costs of organization and the small benefit to the individual.
2. Domestic cotton producers would lobby heavily to protect themselves from the competition, whereas the consumers have little incentive to organize.

24.3 FLAWS IN THE DEMOCRATIC SYSTEM OF GOVERNMENT

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Assess the median voter theory
- Explain the voting cycle
- Analyze the interrelationship between markets and government

Most developed countries today have a democratic system of government: citizens express their opinions through votes and those votes affect the direction of the country. The advantage of democracy over other systems is that it allows everyone in a society an equal say and therefore may reduce the possibility of oppression of the masses by a small group of wealthy oligarchs. There is no such thing as a perfect system, and democracy, for all its popularity, is not without its problems, a few of which we will examine here.

Democracy is sometimes summed up (and oversimplified) in two words: “Majority rule.” When voters face three or more choices, however, then voting may not always be a useful way of determining what the majority prefers.

As one example, consider an election in a state where 60% of the population is liberal and 40% is conservative. If there are only two candidates, one from each side, and if liberals and conservatives vote in the same 60–40 proportions in which they are represented in the population, then the liberal will win. What if the election ends up including two liberal candidates and one conservative? It is possible that the liberal vote will split and victory will go to the minority party. In this case, the outcome does not reflect the majority’s preference.

Does the majority view prevail in the case of sugar **quotas**? Clearly there are more sugar consumers in the United States than sugar producers, but the U.S. domestic sugar lobby (www.sugarcane.org) has successfully argued for protection against imports since 1789. By law, therefore, U.S. makers of cookies and candies must use 85% domestic sugar in their products. Meanwhile quotas on imported sugar restrict supply and keep the domestic price of sugar up—raising prices for companies that use sugar in the production of their goods and for consumers. The European Union allows sugar imports, and prices there are 40% lower than U.S. sugar prices. Sugar-producing countries in the Caribbean repeatedly protest the U.S. quotas at the World Trade Organization meetings, but each bite of cookie, at present, costs you more than if there were no sugar lobby. This case goes against the theory of the “median” voter in a democracy. The **median voter theory** argues that politicians will try to match

policies to what pleases the median voter preferences. If we think of political positions along a spectrum from left to right, the median voter is in the middle of the spectrum. This theory argues that actual policy will reflect “middle of the road.” In the case of sugar lobby politics, the *minority*, not the median, dominates policy.

Sometimes it is not even clear how to define what the majority opinion might be. Step aside from politics for a moment and think about a choice facing three families (the Ortegas, the Schmidts, and the Alexanders) who are planning to celebrate New Year’s Day together. They agree to vote on the menu, choosing from three entrees, and they agree that the majority vote wins. With three families, it seems reasonable that one choice of entree will get a 2–1 majority. What if, however, their vote ends up looking like Table 1?

Clearly, the three families disagree on their first choice. But the problem goes even deeper. Instead of looking at all three choices at once, compare them two at a time. (See Figure 1) In a vote of turkey versus beef, turkey wins by 2–1. In a vote of beef versus lasagna, beef wins 2–1. If turkey beats beef, and beef beats lasagna, then it might seem only logical that turkey must also beat lasagna. However, with the preferences shown, lasagna is preferred to turkey by a 2–1 vote, as well. If lasagna is preferred to turkey, and turkey beats beef, then surely it must be that lasagna also beats beef? Actually, no; beef beats lasagna. In other words, the majority view may not win. Clearly, as any car salesmen will tell you, choices are influenced by the way they are presented.

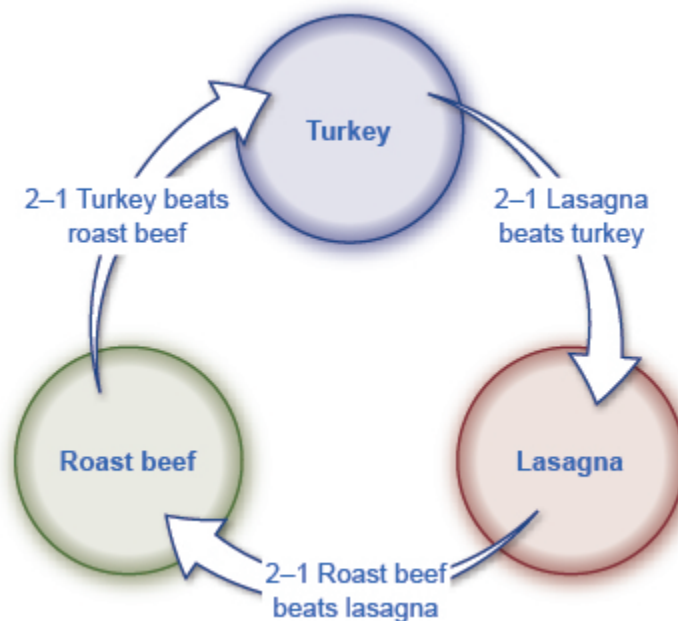


Figure 1. A Voting Cycle. Given these choices, voting will struggle to produce a majority outcome. Turkey is favored over roast beef by 2–1 and roast beef is favored over lasagna by 2–1. If turkey beats roast beef and roast beef beats lasagna, then it might seem that turkey must beat lasagna, too. But given these preferences, lasagna is favored over turkey by 2–1.

	The Ortega Family	The Schmidt Family	The Alexander Family
First Choice	Turkey	Roast beef	Lasagna
Second Choice	Roast beef	Lasagna	Turkey
Third Choice	Lasagna	Turkey	Roast beef

Table 1. Circular Preferences

The situation in which Choice A is preferred by a majority over Choice B, Choice B is preferred by a majority over Choice C, and Choice C is preferred by a majority over Choice A is called a **voting cycle**. It is easy to imagine sets of government choices—say, perhaps the choice between increased defense spending, increased government spending on health care, and a tax cut—in which a voting cycle could occur. The result will be determined by the order in which choices are presented and voted on, not by majority rule, because every choice is both preferred to some alternative and also not preferred to another alternative.

Visit this website to read about instant runoff voting, a preferential voting system.



WHERE IS GOVERNMENT'S SELF-CORRECTING MECHANISM?

When a **firm** produces a product no one wants to buy or produces at a higher cost than its competitors, the firm is likely to suffer losses. If it cannot change its ways, it will go out of business. This self-correcting mechanism in the marketplace can have harsh effects on workers or on local economies, but it also puts pressure on firms for good performance.

Government agencies, on the other hand, do not sell their products in a market; they receive tax dollars instead. They are not challenged by competitors as are private-sector firms. If the U.S. Department of Education or the U.S. Department of Defense is doing a poor job, citizens cannot purchase their services from another provider and drive the existing government agencies into bankruptcy. If you are upset that the Internal Revenue Service is slow in sending you a tax refund or seems unable to answer your questions, you cannot decide to pay your income taxes through a different organization. Of course, elected politicians can assign new leaders to government agencies and instruct them to reorganize or to emphasize a different mission. The pressure government faces, however, to change its bureaucracy, to seek greater efficiency, and to improve customer responsiveness is much milder than the threat of being put out of business altogether.

This insight suggests that when government provides goods or services directly, we might expect it to do so with less efficiency than private firms—except in certain cases where the government agency may compete directly with private firms. At the local level, for example, services like garbage collec-

tion can be provided by government directly, by private firms under contract to the government, or by a mix of government employees competing with private firms.

A BALANCED VIEW OF MARKETS AND GOVERNMENT

The British statesman Sir Winston Churchill (1874–1965) once wrote: “No one pretends that democracy is perfect or all-wise. Indeed, it has been said that democracy is the worst form of government except for all of the other forms which have been tried from time to time.” In that spirit, the theme of this discussion is certainly not that democratic government should be abandoned. A practical student of public policy needs to recognize that in some cases, like the case of well-organized special interests or pork-barrel legislation, a democratic government may seek to enact economically unwise projects or programs. In other cases, by placing a low priority on the problems of those who are not well organized or who are less likely to vote, the government may fail to act when it could do some good. In these and other cases, there is no automatic reason to believe that government will necessarily make economically sensible choices.

“The true test of a first-rate mind is the ability to hold two contradictory ideas at the same time,” wrote the American author F. Scott Fitzgerald (1896–1940). At this point in your study of microeconomics, you should be able to go one better than Fitzgerald and hold three somewhat contradictory ideas about the interrelationship between markets and government in your mind at the same time.

First, markets are extraordinarily useful and flexible institutions through which society can allocate its scarce resources. This idea was introduced with the subjects of international trade and demand and supply in other chapters and reinforced in all the subsequent discussions of how households and firms make decisions.

Second, markets may sometimes produce unwanted results. A short list of the cases in which markets produce unwanted results includes monopoly and other cases of imperfect competition, pollution, poverty and inequality of incomes, discrimination, and failure to provide insurance.

Third, while government may play a useful role in addressing the problems of markets, government action is also imperfect and may not reflect majority views. Economists readily admit that, in settings like monopoly or negative externalities, a potential role exists for government intervention. However, in the real world, it is not enough to point out that government action might be a good idea. Instead, we must have some confidence that the government is likely to identify and carry out the appropriate public policy. To make sensible judgments about economic policy, we must see the strengths and weaknesses of both markets and government. We must not idealize or demonize either unregulated markets or government actions. Instead, consider the actual strengths and weaknesses of real-world markets and real-world governments.

These three insights seldom lead to simple or obvious political conclusions. As the famous British economist **Joan Robinson** wrote some decades ago: “[E]conomic theory, in itself, preaches no doctrines and cannot establish any universally valid laws. It is a method of ordering ideas and formulating questions.” The study of economics is neither politically conservative, nor moderate, nor liberal. There are economists who are Democrats, Republicans, libertarians, socialists, and members of every other political group you can name. Of course, conservatives may tend to emphasize the virtues of markets and the limitations of government, while liberals may tend to emphasize the shortcomings

of markets and the need for government programs. Such differences only illustrate that the language and terminology of economics is not limited to one set of political beliefs, but can be used by all.

CHINESE TIRE TARIFFS

In April 2009, the union representing U.S. tire manufacturing workers filed a request with the U.S. **International Trade Commission (ITC)**, asking it to investigate tire imports from China. Under U.S. trade law, if imports from a country increase to the point that they cause market disruption in the United States, as determined by the ITC, then it can also recommend a remedy for this market disruption. In this case, the ITC determined that from 2004 to 2008, U.S. tire manufacturers suffered declines in production, financial health, and employment as a direct result of increases in tire imports from China. The ITC recommended that an additional tax be placed on tire imports from China. President Obama and Congress agreed with the ITC recommendation, and in June 2009 tariffs on Chinese tires increased from 4% to 39%.

Why would U.S. consumers buy imported tires from China in the first place? Most likely, because they are cheaper than tires produced domestically or in other countries. Therefore, this tariff increase should cause U.S. consumers to pay higher prices for tires, either because Chinese tires are now more expensive, or because U.S. consumers are pushed by the tariff to buy more expensive tires made by U.S. manufacturers or those from other countries. In the end, this tariff made U.S. consumers pay more for tires.

Was this tariff met with outrage expressed via social media, traditional media, or mass protests? Were there “Occupy Wall Street-type” demonstrations? The answer is a resounding “No.” Most U.S. tire consumers were likely unaware of the tariff increase, although they may have noticed the price increase, which was between \$4 and \$13 depending on the type of tire. Tire consumers are also potential voters. Conceivably, a tax increase, even a small one, might make voters unhappy. However, voters probably realized that it was not worth their time to learn anything about this issue or cast a vote based on it. They probably thought their vote would not matter in determining the outcome of an election or changing this policy.

Estimates of the impact of this tariff show it costs U.S. consumers around \$1.11 billion annually. Of this amount, roughly \$817 million ends up in the pockets of foreign tire manufacturers other than in China, and the remaining \$294 million goes to U.S. tire manufacturers. In other words, the tariff increase on Chinese tires may have saved 1,200 jobs in the domestic tire sector, but it cost 3,700 jobs in other sectors, as consumers had to cut down on their spending because they were paying more for tires. Jobs were actually lost as a result of this tariff. Workers in U.S. tire manufacturing firms earned about \$40,000 in 2010. Given the number of jobs saved and the total cost to U.S. consumers, the cost of saving one job amounted to \$926,500!

This tariff caused a net decline in U.S. social surplus. (Total surplus is discussed in the Demand and Supply chapter, and tariffs are discussed in the The International Trade and Capital Flows chapter.) Instead of saving jobs, it cost jobs, and those jobs that it saved cost many times more than the people working in them could ever hope to earn. Why would the government do this?

The chapter answers this question by discussing the influence special interest groups have on economic policy. The steelworkers union, whose members make tires, saw more and more of its members lose their jobs as U.S. consumers consumed more and more cheap Chinese tires. By definition, this union is relatively small but well organized, especially compared to tire consumers. It stands to gain much for each of its members, compared to what each tire consumer may have to give up in terms of higher prices. So the steelworkers union (joined by domestic tire manufacturers) has not only the means but the incentive to lobby economic policymakers and lawmakers. Given that U.S. tire consumers are a large and unorganized group, if they even are a group, it is unlikely they will lobby against higher tire tariffs. In the end, lawmakers tend to listen to those who lobby them, even though the results make for bad economic policy.

KEY CONCEPTS AND SUMMARY

Majority votes can run into difficulties when more than two choices exist. A voting cycle occurs when,

in a situation with at least three choices, choice A is preferred by a majority vote to choice B, choice B is preferred by a majority vote to choice C, and choice C is preferred by a majority vote to choice A. In such a situation, it is impossible to identify what the majority prefers. Another difficulty arises when the vote is so divided that no choice receives a majority.

A practical approach to microeconomic policy will need to take a realistic view of the specific strengths and weaknesses of markets and the specific strengths and weaknesses of government, rather than making the easy but wrong assumption that either the market or government is always beneficial or always harmful.

SELF-CHECK QUESTIONS

1. True or false: Majority rule can fail to produce a single preferred outcome when there are more than two choices.
2. Anastasia, Emma, and Greta are deciding what to do on a weekend getaway. They each suggest a first, second, and third choice and then vote on the options. Their first choice, second choice, and third choice preferences are as shown in Table 2. Explain why they will have a hard time reaching a decision. Does the group prefer mountain biking to canoeing? What about canoeing compared to the beach? What about the beach compared to the original choice of mountain biking?

	Anastasia	Emma	Greta
First Choice	Beach	Mountain biking	Canoeing
Second Choice	Mountain biking	Canoeing	Beach
Third Choice	Canoeing	Beach	Mountain biking

Table 2.

3. Suppose an election is being held for Soft Drink Commissioner. The field consists of one candidate from the Pepsi party and four from the Coca-Cola party. This would seem to indicate a strong preference for Coca-Cola among the voting population, but the Pepsi candidate ends up winning in a landslide. Why does this happen?

REVIEW QUESTIONS

1. Why does a voting cycle make it impossible to decide on a majority-approved choice?
2. How does a government agency raise revenue differently from a private company, and how does that affect the way government decisions are made, compared to business decisions?

CRITICAL THINKING QUESTIONS

1. The United States currently uses a voting system called “first past the post” in elections, meaning that the candidate with the most votes wins. What are some of the problems with a “first past the post” system?
2. What are some alternatives to a “first past the post” system that might reduce the problem of voting cycles?

3. AT&T spent some \$10 million dollars lobbying Congress to block entry of competitors into the telephone market in 1978. Why do you think its efforts failed?
4. Occupy Wall Street was a national (and later global) organized protest against the greed, bank profits, and financial corruption that led to the 2008–2009 recession. The group popularized slogans like “We are the 99%,” meaning it represented the majority against the wealth of the top 1%. Does the fact that the protests had little to no effect on legislative changes support or contradict the chapter?

REFERENCES

Nixon, Ron. “American Candy Makers. Pinched by Inflated Sugar Prices. Look Abroad.” *The New York Times*. Last modified October 30, 2013. http://www.nytimes.com/2013/10/31/us/american-candy-makers-pinched-by-inflated-sugar-prices-look-abroad.html?_r=0.

Hufbauer, Gary Clyde, and Sean Lowry. “U.S. Tire Tariffs: Saving Few Jobs at High Cost (Policy Brief 12-9).” *Peterson Institute for International Economics*. Last modified April 2012.

GLOSSARY

median voter theory theory that politicians will try to match policies to what pleases the median voter preferences

voting cycle the situation in which a majority prefers A over B, B over C, and C over A

SOLUTIONS

Answers to Self-Check Questions

1. True. This is exactly what occurs in a voting cycle. That is, the majority can prefer policy A to policy B, policy B to policy C, but also prefer policy C to policy A. Then, the majority will never reach a conclusive outcome.
2. The problem is an example of a voting cycle. The group will vote for mountain biking over canoeing by 2–1. It will vote for canoeing over the beach by 2–1. If mountain biking is preferred to canoeing and canoeing is preferred to the beach, it might seem that it must be true that mountain biking is the favorite. But in a vote of the beach versus mountain biking, the beach wins by a 2–1 vote. When a voting cycle occurs, choosing a single favorite that is always preferred by a majority becomes impossible.
3. The four Coca-Cola candidates compete with each other for Coca-Cola voters, whereas everyone who prefers Pepsi had only one candidate to vote for. Thus the will of the majority is not satisfied.

CHAPTER 25. MONEY AND THE THEORY OF THE FIRM

INTRODUCTION TO MONEY AND THE THEORY OF THE FIRM



A \$10 National Gold Bank Note — issued by the First National Gold Bank of Oakland, California (c. the 1870s).
(Wikimedia, Public Domain)

CHAPTER OBJECTIVES

In this chapter you will learn about:

- The Metallist and the Barter Myth
- Smith, Marx, Keynes, Chartalism and Modern Money Theory
- The Money Hierarchy and the False Duality of the State and Market
- Local Currency Systems: Social Money and Community Currencies

The standard approach, taken in neoclassical texts, to describe the firm's role in the economy is to ascribe a coordinating function: the firm optimally selects the inputs of labor and capital to efficiently produce output for the market. This functional interpretation (1) of the firm allows the neoclassical economist to maintain the same level of simplicity, or abstraction achieved in consumer theory. This approach limits the scope of value determination to commodity market exchanges, as seen in Chapter 3. In this chapter, we investigate how this abstract methodology of modeling the economy influences our understanding of money.

$$q = f(k, l)$$

where

q is output,

k is capital,

l is labor,

f provides a functional description of the firm's technology.

Understanding the limitation of money in the neoclassical framework will be achieved in three steps. First, the story of money will be told from the neoclassical perspective. In telling this story, we will discover that the neoclassical economist relegates money to a secondary role in economic activity. The neoclassical presentation describes money as a commodity that facilitates the exchange of other commodities. This secondary role for money allows the neoclassical economist to continue to apply, what we will define as real analysis. After our neoclassical story is completed, a second story of money will be told. This story is guided, not by theory, but by history, and will help us to understand money's true origins. In applying a historical and institutional methodology, we will see that money is much more critical to economic activity than neoclassical economists suppose. Money is special, because of how it can be used, and who is able to produce it. In the next section, the historical understanding of money not as a commodity, but as a social relation will break down the traditional perspective of a strict duality between the state and the market. This is accomplished through an analysis of or

economy not as a barter system, but as a monetary production economy that operates within the constraints of the legal and institutional framework of a money hierarchy.

The final section extends our understanding of the monetary production economy by examining local currency systems. Local currency systems are emerging across the world in communities that have been abandoned by mainstream economic policies. These local money systems are supporting single mothers, helping children go to school, cleaning up streets and neighborhoods, and reducing waste. By comparing the monopoly issuing authority of the United States and these micro currency systems, the coordinating function of money in economic activity will become clear. Geoffrey Ingham once argued that money is our most powerful and underemployed social technology[1]. This chapter will help you to understand why.

[1] This insight is drawn from Geoffery Ingham's contribution to John Smithin's 2001 edited volume, *What is Money?*.

25.1 THE METALLIST AND THE BARTER MYTH

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Discuss the importance of barter in modern economic analysis

In what kind of economy do we live and operate our daily lives? This is not a trivial question. As ideas about how human beings can best organize their resources developed into enlightenment philosophy, a prominent method for expressing what a society free of the feudal constraints would look like involved imaginary stories. Philosophers, such as John Locke, Thomas Hobbes, and Jean-Jacques Rousseau asked their audience to imagine a place called the state of nature to consider possible social outcomes when this place is populated with free individuals. In the state of nature, will chaos, barbarism, or a harmonious stable society result?

In Locke's version, individuals needed to follow two constraints or simple rules in order for a harmonious and stable society to emerge. The first of these constraints is the spoilage constraint. In the state of nature, no individual should collect more food, nuts, and berries than they can consume before they spoil. Second, the prejudice constraint cannot be violated. This constraint is "predicated on the right to subsistence: the privatization of land [can]not disadvantage non-property holders, violating their right to subsistence goods." [1] As long as all those in the state of nature follow these two simple moral constraints, then a harmonious and stable society is argued to emerge. Through the evolution of similar stories, political economist would add their own plot twists and establish the setting for neoclassical economics.

What these imaginary stories have in common is the collective illusion that capitalism or free-market economics grows out of a state of nature. The basic plot and stage for this story remain the same in neoclassical economics. How many of you got dressed and ready for your classes today in the Cartesian plane? Like their enlightenment predecessors' use of the state of nature, the neoclassical story also takes place in a space that is not inhabited by human beings. The mathematical universe of maximization and optimal decisions, like Locke's imagined individuals, aggregates to a harmonious and stable solution. This solution is known as equilibrium in the formalized mathematical models of neoclassical economics. While the technical names and the math provide the neoclassical story with a scientific veneer, the underlying plot remains generally the same as those told over the previous four hundred plus years. A strong characteristic of these stories is their staying power, but a significant cost of such longevity is a lack of analysis regarding the actual economy in which we live and conduct our daily lives.

The economy described by Jean-Baptiste Say^[2] and other classical political economists of his era, as well as by modern neoclassical economists, is a **barter economy**. In a barter economy, goods trade for other goods. If we were to open a barter market in class, students could trade their hats, backpacks, books, shoes, and whatever else they had in their possession with other students. In this scenario, would much trade occur? Generally, the answer would be no. This is because it would be rare that the **double coincidence of wants** would be solved very often. This means that for trade to occur I must have what you want and you must have what I want. If and only if this is the case will trade occur because trade must make us both better off than we were before the trade took place. To solve this problem and reduce the amount of time individuals would need to spend seeking out solutions to the double coincidence of wants, society spontaneously agrees to accept a commodity as the facilitator of exchange—or at least this is how the story is told from the state of nature and Cartesian plane origins.

The magical device that solves the double coincidence of wants and sets the economy free is money. Commodity money in the form of gold or silver has an **intrinsic value** that can be measured and all parties understand its value. Intrinsic value is a key concept and can be defined as the value the object has in and of itself. This is why precious metals are central to this story, as it is believed gold has value in and of itself. Thus, through this agreement, the transaction costs for all the producers and consumers in this economy are reduced and are therefore better off. This is the story often told by a group of economists known as the **Metallists**, because of the use of precious metals to denote the value of money. We can model money as the facilitator of a barter transaction using the following notation of C for commodity and M for money.

This is likely the story you have heard before, and it is definitely the one told to your parents in their introductory economics courses. As mentioned above, it is an old story. Roots to its origins go as far back as Plato and Aristotle, but the more modern versions are linked to the 16th and 17th-century enlightenment philosophers discussed above. The staying power and continued adherence to this story are heavily reliant upon metal's intrinsic value. If money's value is connected to the weight and quality of metal, then it, just like the gold, silver, or the other goods it trades for, is a commodity. As a commodity, money preserves the barter system. In a barter framework, economists are able to practice **real analysis**.

Real analysis is described by the 20th-century economist Joseph Schumpeter as:

[all] the essential phenomena of economic life are capable of being described in terms of goods and services, of decisions about them, and of relations between them. Money enters the picture only in the modest role of a technical device that has been adopted in order to facilitate transactions. This device can no doubt get out of order, and if it does it will indeed produce phenomena that are specifically attributable to its *modus operandi*. But so long as it functions normally, it does not affect the economic process, which behaves in the same way as it would in a barter economy.^[3]

In other words, economists can study real changes to output, employment, distribution, and growth without addressing money as anything other than a facilitator of exchange, because in the long-run money is neutral. Money does not generate real changes only nominal changes that create short-run disturbances to the stability of economic equilibrium. In other words, it disrupts the *modus operandi* in Schumpeter's quote. By examining money, within the context of modern economic events, it will become clear that the commitment made by neoclassical economics to the Metallist's story is not

intended to provide an accurate understanding of history, but is rather an effort to preserve their method of economic analysis and the continued application of optimization techniques and mathematical modeling.

While the barter story is intuitively appealing, especially when described as an economy at a time and place far, far away, difficult questions about value and the role of the state complicate the drama. For instance, when gold or silver was spontaneously accepted as the commodity money, did people just trade nuggets of the metal? Where did coins come from? Who made the coins, and how was their value determined? Then finally, as economies incorporate paper money, how is the value of the currency maintained, especially if it cannot be directly converted into gold or silver? These questions are examples of critical analysis. Rather than accepting the story as given, critical thinking and investigation will help economists and policymakers to develop an understanding of money and its evolution in economic systems.

To address these critical inquiries, let us return to Schumpeter's description. Money's value is as a technical device used in the facilitation of exchanges. Its value is intrinsically derived through a precious metal. In this scenario, it is a commodity like all other commodities. Difficulty only emerges, if it gets out of order. If money's *modus operandi* extends beyond the facilitator role, then the economy is no longer functioning as a barter system. It is in this capacity, as something more than a technical device that the **Chartalist** school of thought is introduced here, in the second act.

The complications, regarding money's value, have become increasingly difficult for Medallists to assume away and maintain money's status as a commodity. The most significant of these changes to the global economy has been the development of fiat currency systems that have completely broken monetary systems from the gold standard. An example of fiat money is the U.S. dollar. They are issued by fiat because the issuer will not give us anything other than other dollars in return. It is an inconvertible paper money. The gold standard allowed for the commodity money story to survive the complications associated with the expansions of both credit money lending and paper money, because money remained **convertible** into gold. However, the elimination of the gold standard has ended the convertibility scene and requires a new dialogue. If money cannot be converted into gold, and value is not intrinsically determined, then where does its value come from?

ENDNOTES

[1] The academic journal *The Review of Social Economy* has published a number of works of by Modern Money Theorist including the co-authored piece by John F. Henry, Stephanie Kelton, and L. Randall Wray, "A Chartalist Critique of John Locke's Theory of Peoples Accumulation, and Money: or is it Moral Trade your Nuts for Gold?"

[2] Jean-Baptist Say is an important character in the development of the ideas that support neoclassical economics. One of his most well-known contributions is Say' Law, which is commonly paraphrased as "supply creates its own demand".

[3] Joseph A. Schumpeter is widely regarded as one of the most influential economists of the 20th Century. Students are encouraged to explore his ideas on money in his 1954 text *History of Economic Analysis*.

GLOSSARY

barter economy

a system in which goods trade for other goods, that is, without money as an intermediary

double coincidence of wants

the fundamental problem of barter in which, for trade to occur, I must have what you want and you must have what I want

intrinsic value (of commodity money)

the value the object has in and of itself

metallism

the belief that the value of money derives from the intrinsic value of the commodity (for instance, gold) the physical money is made from

real-analysis

the study of real changes to output, employment, distribution, and growth without addressing money as anything other than a facilitator of exchange

25.2 SMITH, MARX, KEYNES, CHARTALISM AND MODERN MONEY THEORY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain and compare the two views of money's origin and its value

The answers to the questions of value raised in the previous section can be found by examining the following two quotes, from arguably the two most significant characters in the history of political economy. The first comes from Adam Smith's *The Wealth of Nations*:

A prince, who should enact a certain proportion of his taxes be paid in a paper money of a certain kind, might thereby give a certain value to this paper money.[1]

The second quote, similar in theme, comes from Karl Marx's *Capital Vol. I*, and reads:

The only part of the so-called national wealth that actually enters into the collective possessions of modern peoples is—their national debt. Hence as a necessary consequence, the modern doctrine that a nation becomes the richer the more deeply it is in debt. Public credit becomes the credo of capital. And with the rise of national debt-making, want of faith in the national debt takes the place of the blasphemy against the Holy Ghost, which may not be forgiven.[2]

From these two quotes, the primary difference between the Metallists and Chartalists approaches to understanding money is revealed. Money is not a commodity. Money is, in the words of the legal scholar Friedrich Knapp, a creature of the state.

What may seem to be a subtle difference is in reality substantial. The reorientation of money as a state phenomenon, rather than a market solution to the double coincidence of wants requires a completely different framework for analyzing the economy. The economy can no longer be modeled as a barter system. Real analysis fails to capture the complexity of money, and thus even the existence of general equilibrium itself is called into question. Returning to Schumpeter, if the *modus operandi* of money is not simply a facilitator role, then we must conduct **monetary analysis**. Simply put, this requires the abandonment “of the idea that all essential features of economic life can be represented by a barter-economy model” (Schumpeter 1954).

We can begin to appreciate the difference between the analysis of a barter system and a money econ-

omy by applying Marx's notation and modeling of the **circuit of money capital**. This notation is simple and straightforward (see also Chapter 16: "The Megacorp").

$$M \rightarrow C \dots P \dots C' \rightarrow M'$$

$C \dots P \dots C' \rightarrow M'$ width="334" height="49" sizes="(max-width: 334px) 100vw, 334px">

where the first stage $M - C$ is termed purchase and is followed by the production process P culminating with $C' - M'$ sale. This analysis of the economy is a monetary analysis. Returning to the barter notation, from above, the transaction would look like this.

$$C \rightarrow M \rightarrow C$$

$M \rightarrow C$ width="196" height="55" srcset="https://openoregon.pressbooks.pub/app/uploads/sites/14/2017/06/C-M-C.jpg 239w, https://openoregon.pressbooks.pub/app/uploads/sites/14/2017/06/C-M-C-65x18.jpg 65w, https://openoregon.pressbooks.pub/app/uploads/sites/14/2017/06/C-M-C-225x63.jpg 225w" sizes="(max-width: 196px) 100vw, 196px">

Note, there is no C' in this exchange. This is because the value is generated, not by production or labor, but through the act of exchange. Remember, in the barter economy of the neoclassical school, trade does not occur unless both parties benefit. Value is created through the exchange process and relative prices become the measure of those values (this is also described in Chapter 3 On Values). Where in Marx's model, value is created through the production process and comes from labor.

As you can see, the interpretation of how the economy operates is very different. The predictions of outcomes from market activity are also at odds. In the real analysis of neoclassical economics, markets tend towards equilibrium. Real analysis predicts a stable solution where the quantity demanded is equal to the quantity supplied. However, if the economy is investigated from the monetary analysis perspective, simple solutions do not immediately present themselves, as conflict, business cycles, and crisis are all consistently observed in capitalist economies and central topics of analysis in Marx's theoretical framework.

