UNDERSTANDING CAPITALISM

COMPETITION, COMMAND, AND CHANGE

FOURTH EDITION

SAMUEL BOWLES | RICHARD EDWARDS
FRANK ROOSEVELT | MEHRENE LARUDEE

OXFORD
UNIVERSITY PRESS

Not for Profit. All for Education.

Oxford University Press USA is a not-for-profit publisher dedicated to offering the highest quality textbooks at the best possible prices. We believe that it is important to provide everyone with access to superior textbooks at affordable prices. Oxford University Press textbooks are 30%–70% less expensive than comparable books from commercial publishers.

The press is a department of the University of Oxford, and our publishing proudly serves the university's mission: promoting excellence in research, scholarship, and education around the globe. We do not publish in order to generate revenue: we generate revenue in order to publish and also to fund scholarships, provide start-up grants to early-stage researchers, and refurbish libraries.

What does this mean to you?
It means that Oxford University Press USA published this book to best support your studies while also being mindful of your wallet.

***Not* for Profit. *All* for Education.**
As a not-for-profit publisher, Oxford University Press USA is uniquely situated to offer the highest quality scholarship at the best possible prices.

OXFORD
UNIVERSITY PRESS

Understanding Capitalism



Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

https://archive.org/details/understandingcap0000bowl

Understanding Capitalism

COMPETITION, COMMAND, AND CHANGE

FOURTH EDITION

SAMUEL BOWLES
Santa Fe Institute

RICHARD EDWARDS
University of Nebraska – Lincoln

FRANK ROOSEVELT
Sarah Lawrence College (Emeritus)

MEHRENE LARUDEE
Hampshire College

New York Oxford
OXFORD UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trademark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

© 2006, 2018 by Oxford University Press
© 1985, 1993 by HarperCollins

For titles covered by Section 112 of the US Higher Education Opportunity Act, please visit www.oup.com/us/he for the latest information about pricing and alternate formats.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data

Names: Bowles, Samuel, author. | Edwards, Richard, 1944- author. | Roosevelt, Frank, 1938- author.
Title: Understanding capitalism : competition, command, and change / Samuel Bowles, Richard Edwards, Frank Roosevelt, Mehrene Larudee.
Description: Fourth Edition. | New York : Oxford University Press, [2017] | Revised edition of the authors' Understanding capitalism, 2005. | Includes index.
Identifiers: LCCN 2017018189| ISBN 9780190610937 (pbk.) | ISBN 9780190610944 (electronic text)
Subjects: LCSH: Economics. | Capitalism.
Classification: LCC HB171.5 .B6937 2017 | DDC 330.12/2—dc23
LC record available at https://lccn.loc.gov/2017018189

9 8 7 6 5 4 3 2
Printed by Sheridan Books, Inc., United States of America

Brief Contents

PREFACE xix
FIGURES AND TABLES xxiii
ACKNOWLEDGMENTS xxix

PART 1 POLITICAL ECONOMY

1 Capitalism Shakes the World 1
2 People, Preferences, and Society 29
3 A Three-Dimensional Approach to Economics 49
4 The Surplus Product: Conflict and Change 67
5 Capitalism as an Economic System 89
6 Government and the Economy 113
7 U.S. Capitalism: Accumulation and Change 132

PART 2 MICROECONOMICS

8 Supply and Demand: How Markets Work 169
9 Competition and Coordination: The Invisible Hand 181
10 Capitalist Production and Profits 207
11 Competition and Concentration 233
12 Wages and Work 271
13 Technology, Control, and Conflict in the Workplace 299

PART 3 MACROECONOMICS

14 The Mosaic of Inequality 325
15 Progress and Poverty on a World Scale 352
16 Aggregate Demand, Employment, and Unemployment 383
17 The Dilemmas of Macroeconomic Policy 422
18 Financial and Economic Crisis 454
19 Government and the Economy in Transnational Capitalism 491