This is one reason why John Maynard Keynes considered his approach to understanding the economy as *The General Theory of Employment, Interest, and Money*. Similar to Marx, although Keynes denies any Marxian influence, Keynes moves his analysis of the economy beyond the barter system and real analysis and attempts to understand the dynamics of a **monetary production economy**. For Keynes, in a monetary production economy, there is no reason to assume the economy will trend towards full employment equilibrium in the labor market. In fact, given the special properties of money, the economy is more than capable of coming to rest, for long periods of time, at a level of output far below what is necessary to achieve full employment.

Keynes argues that money has two special properties that differentiate it from all other commodities in the economy. The first is that money has a near-zero elasticity of production. This means that, while it might be reasonable to argue that a problem in our local economy is that there is a shortage of money, we cannot go into the business of selling money. So unlike wheat, milk, cars, tables, or any other good or service, as entrepreneurs, we cannot address the unmet demand for money by producing it ourselves. The second of these properties is a near-zero elasticity of substitution. This property relates to the use of money as a means of payment. The state and most private vendors will only accept dollars (in the United States) as a means of payment to satisfy debt obligations. Wow, that is a mouthful, but it simply means that there are very few substitutes for money as a method of payment. The IRS will not take your laptop as payment for your taxes. The gas company is not interested in your tee-shirts to satisfy your bill, they will only accept one thing: cash. As the Wu-Tang Clan so succinctly puts it, C.R.E.A.M. cash rules everything around me.... dollar dollar bill y'all.

Figure 1. \$1 Federal Reserve Note



Image from the Federal Reserve Bank of Philadelphia,
[https://www.philadelphiafed.org/education/teachers/
 publications/symbols-on-american-money](https://www.philadelphiafed.org/education/teachers/publications/symbols-on-american-money)

These two special characteristics of money influence a monetary production economy, by incorporating two missing pieces from the real analysis approach of neoclassical economics, uncertainty, and the state. Hence, it is a more “general theory.” Let’s discuss the state first and then provide a brief discussion of uncertainty. Why can’t we produce dollars? Well, this is against the law. By law, the Treasury Department is the only producer of new U.S. dollars. We will elaborate on this in detail in the next section. Second, why won’t the state take my laptop? For this let’s take a moment to look at the dollar (Figure 1 above). Up in the upper left corner, it says “THIS NOTE IS LEGAL TENDER FOR ALL DEBTS, PUBLIC AND PRIVATE”. This means that there is no reason for the state to accept anything else in payment. If you believe that your laptop is of sufficient value to cover your tax obligation, then you must first convert it into cash to pay your tax bill, and this introduces an important concept in Keynes’s monetary framework, **liquidity**.

Liquidity is a measure of the costs associated with converting an asset into cash. For example, your savings account is an asset. It is very liquid, but to convert it into cash you need to go to the bank and withdraw the funds. This time and effort is a cost, but these costs are relatively low. An example of an asset that is not generally described as being liquid is a home. The sale and conversion of a home into cash is a time consuming and often expensive process. An asset's liquidity is of great importance for understanding economic volatility and how risk and uncertainty differ.

To describe the value of assets in a monetary production economy, Keynes uses what he describes as the own rate. The own rate is a measure of an asset's economic value and is composed of three parts, the expected return on the asset (q), its carrying costs (c), and the liquidity premium (l). This is displayed mathematically as,

$$\text{own rate} = q - c + l.$$

Returning to the example of a home, the expected return, pre-housing crisis, was strong, based on historical data. A home's carrying costs are not cheap. Maintenance, mortgage payments, property taxes, etc. all push up the carrying costs, but you get to live there and conventional wisdom said that a home's value always increases. The final component is its liquidity premium, which is also related to the expected return. A house in a sellers' market in a good neighborhood might sell quickly, whereas a home next to a newly constructed slaughterhouse might not be so easy to unload.

Thus, people's view of the future plays a critical role in determining the own rate. If the housing market is strong, then it is easy to sell because the expected return is positive and the future seller does not think selling in the future has many risks either. On the other hand, if the market is doing poorly, then the expected return is going down, carrying costs might be increasing, as pickier buyers are demanding more and more repairs, and the liquidity premium is disappearing, as a buyer cannot be found.

This equation can be applied to all sorts of commodities. Give it a try in the exercise section below. Think about a car, stocks, and bonds, or a record collection. These assets all change in own rate based on expectations of the future and how quickly they can be converted into cash. Because it is more difficult to convert assets into cash when markets are unstable and uncertain, people tend to hold onto the one asset that is not negatively impacted. You guessed it: money.

The saving money by an individual in times of uncertainty is problematic for the economy. Keynes argues that this virtuous behavior, from the classical economic perspective, aggregates to difficulties for the macroeconomy as a whole. If you are saving, then, yes, this practice can be beneficial for you, but if we are all saving, then we are not spending and this sends negative signals to investors and business owners. So while individual savings are good, the aggregation of this positive behavior generates adverse economic conditions as a whole. This problem is known as the Paradox of Thrift. This paradox creates difficulties for the neoclassical argument that the aggregation of optimal decisions aggregates to a stable and harmonious society or equilibrium conditions.

If we return to Adam Smith's comments above, we can add the contributions by Marx and Keynes to complete the basic framework for understanding money, not as a commodity in a barter system, but as a social relation in a modern capitalist economy. Remember, the value of money in the Metallist's story is derived intrinsically by a precious metal, but Adam Smith seems to be suggesting that value can be generated by the actions of the Prince. This is a prescient observation by Smith, as modern fiat currencies maintain the value by generating demand through the implementation of a tax or other obligation to the money issuing authority. Therefore, an historical and institutional approach

to investigating money reveals fundamentally different dynamics and explanations of the operations of the modern economy. From this interdisciplinary perspective: all economic activity is observed to begin and end with money, money plays this central role because it has special properties, and an authority, such as the state, is responsible for maintaining their currency's value.

By breaking free from the methodological constraints of real analysis, Modern Monetary Theory (MMT) is able to provide insights into economic activity where money serves as more than a contributor to the *modus operandi* in support of barter. In the next section, money is investigated as a social relation. This social relation can be described as a two-sided balance sheet operation. These relations of credits and debits are ordered in the modern economy by a hierarchical structure. As this institutional structure is outlined below, the coordinating function of money and its role in determining value becomes clear.

ENDNOTES

[1] For those looking for an accessible way to introduce themselves to Adam Smith's vast opus, Robert Heilbroner's *The Essential Adam Smith* is an excellent resource. Some may be already familiar with his text *The Worldly Philosophers*.

[2] This insight is drawn from *Capital Vol. I*.

GLOSSARY

chartalism

the belief that the value of money derives, not from the intrinsic value of the commodity, but from its general acceptability for payments and the ability of the state to create demand for it through taxation

liquidity

the ease with which an asset can be converted into cash

25.3 THE MONEY HIERARCHY AND THE FALSE DUALITY OF THE STATE AND MARKET

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Apply the vocabulary and conceptions framework developed in the previous section to the Hierarchy of Money
- Explain where money comes from in a monetary production economy

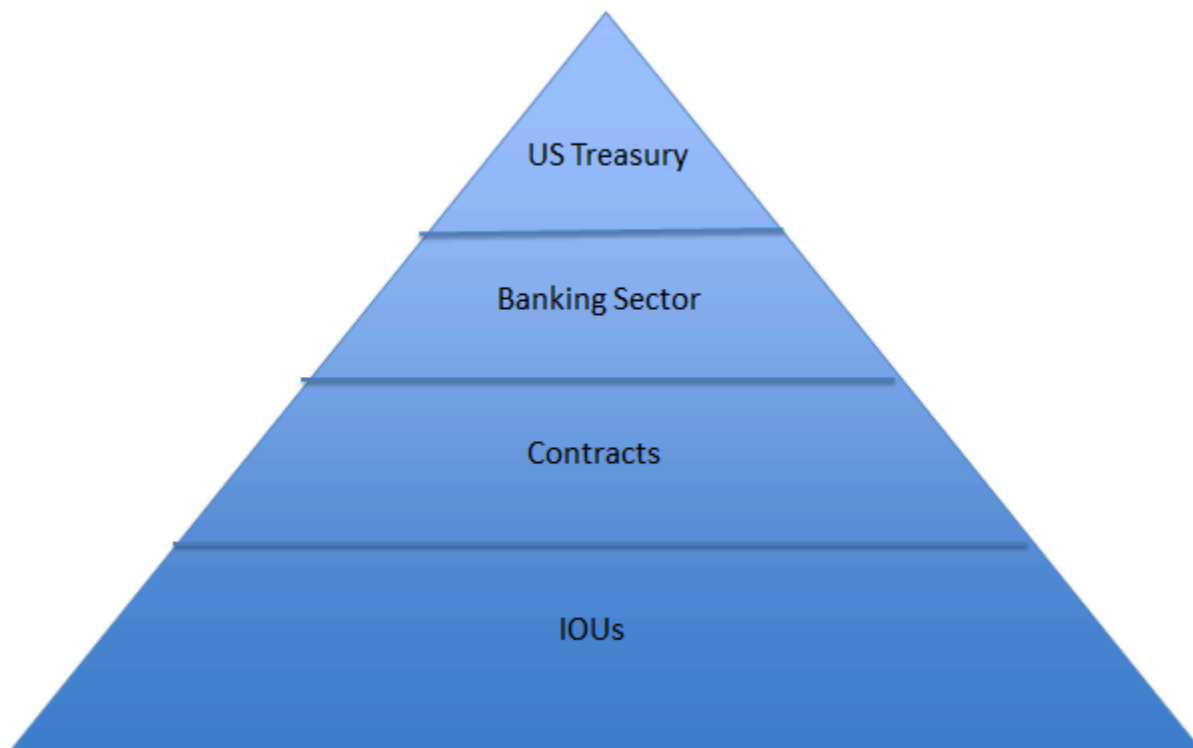
Hyman Minsky once argued that anyone can make money; the real trick is getting people to accept it.[1] For the Metallists, gold or a precious metal's intrinsic value insured money's acceptance as a facilitator of exchange. While simplicity is often a strong characteristic, when it comes to money, a more thorough understanding of money's source of value transforms the relationship between the market and the state. We often think of these two concepts as opposing forces struggling to direct economic activity. From this perspective, one might imagine the market as the strong lead character in pursuit of efficient solutions, and the state as a pesky nemesis taking resources from the market to achieve its own objectives. This plot, however, takes an unexpected twist in Modern Money Theory, as these two characters are revealed to be the same person. The real analysis claims of duality between the state and the market are the product of methodology. The "trick" is not that we accept dollars, but that the true source of their value continues to be largely ignored by economists, policymakers, and the general public.

In 2011, renowned London School of Economics anthropologist David Graeber published a comprehensive examination of the historical origins and development of money, titled *Debt: The First 5000 Years*. While all 5000 years are interesting, and students are encouraged to explore this exemplary work of scholarship, our focus is limited to the current economic system. This narrow focus will allow us to build upon the above ideas of Marx and Keynes and to develop an understanding of key concepts from MMT. The first of these concepts is the hierarchy of money. From this conceptual framework, we will explore the **technostructure** of money (see Chapter 16 The Megacorp). The collection of institutions that regulate money's issuance or production is similar to **market governance** (see Chapter 16 The Megacorp) of business enterprises in that stability is a primary objective. Given the central role of money in economic activity and its origins with the state, the clear delineation between where the market begins and state ends is all but erased.

We begin our analysis with MMT's hierarchy of money. As a social relation, not all money is created equal. For example, by definition as a social relation, you borrowing a shirt from your roommate is a money transaction, as long as you *promise* to give it back. We call this an IOU. I had a friend in high school that literally carried around a notebook listing all of the people that owed him money (he now works in finance). These records represented promises to pay. For some of those entries, repayment was completed and their name removed, for others, I'm sure they are still in that notebook somewhere in Ol' Johnny Cocktail's archives.

These promises to pay were money transactions. They are two-sided balance sheet operations. Johnny extends credit to his friend; the friend takes on a debt to Johnny. When repayment occurs Johnny's asset, and his friend's liability is terminated. Going back to Hyman Minsky's comment, if one regularly borrows and repays Johnny rapidly, then both parties are willing to accept the terms of the social relation. However, if one does not repay, then future credit extension will not occur. Thus, it takes two to tango. This type of transaction between friends and family are so common, we would not generally describe them as money, because that seems odd. Similarly, borrowing a cup of sugar from your neighbor to make cookies and promising to give some of the cookies in exchange is an informal arrangement, but is still one of credit and debt and thus money. The informality of these transactions makes them difficult to enforce; hence they represent the money social relations at the base of our money hierarchy (Figure 2).

Figure 2. The Money Hierarchy



What differentiates the common and largely unregulated IOU transactions from those that are described as we move up the hierarchy is the institutional formality and enforceability of those social

relations. So the second level, contracts, might be promises to pay businesses, such as your network services provider. They promise to provide a service for a year, and you promise to pay a dollar amount for that service. They credit you with service and you pay back your debt. This contract, unlike Johnny's notebook, is legally enforceable in dollars. If you or the network provider breaks your promise legal recourse is available.

Moving on up the hierarchy, we have the financial sector and then the U.S. Treasury at the very top. The formality and institutional structure of the social relations generated by these organizations is what places them at higher levels of this hierarchy. The financial sector is similar to the contracts created below it on the hierarchy, so we will focus our attention on the U.S. Treasury. The U.S. Treasury is unique because by Constitutional Law it is the sole source of new U.S. dollar issue. This institutional reality is one of the most significant aspects of MMT because in stark contrast to the Metallist's story money comes from the state, not the market.

Figure 3.

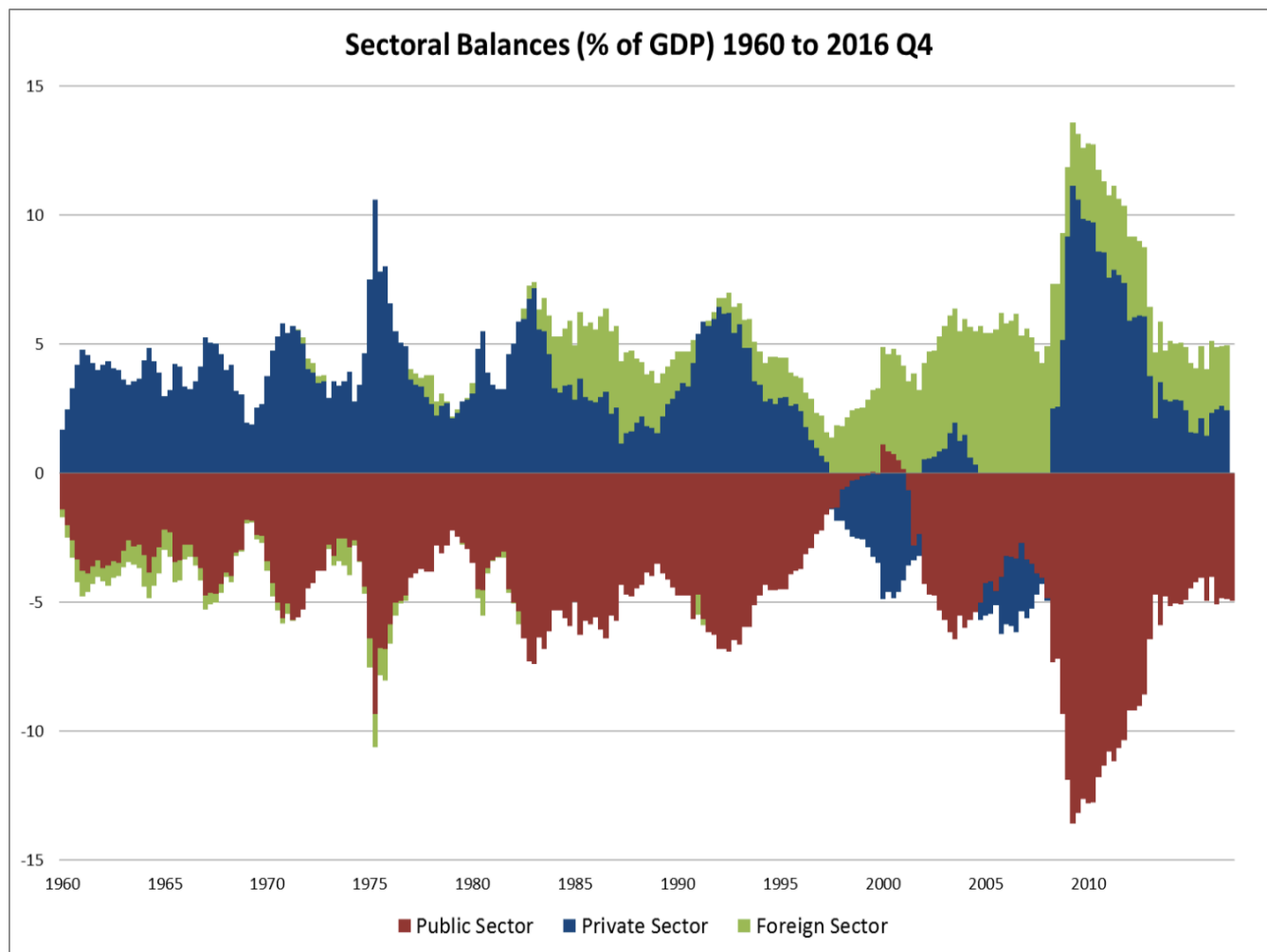


Figure (3) vividly displays the hierarchy in action. As you can see, as the red line goes into the negative (government deficit), the green line increases (private sector wealth), at a nearly 1:1 ratio. This rela-

tionship between public and private spending is described by the accounting identity equation known as **sectoral balances**.

$$0 = (S - I) + (T - G) + (\text{net exports})$$

This equation states that the balance sheets of the private, public, and international current accounts sectors of the economy sum to zero. If we were to close the economy and only have the public and private sector, then when the private sector spends more than it takes in (deficit spends), the public sector, by accounting identity, runs a surplus. This situation rarely occurs. If we examine Figure 3, we see that in the early 2000s a very brief period of government budget surplus took place, only to be followed by recession and a movement back to state deficits and private surpluses.

Before pressing forward, let's take a moment to summarize the hierarchy of money. Money is a social relation. It is a two-sided balance sheet operation or an IOU. It consists of at least two parties who agree to a credit debit relationship. The hierarchy of money is characterized by the institutional constraints and enforcement of these social relations. At the top of the hierarchy is the issuer of the currency. The currency issued in the United States is the dollar. Rather than the dollar emerging in the market to solve the double coincidence of wants, the dollar comes from the state. This leaves us with the question of value. If there is not a precious metal or intrinsic value to the dollar, then where does its value come from?

Value comes from a promise from the state to accept U.S. dollars as payment for taxes. This promise creates a demand for dollars in the United States. The demand for dollars is what maintains the dollar's value. Thus, MMT argues that taxes drive money. An important question that this raises is how does this impact the value of goods and services produced? Remember in the orthodox approach the prices of goods and services are measures of their relative values as they are traded in a barter system, but if we have a monetary production economy, then how are these activities' values determined? Much of this was clearly outlined at the enterprise level in the Costs and Prices chapter and further analyzed in the Megacorp. Here we ask the question, how does our perspective on value change when we approach money as an industry with the United States as the monopoly producer?

As the monopoly producer of the dollar, the United States government is not the pesky nemesis taking money from the market to fund its spending. Instead, private markets and the financial industry are provided with the legal authority to create dollar-denominated assets. This process greatly limits what is viewed to have value to activities that generate profits. This mode of utilizing money-issuing authority makes it very difficult for many in the economic system to participate in the dollar economy. Orthodox economics and the Metallist story argues that value creation is best handled by markets, rather than direct state spending. As the monopoly producer of the currency, it is a political decision about how money enters the system and assigns value. Many MMT theorists argue that the values of the community would be better met through a universal program for full employment. This would mean the monopoly producer would take more control over value creation to meet the needs of the public than it currently does.

The sectoral balances relationship between the public and private sector, suggests that there is an intimate relationship between the state and market activities. The approach taken by neoclassical economics suggests an antagonistic relationship based on real analysis. The foundation for this position is that the state must take money from the private sector to spend. The history and an institutional analysis of where money comes from debunk this foundation. If, as Keynes suggests, we live in a

monetary production economy and new issue comes from the state, then continuing to approach the market and state as hard duals is likely to continue to leave many outside of the dollar economy in states of unemployment and poverty.

To further examine money's ability to activate value, we will now turn to an examination of local currency systems. Local currency systems are experiments. These experiments are attempting to build an understanding of the power of our most important social technology. Remember, we have defined technology as the application of scientific or organized knowledge of practical tasks. What could be more practical a task than to understand money?

ENDNOTES

[1] Stephanie [Bell] Kelton's 2001 *Cambridge Journal of Economics* article "The Role of the State and the Hierarchy of Money" is widely considered to be a seminal piece in the development of Modern Money Theory.

25.4 LOCAL CURRENCY SYSTEMS: SOCIAL MONEY AND COMMUNITY CURRENCIES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Compare to different monetary systems
- Critique money issue and the how money can be utilized for value creation

As an alternative to centralized money issue, communities are exploring their own money issuing strategies. There are several types of local currency systems, and we will describe a couple of the most common ones' general characteristics and institutional structures below. One source that outlines the extent to which currency experimentation has grown during the era of neoliberalism is described by the 2012 *International Journal of Community Currency Research* Special Issue titled "Thirty Years of Community Currencies: A Review of Impacts, Potential and Challenges." This edition contains papers from a two-day academic conference that included scholars from several academic disciplines. Scholars in economics, history, geography, law, sociology, anthropology, and political science, etc. presented research from those disciplines individually and through interdisciplinary methodologies as well. In addition to the diversity of academic perspectives community currencies are attracting, they are also covering diverse political and geographic terrains across the planet. More than 20 different countries were represented at this conference.

One reason that such a diverse set of tools is necessary for studying community currencies is that if the Metallist's story of money is false, then the price mechanism or the invisible hand is not an adequate means of determining value. In *The Picture of Dorian Gray*, Oscar Wilde tells us that; "Nowadays people know the price of everything and the value of nothing." Community currencies are projects that work to allow communities to assign values to activities and goods that central currency schemes are not pricing properly. As such, these activities are experiments. These experiments are guided by general philosophies, such as solidarity and reciprocity. Their philosophies influence goals and the purpose of currency creation, examples include community stability, intergenerational relationship building, and strengthening of territory or stimulating economic activity. From these general goals, communities create real jobs: to promote sustainable production, support the elderly, women's programs, and other civically oriented tasks not being met by the market of fiscal sectors of the economy.[1] Thus, when value is reintroduced into a discussion about the economy, the limited toolset available under real analysis is replaced by a diversity of methodologies from several academic disciplines.

In an effort to organize the analysis of this diverse and growing collection of money systems, three categories of local currency systems have been suggested.[2] The first of these are local currencies are *territorial* projects. These projects are typically defined by geographic boundaries and are designed to strengthen the area's resilience or development. Second, *community* currencies are projects developed to address a specific community need. Such needs are usually determined by a group of local actors attempting to emphasize increased wellbeing, empowerment of a repressed group, or support environmental health. The third type is a *complementary* currency. Complementary currencies are designed to support an economic objective. These systems are generally designed with regard to market principles and work to support activities between production and exchange of goods and services. These types of currency systems are not mutually exclusive in their characteristics. In fact, many take on objectives and characteristics that draw from multiple currency types. One such currency system is the Red del Traque (RT) of Argentina.

Ecologists developed the RT currency system. In this ecological model, the wellbeing of women and families was a primary objective. To accomplish their objective, markets or 'nodos' were set up in economically depressed neighborhoods. These markets used the RT currency system to organize labor efforts and create household stability by encouraging self-production, reducing waste, promoting the resale of household items, providing social welfare labor, and helping to distribute unsold surpluses in local businesses through new trade outlets.[3] Thus, from this description, we could classify the RT in all three of the suggested category types. It is a community currency based on its primary objective. It can also be described as a complementary currency as it facilitated market exchanges, and it targeted particular territories based on economic need, territorial.

By creating a local currency system, many Argentinians who had been abandoned by the peso economy were provided an opportunity to contribute to social activity and find some economic stability. In a sense, Argentinians were able to solve Keynes's elasticity problem. There was a shortage of money in their communities, so they found a creative, socially driven solution to solve the shortage. This local currency system effectively organized available resources and produced outputs that were deemed valuable to the community.

To conclude, money takes on very different roles in our economic system based on the methodology applied to study economic activity. If it is simply the medium of exchange, then the coordination of inputs to produce outputs is the role of individual firms. However, when money is observed in its historical and institutional setting, business enterprises need money to initiate those coordinating functions, as we see in Marx's circuit of money capital. Given this central role in all economic activity, the task of economists and other scholars is to help society create laws and institutions such that money is available to initiate productive value-generating activities because merely assessing value through profits appears to be a woefully inadequate approach.

ENDNOTES

[1] This information can be found in "Community currency research: An analysis of the literature" in Vol. 15 in the 2011 *International Journal of Community Currency*.

[2] The French economists Jerome Blanc has published widely in the area of community and complementary currencies.

[3] Georgina Gomez and Bert Helmsing from the Hague in the Netherlands published a very inter-

esting article about the spatial distribution of the RT in Argentina in the 2008 Vol. 36(11) of *World Development* titled “Selective Spatial Closure and Local Economic Development.”

CHAPTER 26. INTERNATIONAL TRADE

INTRODUCTION TO INTERNATIONAL TRADE



Figure 1. Apple or Samsung iPhone? While the iPhone is readily recognized as an Apple product, 26% of the component costs in it come from components made by rival phone-maker, Samsung. In international trade, there are often “conflicts” like this as each country or company focuses on what it does best. (Credit: modification of work by Yutaka Tsutano Creative Commons)

JUST WHOSE IPHONE IS IT?

The iPhone is a global product. Apple does not manufacture the iPhone components, nor does it assemble them. The assembly is done by Foxconn Corporation, a Taiwanese company, at its factory in Sengzhen, China. But, Samsung, the electronics firm and competitor to Apple, actually supplies many of the parts that make up an iPhone—about 26%. That means, that Samsung is both the biggest supplier and biggest competitor for Apple. Why do these two firms work together to produce the iPhone? To understand the economic logic behind international trade, you have to accept, as these firms do, that trade is about mutually beneficial exchange. Samsung is one of the world’s largest electronics parts suppliers. Apple lets Samsung focus on making the best parts, which allows Apple to concentrate on its strength—designing elegant products

that are easy to use. If each company (and by extension each country) focuses on what it does best, there will be gains for all through trade.

CHAPTER OBJECTIVES

Introduction to International Trade

In this chapter, you will learn about:

- Absolute and Comparative Advantage
- What Happens When a Country Has an Absolute Advantage in All Goods
- Intra-industry Trade between Similar Economies
- The Benefits of Reducing Barriers to International Trade

We live in a global marketplace. The food on your table might include fresh fruit from Chile, cheese from France, and bottled water from Scotland. Your wireless phone might have been made in Taiwan or Korea. The clothes you wear might be designed in Italy and manufactured in China. The toys you give to a child might have come from India. The car you drive might come from Japan, Germany, or Korea. The gasoline in the tank might be refined from crude oil from Saudi Arabia, Mexico, or Nigeria. As a worker, if your job is involved with farming, machinery, airplanes, cars, scientific instruments, or many other technology-related industries, the odds are good that a hearty proportion of the sales of your employer—and hence the money that pays your salary—comes from export sales. We are all linked by international trade, and the volume of that trade has grown dramatically in the last few decades.

The first wave of **globalization** started in the nineteenth century and lasted up to the beginning of World War I. Over that time, global exports as a share of global GDP rose from less than 1% of GDP in 1820 to 9% of GDP in 1913. As the Nobel Prize-winning economist Paul Krugman of Princeton University wrote in 1995:

It is a late-twentieth-century conceit that we invented the global economy just yesterday. In fact, world markets achieved an impressive degree of integration during the second half of the nineteenth century. Indeed, if one wants a specific date for the beginning of a truly global economy, one might well choose 1869, the year in which both the Suez Canal and the Union Pacific railroad were completed. By the eve of the First World War steamships and railroads had created markets for standardized commodities, like wheat and wool, that were fully global in their reach. Even the global flow of information was better than modern observers, focused on electronic technology, tend to realize: the first submarine telegraph cable was laid under the Atlantic in 1858, and by 1900 all of the world's major economic regions could effectively communicate instantaneously.

This first wave of globalization crashed to a halt in the beginning of the twentieth century. World War I severed many economic connections. During the **Great Depression** of the 1930s, many nations misguidedly tried to fix their own economies by reducing foreign trade with others. World War II further hindered international trade. Global flows of goods and financial capital rebuilt themselves only slowly after World War II. It was not until the early 1980s that global economic forces again became as important, relative to the size of the world economy, as they were before World War I.

26.1 ABSOLUTE AND COMPARATIVE ADVANTAGE

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define absolute advantage, comparative advantage, and opportunity costs
- Explain the gains of trade created when a country specializes

The American statesman Benjamin Franklin (1706–1790) once wrote: “No nation was ever ruined by trade.” Many economists would express their attitudes toward international trade in an even more positive manner. The evidence that international trade confers overall benefits on economies is pretty strong. Trade has accompanied economic growth in the United States and around the world. Many of the national economies that have shown the most rapid growth in the last few decades—for example, Japan, South Korea, China, and India—have done so by dramatically orienting their economies toward international trade. There is no modern example of a country that has shut itself off from world trade and yet prospered. To understand the benefits of trade, or why we trade in the first place, we need to understand the concepts of comparative and absolute advantage.

In 1817, **David Ricardo**, a businessman, economist, and member of the British Parliament, wrote a treatise called *On the Principles of Political Economy and Taxation*. In this treatise, Ricardo argued that specialization and free trade benefit all trading partners, even those that may be relatively inefficient. To see what he meant, we must be able to distinguish between absolute and comparative advantage.

A country has an **absolute advantage** in producing a good over another country if it uses fewer resources to produce that good. Absolute advantage can be the result of a country’s natural endowment. For example, extracting oil in Saudi Arabia is pretty much just a matter of “drilling a hole.” Producing oil in other countries can require considerable exploration and costly technologies for drilling and extraction—if indeed they have any oil at all. The United States has some of the richest farmland in the world, making it easier to grow corn and wheat than in many other countries. Guatemala and Colombia have climates especially suited for growing coffee. Chile and Zambia have some of the world’s richest copper mines. As some have argued, “geography is destiny.” Chile will provide copper and Guatemala will produce coffee, and they will trade. When each country has a product others need and it can be produced with fewer resources in one country over another, then it is easy to imagine all parties benefitting from trade. However, thinking about trade just in terms of geography and absolute advantage is incomplete. Trade really occurs because of comparative advantage.

Recall from the chapter *Choice in a World of Scarcity* that a country has a **comparative advantage** when a good can be produced at a lower cost in terms of other goods. The question each country or

company should be asking when it trades is this: “What do we give up to produce this good?” It should be no surprise that the concept of comparative advantage is based on this idea of **opportunity cost** from Choice in a World of Scarcity. For example, if Zambia focuses its resources on producing copper, its labor, land and financial resources cannot be used to produce other goods such as corn. As a result, Zambia gives up the opportunity to produce corn. How do we quantify the cost in terms of other goods? Simplify the problem and assume that Zambia just needs labor to produce copper and corn. The companies that produce either copper or corn tell you that it takes 10 hours to mine a ton of copper and 20 hours to harvest a bushel of corn. This means the opportunity cost of producing a ton of copper is 2 bushels of corn. The next section develops absolute and comparative advantage in greater detail and relates them to trade.

Visit this website for a list of articles and podcasts pertaining to international trade topics.



A NUMERICAL EXAMPLE OF ABSOLUTE AND COMPARATIVE ADVANTAGE

Consider a hypothetical world with two countries, Saudi Arabia and the United States, and two products, oil and corn. Further assume that consumers in both countries desire both these goods. These goods are homogeneous, meaning that consumers/producers cannot differentiate between corn or oil from either country. There is only one resource available in both countries, labor hours. Saudi Arabia can produce oil with fewer resources, while the United States can produce corn with fewer resources. Table 1 illustrates the advantages of the two countries, expressed in terms of how many hours it takes to produce one unit of each good.

Country	Oil (hours per barrel)	Corn (hours per bushel)
Saudi Arabia	1	4
United States	2	1

Table 1. How Many Hours It Takes to Produce Oil and Corn

In Table 1, Saudi Arabia has an absolute advantage in the production of oil because it only takes an hour to produce a barrel of oil compared to two hours in the United States. The United States has an absolute advantage in the production of corn.

To simplify, let's say that Saudi Arabia and the United States each have 100 worker hours (see Table 2). We illustrate what each country is capable of producing on its own using a **production possibility frontier (PPF)** graph, shown in Figure 1. Recall from Choice in a World of Scarcity that the production possibilities frontier shows the maximum amount that each country can produce given its limited resources, in this case workers, and its level of technology.

Country	Oil Production using 100 worker hours (barrels)	Corn Production using 100 worker hours (bushels)
Saudi Arabia	100	or 25
United States	50	or 100

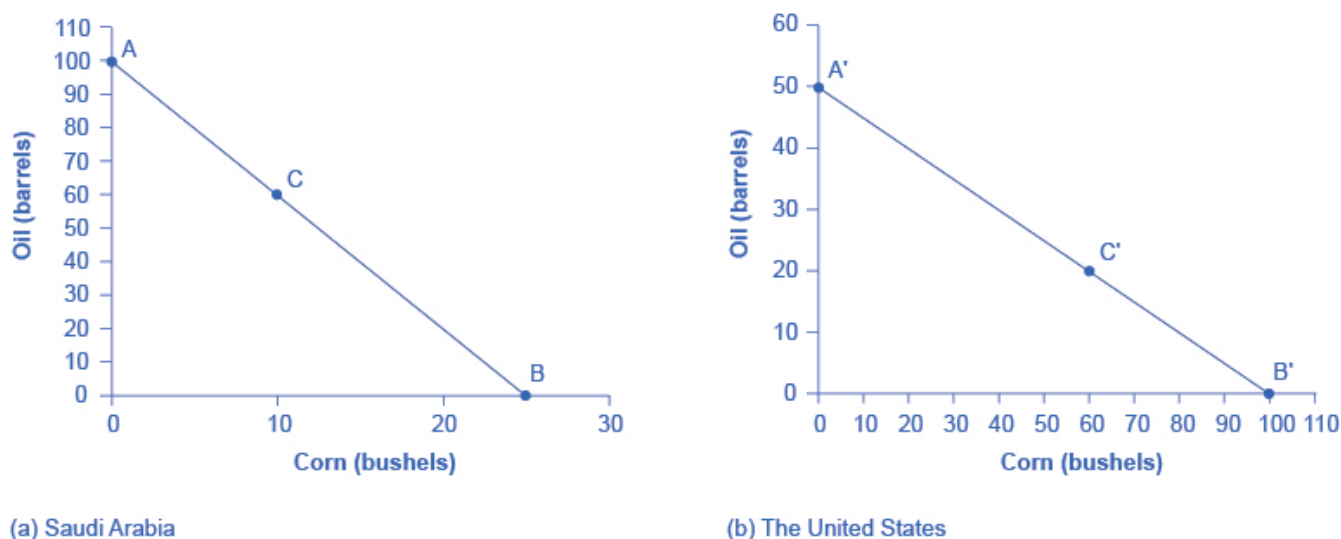
Table 2. Production Possibilities before Trade

Figure 1. Production Possibilities Frontiers. (a) Saudi Arabia can produce 100 barrels of oil at maximum and zero corn (point A), or 25 bushels of corn and zero oil (point B). It can also produce other combinations of oil and corn if it wants to consume both goods, such as at point C. Here it chooses to produce/consume 60 barrels of oil, leaving 40 work hours that can be allocated to producing 10 bushels of corn, using the data in Table 1. (b) If the United States produces only oil, it can produce, at maximum, 50 barrels and zero corn (point A'), or at the other extreme, it can produce a maximum of 100 bushels of corn and no oil (point B'). Other combinations of both oil and corn are possible, such as point C'. All points above the frontiers are impossible to produce given the current level of resources and technology.

Arguably Saudi and U.S. consumers desire both oil and corn to live. Let's say that before trade occurs, both countries produce and consume at point C or C'. Thus, before trade, the Saudi Arabian economy will devote 60 worker hours to produce oil, as shown in Table 3. Given the information in Table 1, this choice implies that it produces/consumes 60 barrels of oil. With the remaining 40 worker hours, since it needs four hours to produce a bushel of corn, it can produce only 10 bushels. To be at point C', the U.S. economy devotes 40 worker hours to produce 20 barrels of oil and the remaining worker hours can be allocated to produce 60 bushels of corn.

Country	Oil Production (barrels)	Corn Production (bushels)
Saudi Arabia (C)	60	10
United States (C')	20	60
Total World Production	80	70

Table 3. Production before Trade

The slope of the production possibility frontier illustrates the opportunity cost of producing oil in terms of corn. Using all its resources, the United States can produce 50 barrels of oil *or* 100 bushels of

corn. So the opportunity cost of one barrel of oil is two bushels of corn—or the slope is $1/2$. Thus, in the U.S. production possibility frontier graph, every increase in oil production of one barrel implies a decrease of two bushels of corn. Saudi Arabia can produce 100 barrels of oil *or* 25 bushels of corn. The opportunity cost of producing one barrel of oil is the loss of $1/4$ of a bushel of corn that Saudi workers could otherwise have produced. In terms of corn, notice that Saudi Arabia gives up the least to produce a barrel of oil. These calculations are summarized in Table 4.

Country	Opportunity cost of one unit — Oil (in terms of corn)	Opportunity cost of one unit — Corn (in terms of oil)
Saudi Arabia	$1/4$	4
United States	2	$1/2$

Table 4. Opportunity Cost and Comparative Advantage

Again recall that comparative advantage was defined as the opportunity cost of producing goods. Since Saudi Arabia gives up the least to produce a barrel of oil, ($1/4 < 2$ in Table 4) it has a comparative advantage in oil production. The United States gives up the least to produce a bushel of corn, so it has a comparative advantage in corn production.

In this example, there is symmetry between absolute and comparative advantage. Saudi Arabia needs fewer worker hours to produce oil (absolute advantage, see Table 1), and also gives up the least in terms of other goods to produce oil (comparative advantage, see Table 4). Such symmetry is not always the case, as we will show after we have discussed gains from trade fully. But first, read the following Clear It Up feature to make sure you understand why the PPF line in the graphs is straight.

CAN A PRODUCTION POSSIBILITY FRONTIER BE STRAIGHT?

When you first met the production possibility frontier (PPF) in the chapter on Choice in a World of Scarcity it was drawn with an outward-bending shape. This shape illustrated that as inputs were transferred from producing one good to another—like from education to health services—there were increasing opportunity costs. In the examples in this chapter, the PPFs are drawn as straight lines, which means that opportunity costs are constant. When a marginal unit of labor is transferred away from growing corn and toward producing oil, the decline in the quantity of corn and the increase in the quantity of oil is always the same. In reality this is possible only if the contribution of additional workers to output did not change as the scale of production changed. The linear production possibilities frontier is a less realistic model, but a straight line simplifies calculations. It also illustrates economic themes like absolute and comparative advantage just as clearly.

GAINS FROM TRADE

Consider the trading positions of the United States and Saudi Arabia after they have specialized and traded. Before trade, Saudi Arabia produces/consumes 60 barrels of oil and 10 bushels of corn. The United States produces/consumes 20 barrels of oil and 60 bushels of corn. Given their current production levels, if the United States can trade an amount of corn fewer than 60 bushels and receives in exchange an amount of oil greater than 20 barrels, it will **gain from trade**. With trade, the United States can consume more of both goods than it did without **specialization** and trade. (Recall that the chapter Welcome to Economics! defined specialization as it applies to workers and firms. Specialization is also used to describe the occurrence when a country shifts resources to focus on producing

a good that offers comparative advantage.) Similarly, if Saudi Arabia can trade an amount of oil less than 60 barrels and receive in exchange an amount of corn greater than 10 bushels, it will have more of both goods than it did before specialization and trade. Table 5 illustrates the range of trades that would benefit both sides.

The U.S. Economy, after Specialization, Will Benefit If It:	The Saudi Arabian Economy, after Specialization, Will Benefit If It:
Exports no more than 60 bushels of corn	Imports at least 10 bushels of corn
Imports at least 20 barrels of oil	Exports less than 60 barrels of oil

Table 5. The Range of Trades That Benefit Both the United States and Saudi Arabia

The underlying reason why trade benefits both sides is rooted in the concept of opportunity cost, as the following Clear It Up feature explains. If Saudi Arabia wishes to expand domestic production of corn in a world without international trade, then based on its opportunity costs it must give up four barrels of oil for every one additional bushel of corn. If Saudi Arabia could find a way to give up less than four barrels of oil for an additional bushel of corn (or equivalently, to receive more than one bushel of corn for four barrels of oil), it would be better off.

WHAT ARE THE OPPORTUNITY COSTS AND GAINS FROM TRADE?

The range of trades that will benefit each country is based on the country's opportunity cost of producing each good. The United States can produce 100 bushels of corn or 50 barrels of oil. For the United States, the opportunity cost of producing one barrel of oil is two bushels of corn. If we divide the numbers above by 50, we get the same ratio: one barrel of oil is equivalent to two bushels of corn, or $(100/50 = 2$ and $50/50 = 1)$. In a trade with Saudi Arabia, if the United States is going to give up 100 bushels of corn in exports, it must import at least 50 barrels of oil to be just as well off. Clearly, to gain from trade it needs to be able to gain more than a half barrel of oil for its bushel of corn—or why trade at all?

Recall that David Ricardo argued that if each country specializes in its comparative advantage, it will benefit from trade, and total global output will increase. How can we show gains from trade as a result of comparative advantage and specialization? Table 6 shows the output assuming that each country specializes in its comparative advantage and produces no other good. This is 100% specialization. Specialization leads to an increase in total world production. (Compare the total world production in Table 3 to that in Table 6.)

Country	Quantity produced after 100% specialization — Oil (barrels)	Quantity produced after 100% specialization — Corn (bushels)
Saudi Arabia	100	0
United States	0	100
Total World Production	100	100

Table 6. How Specialization Expands Output

What if we did not have complete specialization, as in Table 6? Would there still be gains from trade? Consider another example, such as when the United States and Saudi Arabia start at C and C', respectively, as shown in Figure 1. Consider what occurs when trade is allowed and the United States exports 20 bushels of corn to Saudi Arabia in exchange for 20 barrels of oil.

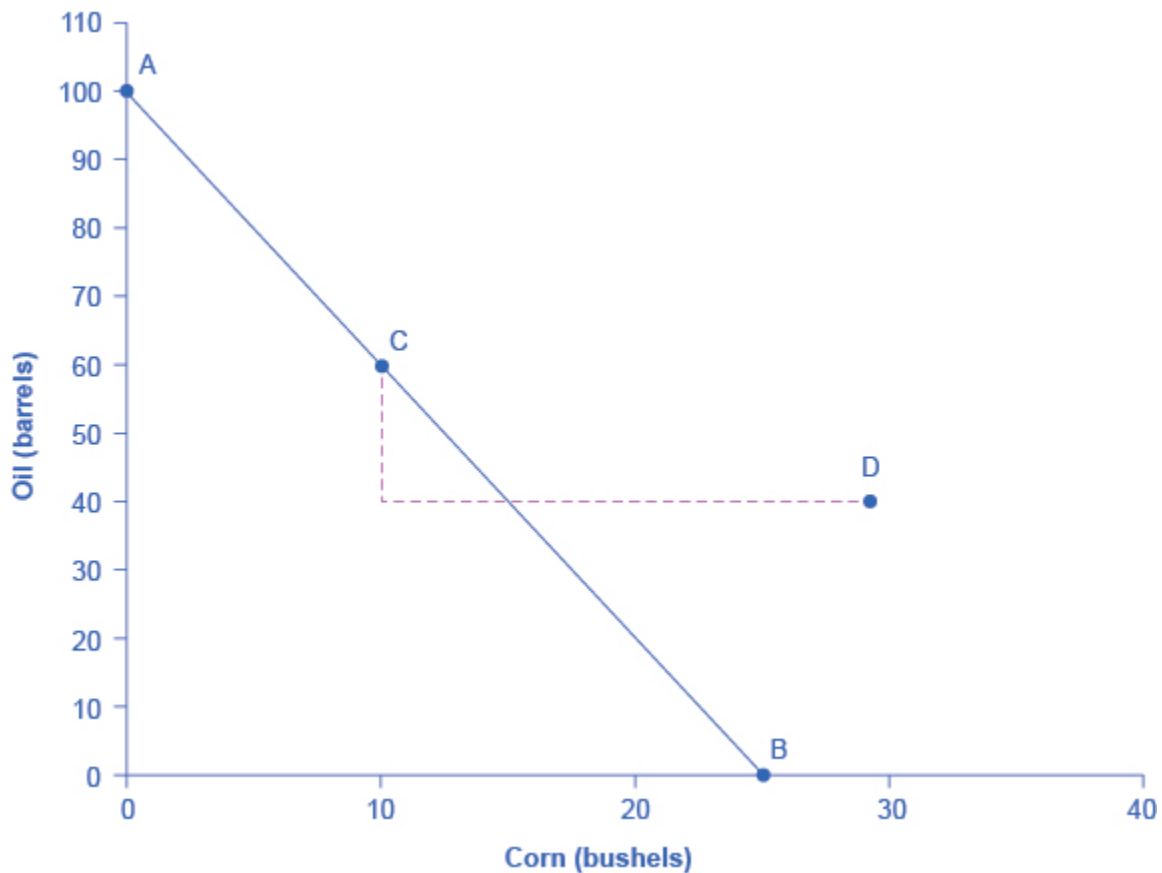


Figure 2. Production Possibilities Frontier in Saudi Arabia. Gains from trade of oil can increase only by achieving less from trade of corn. The opposite is true as well: The more gains from trade of corn, the fewer gains from trade of oil.

Starting at point C, reduce Saudi Oil production by 20 and exchange it for 20 units of corn to reach point D (see Figure 2). Notice that even without 100% specialization, if the “trading price,” in this case 20 barrels of oil for 20 bushels of corn, is greater than the country’s opportunity cost, the Saudis will gain from trade. Indeed both countries consume more of both goods after specialized production and trade occurs.

Visit this website for trade-related data visualizations.



KEY CONCEPTS AND SUMMARY

A country has an absolute advantage in those products in which it has a productivity edge over other countries; it takes fewer resources to produce a product. A country has a comparative advantage when a good can be produced at a lower cost in terms of other goods. Countries that specialize based on comparative advantage gain from trade.

SELF-CHECK QUESTIONS

1. True or False: The source of comparative advantage must be natural elements like climate and mineral deposits. Explain.
2. Brazil can produce 100 pounds of beef or 10 autos; in contrast the United States can produce 40 pounds of beef or 30 autos. Which country has the absolute advantage in beef? Which country has the absolute advantage in producing autos? What is the opportunity cost of producing one pound of beef in Brazil? What is the opportunity cost of producing one pound of beef in the United States?
3. In France it takes one worker to produce one sweater, and one worker to produce one bottle of wine. In Tunisia it takes two workers to produce one sweater, and three workers to produce one bottle of wine. Who has the absolute advantage in production of sweaters? Who has the absolute advantage in the production of wine? How can you tell?

REVIEW QUESTIONS

1. What is absolute advantage? What is comparative advantage?
2. Under what conditions does comparative advantage lead to gains from trade?
3. What factors does Paul Krugman identify that supported the expansion of international trade in the 1800s?

CRITICAL THINKING QUESTIONS

1. Are differences in geography behind the differences in absolute advantages?
2. Why does the United States not have an absolute advantage in coffee?
3. Look at Self-Check Question 3. Compute the opportunity costs of producing sweaters and wine in both France and Tunisia. Who has the lowest opportunity cost of producing sweaters and who has the lowest opportunity cost of producing wine? Explain what it means to have a lower opportunity cost.

PROBLEMS

France and Tunisia both have Mediterranean climates that are excellent for producing/harvesting green beans and tomatoes. In France it takes two hours for each worker to harvest green beans and two hours to harvest a tomato. Tunisian workers need only one hour to harvest the tomatoes but four hours to harvest green beans. Assume there are only two workers, one in each country, and each works 40 hours a week.

- a. Draw a production possibilities frontier for each country. *Hint:* Remember the production possibility frontier is the maximum that all workers can produce at a unit of time which, in this problem, is a week.
- b. Identify which country has the absolute advantage in green beans and which country has the absolute advantage in tomatoes.
- c. Identify which country has the comparative advantage.
- d. How much would France have to give up in terms of tomatoes to gain from trade? How much would it have to give up in terms of green beans?

REFERENCES

Krugman, Paul R. *Pop Internationalism*. The MIT Press, Cambridge. 1996.

Krugman, Paul R. "What Do Undergrads Need to Know about Trade?" *American Economic Review* 83, no. 2. 1993. 23-26.

Ricardo, David. *On the Principles of Political Economy and Taxation*. London: John Murray, 1817.

Ricardo, David. "On the Principles of Political Economy and Taxation." *Library of Economics and Liberty*. <http://www.econlib.org/library/Ricardo/ricP.html>.

GLOSSARY

absolute advantage when one country can use fewer resources to produce a good compared to another country; when a country is more productive compared to another country

gain from trade a country that can consume more than it can produce as a result of specialization and trade

SOLUTIONS

Answers to Self-Check Questions

1. False. Anything that leads to different levels of productivity between two economies can be a source of comparative advantage. For example, the education of workers, the knowledge base of engineers and scientists in a country, the part of a split-up value chain where they have their specialized learning, economies of scale, and other factors can all determine comparative advantage.
2. Brazil has the absolute advantage in producing beef and the United States has the absolute advantage in autos. The opportunity cost of producing one pound of beef is 1/10 of an auto; in the United States it is 3/4 of an auto.
3. In answering questions like these, it is often helpful to begin by organizing the information in a table, such as in the following table. Notice that, in this case, the productivity of the countries is expressed in terms of how many workers it takes to produce a unit of a product.

Country	One Sweater	One Bottle of wine
France	1 worker	1 worker
Tunisia	2 workers	3 workers

Table 7.

In this example, France has an absolute advantage in the production of both sweaters and wine. You can tell because it takes France less labor to produce a unit of the good.

26.2 WHAT HAPPENS WHEN A COUNTRY HAS AN ABSOLUTE ADVANTAGE IN ALL GOODS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Show the relationship between production costs and comparative advantage
- Identify situations of mutually beneficial trade
- Identify trade benefits by considering opportunity costs

What happens to the possibilities for trade if one country has an absolute advantage in everything? This is typical for high-income countries that often have well-educated workers, technologically advanced equipment, and the most up-to-date production processes. These high-income countries can produce all products with fewer resources than a low-income country. If the high-income country is more productive across the board, will there still be gains from trade? Good students of Ricardo understand that trade is about mutually beneficial exchange. Even when one country has an absolute advantage in all products, trade can still benefit both sides. This is because gains from trade come from specializing in one's comparative advantage.

PRODUCTION POSSIBILITIES AND COMPARATIVE ADVANTAGE

Consider the example of trade between the United States and Mexico described in Table 8. In this example, it takes four U.S. workers to produce 1,000 pairs of shoes, but it takes five Mexican workers to do so. It takes one U.S. worker to produce 1,000 refrigerators, but it takes four Mexican workers to do so. The United States has an absolute advantage in productivity with regard to both shoes and refrigerators; that is, it takes fewer workers in the United States than in Mexico to produce both a given number of shoes and a given number of refrigerators.

Country	Number of Workers needed to produce 1,000 units — Shoes	Number of Workers needed to produce 1,000 units — Refrigerators
United States	4 workers	1 worker
Mexico	5 workers	4 workers

Table 8. Resources Needed to Produce Shoes and Refrigerators

Absolute advantage simply compares the **productivity** of a worker between countries. It answers the question, “How many inputs do I need to produce shoes in Mexico?” Comparative advantage asks this

same question slightly differently. Instead of comparing how many workers it takes to produce a good, it asks, “How much am I giving up to produce this good in this country?” Another way of looking at this is that comparative advantage identifies the good for which the producer’s absolute advantage is relatively larger, or where the producer’s absolute productivity disadvantage is relatively smaller. The United States can produce 1,000 shoes with four-fifths as many workers as Mexico (four versus five), but it can produce 1,000 refrigerators with only one-quarter as many workers (one versus four). So, the comparative advantage of the United States, where its absolute productivity advantage is relatively greatest, lies with refrigerators, and Mexico’s comparative advantage, where its absolute productivity disadvantage is least, is in the production of shoes.

MUTUALLY BENEFICIAL TRADE WITH COMPARATIVE ADVANTAGE

When nations increase production in their area of comparative advantage and trade with each other, both countries can benefit. Again, the production possibility frontier is a useful tool to visualize this benefit.

Consider a situation where the United States and Mexico each have 40 workers. For example, as Table 9 shows, if the United States divides its labor so that 40 workers are making shoes, then, since it takes four workers in the United States to make 1,000 shoes, a total of 10,000 shoes will be produced. (If four workers can make 1,000 shoes, then 40 workers will make 10,000 shoes). If the 40 workers in the United States are making refrigerators, and each worker can produce 1,000 refrigerators, then a total of 40,000 refrigerators will be produced.

Country	Shoe Production — using 40 workers	Refrigerator Production — using 40 workers
United States	10,000 shoes	or 40,000 refrigerators
Mexico	8,000 shoes	or 10,000 refrigerators

Table 9. Production Possibilities before Trade with Complete Specialization

As always, the slope of the production possibility frontier for each country is the opportunity cost of one refrigerator in terms of foregone shoe production—when labor is transferred from producing the latter to producing the former (see Figure 1).

Let’s say that, in the situation before trade, each nation prefers to produce a combination of shoes and refrigerators that is shown at point A. Table 10 shows the output of each good for each country and the total output for the two countries.

Country	Current Shoe Production	Current Refrigerator Production
United States	5,000	20,000
Mexico	4,000	5,000
Total	9,000	25,000

Table 10. Total Production at Point A before Trade

Continuing with this scenario, each country transfers some amount of labor toward its area of comparative advantage. For example, the United States transfers six workers away from shoes and toward producing refrigerators. As a result, U.S. production of shoes decreases by 1,500 units ($6/4 \times 1,000$), while its production of refrigerators increases by 6,000 (that is, $6/1 \times 1,000$). Mexico also moves production toward its area of comparative advantage, transferring 10 workers away from refrigerators

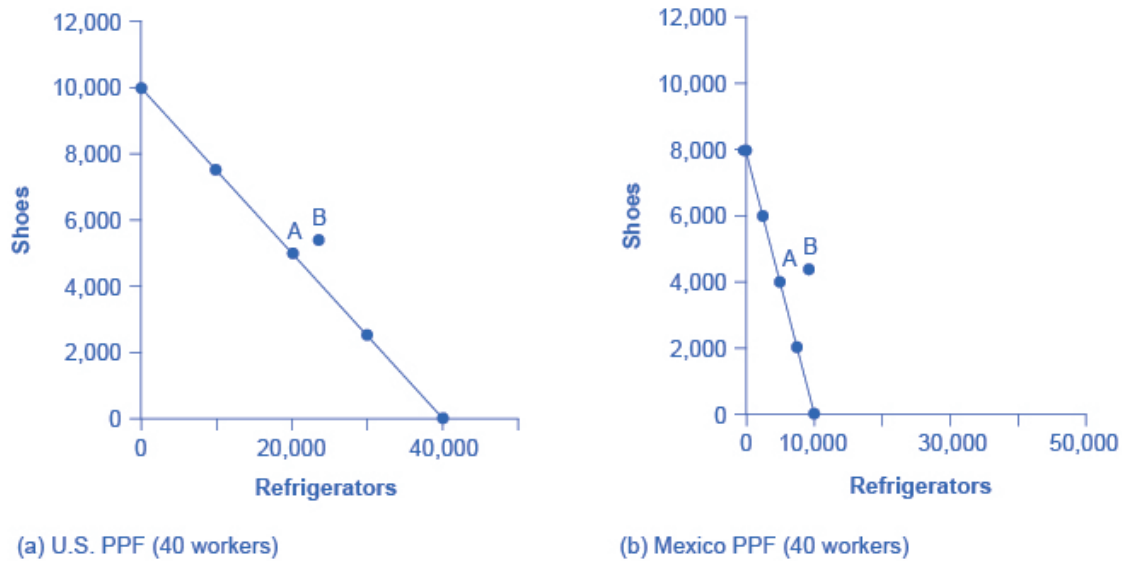


Figure 1. Production Possibility Frontiers. (a) With 40 workers, the United States can produce either 10,000 shoes and zero refrigerators or 40,000 refrigerators and zero shoes. (b) With 40 workers, Mexico can produce a maximum of 8,000 shoes and zero refrigerators, or 10,000 refrigerators and zero shoes. All other points on the production possibility line are possible combinations of the two goods that can be produced given current resources. Point A on both graphs is where the countries start producing and consuming before trade. Point B is where they end up after trade.

and toward production of shoes. As a result, production of refrigerators in Mexico falls by 2,500 ($10/4 \times 1,000$), but production of shoes increases by 2,000 pairs ($10/5 \times 1,000$). Notice that when both countries shift production toward each of their comparative advantages (what they are relatively better at), their combined production of both goods rises, as shown in Table 11. The reduction of shoe production by 1,500 pairs in the United States is more than offset by the gain of 2,000 pairs of shoes in Mexico, while the reduction of 2,500 refrigerators in Mexico is more than offset by the additional 6,000 refrigerators produced in the United States.

Country	Shoe Production	Refrigerator Production
United States	3,500	26,000
Mexico	6,000	2,500
Total	9,500	28,500

Table 11. Shifting Production Toward Comparative Advantage Raises Total Output

This numerical example illustrates the remarkable insight of comparative advantage: even when one country has an absolute advantage in all goods and another country has an absolute disadvantage in all goods, both countries can still benefit from trade. Even though the United States has an absolute advantage in producing both refrigerators and shoes, it makes economic sense for it to specialize in the good for which it has a comparative advantage. The United States will export refrigerators and in return import shoes.

HOW OPPORTUNITY COST SETS THE BOUNDARIES OF TRADE

This example shows that both parties can benefit from specializing in their comparative advantages and trading. By using the opportunity costs in this example, it is possible to identify the range of possible trades that would benefit each country.

Mexico started out, before specialization and trade, producing 4,000 pairs of shoes and 5,000 refrigerators (see Figure 1 and Table 10). Then, in the numerical example given, Mexico shifted production toward its comparative advantage and produced 6,000 pairs of shoes but only 2,500 refrigerators. Thus, if Mexico can **export** no more than 2,000 pairs of shoes (giving up 2,000 pairs of shoes) in exchange for **imports** of at least 2,500 refrigerators (a gain of 2,500 refrigerators), it will be able to consume more of both goods than before trade. Mexico will be unambiguously better off. Conversely, the United States started off, before specialization and trade, producing 5,000 pairs of shoes and 20,000 refrigerators. In the example, it then shifted production toward its comparative advantage, producing only 3,500 shoes but 26,000 refrigerators. If the United States can export no more than 6,000 refrigerators in exchange for imports of at least 1,500 pairs of shoes, it will be able to consume more of both goods and will be unambiguously better off.

The range of trades that can benefit both nations is shown in Table 12. For example, a trade where the U.S. exports 4,000 refrigerators to Mexico in exchange for 1,800 pairs of shoes would benefit both sides, in the sense that both countries would be able to consume more of both goods than in a world without trade.

The U.S. economy, after specialization, will benefit if it:	The Mexican economy, after specialization, will benefit if it:
--	---

Exports fewer than 6,000 refrigerators

Imports at least 2,500 refrigerators

Imports at least 1,500 pairs of shoes

Exports no more than 2,000 pairs of shoes

Table 12. The Range of Trades That Benefit Both the United States and Mexico

Trade allows each country to take advantage of lower opportunity costs in the other country. If Mexico wants to produce more refrigerators without trade, it must face its domestic opportunity costs and reduce shoe production. If Mexico, instead, produces more shoes and then trades for refrigerators made in the United States, where the **opportunity cost** of producing refrigerators is lower, Mexico can in effect take advantage of the lower opportunity cost of refrigerators in the United States. Conversely, when the United States specializes in its comparative advantage of refrigerator production and trades for shoes produced in Mexico, international trade allows the United States to take advantage of the lower opportunity cost of shoe production in Mexico.

The theory of comparative advantage explains why countries trade: they have different comparative advantages. It shows that the gains from international trade result from pursuing comparative advantage and producing at a lower opportunity cost. The following Work It Out feature shows how to calculate absolute and comparative advantage and the way to apply them to a country's production.

CALCULATING ABSOLUTE AND COMPARATIVE ADVANTAGE

In Canada a worker can produce 20 barrels of oil or 40 tons of lumber. In Venezuela, a worker can produce 60 barrels of oil or 30 tons of lumber.

Country	Oil (barrels)		Lumber (tons)
Canada	20	or	40
Venezuela	60	or	30

Table 13.

- Who has the absolute advantage in the production of oil or lumber? How can you tell?
- Which country has a comparative advantage in the production of oil?
- Which country has a comparative advantage in producing lumber?
- In this example, is absolute advantage the same as comparative advantage, or not?
- In what product should Canada specialize? In what product should Venezuela specialize?

Step 1. Make a table like Table 13.

Step 2. To calculate absolute advantage, look at the larger of the numbers for each product. One worker in Canada can produce more lumber (40 tons versus 30 tons), so Canada has the absolute advantage in lumber. One worker in Venezuela can produce 60 barrels of oil compared to a worker in Canada who can produce only 20.