LIST OF VARIABLES 529
SOURCES OF ECONOMIC INFORMATION 532
GLOSSARY 535
INDEX 544

CONTENTS

PREFACE xix

LIST OF ILLUSTRATIONS xxiii

ACKNOWLEDGMENTS xxix

PART 1 POLITICAL ECONOMY

1 Capitalism Shakes the World 1

The Permanent Technological Revolution 3

The Enrichment of Material Life 6

Growing Inequality 11

The Population Explosion and the Growth of Cities 12

The Changing Nature of Work 15

The Transformation of the Family 16

Threats to the Ecosystem 17

The Effects of Climate Change 17

Contamination 21

New Roles for Government 22

Globalization 23

Conclusion 26

2 People, Preferences, and Society 29

Constraints, Preferences, and Beliefs 32

"Economic Man" Reconsidered 34

Human Nature and Cultural Differences 36

The Economy Produces People 39

The Cooperative Species 45
Conclusion 46

3 A Three-Dimensional Approach to Economics 49

Economic Systems and Capitalism 50
Three-Dimensional Economics 51
Competition 51
Command 52
Change 53
Neoclassical Economics 55
Values in Political Economy 57
Efficiency 59
Inputs and Outputs 59
Pareto Efficiency 60
Fairness 60
Democracy 62
Balancing Efficiency, Fairness, and Democracy 63

4 The Surplus Product: Conflict and Change 67

Economic Interdependence, Production, and Reproduction 69
Production 70
Links Between Production and Reproduction 71
A Labor Process Example: Making Pancakes 73
The Surplus Product 74
A Model of Production and Reproduction 78
Who Gets the Surplus? 80
Enlarging the Surplus 80
Application of Model to Labor 83
The Surplus Product and Conflict 84
The Surplus Product and Change 86

5 Capitalism as an Economic System 89

Class and Class Relationships 91
Classes and Economic Systems 93
Slavery 93

Feudalism 94
Distinctions Among Economic Systems 94
Capitalism 96
Commodities 97
Privately Owned Capital Goods 102
Wage Labor 108
Capitalism, the Surplus Product, and Profits 108
Conclusion 110

6 Government and the Economy 113

Rules of the Game: Government and the Capitalist Economy 115
The Emergence of Modern Legal Forms of Corporate Business 115
Limited Liability 115
Corporate Personhood 116
From Competition to Monopoly 117
From National to Transnational 118
Democratic Government as Collective Provision for the General Welfare 119
Government and the Distribution of Income, Wealth, and Welfare in the Nineteenth Century 120
Railroad Financing: Public and Private Gains 121
Other Rules: Winners and Losers 123
Contention Between Capital and Labor over Rules 124
Voter Turnout in the United States and Around the World 126
Democracy in the Twentieth Century 128
Conclusion 130

7 U.S. Capitalism: Accumulation and Change 132

Accumulation as a Source of Change 134
The Accumulation Process 135
Competition for Profits 136
Capitalism Comes to the United States 137
Social Structures of Accumulation 141

Changing Strategies for Profit-Making 142
Consolidation and Decay of an SSA 143
The Stages of Capitalism in the United States 144
Competitive Capitalism (1860s–1898) 144
Corporate Capitalism (1898–1939) 146
Regulated Capitalism (1939–1991) 146
Transnational Capitalism (1991–) 147
U.S. Capitalism: Labor Organizing and Labor Markets 147
The Rise and Fall of Labor Unions 147
The Decline of Labor Unions 149
The Rise and Fall of the Labor Accord 151
Segmented Labor Markets 152
The Independent Primary Labor Market 153
The Subordinate Primary Labor Market 154
The Secondary Labor Market 155
U.S. Capitalism Today is Transnational 155
Immigration 157
Shifts in What U.S. Labor Produces 159
Fragmenting Global Production 161
Rules for the Global Economy 162
Tax Motives for TNCs to Go Global 163
The Transnational Capitalism SSA: Deregulation and Financialization 163
Corporate Stock Buybacks and Falling Productive Investment 166
Conclusion 167

PART 2 MICROECONOMICS

8 Supply and Demand: How Markets Work 169

The Nature of Markets 170
Demand and Supply 171
Demand 171
Supply 173
Marginal Cost 174
Average Cost 175
Demand and Supply Interacting 175