Step 3. To calculate comparative advantage, find the opportunity cost of producing one barrel of oil in both countries. The country with the lowest opportunity cost has the comparative advantage. With the same labor time, Canada can produce either 20 barrels of oil or 40 tons of lumber. So in effect, 20 barrels of oil is equivalent to 40 tons of lumber: $20 \text{ oil} = 40 \text{ lumber}$. Divide both sides of the equation by 20 to calculate the opportunity cost of one barrel of oil in Canada. $20/20 \text{ oil} = 40/20 \text{ lumber}$. $1 \text{ oil} = 2 \text{ lumber}$. To produce one additional barrel of oil in Canada has an opportunity cost of 2 lumber. Calculate the same way for Venezuela: $60 \text{ oil} = 30 \text{ lumber}$. Divide both sides of the equation by 60. One oil in Venezuela has an opportunity cost of $1/2 \text{ lumber}$. Because $1/2 \text{ lumber} < 2 \text{ lumber}$, Venezuela has the comparative advantage in producing oil.

Step 4. Calculate the opportunity cost of one lumber by reversing the numbers, with lumber on the left side of the equation. In Canada, 40 lumber is equivalent in labor time to 20 barrels of oil: $40 \text{ lumber} = 20 \text{ oil}$. Divide each side of the equation by 40. The opportunity cost of one lumber is $1/2 \text{ oil}$. In Venezuela, the equivalent labor time will produce 30 lumber or 60 oil: $30 \text{ lumber} = 60 \text{ oil}$. Divide each side by 30. One lumber has an opportunity cost of two oil. Canada has the lower opportunity cost in producing lumber.

Step 5. In this example, absolute advantage is the same as comparative advantage. Canada has the absolute and comparative advantage in lumber; Venezuela has the absolute and comparative advantage in oil.

Step 6. Canada should specialize in what it has a relative lower opportunity cost, which is lumber, and Venezuela should specialize in oil. Canada will be exporting lumber and importing oil, and Venezuela will be exporting oil and importing lumber.

COMPARATIVE ADVANTAGE GOES CAMPING

To build an intuitive understanding of how comparative advantage can benefit all parties, set aside examples that involve national economies for a moment and consider the situation of a group of friends who decide to go camping together. The six friends have a wide range of skills and experiences, but one person in particular, Jethro, has done lots of camping before and is also a great athlete. Jethro has an absolute advantage in all aspects of camping: he is faster at carrying a backpack, gathering firewood, paddling a canoe, setting up tents, making a meal, and washing up. So here is the question: Because Jethro has an absolute productivity advantage in everything, should he do all the work?

Of course not! Even if Jethro is willing to work like a mule while everyone else sits around, he, like most mortals, only has 24 hours in a day. If everyone sits around and waits for Jethro to do everything,

not only will Jethro be an unhappy camper, but there will not be much output for his group of six friends to consume. The theory of comparative advantage suggests that everyone will benefit if they figure out their areas of comparative advantage—that is, the area of camping where their productivity disadvantage is least, compared to Jethro. For example, it may be that Jethro is 80% faster at building fires and cooking meals than anyone else, but only 20% faster at gathering firewood and 10% faster at setting up tents. In that case, Jethro should focus on building fires and making meals, and others should attend to the other tasks, each according to where their productivity disadvantage is smallest. If the campers coordinate their efforts according to comparative advantage, they can all gain.

KEY CONCEPTS AND SUMMARY

Even when a country has high levels of productivity in all goods, it can still benefit from trade. Gains from trade come about as a result of comparative advantage. By specializing in a good that it gives up the least to produce, a country can produce more and offer that additional output for sale. If other countries specialize in the area of their comparative advantage as well and trade, the highly productive country is able to benefit from a lower opportunity cost of production in other countries.

SELF-CHECK QUESTIONS

In Germany it takes three workers to make one television and four workers to make one video camera. In Poland it takes six workers to make one television and 12 workers to make one video camera.

- a. Who has the absolute advantage in the production of televisions? Who has the absolute advantage in the production of video cameras? How can you tell?
- b. Calculate the opportunity cost of producing one additional television set in Germany and in Poland. (Your calculation may involve fractions, which is fine.) Which country has a comparative advantage in the production of televisions?
- c. Calculate the opportunity cost of producing one video camera in Germany and in Poland. Which country has a comparative advantage in the production of video cameras?
- d. In this example, is absolute advantage the same as comparative advantage, or not?
- e. In what product should Germany specialize? In what product should Poland specialize?

REVIEW QUESTIONS

1. Is it possible to have a comparative advantage in the production of a good but not to have an absolute advantage? Explain.
2. How does comparative advantage lead to gains from trade?

CRITICAL THINKING QUESTIONS

1. You just overheard your friend say the following: “Poor countries like Malawi have no absolute advantages.

They have poor soil, low investments in formal education and hence low-skill workers, no capital, and no natural resources to speak of. Because they have no advantage, they cannot benefit from trade.” How would you respond?

2. Look at Table 10. Is there a range of trades for which there will be no gains?
3. You just got a job in Washington, D.C. You move into an apartment with some acquaintances. All your roommates, however, are slackers and do not clean up after themselves. You, on the other hand, can clean faster than each of them. You determine that you are 70% faster at dishes and 10% faster with vacuuming. All of these tasks have to be done daily. Which jobs should you assign to your roommates to get the most free time overall? Assume you have the same number of hours to devote to cleaning. Now, since you are faster, you seem to get done quicker than your roommate. What sorts of problems may this create? Can you imagine a trade-related analogy to this problem?

PROBLEMS

1. In Japan, one worker can make 5 tons of rubber or 80 radios. In Malaysia, one worker can make 10 tons of rubber or 40 radios.
 - a. Who has the absolute advantage in the production of rubber or radios? How can you tell?
 - b. Calculate the opportunity cost of producing 80 additional radios in Japan and in Malaysia. (Your calculation may involve fractions, which is fine.) Which country has a comparative advantage in the production of radios?
 - c. Calculate the opportunity cost of producing 10 additional tons of rubber in Japan and in Malaysia. Which country has a comparative advantage in producing rubber?
 - d. In this example, does each country have an absolute advantage and a comparative advantage in the same good?
 - e. In what product should Japan specialize? In what product should Malaysia specialize?
2. Review the numbers for Canada and Venezuela from Table 13 which describes how many barrels of oil and tons of lumber the workers can produce. Use these numbers to answer the rest of this question.
 - a. Draw a production possibilities frontier for each country. Assume there are 100 workers in each country. Canadians and Venezuelans desire both oil and lumber. Canadians want at least 2,000 tons of lumber. Mark a point on their production possibilities where they can get at least 3,000 tons.
 - b. Assume that the Canadians specialize completely because they figured out they have a comparative advantage in lumber. They are willing to give up 1,000 tons of lumber. How much oil should they ask for in return for this lumber to be as well off as they were with no trade? How much should they ask for if they want to gain from trading with Venezuela? *Note:* We can think of this “ask” as the relative price or trade price of lumber.
 - c. Is the Canadian “ask” you identified in (b) also beneficial for Venezuelans? Use the production possibilities frontier graph for Venezuela to show that Venezuelans can gain from trade.
3. In Problem 2, is there an “ask” where Venezuelans may say “no thank you” to trading with Canada?

REFERENCES

Bernstein, William J. *A Splendid Exchange: How Trade Shaped the World*. Atlantic Monthly Press. New York. 2008.

SOLUTIONS

Answers to Self-Check Questions

- a. In Germany, it takes fewer workers to make either a television or a video camera. Germany has an absolute advantage in the production of both goods.
- b. Producing an additional television in Germany requires three workers. Shifting those three German workers will reduce video camera production by $\frac{3}{4}$ of a camera. Producing an additional television set in Poland requires six workers, and shifting those workers from the other good reduces output of video cameras by $\frac{6}{12}$ of a camera, or $\frac{1}{2}$. Thus, the opportunity cost of producing televisions is lower in Poland, so Poland has the comparative advantage in the production of televisions. *Note:* Do not let the fractions like $\frac{3}{4}$ of a camera or $\frac{1}{2}$ of a video camera bother you. If either country was to expand television production by a significant amount—that is, lots more than one unit—then we will be talking about whole cameras and not fractional ones. You can also spot this conclusion by noticing that Poland's absolute disadvantage is relatively lower in televisions, because Poland needs twice as many workers to produce a television but three times as many to produce a video camera, so the product with the relatively lower absolute disadvantage is Poland's comparative advantage.
- c. Producing a video camera in Germany requires four workers, and shifting those four workers away from television production has an opportunity cost of $\frac{4}{3}$ television sets. Producing a video camera in Poland requires 12 workers, and shifting those 12 workers away from television production has an opportunity cost of two television sets. Thus, the opportunity cost of producing video cameras is lower in Germany, and video cameras will be Germany's comparative advantage.
- d. In this example, absolute advantage differs from comparative advantage. Germany has the absolute advantage in the production of both goods, but Poland has a comparative advantage in the production of televisions.
- e. Germany should specialize, at least to some extent, in the production of video cameras, export video cameras, and import televisions. Conversely, Poland should specialize, at least to some extent, in the production of televisions, export televisions, and import video cameras.

26.3 INTRA-INDUSTRY TRADE BETWEEN SIMILAR ECONOMIES

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify at least two advantages of intra-industry trading
- Explain the relationship between economies of scale and intra-industry trade

Absolute and comparative advantages explain a great deal about patterns of global trade. For example, they help to explain the patterns noted at the start of this chapter, like why you may be eating fresh fruit from Chile or Mexico, or why lower productivity regions like Africa and Latin America are able to sell a substantial proportion of their exports to higher productivity regions like the European Union and North America. Comparative advantage, however, at least at first glance, does not seem especially well-suited to explain other common patterns of international trade.

THE PREVALENCE OF INTRA-INDUSTRY TRADE BETWEEN SIMILAR ECONOMIES

The theory of comparative advantage suggests that trade should happen between economies with large differences in opportunity costs of production. Roughly half of all world trade involves shipping goods between the fairly similar high-income economies of the United States, Canada, the European Union, Japan, Mexico, and China (see Table 14).

Country	U.S. Exports Go to ...	U.S. Imports Come from ...
European Union	19.0%	21.0%
Canada	22.0%	14.0%
Japan	4.0%	6.0%
Mexico	15.0%	13.0%
China	8.0%	20.0%

Table 14. Where U.S. Exports Go and U.S. Imports Originate (2015) (Source: https://www.census.gov/foreign-trade/Press-Release/current_press_release/ft900.pdf)

Moreover, the theory of comparative advantage suggests that each economy should specialize to a degree in certain products, and then exchange those products. A high proportion of trade, however, is **intra-industry trade**—that is, trade of goods within the same industry from one country to another. For example, the United States produces and exports autos and imports autos. Table 15 shows some of the largest categories of U.S. exports and imports. In all of these categories, the United States is both

a substantial exporter and a substantial importer of goods from the same industry. In 2014, according to the Bureau of Economic Analysis, the United States exported \$159 billion worth of autos, and imported \$327 billion worth of autos. About 60% of U.S. trade and 60% of European trade is intra-industry trade.

Some U.S. Exports	Quantity of Exports (\$ billions)	Quantity of Imports (\$ billions)
Autos	\$146	\$327
Food and beverages	\$144	\$126
Capital goods	\$550	\$551
Consumer goods	\$199	\$558
Industrial supplies	\$507	\$665
Other transportation	\$45	\$55

Table 15. Some Intra-Industry U.S. Exports and Imports in 2014 (Source: <http://www.bea.gov/newsreleases/international/trade/tradnewsrelease.htm>)

Why do similar high-income economies engage in intra-industry trade? What can be the economic benefit of having workers of fairly similar skills making cars, computers, machinery and other products which are then shipped across the oceans to and from the United States, the European Union, and Japan? There are two reasons: (1) The **division of labor** leads to learning, innovation, and unique skills; and (2) economies of scale.

GAINS FROM SPECIALIZATION AND LEARNING

Consider the category of machinery, where the U.S. economy has considerable intra-industry trade. Machinery comes in many varieties, so the United States may be exporting machinery for manufacturing with wood, but importing machinery for photographic processing. The underlying reason why a country like the United States, Japan, or Germany produces one kind of machinery rather than another is usually not related to U.S., German, or Japanese firms and workers having generally higher or lower skills. It is just that, in working on very specific and particular products, firms in certain countries develop unique and different skills.

Specialization in the world economy can be very finely split. In fact, recent years have seen a trend in international trade called **splitting up the value chain**. The **value chain** describes how a good is produced in stages. As indicated in the beginning of the chapter, the production of the iPhone involves the design and engineering of the phone in the United States, parts supplied from Korea, the assembly of the parts in China, and the advertising and marketing done in the United States. Thanks in large part to improvements in communication technology, sharing information, and transportation, it has become easier to split up the value chain. Instead of production in a single large factory, all of these steps can be split up among different firms operating in different places and even different countries. Because firms split up the value chain, international trade often does not involve whole finished products like automobiles or refrigerators being traded between nations. Instead, it involves shipping more specialized goods like, say, automobile dashboards or the shelving that fits inside refrigerators. Intra-industry trade between similar countries produces economic gains because it allows workers and firms to learn and innovate on particular products—and often to focus on very particular parts of the value chain.

Visit this website for some interesting information about the assembly of the iPhone.



ECONOMIES OF SCALE, COMPETITION, VARIETY

A second broad reason that intra-industry trade between similar nations produces economic gains involves economies of scale. The concept of **economies of scale**, as introduced in Cost and Industry Structure, means that as the scale of output goes up, average costs of production decline—at least up to a point. Figure 1 illustrates economies of scale for a plant producing toaster ovens. The horizontal axis of the figure shows the quantity of production by a certain firm or at a certain manufacturing plant. The vertical axis measures the average cost of production. Production plant S produces a small level of output at 30 units and has an average cost of production of \$30 per toaster oven. Plant M produces at a medium level of output at 50 units, and has an average cost of production of \$20 per toaster oven. Plant L produces 150 units of output with an average cost of production of only \$10 per toaster oven. Although plant V can produce 200 units of output, it still has the same unit cost as Plant L.

In this example, a small or medium plant, like S or M, will not be able to compete in the market with a large or a very large plant like L or V, because the firm that operates L or V will be able to produce and sell their output at a lower price. In this example, economies of scale operate up to point L, but beyond point L to V, the additional scale of production does not continue to reduce average costs of production.

The concept of economies of scale becomes especially relevant to international trade when it enables one or two large producers to supply the entire country. For example, a single large automobile factory could probably supply all the cars purchased in a smaller economy like the United Kingdom or Belgium in a given year. However, if a country has only one or two large factories producing cars, and no **international trade**, then consumers in that country would have relatively little choice between kinds of cars (other than the color of the paint and other nonessential options). Little or no competition will exist between different car manufacturers.

International trade provides a way to combine the lower average production costs that come from economies of scale and still have competition and variety for consumers. Large automobile factories in different countries can make and sell their products around the world. If the U.S. automobile market was made up of only General Motors, Ford, and Chrysler, the level of competition and consumer choice would be quite a lot lower than when U.S. carmakers must face competition from Toyota, Honda, Suzuki, Fiat, Mitsubishi, Nissan, Volkswagen, Kia, Hyundai, BMW, Subaru, and others. Greater competition brings with it innovation and responsiveness to what consumers want. America's car producers make far better cars now than they did several decades ago, and much of the reason is competitive pressure, especially from East Asian and European carmakers.

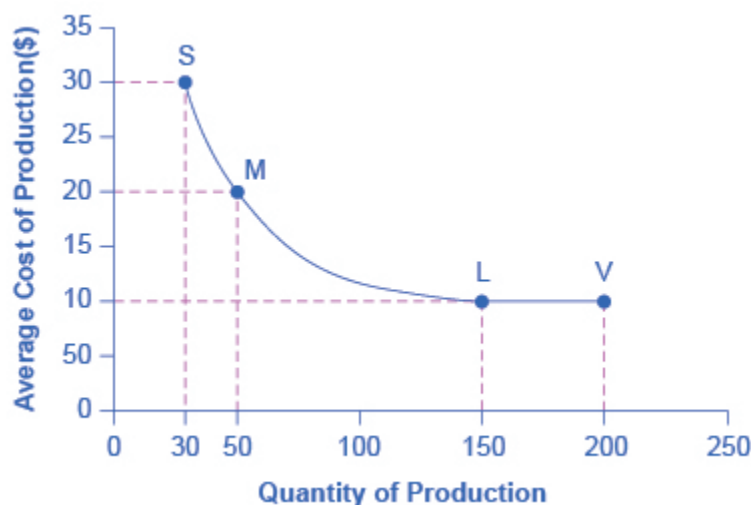


Figure 1. Economies of Scale. Production Plant S, has an average cost of production of \$30 per toaster oven. Production plant M has an average cost of production of \$20 per toaster oven. Production plant L has an average cost of production of only \$10 per toaster oven. Production plant V would still have an average cost of production of \$10 per toaster oven. Thus, production plant M can produce toaster ovens more cheaply than plant S because of economies of scale, and plants L or V can produce more cheaply than S or M because of economies of scale. However, the economies of scale end at an output level of 150. Plant V, despite being larger, cannot produce more cheaply on average than plant L.

DYNAMIC COMPARATIVE ADVANTAGE

The sources of gains from intra-industry trade between similar economies—namely, the learning that comes from a high degree of specialization and splitting up the value chain and from economies of scale—do not contradict the earlier theory of comparative advantage. Instead, they help to broaden the concept.

In intra-industry trade, the level of worker productivity is not determined by climate or geography. It is not even determined by the general level of education or skill. Instead, the level of worker productivity is determined by how firms engage in specific learning about specialized products, including taking advantage of economies of scale. In this vision, comparative advantage can be dynamic—that is, it can evolve and change over time as new skills are developed and as the value chain is split up in new ways. This line of thinking also suggests that countries are not destined to have the same comparative advantage forever, but must instead be flexible in response to ongoing changes in comparative advantage.

KEY CONCEPTS AND SUMMARY

A large share of global trade happens between high-income economies that are quite similar in having well-educated workers and advanced technology. These countries practice intra-industry trade, in

which they import and export the same products at the same time, like cars, machinery, and computers. In the case of intra-industry trade between economies with similar income levels, the gains from trade come from specialized learning in very particular tasks and from economies of scale. Splitting up the value chain means that several stages of producing a good take place in different countries around the world.

SELF-CHECK QUESTIONS

1. How can there be any economic gains for a country from both importing and exporting the same good, like cars?
2. Table 16 shows how the average costs of production for semiconductors (the “chips” in computer memories) change as the quantity of semiconductors built at that factory increases.
 - a. Based on these data, sketch a curve with quantity produced on the horizontal axis and average cost of production on the vertical axis. How does the curve illustrate economies of scale?
 - b. If the equilibrium quantity of semiconductors demanded is 90,000, can this economy take full advantage of economies of scale? What about if quantity demanded is 70,000 semiconductors? 50,000 semiconductors? 30,000 semiconductors?
 - c. Explain how international trade could make it possible for even a small economy to take full advantage of economies of scale, while also benefiting from competition and the variety offered by several producers.

Quantity of Semiconductors	Average Total Cost
10,000	\$8 each
20,000	\$5 each
30,000	\$3 each
40,000	\$2 each
100,000	\$2 each

Table 16.

REVIEW QUESTIONS

1. What is intra-industry trade?
2. What are the two main sources of economic gains from intra-industry trade?
3. What is splitting up the value chain?

CRITICAL THINKING QUESTIONS

1. Does intra-industry trade contradict the theory of comparative advantage?
2. Do consumers benefit from intra-industry trade?
3. Why might intra-industry trade seem surprising from the point of view of comparative advantage?

PROBLEMS

1. From earlier chapters you will recall that technological change shifts the average cost curves. Draw a graph showing how technological change could influence intra-industry trade.
2. Consider two countries: South Korea and Taiwan. Taiwan can produce one million mobile phones per day at the cost of \$10 per phone and South Korea can produce 50 million mobile phones at \$5 per phone. Assume these phones are the same type and quality and there is only one price. What is the minimum price at which both countries will engage in trade?

REFERENCES

U.S. Census Bureau. 2015. "U.S. International Trade in Goods and Services: December 2014." Accessed April 13, 2015. <http://www.bea.gov/newsreleases/international/trade/2015/pdf/trad1214.pdf>.

U.S. Census Bureau. U.S. Bureau of Economic Analysis. 2015. "U.S. International Trade in Goods and Services February 2015." Accessed April 10, 2015. https://www.census.gov/foreign-trade/Press-Release/current_press_release/ft900.pdf.

Vernengo, Matias. "What Do Undergraduates Really Need to Know About Trade and Finance?" in *Political Economy and Contemporary Capitalism: Radical Perspectives on Economic Theory and Policy*, ed. Ron Baiman, Heather Boushey, and Dawn Saunders. M. E. Sharpe Inc, 2000. Armonk. 177-183.

GLOSSARY

intra-industry trade international trade of goods within the same industry

splitting up the value chain many of the different stages of producing a good happen in different geographic locations

value chain how a good is produced in stages

SOLUTIONS

Answers to Self-Check Questions

1. There are a number of possible advantages of intra-industry trade. Both nations can take advantage of extreme specialization and learning in certain kinds of cars with certain traits, like gas-efficient cars, luxury cars, sport-utility vehicles, higher- and lower-quality cars, and so on. Moreover, nations can take advantage of economies of scale, so that large companies will compete against each other across international borders, providing the benefits of competition and variety to customers. This same argument applies to trade between U.S. states, where people often buy products made by people of other states, even though a similar product is made within the boundaries of their own state. All states—and all countries—can benefit from this kind of competition and trade.
2.
 - a. Start by plotting the points on a sketch diagram and then drawing a line through them. The following figure illustrates the average costs of production of semiconductors.

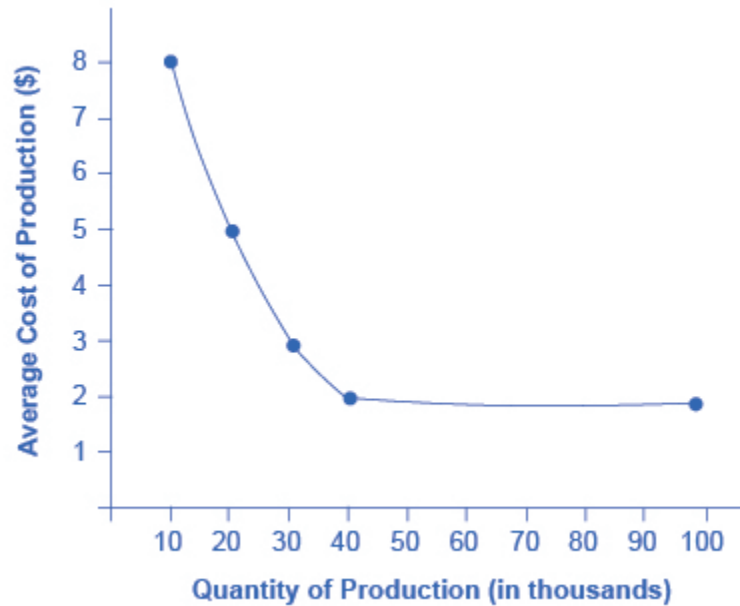


Figure 2.

The curve illustrates economies of scale by showing that as the scale increases—that is, as production at this particular factory goes up—the average cost of production declines. The economies of scale exist up to an output of 40,000 semiconductors; at higher outputs, the average cost of production does not seem to decline any further.

- b. At any quantity demanded above 40,000, this economy can take full advantage of economies of scale; that is, it can produce at the lowest cost per unit. Indeed, if the quantity demanded was quite high, like 500,000, then there could be a number of different factories all taking full advantage of economies of scale and competing with each other. If the quantity demanded falls below 40,000, then the economy by itself, without foreign trade, cannot take full advantage of economies of scale.
- c. The simplest answer to this question is that the small country could have a large enough factory to take full advantage of economies of scale, but then export most of the output. For semiconductors, countries like Taiwan and Korea have recently fit this description. Moreover, this country could also import semiconductors from other countries which also have large factories, thus getting the benefits of competition and variety. A slightly more complex answer is that the country can get these benefits of economies of scale without producing semiconductors, but simply by buying semiconductors made at low cost around the world. An economy, especially a smaller country, may well end up specializing and producing a few items on a large scale, but then trading those items for other items produced on a large scale, and thus gaining the benefits of economies of scale by trade, as well as by direct production.

26.4 THE BENEFITS OF REDUCING BARRIERS TO INTERNATIONAL TRADE

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain tariffs as barriers to trade
- Identify at least two benefits of reducing barriers to international trade

Tariffs are taxes that governments place on imported goods for a variety of reasons. Some of these reasons include protecting sensitive industries, for humanitarian reasons, and protecting against **dumping**. Traditionally, tariffs were used simply as a political tool to protect certain vested economic, social, and cultural interests. The **World Trade Organization (WTO)** is committed to lowering barriers to trade. The world's nations meet through the WTO to negotiate how they can reduce barriers to trade, such as tariffs. WTO negotiations happen in "rounds," where all countries negotiate one agreement to encourage trade, take a year or two off, and then start negotiating a new agreement. The current round of negotiations is called the Doha Round because it was officially launched in Doha, the capital city of Qatar, in November 2001. In 2009, economists from the World Bank summarized recent research and found that the Doha round of negotiations would increase the size of the world economy by \$160 billion to \$385 billion per year, depending on the precise deal that ended up being negotiated.

In the context of a global economy that currently produces more than \$30 trillion of goods and services each year, this amount is not huge: it is an increase of 1% or less. But before dismissing the gains from trade too quickly, it is worth remembering two points.

- First, a gain of a few hundred billion dollars is enough money to deserve attention! Moreover, remember that this increase is not a one-time event; it would persist each year into the future.
- Second, the estimate of gains may be on the low side because some of the gains from trade are not measured especially well in economic statistics. For example, it is difficult to measure the potential advantages to consumers of having a variety of products available and a greater degree of competition among producers. Perhaps the most important unmeasured factor is that trade between countries, especially when firms are splitting up the value chain of production, often involves a transfer of knowledge that can involve skills in production, technology, management, finance, and law.

Low-income countries benefit more from trade than high-income countries do. In some ways, the giant U.S. economy has less need for international trade, because it can already take advantage of internal trade within its economy. However, many smaller national economies around the world, in regions like Latin America, Africa, the Middle East, and Asia, have much more limited possibilities for trade inside their countries or their immediate regions. Without international trade, they may have little ability to benefit from comparative advantage, slicing up the value chain, or economies of scale. Moreover, smaller economies often have fewer competitive firms making goods within their economy, and thus firms have less pressure from other firms to provide the goods and prices that consumers want.

The economic gains from expanding international trade are measured in hundreds of billions of dollars, and the gains from international trade as a whole probably reach well into the trillions of dollars. The potential for gains from trade may be especially high among the smaller and lower-income countries of the world.

Visit this website for a list of some benefits of trade.



FROM INTERPERSONAL TO INTERNATIONAL TRADE

Most people find it easy to believe that they, personally, would not be better off if they tried to grow and process all of their own food, to make all of their own clothes, to build their own cars and houses from scratch, and so on. Instead, we all benefit from living in economies where people and firms can specialize and trade with each other.

The benefits of trade do not stop at national boundaries, either. Earlier we explained that the division of labor could increase output for three reasons: (1) workers with different characteristics can specialize in the types of production where they have a comparative advantage; (2) firms and workers who specialize in a certain product become more productive with learning and practice; and (3) economies of scale. These three reasons apply from the individual and community level right up to the international level. If it makes sense to you that interpersonal, intercommunity, and interstate trade offer economic gains, it should make sense that international trade offers gains, too.

International trade currently involves about \$20 trillion worth of goods and services moving around the globe. Any economic force of that size, even if it confers overall benefits, is certain to cause disruption and controversy. This chapter has only made the case that trade brings economic benefits. Other chapters discuss, in detail, the public policy arguments over whether to restrict international trade.

IT'S APPLE'S (GLOBAL) IPHONE

Apple Corporation uses a global platform to produce the iPhone. Now that you understand the concept of comparative advantage, you can see why the engineering and design of the iPhone is done in the United States. The United States has built up a comparative advantage over the years in designing and marketing products, and sacrifices fewer resources to design high-tech devices relative to other countries. China has a comparative advantage in assembling the phone due to its large skilled labor force. Korea has a comparative advantage in producing components. Korea focuses its production by increasing its scale, learning better ways to produce screens and computer chips, and uses innovation to lower average costs of production. Apple, in turn, benefits because it can purchase these quality products at lower prices. Put the global assembly line together and you have the device with which we are all so familiar.

KEY CONCEPTS AND SUMMARY

Tariffs are placed on imported goods as a way of protecting sensitive industries, for humanitarian reasons, and for protection against dumping. Traditionally, tariffs were used as a political tool to protect certain vested economic, social, and cultural interests. The WTO has been, and continues to be, a way for nations to meet and negotiate through barriers to trade. The gains of international trade are very large, especially for smaller countries, but are beneficial to all.

SELF-CHECK QUESTIONS

If the removal of trade barriers is so beneficial to international economic growth, why would a nation continue to restrict trade on some imported or exported products?

REVIEW QUESTIONS

Are the gains from international trade more likely to be relatively more important to large or small countries?

CRITICAL THINKING QUESTIONS

1. In World Trade Organization meetings, what do you think low-income countries lobby for?
2. Why might a low-income country put up barriers to trade, such as tariffs on imports?
3. Can a nation's comparative advantage change over time? What factors would make it change?

PROBLEMS

If trade increases world GDP by 1% per year, what is the global impact of this increase over 10 years? How does this increase compare to the annual GDP of a country like Sri Lanka? Discuss. *Hint:* To answer this question, here are steps you may want to consider. Go to the World Development Indicators (online) published by the World Bank. Find the

current level of World GDP in constant international dollars. Also, find the GDP of Sri Lanka in constant international dollars. Once you have these two numbers, compute the amount the additional increase in global incomes due to trade and compare that number to Sri Lanka's GDP.

REFERENCES

World Trade Organization. "The Doha Round." Accessed October 2013. http://www.wto.org/english/tratop_e/dda_e/dda_e.htm.

The World Bank. "Data: World Development Indicators." Accessed October 2013. <http://data.worldbank.org/data-catalog/world-development-indicators>.

GLOSSARY

tariffs taxes that governments place on imported goods

SOLUTIONS

Answers to Self-Check Questions

A nation might restrict trade on imported products to protect an industry that is important for national security. For example, nation X and nation Y may be geopolitical rivals, each with ambitions of increased political and economic strength. Even if nation Y has comparative advantage in the production of missile defense systems, it is unlikely that nation Y would seek to export those goods to nation X. It is also the case that, for some nations, the production of a particular good is a key component of national identity. In Japan, the production of rice is culturally very important. It may be difficult for Japan to import rice from a nation like Vietnam, even if Vietnam has a comparative advantage in rice production.

CHAPTER 27. GLOBALIZATION AND PROTECTIONISM

INTRODUCTION TO GLOBALIZATION AND PROTECTIONISM



Figure 1. Flat Screen Competition. The market for flat-panel displays in the United States is huge. The manufacturers of flat screens in the United States must compete against manufacturers from around the world.
(Credit: modification of work by “Jemimus”/Flickr Creative Commons)

WHAT'S THE DOWNSIDE OF PROTECTION?

Governments are motivated to limit and alter market outcomes for political or social ends. While governments can limit the rise in prices of some products, they cannot control how much people want to buy or how much firms are willing to sell. The laws of demand and supply still hold. Trade policy is an example where regulations can redirect economic forces, but it cannot stop them from manifesting themselves elsewhere.

Flat-panel displays, the displays for laptop computers, tablets, and flat screen televisions, are an example of such an enduring principle. In the early 1990s, the vast majority of flat-panel displays used in U.S.-manufactured laptops were imported, primarily from Japan. The small but politically powerful U.S. flat-panel-display industry filed a dumping complaint with the Commerce Department. They argued that Japanese firms were selling displays at “less than fair value,” which made it difficult for U.S. firms to compete. This argument for trade protection is referred to as anti-dumping. Other arguments for protection in this complaint included national security. After a preliminary determination by the Commerce Department that the Japanese firms were dumping, the U.S. International Trade Commission imposed a 63% dumping margin (or tax)

on the import of flat-panel displays. Was this a successful exercise of U.S. trade policy? See what you think after reading the chapter.

CHAPTER OBJECTIVES

Introduction to Globalization and Protectionism

In this chapter, you will learn about:

- Protectionism: An Indirect Subsidy from Consumers to Producers
- International Trade and Its Effects on Jobs, Wages, and Working Conditions
- Arguments in Support of Restricting Imports
- How Trade Policy Is Enacted: Globally, Regionally, and Nationally
- The Tradeoffs of Trade Policy

The world has become more connected on multiple levels, especially economically. In 1970, imports and exports made up 11% of U.S. GDP, while now they make up 32%. However, the United States, due to its size, is less internationally connected than most countries. For example, according to the World Bank, 97% of Botswana's economic activity is connected to trade. This chapter explores trade policy—the laws and strategies a country uses to regulate international trade. This topic is not without controversy.

As the world has become more globally connected, firms and workers in high-income countries like the United States, Japan, or the nations of the European Union, perceive a competitive threat from firms in medium-income countries like Mexico, China, or South Africa, that have lower costs of living and therefore pay lower wages. Firms and workers in low-income countries fear that they will suffer if they must compete against more productive workers and advanced technology in high-income countries.

On a different tack, some environmentalists worry that multinational firms may evade environmental protection laws by moving their production to countries with loose or nonexistent pollution standards, trading a clean environment for jobs. Some politicians worry that their country may become overly dependent on key imported products, like oil, which in a time of war could threaten national security. All of these fears influence governments to reach the same basic policy conclusion: to protect national interests, whether businesses, jobs, or security, imports of foreign products should be restricted. This chapter analyzes such arguments. First, however, it is essential to learn a few key concepts and understand how the demand and supply model applies to international trade.

27.1 PROTECTIONISM: AN INDIRECT SUBSIDY FROM CONSUMERS TO PRODUCERS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain protectionism and its three main forms
- Analyze protectionism through concepts of demand and supply, noting its effects on equilibrium
- Calculate the effects of trade barriers

When a government legislates policies to reduce or block international trade it is engaging in **protectionism**. Protectionist policies often seek to shield domestic producers and domestic workers from foreign competition. Protectionism takes three main forms: tariffs, import quotas, and nontariff barriers.

Recall from International Trade that **tariffs** are taxes imposed on imported goods and services. They make imports more expensive for consumers, discouraging imports. For example, in recent years large, flat-screen televisions imported from China have faced a 5% tariff rate.

Another way to control trade is through **import quotas**, which are numerical limitations on the quantity of products that can be imported. For instance, during the early 1980s, the Reagan Administration imposed a quota on the import of Japanese automobiles. In the 1970s, many developed countries, including the United States, found themselves with declining textile industries. Textile production does not require highly skilled workers, so producers were able to set up lower-cost factories in developing countries. In order to “manage” this loss of jobs and income, the developed countries established an international Multifiber Agreement that essentially divided up the market for textile exports between importers and the remaining domestic producers. The agreement, which ran from 1974 to 2004, specified the exact quota of textile imports that each developed country would accept from each low-income country. A similar story exists for sugar imports into the United States, which are still governed by quotas.

Nontariff barriers are all the other ways that a nation can draw up rules, regulations, inspections, and paperwork to make it more costly or difficult to import products. A rule requiring certain safety standards can limit imports just as effectively as high tariffs or low import quotas, for instance. There are also nontariff barriers in the form of “rules-of-origin” regulations- these rules describe the “Made in Country X” label as the one in which the last substantial change in the product took place. A manufacturer wishing to evade import restrictions may try to change the production process so that the last

big change in the product happens in his or her own country. For example, certain textiles are made in the United States, shipped to other countries, combined with textiles made in those other countries to make apparel—and then re-exported back to the United States for a final assembly, to escape paying tariffs or to obtain a “Made in the USA” label.

Despite import quotas, tariffs, and nontariff barriers, the share of apparel sold in the United States that is imported rose from about half in 1999 to about three-quarters today. The U.S. Bureau of Labor Statistics (BLS), estimated the number of U.S. jobs in textiles and apparel fell from 666,360 in 2007 to 385,240 in 2012, a 42% decline. Even more U.S. textile industry jobs would have been lost without tariffs, however, domestic jobs that are saved by import quotas come at a cost. Because textile and apparel protectionism adds to the costs of imports, consumers end up paying billions of dollars more for clothing each year.

When the United States eliminates trade barriers in one area, consumers spend the money they save on that product elsewhere in the economy—so there is no overall loss of jobs for the economy as a whole. Of course, workers in some of the poorest countries of the world who would otherwise have jobs producing textiles, would gain considerably if the United States reduced its barriers to trade in textiles. That said, there are good reasons to be wary about reducing barriers to trade. The 2012 and 2013 Bangladeshi fires in textile factories, which resulted in a horrific loss of life, present complications that our simplified analysis in the chapter will not capture.

Realizing the compromises between nations that come about due to trade policy, many countries came together in 1947 to form the General Agreement on Tariffs and Trade (GATT). (We’ll cover the GATT in more detail later in the chapter.) This agreement has since been superseded by the **World Trade Organization (WTO)**, whose membership includes about 150 nations and most of the economies of the world. It is the primary international mechanism through which nations negotiate their trade rules—including rules about tariffs, quotas, and nontariff barriers. The next section examines the results of such protectionism and develops a simple model to show the impact of trade policy.

DEMAND AND SUPPLY ANALYSIS OF PROTECTIONISM

To the non-economist, restricting imports may appear to be nothing more than taking sales from foreign producers and giving them to domestic producers. Other factors are at work, however, because firms do not operate in a vacuum. Instead, firms sell their products either to consumers or to other firms (if they are business suppliers), who are also affected by the trade barriers. A demand and supply analysis of protectionism shows that it is not just a matter of domestic gains and foreign losses, but a policy that imposes substantial domestic costs as well.

Consider two countries, Brazil and the United States, who produce sugar. Each country has a domestic supply and demand for sugar, as detailed in Table 1 and illustrated in Figure 1. In Brazil, without trade, the **equilibrium** price of sugar is 12 cents per pound and the equilibrium output is 30 tons. When there is no trade in the United States, the equilibrium price of sugar is 24 cents per pound and the equilibrium quantity is 80 tons. These equilibrium points are labeled with the point E.

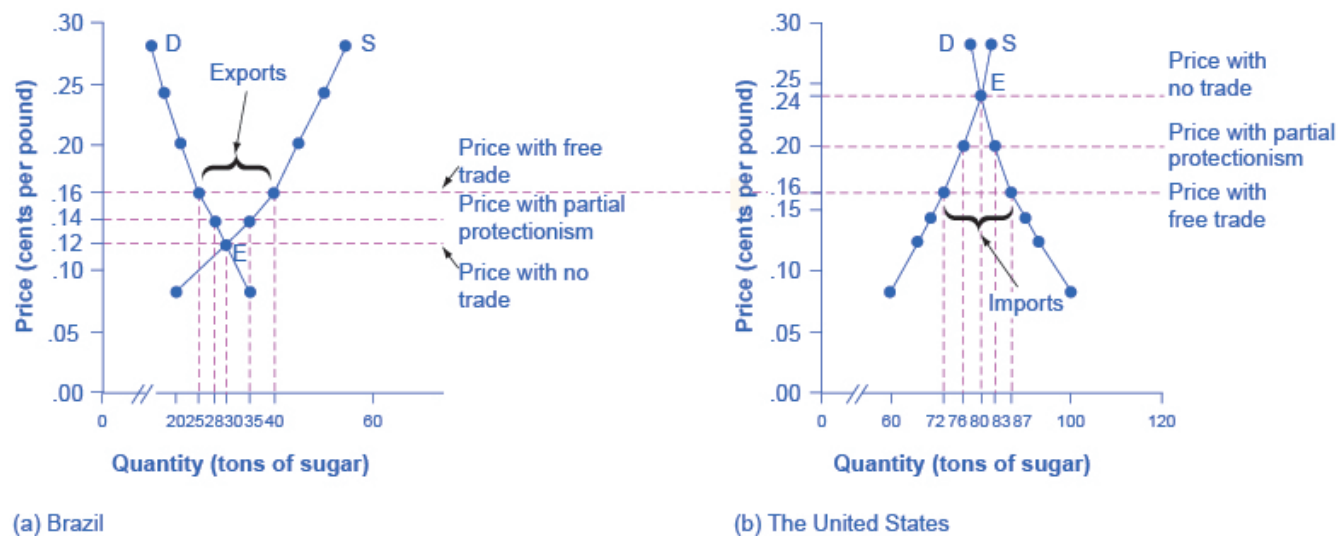


Figure 1. The Sugar Trade between Brazil and the United States. Before trade, the equilibrium price of sugar in Brazil is 12 cents a pound and for 24 cents per pound in the United States. When trade is allowed, businesses will buy cheap sugar in Brazil and sell it in the United States. This will result in higher prices in Brazil and lower prices in the United States. Ignoring transaction costs, prices should converge to 16 cents per pound, with Brazil exporting 15 tons of sugar and the United States importing 15 tons of sugar. If trade is only partly open between the countries, it will lead to an outcome between the free-trade and no-trade possibilities.

The Sugar Trade between Brazil and the United States

Price	Brazil: Quantity Supplied (tons)	Brazil: Quantity Demanded (tons)	U.S.: Quantity Supplied (tons)	U.S.: Quantity Demanded (tons)
8 cents	20	35	60	100
12 cents	30	30	66	93
14 cents	35	28	69	90
16 cents	40	25	72	87
20 cents	45	21	76	83
24 cents	50	18	80	80
28 cents	55	15	82	78

Table 1. The Sugar Trade between Brazil and the United States

If international trade between Brazil and the United States now becomes possible, profit-seeking firms will spot an opportunity: buy sugar cheaply in Brazil, and sell it at a higher price in the United States. As sugar is shipped from Brazil to the United States, the quantity of sugar produced in Brazil will be greater than Brazilian consumption (with the extra production being exported), and the amount produced in the United States will be less than the amount of U.S. consumption (with the extra consumption being imported). Exports to the United States will reduce the supply of sugar in

Brazil, raising its price. Imports into the United States will increase the supply of sugar, lowering its price. When the price of sugar is the same in both countries, there is no incentive to trade further. As Figure 1 shows, the equilibrium with trade occurs at a price of 16 cents per pound. At that price, the sugar farmers of Brazil supply a quantity of 40 tons, while the consumers of Brazil buy only 25 tons.

The extra 15 tons of sugar production, shown by the horizontal gap between the **demand curve** and the **supply curve** in Brazil, is exported to the United States. In the United States, at a price of 16 cents, the farmers produce a quantity of 72 tons and consumers demand a quantity of 87 tons. The excess demand of 15 tons by American consumers, shown by the horizontal gap between demand and domestic supply at the price of 16 cents, is supplied by imported sugar. Free trade typically results in income distribution effects, but the key is to recognize the overall gains from trade, as shown in Figure 2. Building on the concepts outlined in Demand and Supply and Demand, Supply, and Efficiency in terms of consumer and producer surplus, Figure 2 (a) shows that producers in Brazil gain by selling more sugar at a higher price, while Figure 2 (b) shows consumers in the United States benefit from the lower price and greater availability of sugar. Consumers in Brazil are worse off (compare their no-trade consumer surplus with the free-trade consumer surplus) and U.S. producers of sugar are worse off. There are gains from trade—an increase in social surplus in each country. That is, both the United States and Brazil are better off than they would be without trade. The following Clear It Up feature explains how trade policy can influence low-income countries.

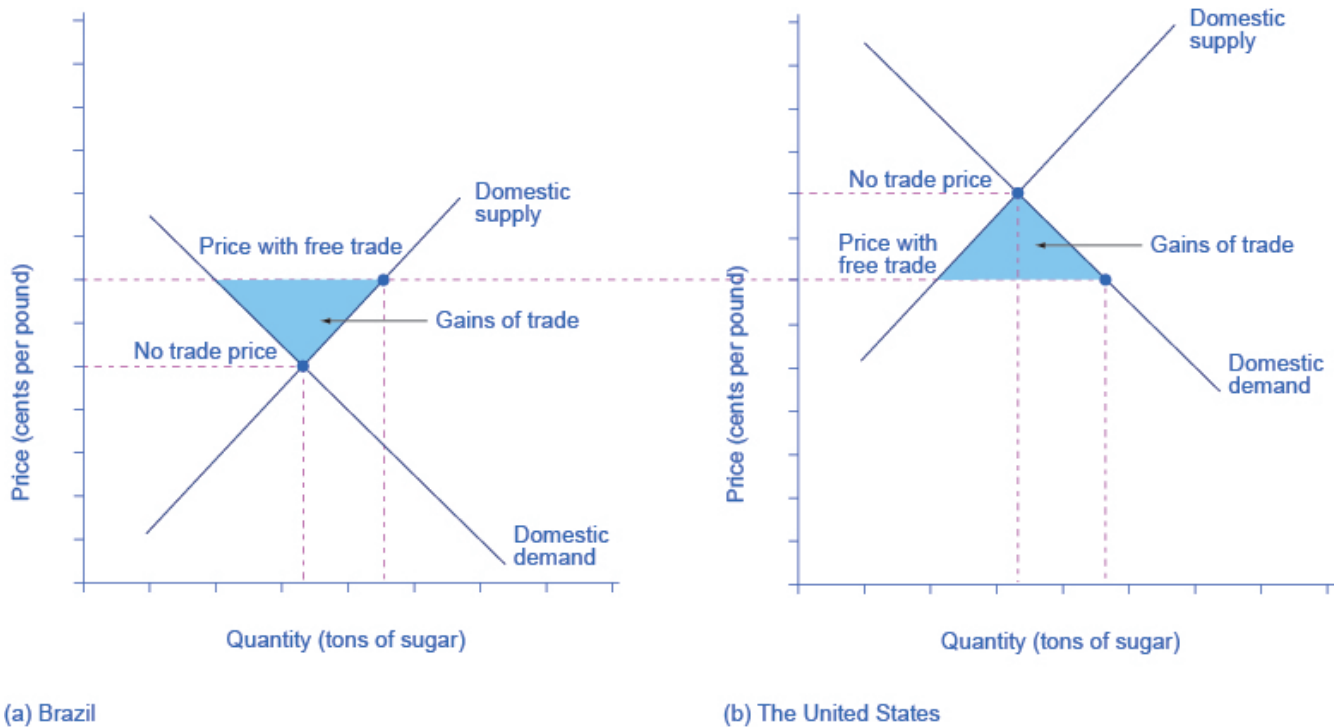


Figure 2. Free Trade of Sugar. Free trade results in gains from trade. Total surplus increases in both countries. However, there are clear income distribution effects.

Visit this website to read more about the global sugar trade.



WHY ARE THERE LOW-INCOME COUNTRIES?

Why are the poor countries of the world poor? There are a number of reasons, but one of them will surprise you: the trade policies of the high-income countries. Following is a stark review of social priorities which has been widely publicized by the international aid organization, **Oxfam International**.

High-income countries of the world—primarily the United States, Canada, countries of the European Union, and Japan—subsidize their domestic farmers collectively by about \$360 billion per year. By contrast, the total amount of foreign aid from these same high-income countries to the poor countries of the world is about \$70 billion per year, or less than 20% of the farm subsidies. Why does this matter?

It matters because the support of farmers in high-income countries is devastating to the livelihoods of farmers in **low-income countries**. Even when their climate and land are well-suited to products like cotton, rice, sugar, or milk, farmers in low-income countries find it difficult to compete. Farm **subsidies** in the high-income countries cause farmers in those countries to increase the amount they produce. This increase in supply drives down world prices of farm products below the costs of production. As Michael Gerson of the *Washington Post* describes it: “[T]he effects in the cotton-growing regions of West Africa are dramatic . . . keep[ing] millions of Africans on the edge of malnutrition. In some of the poorest countries on Earth, cotton farmers are some of the poorest people, earning about a dollar a day. . . . Who benefits from the current system of subsidies? About 20,000 American cotton producers, with an average annual income of more than \$125,000.”

As if subsidies were not enough, often, the high-income countries block agricultural exports from low-income countries. In some cases, the situation gets even worse when the governments of high-income countries, having bought and paid for an excess supply of farm products, give away those products in poor countries and drive local farmers out of business altogether.

For example, shipments of excess milk from the European Union to Jamaica have caused great hardship for Jamaican dairy farmers. Shipments of excess rice from the United States to Haiti drove thousands of low-income rice farmers in Haiti out of business. The opportunity costs of protectionism are not paid just by domestic consumers, but also by foreign producers—and for many agricultural products, those foreign producers are the world’s poor.

Now, let’s look at what happens with **protectionism**. U.S. sugar farmers are likely to argue that, if only they could be protected from sugar imported from Brazil, the United States would have higher domestic sugar production, more jobs in the sugar industry, and American sugar farmers would receive a higher price. If the United States government sets a high-enough tariff on imported sugar, or sets an import quota at zero, the result will be that the quantity of sugar traded between countries could be reduced to zero, and the prices in each country will return to the levels before trade was allowed.

Blocking only some trade is also possible. Suppose that the United States passed a sugar import quota of seven tons. The United States will import no more than seven tons of sugar, which means that

Brazil can export no more than seven tons of sugar to the United States. As a result, the price of sugar in the United States will be 20 cents, which is the price where the quantity demanded is seven tons greater than the domestic quantity supplied. Conversely, if Brazil can export only seven tons of sugar, then the price of sugar in Brazil will be 14 cents per pound, which is the price where the domestic quantity supplied in Brazil is seven tons greater than domestic demand.

In general, when a country sets a low or medium tariff or import quota, the equilibrium price and quantity will be somewhere between no trade and completely free trade. The following Work It Out explores the impact of these trade barriers.

EFFECTS OF TRADE BARRIERS

Let's look carefully at the effects of tariffs or quotas. If the U.S. government imposes a tariff or quota sufficient to eliminate trade with Brazil, two things occur: U.S. consumers pay a higher price and therefore buy a smaller quantity of sugar. U.S. producers obtain a higher price so they sell a larger quantity of sugar. The effects of a tariff on producers and consumers in the United States can be measured using two concepts developed in Demand, Supply, and Efficiency: **consumer surplus** and **producer surplus**.

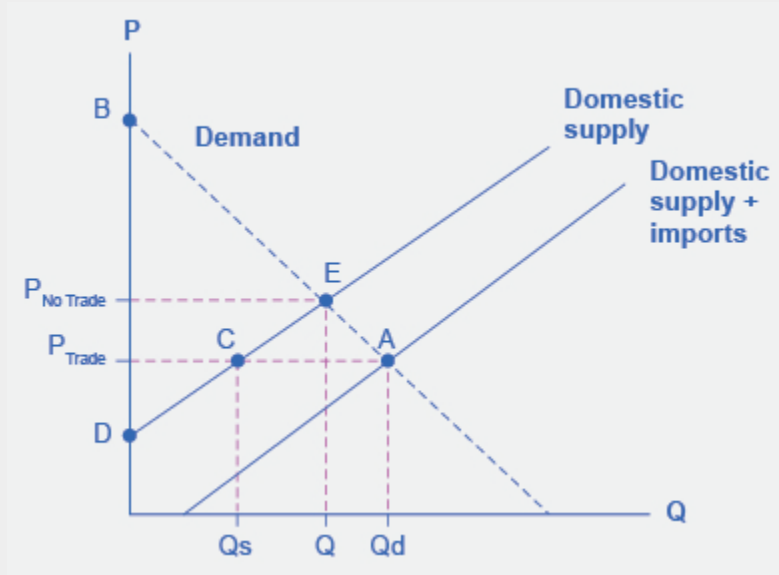


Figure 3. U.S. Sugar Supply and Demand. When there is free trade, the equilibrium is at point A. When there is no trade, the equilibrium is at point E.

Step 1. Look at Figure 3, which shows a hypothetical version of the demand and supply of sugar in the United States.

Step 2. Note that the sugar market is in equilibrium at point A where Domestic Quantity Demanded (Q_d) = Quantity Supplied (Domestic Q_s + Imports from Brazil) at a price of P_{Trade} when there is free trade.

Step 3. Note, also, that imports are equal to the distance between points C and A.

Step 4. Recall that consumer surplus is the value a consumer gets beyond what they paid for when they buy a product. Graphically, it is the area under a demand curve but above the price. In this case, the consumer surplus in the United States is the area of the triangle formed by the points P_{Trade} , A, and B.

Step 5. Recall, also, that producer surplus is another name for profit—it is the income producers get above the cost of production, which is shown by the supply curve here. In this case, the producer surplus with trade is the area of the triangle formed by the points P_{Trade} , C, and D.

Step 6. Suppose that the barriers to trade are imposed, imports are excluded, and the price rises to P_{NoTrade} . Look what happens to producer surplus and consumer surplus. At the higher price, the domestic quantity supplied increases from Q_s to Q at point E. Because producers are selling more quantity at a higher price, the producer surplus increases to the area of the triangle P_{NoTrade} , E, and D.

Step 7. Compare the areas of the two triangles and you will see the increase in the producer surplus.

Step 8. Examine the consumer surplus. Consumers are now paying a higher price to get a lower quantity (Q instead of Q_d). Their consumer surplus shrinks to the area of the triangle P_{NoTrade} , E, and B.

Step 9. Determine the net effect. The producer surplus increases by the area P_{trade} , C, E, P_{NoTrade} . The loss of consumer surplus, however, is larger. It is the area P_{trade} , A, E, P_{NoTrade} . In other words, consumers lose more than producers gain as a result of the trade barriers and the United States has a lower social surplus.

WHO BENEFITS AND WHO PAYS?

Using the demand and supply model, consider the impact of protectionism on producers and consumers in each of the two countries. For protected producers like U.S. sugar farmers, restricting imports is clearly positive. Without a need to face imported products, these producers are able to sell more, at a higher price. For consumers in the country with the protected good, in this case U.S. sugar consumers, restricting imports is clearly negative. They end up buying a lower quantity of the good and paying a higher price for what they do buy, compared to the equilibrium price and quantity without trade. The following Clear It Up feature considers why a country might outsource jobs even for a domestic product.

WHY ARE LIFE SAVERS, AN AMERICAN PRODUCT, NOT MADE IN AMERICA?

Life Savers, the hard candy with the hole in the middle, were invented in 1912 by Clarence Crane in Cleveland, Ohio. Starting in the late 1960s and for 35 years afterward, 46 billion Life Savers a year, in 200 million rolls, were produced by a plant in Holland, Michigan. But in 2002, the Kraft Company announced that the Michigan plant would be closed and Life Saver production moved across the border to Montreal, Canada.

One reason is that Canadian workers are paid slightly less, especially in healthcare and insurance costs that are not linked to employment there. Another main reason is that the United States government keeps the price of sugar high for the benefit of sugar farmers, with a combination of a government price floor program and strict quotas on imported sugar. According to the Coalition for Sugar Reform, from 2009 to 2012, the price of refined sugar in the United States ranged from 64% to 92% higher than the world price. Life Saver production uses over 100 tons of sugar each day, because the candies are 95% sugar.

A number of other candy companies have also reduced U.S. production and expanded foreign production. Indeed, from 1997 to 2011, some 127,000 jobs in the sugar-using industries, or more than seven times the total employment in sugar production, were eliminated. While the candy industry is especially affected by the cost of sugar, the costs are spread more broadly. U.S. consumers pay roughly \$1 billion per year in higher food prices because of elevated sugar costs. Meanwhile, sugar producers in low-income countries are driven out of business. Because of the sugar subsidies to domestic producers and the quotas on imports, they cannot sell their output profitably, or at all, in the United States market.

The fact that protectionism pushes up prices for consumers in the country enacting such protectionism is not always acknowledged openly, but it is not disputed. After all, if protectionism did not benefit domestic producers, there would not be much point in enacting such policies in the first place. Protectionism is simply a method of requiring consumers to subsidize producers. The subsidy is indi-

rect, since it is paid by consumers through higher prices, rather than a direct subsidy paid by the government with money collected from taxpayers. But protectionism works like a subsidy, nonetheless. The American satirist Ambrose Bierce defined “tariff” this way in his 1911 book, *The Devil’s Dictionary*: “Tariff, n. A scale of taxes on imports, designed to protect the domestic producer against the greed of his consumer.”

The effect of protectionism on producers and consumers in the foreign country is complex. When an import quota is used to impose partial protectionism, the sugar producers of Brazil receive a lower price for the sugar they sell in Brazil—but a higher price for the sugar they are allowed to export to the United States. Indeed, notice that some of the burden of protectionism, paid by domestic consumers, ends up in the hands of foreign producers in this case. Brazilian sugar consumers seem to benefit from U.S. protectionism, because it reduces the price of sugar that they pay. On the other hand, at least some of these Brazilian sugar consumers also work as sugar farmers, so their incomes and jobs are reduced by protectionism. Moreover, if trade between the countries vanishes, Brazilian consumers would miss out on better prices for imported goods—which do not appear in our single-market example of sugar protectionism.

The effects of protectionism on foreign countries notwithstanding, protectionism requires domestic consumers of a product (consumers may include either households or other firms) to pay higher prices to benefit domestic producers of that product. In addition, when a country enacts protectionism, it loses the economic gains it would have been able to achieve through a combination of comparative advantage, specialized learning, and economies of scale, concepts discussed in International Trade.

KEY CONCEPTS AND SUMMARY

There are three tools for restricting the flow of trade: tariffs, import quotas, and nontariff barriers. When a country places limitations on imports from abroad, regardless of whether it uses tariffs, quotas, or nontariff barriers, it is said to be practicing protectionism. Protectionism will raise the price of the protected good in the domestic market, which causes domestic consumers to pay more, but domestic producers to earn more.

SELF-CHECK QUESTIONS

1. Explain how a tariff reduction causes an increase in the equilibrium quantity of imports and a decrease in the equilibrium price. *Hint: Consider the Work It Out “Effects of Trade Barriers.”*
2. Explain how a subsidy on agricultural goods like sugar adversely affects the income of foreign producers of imported sugar.

REVIEW QUESTIONS

1. Who does protectionism protect? What does it protect them from?
2. Name and define three policy tools for enacting protectionism.
3. How does protectionism affect the price of the protected good in the domestic market?

CRITICAL THINKING QUESTIONS

1. Show graphically that for any tariff, there is an equivalent quota that would give the same result. What would be the difference, then, between the two types of trade barriers? *Hint:* It is not something you can see from the graph.
2. From the Work It Out “Effects of Trade Barriers,” you can see that a tariff raises the price of imports. What is interesting is that the price rises by less than the amount of the tariff. Who pays the rest of the tariff amount? Can you show this graphically?

PROBLEMS

Assume two countries, Thailand (T) and Japan (J), have one good: cameras. The demand (d) and supply (s) for cameras in Thailand and Japan is described by the following functions:

$$Qd^T = 60 - P$$

$$Qs^T = -5 + \frac{1}{4}P$$

$$Qd^J = 80 - P$$

$$Qs^J = -10 + \frac{1}{2}P$$

P is the price measured in a common currency used in both countries, such as the Thai Baht.

- a. Compute the equilibrium price (P) and quantities (Q) in each country without trade.
- b. Now assume that free trade occurs. The free-trade price goes to 56.36 Baht. Who exports and imports cameras and in what quantities?

REFERENCES

Bureau of Labor Statistics. “Industries at a Glance.” Accessed December 31, 2013. <http://www.bls.gov/iag/>.

Oxfam International. Accessed January 6, 2014. <http://www.oxfam.org/>.

GLOSSARY

import quotas numerical limits on the quantity of products that can be imported

nontariff barriers ways a nation can draw up rules, regulations, inspections, and paperwork to make it more costly or difficult to import products

protectionism government policies to reduce or block imports

World Trade Organization (WTO) organization that seeks to negotiate reductions in barriers to

trade and to adjudicate complaints about violations of international trade policy; successor to the General Agreement on Tariffs and Trade (GATT)

SOLUTIONS

Answers to Self-Check Questions

1. This is the opposite case of the Work It Out feature. A reduced tariff is like a decrease in the cost of production, which is shown by a downward (or rightward) shift in the supply curve.
2. A subsidy is like a reduction in cost. This shifts the supply curve down (or to the right), driving the price of sugar down. If the subsidy is large enough, the price of sugar can fall below the cost of production faced by foreign producers, which means they will lose money on any sugar they produce and sell.

27.2 INTERNATIONAL TRADE AND ITS EFFECTS ON JOBS, WAGES, AND WORKING CONDITIONS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Discuss how international trade influences the job market
- Analyze the opportunity cost of protectionism
- Explain how international trade impacts wages, labor standards, and working conditions

In theory at least, imports might injure workers in several different ways: fewer jobs, lower wages, or poor working conditions. Let's consider these in turn.

FEWER JOBS?

In the early 1990s, the United States was negotiating the **North American Free Trade Agreement (NAFTA)** with Mexico, an agreement that reduced tariffs, import quotas, and nontariff barriers to trade between the United States, Mexico, and Canada. H. Ross Perot, a 1992 candidate for U.S. president, claimed, in prominent campaign arguments, that if the United States expanded trade with Mexico, there would be a “giant sucking sound” as U.S. employers relocated to Mexico to take advantage of lower wages. After all, average wages in Mexico were, at that time, about one-eighth of those in the United States. NAFTA passed Congress, President Bill Clinton signed it into law, and it took effect in 1995. For the next six years, the United States economy had some of the most rapid job growth and low unemployment in its history. Those who feared that open trade with Mexico would lead to a dramatic decrease in jobs were proven wrong.

This result was no surprise to economists. After all, the trend toward globalization has been going on for decades, not just since NAFTA. If trade did reduce the number of available jobs, then the United States should have been seeing a steady loss of jobs for decades. While the United States economy does experience rises and falls in unemployment rates—according to the Bureau of Labor Statistics, from spring 2008 to late 2009, the unemployment rate rose from 4.4% to 10%; it has since fallen back to 5.5% in spring 2015—the number of jobs is not falling over extended periods of time. The number of U.S. jobs rose from 71 million in 1970 to 138 million in 2012.