Shifts in Demand or Supply 178
Conclusion 180

9 Competition and Coordination: The Invisible Hand 181

Adam Smith and Laissez-Faire Economics 182
Coordination 183
Coordination by Rules and by Command 184
The Limits of Coordination by Command 185
The Invisible Hand 187
The Invisible Hand in Action 188
Problems with the Invisible Hand 191
Market Failure 192
Negative Externalities as Market Failures 193
Positive Externalities as Market Failures 194
Monopoly as a Market Failure 196
Economies of Scale as a Market Failure 196
Coordination Failure 200
The Prisoner's Dilemma and the Benefits of Cooperation 200
The Prisoner's Dilemma and Global Warming 202
The Tragedy of the Commons 203

10 Capitalist Production and Profits 207

What Are Profits? 208
Profits from a Production Process 209
Calculating Profit and Other Property Incomes in the Whole Economy 211
Profit in a Grain-Growing Capitalist Economy 213
Calculating the Rate of Profit 213
The Corporate Profit Rate in the United States 215
Corporations and Other Businesses: Who Gets the Profit or Bears the Loss? 216
Corporations 217
Limited Liability 217
Financing by Issuing Bonds or Stock 218
Management Separated from Ownership 219
Determinants of Total Profit and the Profit Rate 220

Example: The Good Cod Fishing Company 223
Profit per Unit of Output 224
Conflict Over the Profit Rate 225
Intermediate Inputs per Unit of Output as Profit Rate Determinants 226
Unit Labor Cost as a Profit Rate Determinant 227
Work Effort and the Efficiency of Labor 228
Understanding the Profit Rate 229
Conclusion 232

11 Competition and Concentration 233

Competition for Profits 235
The Forms of Competition 237
Price Competition 238
Price as a Markup over Cost 241
Economies of Scale and Price Competition 243
Other Advantages of Large Size 247
Capacity Utilization 248
Breakthroughs 250
Monopoly Power 252
Investing to Compete 254
The Dynamics of Competition 258
Toward Equal Profit Rates? 259
Toward Economic Concentration? 261
Large Size and Monopoly Power 262
Collusion 262
Exploiting Breakthroughs 262
Government Subsidies and Contracts 263
Trends in Concentration 264
Antitrust Enforcement 267
Conclusion 268
Appendix to Chapter 11 270

12 Wages and Work 271

Work, Sloth, and Social Organization 273
The Capitalist Firm as a Command Economy 275

The Conflict between Workers and Employers 276
Labor Discipline: Carrots and Sticks 282
The Labor Market, the Wage, and the Intensity of Labor 285
Cost of Job Loss 286
Avoiding the Cost of Job Loss 288
Additional Implications of the Labor Extraction Curve 295

13 Technology, Control, and Conflict in the Workplace 299

The Social Organization of the Workplace 301
Simple Control 304
Technical Control 304
Bureaucratic Control 305
Technology and the Labor Process 307
Conflict in the Workplace 309
Technical Change and Workplace Conflict 309
Raising the Efficiency of Labor 309
Speedup 310
Deskilling 312
Unions 314
Recent Trends in Union Membership 314
Union Activities 315
Managing the Threat of Replacement 315
Promoting Social Changes to Benefit Workers 317
Discrimination in the Workplace 317
Profitability Versus Efficiency 319
Markets and Hierarchies 321
Democratic Firms 322

PART 3 MACROECONOMICS

14 The Mosaic of Inequality 325

Well-Being and Inequality 328
Influences on Well-Being and the Economy 328
Measuring Living Standards and Inequality 329

Growing Inequality 331
Wealth Inequality 334
Unequal Chances 338
Race and Inequality 341
Discrimination in Hiring 342
Problem of Differential Pay 342
Women's Work, Women's Wages 346
Conclusion 350

15 Progress and Poverty on a World Scale 352

Poverty and Progress 355
Productivity and Income 359
Vicious Circles of Low Productivity 359
Economies of Scale 361
Technology Gap 363
Intellectual Property Rights 364
Other Government Interventions to Promote Development 366
Requisites for Economic Development 366
Eight Essential Elements 366
Stability and Peace 367
Financing 367
A Well-Functioning Government 368
Skills and Education 368
Infrastructure 369
Public Health 369
Access to Foreign Markets 369
Other Factors in Development 369
Variations in Rate of Growth of Productivity 371
Foreign Investment and Development 373
Investment Decisions by Transnational Corporations 375
Transnational Investment and Tax Havens 377
Conclusion 381