Protectionism certainly saves jobs in the specific industry being protected but, for two reasons, it costs jobs in other unprotected industries. First, if consumers are paying higher prices to the protected industry, they inevitably have less money to spend on goods from other industries, and so jobs

are lost in those other industries. Second, if the protected product is sold to other firms, so that other firms must now pay a higher price for a key input, then those firms will lose sales to foreign producers who do not need to pay the higher price. Lost sales translate into lost jobs. The hidden **opportunity cost** of using protectionism to save jobs in one industry is jobs sacrificed in other industries. This is why the United States International Trade Commission, in its study of barriers to trade, predicts that reducing trade barriers would not lead to an overall loss of jobs. Protectionism reshuffles jobs from industries without import protections to industries that are protected from imports, but it does not create more jobs.

Moreover, the costs of saving jobs through protectionism can be very high. A number of different studies have attempted to estimate the cost to consumers in higher prices per job saved through protectionism. Table 2 shows a sample of results, compiled by economists at the Federal Reserve Bank of Dallas. Saving a job through protectionism typically costs much more than the actual worker's salary. For example, a study published in 2002 compiled evidence that using protectionism to save an average job in the textile and apparel industry would cost \$199,000 per job saved. In other words, those workers could have been paid \$100,000 per year to be unemployed and the cost would only be half of what it is to keep them working in the textile and apparel industry. This result is not unique to textiles and apparel.

Industry Protected with Import Tariffs or Quotas	Annual Cost per Job Saved
Sugar	\$826,000
Polyethylene resins	\$812,000
Dairy products	\$685,000
Frozen concentrated orange juice	\$635,000
Ball bearings	\$603,000
Machine tools	\$479,000
Women's handbags	\$263,000
Glassware	\$247,000
Apparel and textiles	\$199,000
Rubber footwear	\$168,000
Women's nonathletic footwear	\$139,000

Table 2. Cost to U.S. Consumers of Saving a Job through Protectionism (Source: Federal Reserve Bank of Dallas)

Why does it cost so much to save jobs through protectionism? The basic reason is that not all of the extra money paid by consumers because of tariffs or quotas goes to save jobs. For example, if tariffs are imposed on steel imports so that buyers of steel pay a higher price, U.S. steel companies earn greater profits, buy more equipment, pay bigger bonuses to managers, give pay raises to existing employees—and also avoid firing some additional workers. Only part of the higher price of protected steel goes toward saving jobs. Also, when an industry is protected, the economy as a whole loses the benefits of playing to its comparative advantage—in other words, producing what it is best at. So, part of the higher price that consumers pay for protected goods is lost economic efficiency, which can be measured as another deadweight loss, like that discussed in Labor and Financial Markets.

There's a bumper sticker that speaks to the threat some U.S. workers feel from imported products: "Buy American—Save U.S. Jobs." If the car were being driven by an economist, the sticker might

declare: “Block Imports—Save Jobs for Some Americans, Lose Jobs for Other Americans, and Also Pay High Prices.”

TRADE AND WAGES

Even if trade does not reduce the number of jobs, it could affect wages. Here, it is important to separate issues about the average level of wages from issues about whether the wages of certain workers may be helped or hurt by trade.

Because trade raises the amount that an economy can produce by letting firms and workers play to their comparative advantage, trade will also cause the average level of wages in an economy to rise. Workers who can produce more will be more desirable to employers, which will shift the demand for their labor out to the right, and increase wages in the labor market. By contrast, barriers to trade will reduce the average level of wages in an economy.

However, even if trade increases the overall wage level, it will still benefit some workers and hurt others. Workers in industries that are confronted by competition from imported products may find that demand for their labor decreases and shifts back to the left, so that their wages decline with a rise in international trade. Conversely, workers in industries that benefit from selling in global markets may find that demand for their labor shifts out to the right, so that trade raises their wages.

View this website to read an article on the issues surrounding fair trade coffee.



One concern is that while **globalization** may be benefiting high-skilled, high-wage workers in the United States, it may also impose costs on low-skilled, low-wage workers. After all, high-skilled U.S. workers presumably benefit from increased sales of sophisticated products like computers, machinery, and pharmaceuticals in which the United States has a **comparative advantage**. Meanwhile, low-skilled U.S. workers must now compete against extremely low-wage workers worldwide for making simpler products like toys and clothing. As a result, the wages of low-skilled U.S. workers are likely to fall. There are, however, a number of reasons to believe that while globalization has helped some U.S. industries and hurt others, it has not focused its negative impact on the wages of low-skilled Americans. First, about half of U.S. trade is intra-industry trade. That means the U.S. trades similar goods with other high-wage economies like Canada, Japan, Germany, and the United Kingdom. For instance, in 2014 the U.S. exported over 2 million cars, from all the major automakers, and also imported several million cars from other countries.

Most U.S. workers in these industries have above-average skills and wages—and many of them do quite well in the world of globalization. Some evidence suggested that intra-industry trade between

similar countries had a small impact on domestic workers but later evidence indicates that it all depends on how flexible the labor market is. In other words, the key is how flexible workers are in finding jobs in different industries. Trade on low-wage workers depends a lot on the structure of labor markets and indirect effects felt in other parts of the economy. For example, in the United States and the United Kingdom, because labor market frictions are low, the impact of trade on low income workers is small.

Second, many low-skilled U.S. workers hold service jobs that cannot be replaced by imports from low-wage countries. For example, lawn care services or moving and hauling services or hotel maids cannot be imported from countries long distances away like China or Bangladesh. Competition from imported products is not the primary determinant of their wages.

Finally, while the focus of the discussion here is on wages, it is worth pointing out that low-wage U.S. workers suffer due to protectionism in all the industries—even those that they do not work in the U.S. For example, food and clothing are protected industries. These low-wage workers therefore pay higher prices for these basic necessities and as such their dollar stretches over fewer goods.

The benefits and costs of increased trade in terms of its effect on wages are not distributed evenly across the economy. However, the growth of international trade has helped to raise the productivity of U.S. workers as a whole—and thus helped to raise the average level of wages.

LABOR STANDARDS AND WORKING CONDITIONS

Workers in many low-income countries around the world labor under conditions that would be illegal for a worker in the United States. Workers in countries like China, Thailand, Brazil, South Africa, and Poland are often paid less than the United States **minimum wage**. For example, in the United States, the minimum wage is \$7.25 per hour; a typical wage in many low-income countries might be more like \$7.25 per day, or often much less. Moreover, working conditions in low-income countries may be extremely unpleasant, or even unsafe. In the worst cases, production may involve the labor of small children or even workers who are treated nearly like slaves. These concerns over standards of foreign labor do not affect most of U.S. trade, which is intra-industry and carried out with other high-income countries that have labor standards similar to the United States, but it is, nonetheless, morally and economically important.

In thinking about labor standards in other countries, it is important to draw some distinctions between what is truly unacceptable and what is painful to think about. Most people, economists included, have little difficulty with the idea that production by six-year-olds confined in factories or by slave labor is morally unacceptable. They would support aggressive efforts to eliminate such practices—including shutting out imported products made with such labor. Many cases, however, are less clear-cut. An opinion article in the *New York Times* several years ago described the case of Ahmed Zia, a 14-year-old boy from Pakistan. He earned \$2 per day working in a carpet factory. He dropped out of school in second grade. Should the United States and other countries refuse to purchase rugs made by Ahmed and his co-workers? If the carpet factories were to close, the likely alternative job for Ahmed is farm work, and as Ahmed says of his carpet-weaving job: “This makes much more money and is more comfortable.”

Other workers may have even less attractive alternative jobs, perhaps scavenging garbage or prostitution. The real problem for Ahmed and many others in low-income countries is not that globalization

has made their lives worse, but rather that they have so few good life alternatives. The United States went through similar situations during the nineteenth and early twentieth centuries.

In closing, there is some irony when the United States government or U.S. citizens take issue with labor standards in low-income countries, because the United States is not a world leader in government laws to protect employees. In Western European countries and Canada, all citizens are guaranteed some form of national healthcare by the government; the United States does not offer such a guarantee but has moved in the direction of universal health insurance coverage under the recent Affordable Care Act. Many European workers receive six weeks or more of paid vacation per year; in the United States, vacations are often one to three weeks per year. If European countries accused the United States of using unfair labor standards to make U.S. products cheaply, and announced that they would shut out all U.S. imports until the United States adopted guaranteed national healthcare, added more national holidays, and doubled vacation time, Americans would be outraged. Yet when U.S. protectionists start talking about restricting imports from poor countries because of low wage levels and poor working conditions, they are making a very similar argument. This is not to say that labor conditions in low-income countries are not an important issue. They are. However, linking labor conditions in low-income countries to trade deflects the emphasis from the real question to ask: “What are acceptable and enforceable minimum labor standards and protections to have the world over?”

KEY CONCEPTS AND SUMMARY

As international trade increases, it contributes to a shift in jobs away from industries where that economy does not have a comparative advantage and toward industries where it does have a comparative advantage. The degree to which trade affects labor markets has a lot to do with the structure of the labor market in that country and the adjustment process in other industries. Global trade should raise the average level of wages by increasing productivity. However, this increase in average wages may include both gains to workers in certain jobs and industries and losses to others.

In thinking about labor practices in low-income countries, it is useful to draw a line between what is unpleasant to think about and what is morally objectionable. For example, low wages and long working hours in poor countries are unpleasant to think about, but for people in low-income parts of the world, it may well be the best option open to them. Practices like child labor and forced labor are morally objectionable and many countries refuse to import products made using these practices.

SELF-CHECK QUESTIONS

1. Explain how trade barriers save jobs in protected industries, but only by costing jobs in other industries.
2. Explain how trade barriers raise wages in protected industries by reducing average wages economy-wide.
3. How does international trade affect working conditions of low-income countries?
4. Do the jobs for workers in low-income countries that involve making products for export to high-income countries typically pay these workers more or less than their next-best alternative?
5. How do trade barriers affect the average income level in an economy?
6. How does the cost of “saving” jobs in protected industries compare to the workers’ wages and salaries?

REVIEW QUESTIONS

1. Does international trade, taken as a whole, increase the total number of jobs, decrease the total number of jobs, or leave the total number of jobs about the same?
2. Is international trade likely to have roughly the same effect on the number of jobs in each individual industry?
3. How is international trade, taken as a whole, likely to affect the average level of wages?
4. Is international trade likely to have about the same effect on everyone's wages?

CRITICAL THINKING QUESTIONS

1. If trade barriers hurt the average worker in an economy (due to lower wages), why does government create trade barriers?
2. Why do you think labor standards and working conditions are lower in the low-income countries of the world than in countries like the United States?

REFERENCES

Bureau of Labor Statistics. "Data Retrieval: Employment, Hours, and Earnings (CES)." Last modified February 1, 2013. <http://www.bls.gov/webapps/legacy/ceshtab1.htm>.

Dhillon, Kiran. 2015. "Why Are U.S. Oil Imports Falling?" Time.com. Accessed April 1, 2015. <http://time.com/67163/why-are-u-s-oil-imports-falling/>.

Kristof, Nicholas. "Let Them Sweat." *The New York Times*, June 25, 2002. <http://www.nytimes.com/2002/06/25/opinion/let-them-sweat.html>.

SOLUTIONS

Answers to Self-Check Questions

1. Trade barriers raise the price of goods in protected industries. If those products are inputs in other industries, it raises their production costs and then prices, so sales fall in those other industries. Lower sales lead to lower employment. Additionally, if the protected industries are consumer goods, their customers pay higher prices, which reduce demand for other consumer products and thus employment in those industries.
2. Trade based on comparative advantage raises the average wage rate economy-wide, though it can reduce the incomes of import-substituting industries. By moving away from a country's comparative advantage, trade barriers do the opposite: they give workers in protected industries an advantage, while reducing the average wage economy-wide.
3. By raising incomes, trade tends to raise working conditions also, even though those conditions may not (yet) be equivalent to those in high-income countries.

4. They typically pay more than the next-best alternative. If a Nike firm did not pay workers at least as much as they would earn, for example, in a subsistence rural lifestyle, they many never come to work for Nike.
5. Since trade barriers raise prices, real incomes fall. The average worker would also earn less.
6. Workers working in other sectors and the protected sector see a decrease in their real wage.

27.3 ARGUMENTS IN SUPPORT OF RESTRICTING IMPORTS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain and analyze various arguments that are in support of restricting imports, including the infant industry argument, the anti-dumping argument, the environmental protection argument, the unsafe consumer products argument, and the national interest argument
- Explain dumping and race to the bottom
- Evaluate the significance of countries' perceptions on the benefits of growing trade

As previously noted, protectionism requires domestic consumers of a product to pay higher prices to benefit domestic producers of that product. Countries that institute protectionist policies lose the economic gains achieved through a combination of comparative advantage, specialized learning, and economies of scale. With these overall costs in mind, let us now consider, one by one, a number of arguments that support restricting imports.

THE INFANT INDUSTRY ARGUMENT

Imagine Bhutan wants to start its own computer industry, but it has no computer firms that can produce at a low enough price and high enough quality to compete in world markets. However, Bhutanese politicians, business leaders, and workers hope that if the local industry had a chance to get established, before it needed to face international competition, then a domestic company or group of companies could develop the skills, management, technology, and economies of scale that it needs to become a successful profit-earning domestic industry. Thus, the **infant industry argument** for protectionism is to block imports for a limited time, to give the infant industry time to mature, before it starts competing on equal terms in the global economy. (Revisit Macroeconomic Policy Around the World for more information on the infant industry argument.)

The infant industry argument is theoretically possible, even sensible: give an industry a short-term indirect subsidy through protection, and then reap the long-term economic benefits of having a vibrant, healthy industry. Implementation, however, is tricky. In many countries, infant industries have gone from babyhood to senility and obsolescence without ever having reached the profitable maturity stage. Meanwhile, the protectionism that was supposed to be short-term often took a very long time to be repealed.

As one example, Brazil treated its computer industry as an infant industry from the late 1970s until

about 1990. In an attempt to establish its computer industry in the global economy, Brazil largely barred imports of computer products for several decades. This policy guaranteed increased sales for Brazilian computers. However, by the mid-1980s, due to lack of international competition, Brazil had a backward and out-of-date industry, typically lagging behind world standards for price and performance by three to five years—a long time in this fast-moving industry. After more than a decade, during which Brazilian consumers and industries that would have benefited from up-to-date computers paid the costs and Brazil's computer industry never competed effectively on world markets, Brazil phased out its infant industry policy for the computer industry.

Protectionism for infant industries always imposes costs on domestic users of the product, and typically has provided little benefit in the form of stronger, competitive industries. However, several countries in East Asia offer an exception. Japan, Korea, Thailand, and other countries in this region have sometimes provided a package of indirect and direct subsidies targeted at certain industries, including protection from foreign competition and government loans at interest rates below the market equilibrium. In Japan and Korea, for example, subsidies helped get their domestic steel and auto industries up and running.

Why did the infant industry policy of protectionism and other subsidies work fairly well in East Asia? A study by the World Bank in the early 1990s offered three guidelines to countries thinking about infant industry protection:

1. Do not hand out protectionism and other subsidies to all industries, but focus on a few industries where your country has a realistic chance to be a world-class producer.
2. Be very hesitant about using protectionism in areas like computers, where many other industries rely on having the best products available, because it is not useful to help one industry by imposing high costs on many other industries.
3. Have clear guidelines for when the infant industry policy will end.

In Korea in the 1970s and 1980s, a common practice was to link protectionism and subsidies to export sales in global markets. If export sales rose, then the infant industry had succeeded and the protectionism could be phased out. If export sales did not rise, then the infant industry policy had failed and the protectionism could be phased out. Either way, the protectionism would be temporary.

Following these rules is easier said than done. Politics often intrudes, both in choosing which industries will receive the benefits of being treated as “infants” and when to phase out import restrictions and other subsidies. Also, if the government of a country wishes to impose costs on its citizens so that it can provide subsidies to a few key industries, it has many tools for doing so: direct government payments, loans, targeted tax reductions, government support of research and development of new technologies, and so on. In other words, protectionism is not the only or even the best way to support key industries.

Visit this website to view a presentation by Pankaj Ghemawat questioning how integrated the world really is.



THE ANTI-DUMPING ARGUMENT

Dumping refers to selling goods below their cost of production. **Anti-dumping laws** block imports that are sold below the cost of production by imposing tariffs that increase the price of these imports to reflect their cost of production. Since dumping is not allowed under the rules of the **World Trade Organization (WTO)**, nations that believe they are on the receiving end of dumped goods can file a complaint with the WTO. Anti-dumping complaints have risen in recent years, from about 100 cases per year in the late 1980s to about 200 new cases each year by the late 2000s. Note that dumping cases are countercyclical. During recessions, case filings increase. During economic booms, case filings go down. Individual countries have also frequently started their own anti-dumping investigations. The U.S. government has dozens of anti-dumping orders in place from past investigations. In 2009, for example, some U.S. imports that were under anti-dumping orders included pasta from Turkey, steel pipe fittings from Thailand, pressure-sensitive plastic tape from Italy, preserved mushrooms and lined paper products from India, and cut-to-length carbon steel and non-frozen apple juice concentrate from China.

Why Might Dumping Occur?

Why would foreign firms export a product at less than its cost of production—which presumably means taking a loss? This question has two possible answers, one innocent and one more sinister.

The innocent explanation is that market prices are set by demand and supply, not by the cost of production. Perhaps demand for a product shifts back to the left or supply shifts out to the right, which drives the market price to low levels—even below the cost of production. When a local store has a going-out-of-business sale, for example, it may sell goods at below the cost of production. If international companies find that there is excess supply of steel or computer chips or machine tools that is driving the market price down below their cost of production—this may be the market in action.

The sinister explanation is that dumping is part of a long-term strategy. Foreign firms sell goods at prices below the cost of production for a short period of time, and when they have driven out the domestic U.S. competition, they then raise prices. This scenario is sometimes called predatory pricing, which is discussed in the Monopoly chapter.

Should Anti-Dumping Cases Be Limited?

Anti-dumping cases pose two questions. How much sense do they make in economic theory? How much sense do they make as practical policy?

In terms of economic theory, the case for anti-dumping laws is weak. In a market governed by demand

and supply, the government does not guarantee that firms will be able to make a profit. After all, low prices are difficult for producers, but benefit consumers. Moreover, although there are plenty of cases in which foreign producers have driven out domestic firms, there are zero documented cases in which the foreign producers then jacked up prices. Instead, foreign producers typically continue competing hard against each other and providing low prices to consumers. In short, it is difficult to find evidence of predatory pricing by foreign firms exporting to the United States.

Even if one could make a case that the government should sometimes enact anti-dumping rules in the short term, and then allow free trade to resume shortly thereafter, there is a growing concern that anti-dumping investigations often involve more politics than careful analysis. The U.S. Commerce Department is charged with calculating the appropriate “cost of production,” which can be as much an art as a science.

For example, if a company built a new factory two years ago, should part of the factory’s cost be counted in this year’s cost of production? When a company is in a country where prices are controlled by the government, like China for example, how can one measure the true cost of production? When a domestic industry complains loudly enough, government regulators seem very likely to find that unfair dumping has occurred. Indeed, a common pattern has arisen where a domestic industry files an anti-dumping complaint, the governments meet and negotiate a reduction in imports, and then the domestic producers drop the anti-dumping suit. In such cases, anti-dumping cases often appear to be little more than a cover story for imposing tariffs or import quotas.

In the 1980s, almost all of the anti-dumping cases were initiated by the United States, Canada, the European Union, Australia, and New Zealand. By the 2000s, countries like Argentina, Brazil, South Korea, South Africa, Mexico, and India were filing the majority of the anti-dumping cases before the WTO. As the number of anti-dumping cases has increased, and as countries such as the United States and the European Union feel targeted by the anti-dumping actions of others, the WTO may well propose some additional guidelines to limit the reach of anti-dumping laws.

THE ENVIRONMENTAL PROTECTION ARGUMENT

The potential for global trade to affect the environment has become controversial. A president of the Sierra Club, an environmental lobbying organization, once wrote: “The consequences of globalization for the environment are not good. ... Globalization, if we are lucky, will raise average incomes enough to pay for cleaning up some of the mess that we have made. But before we get there, globalization could also destroy enough of the planet’s basic biological and physical systems that prospects for life itself will be radically compromised.”

If **free trade** meant the destruction of life itself, then even economists would convert to protectionism! While globalization—and economic activity of all kinds—can pose environmental dangers, it seems quite possible that, with the appropriate safeguards in place, the environmental impacts of trade can be minimized. In some cases, trade may even bring environmental benefits.

In general, **high-income countries** such as the United States, Canada, Japan, and the nations of the European Union have relatively strict environmental standards. In contrast, middle- and **low-income countries** like Brazil, Nigeria, India, and China have lower environmental standards. The general view of the governments of such countries is that environmental protection is a luxury: as soon as their people have enough to eat, decent healthcare, and longer life expectancies, then they will spend

more money on sewage treatment plants, scrubbers to reduce air pollution from factory smokestacks, national parks to protect wildlife, and so on.

This gap in environmental standards between high-income and low-income countries raises two worrisome possibilities in a world of increasing global trade: the “race to the bottom” scenario and the question of how quickly environmental standards will improve in low-income countries.

The Race to the Bottom Scenario

The **race to the bottom** scenario of global environmental degradation runs like this. Profit-seeking multinational companies shift their production from countries with strong environmental standards to countries with weak standards, thus reducing their costs and increasing their profits. Faced with such behavior, countries reduce their environmental standards to attract multinational firms, which, after all, provide jobs and economic clout. As a result, global production becomes concentrated in countries where it can pollute the most and environmental laws everywhere “race to the bottom.”

Although the race-to-the-bottom scenario sounds plausible, it does not appear to describe reality. In fact, the financial incentive for firms to shift production to poor countries to take advantage of their weaker environmental rules does not seem especially powerful. When firms decide where to locate a new factory, they look at many different factors: the costs of labor and financial capital; whether the location is close to a reliable suppliers of the inputs that they need; whether the location is close to customers; the quality of transportation, communications, and electrical power networks; the level of taxes; and the competence and honesty of the local government. The cost of environmental regulations is a factor, too, but typically environmental costs are no more than 1 to 2% of the costs faced by a large industrial plant. The other factors that determine location are much more important to these companies than trying to skimp on environmental protection costs.

When an international company does choose to build a plant in a low-income country with lax environmental laws, it typically builds a plant similar to those that it operates in high-income countries with stricter environmental standards. Part of the reason for this decision is that designing an industrial plant is a complex and costly task, and so if a plant works well in a high-income country, companies prefer to use the same design everywhere. Also, companies realize that if they create an environmental disaster in a low-income country, it is likely to cost them a substantial amount of money in paying for damages, lost trust, and reduced sales—by building up-to-date plants everywhere they minimize such risks. As a result of these factors, foreign-owned plants in low-income countries often have a better record of compliance with environmental laws than do locally-owned plants.

Pressuring Low-Income Countries for Higher Environmental Standards

In some cases, the issue is not so much whether globalization will pressure low-income countries to reduce their environmental standards, but instead whether the threat of blocking international trade can pressure these countries into adopting stronger standards. For example, restrictions on ivory imports in high-income countries, along with stronger government efforts to catch elephant poachers, have been credited with helping to reduce the illegal poaching of elephants in certain African countries.

However, it would be highly undemocratic for the well-fed citizens of high-income countries to attempt to dictate to the ill-fed citizens of low-income countries what domestic policies and priorities they must adopt, or how they should balance environmental goals against other priorities for their cit-

izens. Furthermore, if high-income countries want stronger environmental standards in low-income countries, they have many options other than the threat of protectionism. For example, high-income countries could pay for anti-pollution equipment in low-income countries, or could help to pay for national parks. High-income countries could help pay for and carry out the scientific and economic studies that would help environmentalists in low-income countries to make a more persuasive case for the economic benefits of protecting the environment.

After all, environmental protection is vital to two industries of key importance in many low-income countries—agriculture and tourism. Environmental advocates can set up standards for labeling products, like “this tuna caught in a net that kept dolphins safe” or “this product made only with wood not taken from rainforests,” so that consumer pressure can reinforce environmentalist values. These values are also reinforced by the United Nations, which sponsors treaties to address issues such as climate change and global warming, the preservation of biodiversity, the spread of deserts, and the environmental health of the seabed. Countries that share a national border or are within a region often sign environmental agreements about air and water rights, too. The WTO is also becoming more aware of environmental issues and more careful about ensuring that increases in trade do not inflict environmental damage.

Finally, it should be noted that these concerns about the race to the bottom or pressuring low-income countries for more strict environmental standards do not apply very well to the roughly half of all U.S. trade that occurs with other high-income countries. Indeed, many European countries have stricter environmental standards in certain industries than the United States.

THE UNSAFE CONSUMER PRODUCTS ARGUMENT

One argument for shutting out certain imported products is that they are unsafe for consumers. Indeed, consumer rights groups have sometimes warned that the World Trade Organization would require nations to reduce their health and safety standards for imported products. However, the WTO explains its current agreement on the subject in this way: “It allows countries to set their own standards.” But it also says “regulations must be based on science. . . . And they should not arbitrarily or unjustifiably discriminate between countries where identical or similar conditions prevail.” Thus, for example, under WTO rules it is perfectly legitimate for the United States to pass laws requiring that *all* food products or cars sold in the United States meet certain safety standards approved by the United States government, whether or not other countries choose to pass similar standards. However, such standards must have some scientific basis. It is improper to impose one set of health and safety standards for domestically produced goods but a different set of standards for imports, or one set of standards for imports from Europe and a different set of standards for imports from Latin America.

In 2007, Mattel recalled nearly two million toys imported from China due to concerns about high levels of lead in the paint, as well as some loose parts. It is unclear if other toys were subject to similar standards. More recently, in 2013, Japan blocked imports of U.S. wheat because of concerns that genetically modified (GMO) wheat might be included in the shipments. The science on the impact of GMOs on health is still developing.

THE NATIONAL INTEREST ARGUMENT

Some argue that a nation should not depend too heavily on other countries for supplies of certain key

products, such as oil, or for special materials or technologies that might have national security applications. On closer consideration, this argument for protectionism proves rather weak.

As an example, in the United States, oil provides about 40% of all the energy and 32% of the oil used in the United States economy is imported. Several times in the last few decades, when disruptions in the Middle East have shifted the supply curve of oil back to the left and sharply raised the price, the effects have been felt across the United States economy. This is not, however, a very convincing argument for restricting imports of oil. If the United States needs to be protected from a possible cutoff of foreign oil, then a more reasonable strategy would be to import 100% of the petroleum supply now, and save U.S. domestic oil resources for when or if the foreign supply is cut off. It might also be useful to import extra oil and put it into a stockpile for use in an emergency, as the United States government did by starting a Strategic Petroleum Reserve in 1977. Moreover, it may be necessary to discourage people from using oil, and to start a high-powered program to seek out alternatives to oil. A straightforward way to do this would be to raise taxes on oil. What's more, it makes no sense to argue that because oil is highly important to the United States economy, then the United States should shut out oil imports and use up its domestic supplies of oil more quickly. U.S. domestic production of oil is increasing. Shale oil is adding to domestic supply using fracking extraction techniques.

Whether or not to limit certain kinds of imports of key technologies or materials that might be important to national security and weapons systems is a slightly different issue. If weapons' builders are not confident that they can continue to obtain a key product in wartime, they might decide to avoid designing weapons that use this key product, or they can go ahead and design the weapons and stockpile enough of the key high-tech components or materials to last through an armed conflict. Indeed, there is a U.S. Defense National Stockpile Center that has built up reserves of many materials, from aluminum oxides, antimony, and bauxite to tungsten, vegetable tannin extracts, and zinc (although many of these stockpiles have been reduced and sold in recent years). Think every country is pro-trade? How about the U.S.? The following Clear it Up might surprise you.

HOW DOES THE UNITED STATES REALLY FEEL ABOUT EXPANDING TRADE?

How do people around the world feel about expanding trade between nations? In summer 2007, the Pew Foundation surveyed 45,000 people in 47 countries. One of the questions asked about opinions on growing trade ties between countries. Table 3 shows the percentages who answered either "very good" or "somewhat good" for some of countries surveyed. For those who think of the United States as the world's leading supporter of expanding trade, the survey results may be perplexing. When adding up the shares of those who say that growing trade ties between countries is "very good" or "somewhat good," Americans had the least favorable attitude toward increasing globalization, while the Chinese and South Africans ranked highest. In fact, among the 47 countries surveyed, the United States ranked by far the lowest on this measure, followed by Egypt, Italy, and Argentina.

Country	Very Good	Somewhat Good	Total
China	38%	53%	91%
South Africa	42%	43%	87%
South Korea	24%	62%	86%
Germany	30%	55%	85%
Canada	29%	53%	82%
United Kingdom	28%	50%	78%
Mexico	22%	55%	77%
Brazil	13%	59%	72%
Japan	17%	55%	72%
United States	14%	45%	59%

Table 3. The Status of Growing Trade Ties between Countries (Source: <http://www.pewglobal.org/files/pdf/258.pdf>)

One final reason why economists often treat the **national interest argument** skeptically is that almost any product can be touted by lobbyists and politicians as vital to national security. In 1954, the United States became worried that it was importing half of the wool required for military uniforms, so it declared wool and mohair to be “strategic materials” and began to give subsidies to wool and mohair farmers. Although wool was removed from the official list of “strategic” materials in 1960, the subsidies for mohair continued for almost 40 years until they were repealed in 1993, and then were reinstated in 2002. All too often, the national interest argument has become an excuse for handing out the indirect subsidy of protectionism to certain industries or companies. After all, decisions about what constitutes a key strategic material are made by politicians, not nonpartisan analysts.

KEY CONCEPTS AND SUMMARY

There are a number of arguments that support restricting imports. These arguments are based around industry and competition, environmental concerns, and issues of safety and security.

The infant industry argument for protectionism is that small domestic industries need to be temporarily nurtured and protected from foreign competition for a time so that they can grow into strong competitors. In some cases, notably in East Asia, this approach has worked. Often, however, the infant industries never grow up. On the other hand, arguments against dumping (which is setting prices below the cost of production to drive competitors out of the market), often simply seem to be a convenient excuse for imposing protectionism.

Low-income countries typically have lower environmental standards than high-income countries because they are more worried about immediate basics such as food, education, and healthcare. However, except for a small number of extreme cases, shutting off trade seems unlikely to be an effective method of pursuing a cleaner environment.

Finally, there are arguments involving safety and security. Under the rules of the World Trade Organization, countries are allowed to set whatever standards for product safety they wish, but the standards must be the same for domestic products as for imported products and there must be a scientific basis for the standard. The national interest argument for protectionism holds that it is unwise to import certain key products because if the nation becomes dependent on key imported supplies, it could be

vulnerable to a cutoff. However, it is often wiser to stockpile resources and to use foreign supplies when available, rather than preemptively restricting foreign supplies so as not to become dependent on them.

SELF-CHECK QUESTIONS

1. Explain how predatory pricing could be a motivation for dumping.
2. Why do low-income countries like Brazil, Egypt, or Vietnam have lower environmental standards than high-income countries like Germany, Japan, or the United States?
3. Explain the logic behind the “race to the bottom” argument and the likely reason it has not occurred.
4. What are the conditions under which a country may use the unsafe products argument to block imports?
5. Why is the national security argument not convincing?
6. Assume a perfectly competitive market and the exporting country is small. Using a demand and supply diagram, show the impact of increasing standards on a low-income exporter of toys. Show the impact of a tariff. Is the effect on the price of toys the same or different? Why is a standards policy preferred to tariffs?