16 Aggregate Demand, Employment, and Unemployment 383

Counting the Unemployed 386
What Determines Employment and Output? 388
Aggregate Supply and Aggregate Demand 389
Measuring Total Output 391
Terms for Measuring the Macroeconomy 391
Analyzing Aggregate Supply and Aggregate Demand 392
A Basic Macroeconomic Model 395
Unemployment and Government Fiscal Policy 400
Effects of Deficit Spending on Employment 401
Multiplier Effect of Deficit Spending 404
The Business Cycle and the Built-in Stabilizers 407
Conscious Policy Changes 409
Automatic or Built-in Stabilizers 409
Investment, Aggregate Demand, and Monetary Policy 410
Wages, Aggregate Demand, and Unemployment 417
When is an Employment Situation Wage-Led? 419
Implications for Economic Policy 420
Conclusion 421

17 The Dilemmas of Macroeconomic Policy 422

The High-Employment Profit Squeeze 426
The High-Employment Wage Push 426
The Materials Cost Push 430
Rising Costs Squeeze Profits 431
Exports, Imports, and Aggregate Demand 435
The Demand for Exports and Imports 436
The Foreign Exchange Rate 437
International Trade and Macroeconomic Policy 439
Promoting Net Exports 439
Competing in Global Markets 441
Monetary and Fiscal Policy at Odds 442

What Determines the Interest Rate? 443
Borrowing and the Exchange Rate 446
The Conflict between Monetary and Fiscal Policy 447
Institutions for Achieving Full Employment 447
Institutional Obstacles to Full Employment 448
The Handshake: Ways to Reach Full Employment 449
Conclusion 452

18 Financial and Economic Crisis 454

The Great Recession and the Subprime Crisis 456
Understanding the Great Recession 458
Why to Buy a Home 458
Home Prices Start to Rise 459
How to Buy a Home 462
Securitization of Mortgages 462
Credit Rating Agencies 462
Prime and Subprime Mortgages 463
Refinancing 463
Deregulation 465
Liars' Loans 467
Collateral Damage 470
Derivatives of Other Kinds 471
Bailouts and Buyouts 473
Unemployment and Poverty 473
Lessons from the History of Economic Crises 474
Nonfinancial Causes of Crisis: Overinvestment, Underconsumption 476
Underconsumption as a Cause of Stagnation or Crisis 476
Asset Markets Differ from Goods Markets 477
Why Are Asset Markets Prone to Price Bubbles? 478
Feedback Loops 479
Reflexivity 480
Misinformation 481
Deregulation and Financial Fragility 482
Large Firms Exacerbate the Problem 483
Weakening the Social Safety Net 485

Regulation in a Capitalist Economy 486
Regulating the Financial Sector 486
The Dodd-Frank Act 488
Inequality, Concentration and Crisis 488

19 Government and the Economy in Transnational Capitalism 491

More Spending, Less Health 494
Government in the United States: Too Big? 496
Government Spending: Comparing the United States with Other Countries 496
Old-Age Pensions 498
Influencing the Rules 500
Taxation 502
Tax Rates and Tax Payments to the Federal Government 503
Tax Rules and Enforcement 504
Weakening Enforcement 504
Rules that Affect Business Costs and Profits 506
Contracting Out Government Services 506
Government Contracting: Profits and the Public Interest 507
Decline in Competitive Bidding for Government Contracts 508
Government Contracting for Private Prison Services 508
Laws and Norms on Sentencing 509
Rethinking Private Prisons and Mass Incarceration 511
Selling to Government, Maximizing Profit 512
Medical Services in Private Prisons 514
Changing the Rules 515
Rules Written into Contracts 515
Taxing Transnational Corporations: Who Is In Control? 516
Democracy and Inequality 517
Feedback Loops Between Inequality and Power 519
Supranational Rule Changes 519
TNC Power over Foreign Government Decisions 520
Conflict over Rules 521
Proposals for Alternative Rules 522

Capitalism and Inequality 524
Democracy and Inequality 525
Conclusion 526

LIST OF VARIABLES 529
SOURCES OF ECONOMIC INFORMATION 532
GLOSSARY 535
INDEX 544

Preface

Understanding Capitalism: Competition, Command, and Change is an introduction to economics that explains how capitalism works, why it sometimes does not work as well as we would like, and how over time it not only changes its own functioning but also revolutionizes the world around us. The book does not assume that the reader has any prior knowledge of economics.