REVIEW QUESTIONS

1. What are main reasons for protecting “infant industries”? Why is it difficult to stop protecting them?
2. What is dumping? Why does prohibiting it often work better in theory than in practice?
3. What is the “race to the bottom” scenario?
4. Do the rules of international trade require that all nations impose the same consumer safety standards?
5. What is the national interest argument for protectionism with regard to certain products?

CRITICAL THINKING QUESTIONS

1. How would direct subsidies to key industries be preferable to tariffs or quotas?
2. How can governments identify good candidates for infant industry protection? Can you suggest some key characteristics of good candidates? Why are industries like computers not good candidates for infant industry protection?
3. Microeconomic theory argues that it is economically rational (and profitable) to sell additional output as long as the price covers the variable costs of production. How is this relevant to the determination of whether dumping has occurred?
4. How do you think Americans would feel if other countries began to urge the United States to increase environmental standards?
5. Is it legitimate to impose higher safety standards on imported goods than exist in the foreign country where the goods were produced?
6. Why might the unsafe consumer products argument be a more effective strategy (from the perspective of the importing country) than using tariffs or quotas to restrict imports?
7. Why might a tax on domestic consumption of resources critical for national security be a more efficient approach than barriers to imports?

PROBLEMS

1. You have just been put in charge of trade policy for Malawi. Coffee is a recent crop that is growing well and the Malawian export market is developing. As such, Malawi coffee is an infant industry. Malawi coffee producers come to you and ask for tariff protection from cheap Tanzanian coffee. What sorts of policies will you enact? Explain.
2. The country of Pepperland exports steel to the Land of Submarines. Information for the quantity demanded (Q_d) and quantity supplied (Q_s) in each country, in a world without trade, are given in Table 4 and Table 5.

Price (\$)	Q_d	Q_s
60	230	180
70	200	200
80	170	220
90	150	240
100	140	250

Table 4. Pepperland

Price (\$)	Q_d	Q_s
60	430	310
70	420	330
80	410	360
90	400	400
100	390	440

Table 5. Land of Submarines

- a. What would be the equilibrium price and quantity in each country in a world without trade? How can you tell?
- b. What would be the equilibrium price and quantity in each country if trade is allowed to occur? How can you tell?
- c. Sketch two supply and demand diagrams, one for each country, in the situation before trade.
- d. On those diagrams, show the equilibrium price and the levels of exports and imports in the world after trade.
- e. If the Land of Submarines imposes an anti-dumping import quota of 30, explain in general terms whether it will benefit or injure consumers and producers in each country.
- f. Does your general answer change if the Land of Submarines imposes an import quota of 70?

REFERENCES

- Kohut, Andrew, Richard Wike, and Juliana Horowitz. "The Pew Global Attitudes Project." *Pew Research Center*. Last modified October 4, 2007. <http://www.pewglobal.org/files/pdf/258.pdf>.
- Lutz, Hannah. 2015. "U.S. Auto Exports Hit Record in 2014." *Automotive News*. Accessed April 1, 2015. <http://www.autonews.com/article/20150206/OEM01/150209875/u.s.-auto-exports-hit-record-in-2014>.

GLOSSARY

anti-dumping laws laws that block imports sold below the cost of production and impose tariffs that would increase the price of these imports to reflect their cost of production

dumping selling internationally traded goods below their cost of production

national interest argument the argument that there are compelling national interests against depending on key imports from other nations

race to the bottom when production locates in countries with the lowest environmental (or other) standards, putting pressure on all countries to reduce their environmental standards

SOLUTIONS

Answers to Self-Check Questions

1. If imports can be sold at extremely low prices, domestic firms would have to match those prices to be competitive. By definition, matching prices would imply selling under cost and, therefore, losing money. Firms cannot sustain losses forever. When they leave the industry, importers can “take over,” raising prices to monopoly levels to cover their short-term losses and earn long-term profits.
2. Because low-income countries need to provide necessities—food, clothing, and shelter—to their people. In other words, they consider environmental quality a luxury.
3. Low-income countries can compete for jobs by reducing their environmental standards to attract business to their countries. This could lead to a competitive reduction in regulations, which would lead to greater environmental damage. While pollution management is a cost for businesses, it is tiny relative to other costs, like labor and adequate infrastructure. It is also costly for firms to locate far away from their customers, which many low-income countries are.
4. The decision should not be arbitrary or unnecessarily discriminatory. It should treat foreign companies the same way as domestic companies. It should be based on science.
5. Restricting imports today does not solve the problem. If anything, it makes it worse since it implies using up domestic sources of the products faster than if they are imported. Also, the national security argument can be used to support protection of nearly any product, not just things critical to our national security.
6. The effect of increasing standards may increase costs to the small exporting country. The supply curve of toys will shift to the left. Exports will decrease and toy prices will rise. Tariffs also raise prices. So the effect on the price of toys is the same. A tariff is a “second best” policy and also affects other sectors. However, a common standard across countries is a “first best” policy that attacks the problem at its root.

27.4 HOW TRADE POLICY IS ENACTED: GLOBALLY, REGIONALLY, AND NATIONALLY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the origin and role of the World Trade Organization (WTO) and General Agreement on Tariffs and Trade (GATT)
- Discuss the significance and provide examples of regional trading agreements
- Analyze trade policy at the national level
- Evaluate long-term trends in barriers to trade

These public policy arguments about how nations should react to globalization and trade are fought out at several levels: at the global level through the World Trade Organization and through regional trade agreements between pairs or groups of countries.

THE WORLD TRADE ORGANIZATION

The **World Trade Organization (WTO)** was officially born in 1995, but its history is much longer. In the years after the Great Depression and World War II, there was a worldwide push to build institutions that would tie the nations of the world together. The United Nations officially came into existence in 1945. The World Bank, which assists the poorest people in the world, and the International Monetary Fund, which addresses issues raised by international financial transactions, were both created in 1946. The third planned organization was to be an International Trade Organization, which would manage international trade. The United Nations was unable to agree to this. Instead, the **General Agreement on Tariffs and Trade (GATT)**, was established in 1947 to provide a forum in which nations could come together to negotiate reductions in tariffs and other barriers to trade. In 1995, the GATT was transformed into the WTO.

The GATT process was to negotiate an agreement to reduce barriers to trade, sign that agreement, pause for a while, and then start negotiating the next agreement. The rounds of talks in the GATT, and now the WTO, are shown in Table 6. Notice that the early rounds of GATT talks took a relatively short time, included a small number of countries, and focused almost entirely on reducing tariffs. Since the 1970s, however, rounds of trade talks have taken years, included a large number of countries, and an ever-broadening range of issues.

Year	Place or Name of Round	Main Subjects	Number of Countries Involved
1947	Geneva	Tariff reduction	23
1949	Annecy	Tariff reduction	13
1951	Torquay	Tariff reduction	38
1956	Geneva	Tariff reduction	26
1960–61	Dillon round	Tariff reduction	26
1964–67	Kennedy round	Tariffs, anti-dumping measures	62
1973–79	Tokyo round	Tariffs, nontariff barriers	102
1986–94	Uruguay round	Tariffs, nontariff barriers, services, intellectual property, dispute settlement, textiles, agriculture, creation of WTO	123
2001–	Doha round	Agriculture, services, intellectual property, competition, investment, environment, dispute settlement	147

Table 6. The Negotiating Rounds of GATT and the World Trade Organization

The sluggish pace of GATT negotiations led to an old joke that GATT really stood for Gentleman's Agreement to Talk and Talk. The slow pace of international trade talks, however, is understandable, even sensible. Having dozens of nations agree to any treaty is a lengthy process. GATT often set up separate trading rules for certain industries, like agriculture, and separate trading rules for certain countries, like the low-income countries. There were rules, exceptions to rules, opportunities to opt out of rules, and precise wording to be fought over in every case. Like the GATT before it, the WTO is not a world government, with power to impose its decisions on others. The total staff of the WTO in 2014 is 640 people and its annual budget (as of 2014) is \$197 million, which makes it smaller in size than many large universities.

REGIONAL TRADING AGREEMENTS

There are different types of economic integration across the globe, ranging from **free trade agreements**, in which participants allow each other's imports without tariffs or quotas, to **common markets**, in which participants have a common external trade policy as well as free trade within the group, to full **economic unions**, in which, in addition to a common market, monetary and fiscal policies are coordinated. Many nations belong both to the World Trade Organization and to regional trading agreements.

The best known of these regional trading agreements is the **European Union**. In the years after World War II, leaders of several European nations reasoned that if they could tie their economies together more closely, they might be more likely to avoid another devastating war. Their efforts began with a free trade association, evolved into a common market, and then transformed into what is now a full economic union, known as the European Union. The EU, as it is often called, has a number of goals. For example, in the early 2000s it introduced a common currency for Europe, the euro, and phased out most of the former national forms of money like the German mark and the French franc, though a few have retained their own currency. Another key element of the union is to eliminate barriers to the mobility of goods, labor, and capital across Europe.

For the United States, perhaps the best-known regional trading agreement is the **North American Free Trade Agreement (NAFTA)**. The United States also participates in some less-prominent regional

trading agreements, like the Caribbean Basin Initiative, which offers reduced tariffs for imports from these countries, and a free trade agreement with Israel.

The world has seen a flood of regional trading agreements in recent years. About 100 such agreements are now in place. A few of the more prominent ones are listed in Table 7. Some are just agreements to continue talking; others set specific goals for reducing tariffs, import quotas, and nontariff barriers. One economist described the current trade treaties as a “spaghetti bowl,” which is what a map with lines connecting all the countries with trade treaties looks like.

There is concern among economists who favor free trade that some of these regional agreements may promise free trade, but actually act as a way for the countries within the regional agreement to try to limit trade from anywhere else. In some cases, the regional trade agreements may even conflict with the broader agreements of the World Trade Organization.

Trade Agreements	Participating Countries
Asia Pacific Economic Cooperation (APEC)	Australia, Brunei, Canada, Chile, People's Republic of China, Hong Kong, China, Indonesia, Japan, Republic of Korea, Malaysia, Mexico, New Zealand, Papua New Guinea, Peru, Philippines, Russia, Singapore, Chinese Taipei, Thailand, United States, Vietnam
European Union (EU)	Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom
North America Free Trade Agreement (NAFTA)	Canada, Mexico, United States
Latin American Integration Association (LAIA)	Argentina, Bolivia, Brazil, Chile, Columbia, Ecuador, Mexico, Paraguay, Peru, Uruguay, Venezuela
Association of Southeast Asian Nations (ASEAN)	Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Vietnam
Southern African Development Community (SADC)	Angola, Botswana, Congo, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Zambia, Zimbabwe

Table 7. Some Regional Trade Agreements

TRADE POLICY AT THE NATIONAL LEVEL

Yet another dimension of trade policy, along with international and regional trade agreements, happens at the national level. The United States, for example, imposes import quotas on sugar, because of a fear that such imports would drive down the price of sugar and thus injure domestic sugar producers. One of the jobs of the United States Department of Commerce is to determine if imports from other countries are being dumped. The United States International Trade Commission—a government agency—determines whether domestic industries have been substantially injured by the dumping, and if so, the president can impose tariffs that are intended to offset the unfairly low price.

In the arena of trade policy, the battle often seems to be between national laws that increase protectionism and international agreements that try to reduce protectionism, like the WTO. Why would a country pass laws or negotiate agreements to shut out certain foreign products, like sugar or textiles,

while simultaneously negotiating to reduce trade barriers in general? One plausible answer is that international trade agreements offer a method for countries to restrain their own special interests. A member of Congress can say to an industry lobbying for tariffs or quotas on imports: “Sure would like to help you, but that pesky WTO agreement just won’t let me.”

If consumers are the biggest losers from trade, why do they not fight back? The quick answer is because it is easier to organize a small group of people around a narrow interest versus a large group that has diffuse interests. This is a question about trade policy theory. Visit this website and read the article by Jonathan Rauch.



LONG-TERM TRENDS IN BARRIERS TO TRADE

In newspaper headlines, trade policy appears mostly as disputes and acrimony. Countries are almost constantly threatening to challenge the “unfair” trading practices of other nations. Cases are brought to the dispute settlement procedures of the WTO, the European Union, NAFTA, and other regional trading agreements. Politicians in national legislatures, goaded on by lobbyists, often threaten to pass bills that will “establish a fair playing field” or “prevent unfair trade”—although most such bills seek to accomplish these high-sounding goals by placing more restrictions on trade. Protesters in the streets may object to specific trade rules or to the entire practice of international trade.

Through all the controversy, the general trend in the last 60 years is clearly toward lower barriers to trade. The average level of tariffs on imported products charged by industrialized countries was 40% in 1946. By 1990, after decades of GATT negotiations, it was down to less than 5%. Indeed, one of the reasons that GATT negotiations shifted from focusing on tariff reduction in the early rounds to a broader agenda was that tariffs had been reduced so dramatically there was not much more to do in that area. U.S. tariffs have followed this general pattern: After rising sharply during the Great Depression, tariffs dropped off to less than 2% by the end of the century. Although measures of import quotas and nontariff barriers are less exact than those for tariffs, they generally appear to be at lower levels, too.

Thus, the last half-century has seen both a dramatic reduction in government-created barriers to trade, such as tariffs, import quotas, and nontariff barriers, and also a number of technological developments that have made international trade easier, like advances in transportation, communication, and information management. The result has been the powerful surge of international trade.

KEY CONCEPTS AND SUMMARY

Trade policy is determined at many different levels: administrative agencies within government, laws passed by the legislature, regional negotiations between a small group of nations (sometimes just two),

and global negotiations through the World Trade Organization. During the second half of the twentieth century, trade barriers have, in general, declined quite substantially in the United States economy and in the global economy. One reason why countries sign international trade agreements to commit themselves to free trade is to give themselves protection against their own special interests. When an industry lobbies for protection from foreign producers, politicians can point out that, because of the trade treaty, their hands are tied.

SELF-CHECK QUESTIONS

1. What is the difference between a free trade association, a common market, and an economic union?
2. Why would countries promote protectionist laws, while also negotiate for freer trade internationally?
3. What might account for the dramatic increase in international trade over the past 50 years?

REVIEW QUESTIONS

1. Name several of the international treaties where countries negotiate with each other over trade policy.
2. What is the general trend of trade barriers over recent decades: higher, lower, or about the same?
3. If opening up to free trade would benefit a nation, then why do nations not just eliminate their trade barriers, and not bother with international trade negotiations?

CRITICAL THINKING QUESTIONS

1. Why do you think that the GATT rounds and, more recently, WTO negotiations have become longer and more difficult to resolve?
2. An economic union requires giving up some political autonomy to succeed. What are some examples of political power countries must give up to be members of an economic union?

REFERENCES

- United States Department of Labor. Bureau of Labor Statistics. 2015. "Employment Situation Summary." Accessed April 1, 2015. <http://www.bls.gov/news.release/empstat.nr0.htm>.
- United States Department of Commerce. "About the Department of Commerce." Accessed January 6, 2014. <http://www.commerce.gov/about-department-commerce>.
- United States International Trade Commission. "About the USITC." Accessed January 6, 2014. http://www.usitc.gov/press_room/about_usitc.htm.

GLOSSARY

common market economic agreement between countries to allow free trade in goods, services, labor, and financial capital between members while having a common external trade policy

economic union economic agreement between countries to allow free trade between members, a common external trade policy, and coordinated monetary and fiscal policies

free trade agreement economic agreement between countries to allow free trade between members

General Agreement on Tariffs and Trade (GATT) forum in which nations could come together to negotiate reductions in tariffs and other barriers to trade; the precursor to the World Trade Organization

SOLUTIONS

Answers to Self-Check Questions

1. A free trade association offers free trade between its members, but each country can determine its own trade policy outside the association. A common market requires a common external trade policy in addition to free trade within the group. An economic union is a common market with coordinated fiscal and monetary policy.
2. International agreements can serve as a political counterweight to domestic special interests, thereby preventing stronger protectionist measures.
3. Reductions in tariffs, quotas, and other trade barriers, improved transportation, and communication media have made people more aware of what is available in the rest of the world.

27.5 THE TRADEOFFS OF TRADE POLICY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Assess the complexity of international trade
- Discuss why a market-oriented economy is so affected by international trade
- Explain disruptive market change

Economists readily acknowledge that international trade is not all sunshine, roses, and happy endings. Over time, the average person gains from international trade, both as a worker who has greater productivity and higher wages because of the benefits of specialization and comparative advantage, and as a consumer who can benefit from shopping all over the world for a greater variety of quality products at attractive prices. The “average person,” however, is hypothetical, not real—representing a mix of those who have done very well, those who have done all right, and those who have done poorly. It is a legitimate concern of public policy to focus not just on the average or on the success stories, but also on those who have not been so fortunate. Workers in other countries, the environment, and prospects for new industries and materials that might be of key importance to the national economy are also all legitimate issues.

The common belief among economists is that it is better to embrace the gains from trade, and then deal with the costs and tradeoffs with other policy tools, than it is to cut off trade to avoid the costs and tradeoffs.

To gain a better intuitive understanding for this argument, consider a hypothetical American company called Technotron. Technotron invents a new scientific technology that allows the firm to increase the output and quality of its goods with a smaller number of workers at a lower cost. As a result of this technology, other U.S. firms in this industry will lose money and will also have to lay off workers—and some of the competing firms will even go bankrupt. Should the United States government protect the existing firms and their employees by making it illegal for Technotron to use its new technology? Most people who live in market-oriented economies would oppose trying to block better products that lower the cost of services. Certainly, there is a case for society providing temporary support and assistance for those who find themselves without work. Many would argue for government support of programs that encourage retraining and acquiring additional skills. Government might also support research and development efforts, so that other firms may find ways of outdoing Technotron. Blocking the new technology altogether, however, seems like a mistake. After all, few people would advocate giving up electricity because it caused so much disruption to the kerosene and

candle business. Few would suggest holding back on improvements in medical technology because they might cause companies selling leeches and snake oil to lose money. In short, most people view disruptions due to technological change as a necessary cost that is worth bearing.

Now, imagine that Technotron's new "technology" is as simple as this: the company imports what it sells from another country. In other words, think of foreign trade as a type of innovative technology. The objective situation is now exactly the same as before. Because of Technotron's new technology—which in this case is importing goods from another country—other firms in this industry will lose money and lay off workers. Just as it would have been inappropriate and ultimately foolish to respond to the disruptions of new scientific technology by trying to shut it down, it would be inappropriate and ultimately foolish to respond to the disruptions of **international trade** by trying to restrict trade.

Some workers and firms will suffer because of international trade. In a living, breathing market-oriented economy, some workers and firms will always be experiencing disruptions, for a wide variety of reasons. Corporate management can be better or worse. Workers for a certain firm can be more productive or less. Tough domestic competitors can create just as much disruption as tough foreign competitors. Sometimes a new product is a hit with consumers; sometimes it is a flop. Sometimes a company is blessed by a run of good luck or stricken with a run of bad luck. For some firms, international trade will offer great opportunities for expanding productivity and jobs; for other firms, trade will impose stress and pain. The disruption caused by international trade is not fundamentally different from all the other disruptions caused by the other workings of a market economy.

In other words, the economic analysis of free trade does not rely on a belief that foreign trade is not disruptive or does not pose tradeoffs; indeed, the story of Technotron begins with a particular **disruptive market change**—a new technology—that causes real tradeoffs. In thinking about the disruptions of foreign trade, or any of the other possible costs and tradeoffs of foreign trade discussed in this chapter, the best public policy solutions typically do not involve protectionism, but instead involve finding ways for public policy to address the particular issues, while still allowing the benefits of international trade to occur.

WHAT'S THE DOWNSIDE OF PROTECTION?

The domestic flat-panel display industry employed many workers before the ITC imposed the dumping margin tax. Flat-panel displays make up a significant portion of the cost of producing laptop computers—as much as 50%. Therefore, the antidumping tax would substantially increase the cost, and thus the price, of U.S.-manufactured laptops. As a result of the ITC's decision, Apple moved its domestic manufacturing plant for Macintosh computers to Ireland (where it had an existing plant). Toshiba shut down its U.S. manufacturing plant for laptops. And IBM cancelled plans to open a laptop manufacturing plant in North Carolina, instead deciding to expand production at its plant in Japan. In this case, rather than having the desired effect of protecting U.S. interests and giving domestic manufacturing an advantage over items manufactured elsewhere, it had the unintended effect of driving the manufacturing completely out of the country. Many people lost their jobs and most flat-panel display production now occurs in countries other than the United States.

KEY CONCEPTS AND SUMMARY

International trade certainly has income distribution effects. This is hardly surprising. All domestic or international competitive market forces are disruptive. They cause companies and industries to rise

and fall. Government has a role to play in cushioning workers against the disruptions of the market. However, just as it would be unwise in the long term to clamp down on new technology and other causes of disruption in domestic markets, it would be unwise to clamp down on foreign trade. In both cases, the disruption brings with it economic benefits.

SELF-CHECK QUESTIONS

1. How does competition, whether domestic or foreign, harm businesses?
2. What are the gains from competition?

REVIEW QUESTIONS

1. Who gains and who loses from trade?
2. Why is trade a good thing if some people lose?
3. What are some ways that governments can help people who lose from trade?

CRITICAL THINKING QUESTIONS

1. What are some examples of innovative products that have disrupted their industries for the better?
2. In principle, the benefits of international trade to a country exceed the costs, no matter whether the country is importing or exporting. In practice, it is not always possible to compensate the losers in a country, for example, workers who lose their jobs due to foreign imports. In your opinion, does that mean that trade should be inhibited to prevent the losses?
3. Economists sometimes say that protectionism is the “second-best” choice for dealing with any particular problem. What they mean is that there is often a policy choice that is more direct or effective for dealing with the problem—a choice that would still allow the benefits of trade to occur. Explain why protectionism is a “second-best” choice for:
 - a. helping workers as a group
 - b. helping industries stay strong
 - c. protecting the environment
 - d. advancing national defense
4. Trade has income distribution effects. For example, suppose that because of a government-negotiated reduction in trade barriers, trade between Germany and the Czech Republic increases. Germany sells house paint to the Czech Republic. The Czech Republic sells alarm clocks to Germany. Would you expect this pattern of trade to increase or decrease jobs and wages in the paint industry in Germany? The alarm clock industry in Germany? The paint industry in Czech Republic? The alarm clock industry in Czech Republic? What has to happen for there to be no increase in total unemployment in both countries?

REFERENCES

E. Helpman, and O. Itskhoki, "Labour Market Rigidities, Trade and Unemployment," *The Review of Economic Studies*, 77. 3 (2010): 1100-1137.

M.J. Melitz, and D. Trefler. "Gains from Trade when Firms Matter." *The Journal of Economic Perspectives*, 26.2 (2012): 91-118.

Rauch, J. "Was Mancur Olson Wrong?" *The American*, February 15, 2013. <http://www.american.com/archive/2013/february/was-mancur-olson-wrong>.

Office of the United States Trade Representative. "U.S. Trade Representative Froman Announces FY 2014 WTO Tariff-Rate Quota Allocations for Raw Cane Sugar, Refined and Specialty Sugar and Sugar-Containing Products." Accessed January 6, 2014. <http://www.ustr.gov/about-us/press-office/press-releases/2013/september/WTO-trq-for-sugar>.

The World Bank. "Merchandise trade (% of GDP)." Accessed January 4, 2014. <http://data.world-bank.org/indicator/TG.VAL.TOTL.GD.ZS>.

World Trade Organization. 2014. "Annual Report 2014." Accessed April 1, 2015. https://www.wto.org/english/res_e/booksp_e/anrep_e/anrep14_chap10_e.pdf.

GLOSSARY

disruptive market change innovative new product or production technology which disrupts the status quo in a market, leading the innovators to earn more income and profits and the other firms to lose income and profits, unless they can come up with their own innovations

SOLUTIONS

Answers to Self-Check Questions

1. Competition from firms with better or cheaper products can reduce a business's profits, and may drive it out of business. Workers would similarly lose income or even their jobs.
2. Consumers get better or less expensive products. Businesses with the better or cheaper products increase their profits. Employees of those businesses earn more income. On balance, the gains outweigh the losses to a nation.

CHAPTER 28. THE ECONOMICS OF GLOBALIZATION AND TRADE: A PLURALISTIC APPROACH

INTRODUCTION TO GLOBALIZATION AND TRADE FROM A PLURALISTIC PERSPECTIVE

CHAPTER OBJECTIVES

In this chapter, you will learn about:

- Review the Orthodox Economics Comparative Advantage and Trade Story.
- Examine Critiques of Orthodox Free Trade Story.
- Identify an Alternative Approach for Evaluating the Impact of International Trade.

In Chapter 27: Globalization and Protectionism several concerns associated with free trade are tackled and, presumably, debunked. In 27.2 International Trade and Its Effects on Jobs, Wages, and Working Conditions, three issues are raised.

1. Fewer Jobs?
2. Trade and Wages
3. Labor Standards and Working Conditions

In 20.3 Arguments in Support of Restricting Imports, another five issues are tackled.

1. Infant Industry Argument
2. Anti-Dumping
3. Environmental Protection
4. Low Foreign Wages
5. National Interest Argument

But, are these issues really non-starters? Does orthodox economics effectively minimize the above outlined issues? From the standpoint of heterodox economists, the answer to both of these questions is no. There is a significant body of research available, within the context of a more pluralistic traditions in economics, that point to other interpretations of international trade in which the standard orthodox economic free trade story has very real weaknesses.

The intention of this section is to provide a counter-perspective to that of the orthodox comparative advantage story. The critique is directed toward the orthodox presentation of its theoretical ideas, specifically how orthodox theory promotes the merits and benefits of free trade. The critique will take two forms. The first and primary criticism will focus on the wide array of ways in the orthodox theoretical story fails as both a technical and functional theoretical presentation. The second element

of criticism will, in periodic instances, include references to history as well as empirical evidence to document instances in which the orthodox comparative advantage story appears to not fulfill its claims.

28.1 THE ORTHODOX STORY OF TRADE: A SYNOPSIS

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define the term comparative advantage.
- Explain the terms and conditions of trade from an orthodox economic perspective.

If ever there was a sacrosanct idea within orthodox economic thought, it is the belief, faith even, that orthodox economists have regarding the merits and benefits of economic actors being free to trade. The basis for the orthodox economists' faith has much to do with the theoretical explanation orthodox economists utilize to explain how trade works and why it is beneficial. Consider the ideas presented in the chapter entitled **International Trade**.

As presented, orthodox economics organizes its free trade story around several assumptions. The assumptions are designed to constrain the scope of the analysis so as to simplify the story. In the typical introductory textbook presentation, orthodox economists utilize production possibilities frontiers to depict their story.

In support of the production possibilities frontier presentation, orthodox economists apply eight assumptions. The following represents the eight, most common, assumptions.

1. Two producers
2. Two products
3. Fixed resources.
4. Fixed technology.
5. Full resource utilization.
6. Zero transactions costs
7. No external costs or benefits
8. Autarky

Let's take a moment to reflect upon these assumptions and what they mean. For starters, it is typically to assume that the two producers in the model will be countries. Within the two countries, only two products will be produced. Each country will have a fixed number of resources available to produce those products. Limited resources will subsequently limit the amount of production of the two products that is possible within the two countries. Additionally, the fixed technology assumption further limits the extent of possible production. Given the limits to production, all resources and technology are assumed to be used, fully utilized, meaning that production will take place at some point on the

production possibilities frontier. Once products are produced, if the two countries seek to trade with one another, there will be no transactions costs associated with trade. Transactions costs are costs such as the legal costs of creating contracts or the transportation costs associated with shipping a product from one part of the world to another. In addition to no transactions costs, external costs or benefits are assumed to be absent. As a reminder, external costs or benefits are the possible spillover effects that occur as a result of economic activity such as production or consumption. As one can imagine, because trade requires transportation, the social costs associated with the wastes generated by transportation are very real. However, for the sake of simplicity, the model assumes the costs and/or benefits of externalities are zero. Finally, autarky means that each producer is self-sufficient. Self-sufficient producers do not require trade in order to survive. If trade is a luxury rather than a necessity, then when trade occurs it is voluntary trade and only to the benefit of the trading partners.

Given the model's assumptions, the orthodox free trade story then showcases three essential components to the trade story. First, the initial step in facilitating trade is for producers to identify their comparative advantage. Recall, comparative advantage means that producers are capable of producing some product(s) at a lower opportunity cost than another producer is capable of producing the same product(s). As a producer, regardless of whether the producer is an individual, firm, or nation, as long as the producer is producing the product at a lower opportunity cost than the other producer, the producer with the lower opportunity cost has a comparative advantage. The second step in the trade story is the specialization of production. Upon identifying its comparative advantage, orthodox economics argues that a producer should specialize in the production of the product that which it has a comparative advantage. Specialization means that a producer should direct their available resources toward the production of a specific product. The third component of the trade story is, trade. Specialized producers, producing a product of which they have a comparative advantage, should now seek to identify trading partners that are also specialized producers of desirable products.

1	→	2	→	3
Producer		Producer Specializes in		Trade with
Identifies its		the Production		other Producers
Comparative		of the Comparatively		of other
Advantage		Advantageous Product		Products

THE BENEFITS OF TRADE

Orthodox economics sees the outcomes of trade as having predominantly positive impacts for the economic well-being of those people participating in the trade. Trade is deemed beneficial for several reasons. Orthodox economics concludes that trade will not only stimulate economic growth, increase efficiency, and enhance economic development, but that trade is also the best path toward pursuing those goals.

The story of growth is relatively straightforward. The economic growth, when examined more specifically, such as in the case of international trade as producers get access to consumers beyond their domestic borders, the number of available consumers should grow. In the face of a larger poten-

tial consumer market place, domestic producers will seek to expand their productive capacities, otherwise known as growth, in order to produce more in an effort to take advantage of having more potential consumers.

The story of efficiency is less obvious, but is still relatively straightforward, particularly if viewed through the lens of international trade. In an international trade environment, producers that may have otherwise been subject to their own country's domestic competition, are now confronted by international competition. Given more competition, producers are forced to seek ever more efficient ways to produce products or risk being driven out of business. More trade equates to more competition and, through the eyes of an orthodox economist,

Finally, in the case of the economic development story, the relationship between trade and development is a bit murky but still evident. By focusing on the specialized production product(s) that producers are comparatively good at producing, the total amount of available products should grow. As two or more parties trade, each producer's relative abundance of production becomes available to more people to be consumed. Essentially, with the growth of products comes an increased ability to meet people's wants. As more wants are met, facilitated by trade, the sum total of utility grows for all and everyone is made better off. In this case, economic development is being defined as raised living standards by virtue of access to consumer products.

28.2 A CRITICAL EXAMINATION OF THE ORTHODOX DEPICTION OF FREE TRADE

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain ways in which the Orthodox Economic Trade story is limited as a technique for analyzing trade.
- Define the different types of technical critiques of the orthodox free trade model
- Explain that critiques of the orthodox free trade model are of differing severity.
- Analyze applied examples of the limitations of the orthodox economic model.

Within the discipline of economics not all economists agree with the rather rosy picture presented by orthodox economists in their defense of free trade. Critiques of the orthodox free trade story tend to focus on two areas. Some heterodox economists emphasize what they see as the theoretical failures of the orthodox story, namely its disassociation with the “real world.” Other heterodox economists focus their criticisms toward the orthodoxy’s depiction of the outcomes of trade, particularly the orthodoxy’s conclusion that trade stimulates economic growth and economic development

Upon examination it can be argued that both paths of criticism really represent the two different sides of the same coin. For example, if the theoretical model presented by orthodox economics is, in fact, inaccurate because it is in some way divorced of “real world” context, then the theoretical conclusions drawn by the theory will likely also be flawed. Looked at the other way, if “real world” evidence points to outcomes different from the outcomes predicted or perceived by orthodox economics, then it stands to reason that the theoretical model utilized to predict outcomes is failing in some capacity.

In the following sub-sections each of the two critical paths will be explored. The careful reader will note that the two paths frequently intersect.

TECHNICAL CRITIQUES: THE ORTHODOX MODEL IS POORLY CONSTRUCTED

In this section of the chapter we will explore the model imperfections approach to critiquing the orthodox presentation of free trade. The model imperfections critique is, essentially, a failure of the techniques employed by orthodox economists. There are two main types of model imperfections.

1. Type One – Mild
2. Type Two – More Severe to Very Severe

For Type I critics of the orthodox economic free trade story, the shortcomings associated with the orthodox economic presentation of free trade is strictly the result of a failure of technique, but not necessarily a failure of the overall theory. In other words, the theory does generalize to the real world, but due to model imperfections such as the model's choice of assumptions being overly simplified, the model is unable to fully capture actual real world circumstances. The degree of seriousness associated with these critiques can be classified as mild.

For Type II critics, the model imperfections critique represents a much more serious indictment of the standard orthodox free trade story. In this case, comparative advantage, specialization, and trade does not actually generalize well to the real world. The socially positive results that are depicted in the orthodox theoretical presentation, when these imperfections are accounted for, will not emerge. In other words, because the real world will, and does, contain imperfections that do not conform to the orthodox economic model, the orthodox theoretical presentation will often deviate from actual events in the real world causing the model to blatantly misrepresent the benefits of trade. Importantly, with modifications designed to correct the imperfections, the model can be corrected enough so as to reflect conditions in the real world.

A NOTE ON THE FAILURE OF TECHNIQUE

A failure of technique means that by making alterations to the choice of assumptions, the comparative advantage story can be modified to present theoretical outcomes that are more accurate to the real world. As a general rule the changes in assumptions are applied on an ad hoc basis, meaning that assumptions are altered from situation to situation. The goal becomes to modify assumptions in order to reflect phenomena experienced in the actual global economy. The modification of assumptions suggests that there are limits to the usefulness of the techniques utilized within orthodox economic theory to explain.

TYPE I MODEL IMPERFECTIONS

The following represent examples of instances in which the restrictive features of the assumptions do not provide a "real world" depiction, but also do not fundamentally undermine the free trade story.

EXCLUSION OF TRANSACTIONS COSTS

In constructing their theoretical presentation of international trade, orthodox economics excludes transactions costs. Transactions costs include factors such as transportation, legal, and communications costs. As one can imagine, in the actual world in which we live the trade of products will seemingly always include significant transaction costs. When transactions costs are included some trades that would have otherwise been beneficial will not be worthwhile as the costs associated with making the trade become too high to justify the trade. While this critique certainly does point to instances in which some trades will not happen as would be the case in the standard orthodox trade story, the critique is not a deal breaker as most of the orthodox trade story remains intact.

NON-LINEAR PRODUCTION POSSIBILITIES FRONTIERS

Production possibilities curves are not linear (straight line) but rather concave (bowed outward). A bowed outward PPF reflects the principle of increasing opportunity cost. Increasing opportunity cost implies that as a nation moves closer and closer to specializing in the production of one good or one type of good, the opportunity cost of specialization will rise. For example, for countries producing

agricultural commodities, there tend to be land or ecology limits on their productive capacity. As a result, to specialize in an agricultural product and then expanding production in an agricultural producing sector may cause there to be an increase in the average cost of production due to declining productivity associated with the diseconomies of scale that accompanies introduction of less and less fertile land.

Since the free trade story is told on the basis of comparative advantage such that a nation produces those goods that require a low opportunity cost of production, if opportunity costs are changing (specifically, rising), then production of a good that may have a low opportunity cost at one point, may have a significantly higher opportunity cost after specialization. The following table illustrates this point.

		CANADA	
SILVER (OUNCES)	0	500,000	600,000
LUMBER (BOARD – FEET)	1,000,000	500,000	0
	1L : 1S	1L : 0.2S	
		1S : 1L	1S : 5L
		MEXICO	
SILVER (OUNCES)	0	40,000	80,000
LUMBER (BOARD – FEET)	80,000	50,000	0
	1L : 1.33S	1L : 0.8S	
		1S : 0.75L	1S : 1.25L

The ratios in red illustrate the lowest opportunity costs of production for both Canada and Mexico. For Canada, the numbers indicate that they have a comparative advantage in producing lumber because they can produce 1 unit of lumber at the expense of only 0.2 units of silver. Canada's comparative advantage outcome emerges when Canadian production changes from 0 lumber and 600,000 silver to 500,000 lumber and 500,000 silver. For Mexico, the numbers indicate that they have a comparative advantage in producing silver because they are capable of producing 1 unit of silver for 0.75 units of lumber. Mexico's comparative advantage emerges when Mexican production changes from 0 silver and 80,000 lumber to 40,000 silver and 40,000 lumber.

According to the logic of orthodox economic theory, now that the Canadian and Mexican comparative advantages have been identified, with Canada having the advantage in lumber and Mexico the advantage in silver, Mexico and Canada should then specialize and trade. Specialization means that Canada should now only produce lumber, moving their production outcome to 0 silver and 1,000,000 lumber while Mexico should now only produce silver, moving their production to 0 lumber and 80,000 silver. It is at this point that a problem appears. Once each country specializes, what was a comparative advantage dissolves into a comparative disadvantage.

Examining the table, as a result of the principle of increasing opportunity cost, the opportunity cost in silver of specializing in lumber increases as Canada moves to maximum lumber production becom-

ing a 1 lumber to 1 silver ratio. Mexico experiences as similar problem except theirs is the problem of a rising opportunity cost associated with forgone lumber. For Mexico the ratio changes to 1 silver for 1.25 lumber (which is also 1 lumber for 0.8 silver). In fact, as is demonstrated by this table, once the country's specialize, each country now has a comparative advantage in the opposite product from what it previously had a comparative advantage. Post-specialization, Mexico is now capable of producing lumber at a lower opportunity cost than Canada and Canada is now able to produce silver at a lower opportunity cost than Mexico. Clearly the orthodox economic story does not fit this circumstance. This is not to say that free trade necessarily breaks down as a result of this scenario, but it does demonstrate that in a more realistic setting, specialization and trade are not appropriate. In fact, one thing the table clearly demonstrates is that both countries are better off producing a combination, a diversity, of products. At this point, based on a diversity of products produced, if the countries trade, then trade will allow both countries to acquire even more of the respective product that they do not have a comparative advantage in producing.

TYPE II MODEL IMPERFECTIONS

The following examples provide an overview of some of the Type II imperfections that appear to generate a divide between real, actual, world circumstances and the orthodox economic trade story. Each imperfection represents a more fundamental challenge to the orthodox model of trade than did the Type I imperfections. As such, the imperfections are presented from least significant to most significant in terms of their implications for the orthodox model.

SPECIALIZATION AND ECONOMIC VULNERABILITY

Any time production becomes highly specialized, the producer confronts potential risks. Here are two potential risks associated with specialization.

1. Any country that chooses to completely specialize in the production of one product runs the risk of a decline in the market for the specialized good. For example, consider a circumstance in which consumers lose interest in the product. In this case market demand declines and the country specializing in the production of this product will necessarily suffer significant economic damage.
2. Any country that specializes in the production of an agricultural good, beyond losses in consumer demand also face production risks. A change in weather or a natural disaster has the potential to completely undermine the economy of the nation(s) producing the single agricultural product and devastate an economy.

In either of the above stated situations, the orthodox economic advice of specialization and trade has the potential to cause negative, not positive effects, for the participants. Clearly there are potentially economically devastating risks associated with a program of rigid specialization.

Although the issue of risks does present a more pressing critique of the standard orthodox free trade story, because potential risks are not definite risks, absent the emergence of these events, the orthodox free trade story is still applicable. Additionally, if the above stated risks are a concern, one can argue that the orthodox model can alter its assumptions to acknowledge heightened risks and, subsequently, advocate for more diversified economies. Diversified economies can still trade and, within the orthodox economic story, these economies would still be encouraged to open their borders to the free flow of products.

SPECIALIZATION AND LESS THAN FULL EMPLOYMENT

While the orthodox economic free trade story presents the idea that trade may produce increases in available goods and services to the trading partners, as a nation or nations specializes the industries suffering from a comparative disadvantage will, ultimately, displace laborers. Because the losses in one industry may not be fully compensated in another industry or industries, more laborers may be displaced than necessarily employed, thus causing unemployment.

An example of the above situation, consider the trading situation of two countries where one country has an abundance of labor and the other country has an abundance of capital (tools, equipment, machinery). In the country that tends to have an abundance of labor the firms within that country would tend to use a lot of labor, and less capital, for their production. In the country that tends to have an abundance of capital, the firms within the capital intensive country will tend to use a lot of capital, rather than labor, in their production processes.

According to orthodox economic theory, in identifying their respective comparative advantages, the labor intensive country would determine that it has a comparative advantage in producing products that require a lot of labor while the capital intensive country would find that it has an advantage in producing products that require a lot of tools, equipment and machinery. The challenge that emerges in this situation is that not all firms in the capital intensive country will be capital intensive producers, some will be labor intensive producers. The labor intensive firms in the capital intensive country will discover that they are at a comparative disadvantage against labor intensive producers from the labor intensive country. The resulting outcome will be that the labor intensive firms, with an abundance of labor, from the capital intensive country, will begin laying off workers. At the same time, while the capital intensive firms in the capital intensive country will likely grow and expand with trade, thus hiring more laborers, it is likely that their hiring will be less than the number of layoffs from the labor intensive firms. In summary, because the firm using a lot of machines tends to use capital rather than labor for its production the number of laborers it absorbs will likely be less than the number of workers the labor intensive firm fires. The overall outcome would be a rise in unemployment in the capital intensive country.

In this instance, the increase in unemployment being described would imply that trading countries will frequently not be operating at their maximum point of productive possibilities. The notion of less than full employment is a concept that is, more often than not, the case across the entire global economy. The evidence of less than full employment is overwhelming. For example, over the last three decades (as global free trade policies and agreements have expanded, not contracted) unemployment rates across countries identified as economically advanced and developed (countries such as the United States, Canada, Japan, Germany, France, the United Kingdom, Australia, Sweden, and others), unemployment rates have ranged from between a low of 3% to as high as 25%. Additionally, in the rest of world, where the majority of the global population resides the absolute number of unemployed or underemployed laborers exceeds 1 billion people. Clearly less than full employment is the norm, not the exception.

A less than full employment outcome undermines the standard orthodox free trade story because countries at less than full employment can produce more of any two products while opportunity costs are zero. In the case of less than full employment, determine a comparative advantage is impossible because both products reflect the lowest possible opportunity cost, zero. In a circumstance such as this, the orthodox free trade story is not capable of explaining the very real world situation of trade under conditions of less than full employment. Correcting the model to account for this imperfection

would require that the model assume that an action will be taken to push the economy back to full employment. Such actions could include government (public) policy measures designed to alleviate the unemployment that emerges as a result of trade. Even for many orthodox economists acknowledging the less than full employment scenario is a justification for government action to mitigate the negative consequences associated with trade.

28.3 CHALLENGING FUNCTIONALITY: A MORE PENETRATING CRITIQUE

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify different radical critiques of the orthodox free trade model
- Define the phrase factor endowments.
- Define the concept of an export-led growth model.
- Explain that critiques of the orthodox free trade model are not only technical but also functional.
- Analyze ways in which the orthodox economic trade does not explain the “real world.”

Whereas the model imperfections critiques often cite the idea that the orthodox theory, while too simple, still provides an approximation of the real world. More radical critiques of the orthodox free trade story argue that not only does the orthodox story not generalize to the actual world in which we live, but that the model also cannot be modified to explain certain events or phenomena. For many more radical economists the usefulness of any economic model is how well it can explain the basic functioning of the economy. Failure to account for the actual functional behavior of the economy cannot be corrected by ad hoc assumptions. Instead, what is required, is an altogether different theoretical structure.

The following critiques reflect functional limitations associated with the orthodox trade story. Functional limitations imply the failure of theory to adequately explain events relevant to the real world. Perhaps the theory is simply too narrow in scope and fails to account for other important economic considerations, such as economic growth or economic development. For example, the whole point of the orthodox comparative advantage story is to demonstrate trade between two countries improves the overall economic plight of both countries. Paradoxically, even as the immediate exchange benefits appear, the same theory can be utilized to also explain why the long term economic growth within both economies may be harmed as a result of free trade. In other words, trade is shown to make countries both better off and worse off? How is this possible? The answer is that it is probably not possible, therefore the functional usefulness of the orthodox economic trade story is called into question.

THREE COUNTRIES AND THREE PRODUCTS

Perhaps no example of the limits to the applicability of the orthodox free trade story is more direct than the curious situation of three countries and three products. Given three countries and three commodities, it is not obvious that each country will have a clear lowest opportunity cost product to

produce. Absent a clear comparative advantage, what does a country do? If a country doesn't have a comparative advantage does it produce nothing? Does it produce all three products? Can the country without a comparative advantage refuse to engage in trade? Should the country just export its entire population, put up a border wall, and stop functioning? The orthodox presentation of the trade story cannot answer any of these questions. The absence of comparative advantage is the absence of a theoretical explanation.

Clearly this issue of three countries and three products is not the exception, but rather the rule. In the actual world where there are over 200 hundred countries in the world and, potentially, millions of products. The comparative advantage story is not at all relevant to the actual world, and this proves to be a drastic failure of the orthodox economic model.

Consider the table below:

	COLOMBIA	SCOTLAND	PERU
COFFEE	100	10	30
WOOL	50	40	30
SOYBEANS	60	20	30
	1C : .5W	1C : 4W	1C : 1W
	1C : .6S	1C : 2S	1C : 1S
	1W : 2C	1W : .25C	1W : 1C
	1W : 1.2S	1W : .5S	1W : 1S
	1S : 1.67C	1S : .5C	1S : 1C
	1S : .83W	1S : 2W	1S : 1W

The table represents a scenario in which there are three countries and three products. In the above table, one country will not have a comparative advantage in producing any of the products. The numbers and letters highlighted in red indicate the comparison between two products and the point of lowest opportunity cost production which, recall, is the basis for comparative advantage. For example, as can be seen above, Colombia has a comparative advantage in producing coffee relative to both wool and soybeans. In the case of coffee to wool, every 1 unit of coffee produced requires that Colombia foregoes producing $\frac{1}{2}$ unit of wool. In the case of coffee to soybeans, for every 1 unit of coffee produced $\frac{3}{5}$ of a unit of soybeans are foregone. No other producer of coffee can produce coffee at such a low opportunity cost. The reader should note, Colombia also has a comparative advantage in producing soybeans relative to wool, while Scotland has a comparative advantage in producing wool to soybeans and wool to coffee as well as soybeans to coffee. At no point does Peru ever have a comparative advantage. So what should Peru do? The answers really are not remotely obvious with possible solutions becoming increasingly complex. Consider just one possible scenario.

In one scenario, it can be argued that the only viable product that Peru could specialize in producing is soybeans. After all, because Colombia has a comparative advantage in coffee to wool and coffee to soybeans, and its coffee to wool ratio is better than its soybeans to wool ratio, Colombia may find it desirable to specialize in coffee production. Additionally, because Scotland has a comparative advan-

tage in wool to coffee and wool to soybeans, and its wool to coffee ratio is better than its wool to soybeans ratio, Scotland may find it desirable to specialize in wool production. In this case, Peru could, theoretically specialize in soybean production as neither Colombia nor Scotland are producing soybeans but the residents of those countries may want to consume soybeans so Colombia and Scotland could then trade for soybeans and, as Peru is the only soybean producer, Colombia and Scotland would be forced to trade with Peru.

The trouble with the Peru specializing in soybeans scenario is that it wouldn't take Colombia and Scotland too long to realize that they are each better off producing some soybeans for themselves, because each can produce soybeans at a lower opportunity cost than can Peru. For example, if Colombia produced coffee and traded with Peru for Soybeans, the minimum price that Peru could charge Colombia for the soybeans would be 1 unit of soybeans for 1 unit of coffee. If Scotland decided to produce some soybeans to trade with Colombia for coffee, the minimum price Scotland could charge Colombia for soybeans is 1 unit of soybeans for a $\frac{1}{2}$ unit of coffee. In this case Peru would simply not be able to compete with Scotland's price. The same circumstance emerges with Scotland trading wool for soybeans with Peru. The lowest price Peru can charge Scotland for 1 unit of soybeans is 1 unit of wool. If Colombia decided to produce some soybeans to trade with Scotland for wool, the lowest price Colombia could charge is 1 soybean for .83 wool. Peru is, once again, not price competitive and Scotland and Colombia are better off trading with one another for soybeans than they would be trading with Peru for soybeans.

COMPARATIVE ADVANTAGE AND CONSTANT RETURNS TO SCALE

The comparative advantage story is based on an assumption of constant returns to scale. As a reminder, constant returns to scale implies that as a productive process grows, the returns in total output generated grows proportionally to the expansion of the productive process. In other words, if a production process doubles in size, then the amount of productive output will also double. The constant returns to scale concept is closely tied to the linear production possibilities frontier concept discussed above. After all, if a production possibilities frontier is linear, and all productive inputs double, then the production possibilities frontier will shift outward proportionally, remaining parallel to the previous production possibilities frontier. The essential point to be grasped here is this, constant returns to scale imply that as production of one product increases

The challenge associated with the concept of constant returns to scale is, what happens if returns to scale are not constant for producers? Orthodox economic theory provides a theoretical foundation for the idea that sometimes production processes experience increasing returns to scale. As a reminder, increasing returns to scale implies that as a production process expands, the associated increase in production expands by more than the growth of the production process. In other words, as a production process doubles its size it will more than double its output yielding a declining average total cost for the output as more output is produced.

Under conditions of increasing returns to scale in production, the country that is the first to produce a commodity will develop a comparative advantage in the production of the commodity. The reason why the first producer of the product will have the comparative advantage is because as that country increases its production of the product, its costs of doing so will decline making it difficult for another country to compete with the first country to produce the product. In other words, one country will have a head start that other countries will not be able to competitively overcome. Presumably, where increasing returns to scale hold for production processes, the first producer of any product will have

a competitive advantage over all producers. In this situation, one country could come to dominate production throughout many industries leaving other countries to produce a limited range of other products.

Of course, where increasing returns to scale applies the thought that one country can dominate production may not be acceptable to other countries, particularly for countries attempting to develop economically. The implication is that countries need to protect their industries until those industries have increased their production to the point of increasing returns to scale and, thus, declining average costs of production emerge. Only when a producer has achieved low enough average total costs will it be able to compete effectively in global markets.

ABSOLUTE COSTS ARE MORE IMPORTANT THAN FACTOR ENDOWMENTS

One theme that emerged above when describing model imperfections is the notion that due to **factor endowments**, some countries are going to be more prone to a greater loss of employment from specialization and trade than would other countries. The basis for the less than full employment argument was a byproduct of the concept of factor endowments. The concept of **factor endowments** is an idea that argues a country's comparative advantage may be influenced by which factors of production the country has an abundance of and which are relatively scarce. A country is said to be factor endowed when it has a relative abundance of a given factor of production. When a country has an abundance of a factor of production, orthodox economics argues that the country will tend to produce products whose production processes are organized to require a lot of the factor of which the country has an abundance. Due to the relative low cost of the abundant factor, the country will likely have a comparative advantage in producing the product that which they have an abundance of the factor of production. So, for example, a country with an abundance of land will likely produce agricultural products as land is the relatively low cost factor of production that the country is more abundantly endowed.

Taken further, the factor endowment story goes so far as to argue that production processes will be modified to adjust to a country's factor abundance and factor scarcity. For orthodox economics what this means is that if a country has an abundance of capital, then the products it produces will utilize an abundance of capital even if the same product can be produced elsewhere with an abundance of labor. Should labor be the factor of production that a country is relatively well endowed, then production processes will be modified to include more labor and less capital, and so on, and so forth.

Certainly, for some types of production, the factor endowment story will be relevant, which also makes the idea of a less than fully employed economy, as argued in the sub-section above, also relevant. However, what happens when the product and the production process are inextricably linked? It is not at all uncommon for a production process to require the same combination of land, labor, and capital, regardless of where production takes place. So even if a country has a large supply of labor, and relatively cheap labor, this does not mean that firms will suddenly change their production processes to take advantage of the cheap factor endowment. An example of this concept can be seen with a company like Boeing. Boeing produces commercial and military aircraft. The production processes require a significant amount of technological know-how, with capital and labor being very specialized and highly skilled. Furthermore, the capital and technology employed in the production of an airplane are very expensive. It is also a well-known fact that Boeing produces airplanes in the United States and China. In comparison to one-another, the United States has an abundance of capital and China has an abundance of labor. If factor endowments were relevant, Boeing would

produce its aircraft differently in China than in the United States. In China, Boeing would use more labor and in the United States Boeing would use more capital. In fact, Boeing uses the same combination of capital and labor in its production processes regardless of location of production. The advantage to producing in China is, then, some reductions in absolute labor costs. In this scenario, absolute costs are far more important than some vague notion of opportunity cost. Furthermore, the labor cost reduction is not, however, necessarily significant because the overall costs of producing airplanes tend to be more closely tied to capital, due to the advanced capital and technology used in production, as opposed to overall labor costs. As such, if all of the other costs, such as transportation costs, are not favorable, then it may not be cost effective to produce airplanes in China as opposed to the United States. Either way, absolute costs, not opportunity costs, are often times the most important determinant of what is produced and where it will be produced.

MARKETS ARE NOT STATIC

So much of the orthodox free trade story seems to be built on a model that yields a surprising lack of choices for producers. After all a country is deemed to be endowed with an allotment of land, labor, and capital, and among those three factors of production the choice of what is produced is determined by which products can be produced at the lowest opportunity cost of production. In this scenario, the country does not choose its initial endowments of factors of production nor, at that point, is the country really making a choice as to what products to produce as the opportunity cost of production will depend on the country's endowment of factors of production relatively to all other countries' endowments of factors of production. In essence, any country is relegated to being at the mercy of global circumstances and global markets.

Being at the mercy of the market is an awfully risky proposition for many countries. A country better hope that what it specializes to produce and then trade will always be a desirable commodity. Unfortunately for many countries, the reality of global markets, where new technologies and changes in tastes and preferences arise on a regular basis, often times causes countries to suffer significant economic hardship and economic development setbacks.

But, are countries and their economies really at the mercy of a global market? If history is a guide, the answer to this question appears to be unequivocally, no. Many students evaluating the factor endowment story outlined above might, for example, question why a given country would accept its factor endowment position? After all, factor endowments such as labor and capital are not, over time, static. Given time, populations tend to grow and countries will find that their allotment of capital stock will also tend to grow. It is not as though the United States was endowed with capital stock from some divine being and, as a result, is a capital intensive country. Any country could, presumably, develop a capital intensive economy. If so desired, countries have the power to influence where capital stock accumulates and toward what end that capital stock can and should be used. Under conditions such as this it appears as though countries have the power to direct their place in the global economy.

Among advanced developed economies there are countless examples of industrial planning and trade protections, essentially market manipulations, being utilized to promote domestic economic objectives and steer a country's economic place in the global economy. In the 18th and 19th centuries countries such as Great Britain and the United States both utilized trade restrictions in an effort to support the development of domestic, high value added, industries, particularly manufacturing. Into and throughout the 20th century, the United States as well as countries such as Germany, Japan, and South Korea have all also utilized trade restrictions and industrial planning policies. In the case of

post-World War 2 Japan and Germany, trade restrictions and planning policies were utilized to allow domestic industries that had been destroyed during the war to re-establish themselves in their respective domestic markets. By the 1970s, the protective policies utilized by Japan and Germany allowed both countries to engage in international export of products such as automobiles and electronics, stimulating an **export-led growth model**. An export-led growth model is the idea that a country can use trade surpluses, exporting more products than they import, to boost domestic economic activity, generate economic development, and improve the country's international competitiveness in the global economy.

Not to be left out, other countries, particularly in Asia, also recognized the benefit of export-led economic growth. South Korea, as one of the Asian Tigers, used trade protections to bolster and create a domestic electronics and automobile industry. Over time South Korean producers such as Samsung in electronics as well as Hyundai and Kia in the automobile industry have become internationally competitive and internationally recognizable producers. Another example is China. The Chinese economy, through use of targeted industrial planning, trade restrictions, and currency adjustments, has utilized an export-led growth model. China has run consistent trade surpluses with the countries within the European Union as well as the United States. The results, since China's "Great Opening Up" in 1978, which began under then Chinese leader Deng Xiaoping, have been an economic boon for China. In the nearly forty years since 1978, China has experienced consistent annualized growth rates of 7-10%.

A NOTE ON CHINA'S INCLUSION IN THE WORLD TRADE ORGANIZATION

As of December 11, 2001, China has been a member of the World Trade Organization (WTO). As noted by the policy think tank, the Economic Policy Institute (EPI), China's inclusion in the WTO has not been benign in terms of economic consequences. An EPI study suggests that 2001 and 2013, China's export-led growth model and subsequent trade surpluses with the United States (or U.S. trade deficits with China) has caused the United States to lose 3.2 million jobs, including 2.4 million job losses in the manufacturing sector.

GLOSSARY

factor endowments

a country's comparative advantage may be influenced by which factors of production the country has an abundance of and which are relatively scarce

export-led growth model

the idea that a country can use trade surpluses, exporting more products than they import, to boost domestic economic activity, generate economic development, and improve the country's international competitiveness in the global economy

28.4 AN ALTERNATIVE PRESENTATION OF INTERNATIONAL TRADE: PATH DEPENDENCY

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define the term path dependence.
- Explain how a path dependency analysis of trade is different from the comparative advantage approach.
- Define the phrase factor endowments.
- Define the concept of an export-led growth model.
- Analyze different “real world” examples that support the path dependency approach to understanding international trade.

Given the critiques outlined above, many progressive and radical (heterodox) economists believe that the orthodox economic free trade theory is chronically incapable of addressing important, real world, concerns. Because of its failure to explain real world conditions, the orthodox model is also incapable of suggesting policy approaches that may be helpful in correcting many of the problems that emerge when trade between countries expands and grows. For example, in the orthodox world of free trade, any, and all, trade protections are perceived as inefficient and costly. As a result, if the real world provides evidence of the effectiveness of trade restrictions, a story built on the benefits of free trade cannot and will not be functionally applicable. Using the free trade story in this regard would generate inaccurate analyses, failing to recognize that trade protections have had historic usefulness toward assisting an economy’s economic development. In this instance, the free trade story is of no practical/functional use. What is needed, instead, is a different theory of how trade works. A sound theory of regarding trade should be able to demonstrate how trade restrictions may stimulate more economic growth and/or improve a country’s economic development prospects.

Because theorists outside of the orthodoxy view the orthodox economic trade story as being impossibly limited in its real world applications, heterodox economists generally begin their trade story from a different perspective. For many heterodox economists, in the real world, specialization and trade are **path dependent**. Path dependence means that history matters. A given country’s productive situation is the byproduct of past circumstances, past decision making (with respect to domestic economic choices as well as the country’s place in the global economy). Further a country’s economic development, its future productive possibilities and future economic position in the global economy, depends on the choices it makes in the present.

To follow the logic of comparative advantage, a country identifies its comparative advantage, specializes and then trades. So if agricultural products are the only products that a country has a comparative advantage producing, then the country will specialize in the production of agricultural products only and then trade for the other non-agriculture products that are desired. In response to this scenario, heterodox economists ask, for how long does a country only produce agricultural products? After all, from an economic development standpoint it is well known that there is not a single advanced developed economy that only produces agricultural products. In fact, the production of products in all advanced developed economies are overwhelming diverse.

ECONOMIC DEVELOPMENT AND THE CASE FOR CONSCIOUS DIVERSIFICATION

The United States is widely recognized for having one of the most diverse economies in the world. U.S. based firms produce agricultural products, light and heavy manufactured products, industrial products, pharmaceutical products, and so on and so forth. Aside from no longer being de-stabilized due to weather volatility or changes in consumer tastes and preferences, a diversified economy can better endure overall market volatility because some sectors may grow when others contract. Additionally, as markets change, as new technologies emerge, a diversified economy is better apt to be able to find uses for new technologies and integrate those technologies into their economy. As a result, in the path-dependency-trade-story, diversification, rather than simple specialization and trade, is encouraged because diversification is an economic development strategy. Importantly, as is noted by the critiques provided above, *diversification cannot be achieved by comparative advantage and a free trade environment*.

By evaluating trade through a lens of path dependency, heterodox economists are able to correct for some of real world limitations associated with the orthodox free trade story. For example, the very nature of path dependency requires strategic thinking and public policy planning on the part of countries and their governments. Countries will often target industries for protection from international competition because they see an industry as generating a high-value added, high income, and significant employment outcome for their economy. The automobile industry is an example of industry where countries have been apt to enact policies to encourage both domestic production as well as foreign direct investment into the production of automobiles. Countries such as Azerbaijan, Egypt, and Malaysia all have developing internal domestic automobile producers. Additionally, China is currently the world's largest producer and consumer of automobiles with production being generated by a large number of foreign firms such as Volkswagen and General Motors as well as the domestic Chinese producer, Geely. The benefit of having a viable domestic automobile industry is that automobiles are one of the highest value added mass-consumed products in the world. If a country can produce automobiles to meet even just some of its internal consumer demand for automobiles, then the country will be retaining valuable domestic expenditures in the form of valuable domestic income.

Finally, a path dependency approach to examining trade between countries empowers the heterodox economist to understand why some countries have had more relative success than others with respect to economic development. By thinking of trade as a result of path dependency, the role of trade restrictions, industrial planning, and targeted competition can reveal to the theorist tactics and approaches that have successfully enhanced economic development or, alternatively stifled economic development. As a result, the fact that advanced developed economies have a history of managed trade, rather than free trade, can be understood as strategic decision making. Undoubtedly, advanced developed economies owe some of their economic position to the strategic trade choices that they made in the past. Furthermore, the recent success of a country like China can also be explained as

being related to the strategic, State planned and managed, trade decisions that the Chinese government has implemented. Finally, the path dependency approach can also be utilized to explain why persistent free trade policies have also been use to undermine the economic development of many countries, exacerbating persistent inequalities between countries.