The *three-dimensional approach* to economics offered in this book focuses not just on market *competition*, as highlighted in conventional economics textbooks, but also on relationships of *command*—the exercise of power in firms, among nations, and between social groups—and on processes of historical *change*. The approach is multidisciplinary, making extensive use of examples from history, anthropology, and the other behavioral sciences as well as economics.

The core idea uniting the three dimensions of competition, command, and change is the pursuit of profits by firms. Using this central idea, we analyze competition among firms, the search for profits as the driving force of investment and technical change, and profit-seeking as a source of conflict among owners, workers, governments, employers, and consumers.

The book covers the standard topics of supply and demand, market competition, imperfect competition, aggregate demand, and unemployment. In addition, we give special attention to the extraordinary dynamism and material productivity of the capitalist economy, the psychological foundations of human behavior and the importance of Adam Smith's invisible hand, technical change and the new information-based economy, global economic integration and its impact on national economies, and inequality both within and among nations. We also provide a critical evaluation of the tenets of neoclassical economics and a clear introduction to contract theory as well as to new research in behavioral, institutional, and information economics.

We titled the book *Understanding Capitalism* (rather than, say, *Understanding Economics*) to stress that our subject matter is a real economic system—not just the ideas and models of economists. Considered in this way, economic theory is a body of knowledge that, along with history, politics, sociology, and the other social sciences, can help us understand capitalism.

Structure of the Book

Part 1, "Political Economy," introduces the three-dimensional approach to economics, explains its relationship to other approaches, develops its fundamental

concepts, and summarizes some of the pertinent facts relating to life in a capitalist economy.

Part 2, "Microeconomics," develops the theory of the firm and of markets, including the labor market, and analyzes technological change using a Schumpeterian model of dynamic monopolistic competition. Most important, it develops the analysis of profits and the profit rate that provides the integrating principle of three-dimensional economics and establishes the link between micro- and macroeconomics.

Part 3, "Macroeconomics," deals with the workings of the economy as a whole. To provide a foundation for our study of macroeconomics, we examine inequality both within and among nations, focusing in particular on uneven development on a global scale. Here we introduce the concepts of aggregate demand and aggregate supply and explain why unemployment is a more or less permanent feature of capitalist labor markets. This part of the book ends with a chapter on economic crisis and one on the government's role in the economy.

Instructors using the book in a microeconomics course may want to assign Chapters 1–13 and Chapter 19. Those using it in a macroeconomics course might want to focus on Chapters 1–7, 10, 12, and 14–19. For a one-semester course, Chapters 1–10 and 19 might be appropriate, while other chapters could be included as time and student interest permit.

Changes in the Fourth Edition

In this fourth edition of *Understanding Capitalism*, the data and figures have been thoroughly updated (text, figures, and boxes), and there are three new chapters (6, 18, and 19) and four substantially rewritten chapters. Chapters that have been replaced are available on a password-protected website.

- The international aspect of the book has been further strengthened with an updated Chapter 7, substantial improvements to Chapter 15, and some international content in a rewritten Chapter 19.
- In Chapter 10, the algebra of how the profit rate is determined has been simplified, and the simpler algebra has been seamlessly integrated into Chapters 11 and 12.
- Added discussion of the rules that govern the national and world economy, and how they may affect growth and the public interest, now appears in new or rewritten material in Chapters 6 and 19. The new Chapter 6 explains how establishing rules such as limited liability and corporate personhood increased investment, growth and consolidation of enterprises in the U.S., while laying the foundation for the development of transnational corporations. Chapter 19 explores current public debates about what economic rules best serve the public interest.
- A new chapter 18 describes the unfolding of the Great Recession, analyzes changes in the financial sector, and discusses the causes of economic crises in general.
- Chapter 19 newly considers issues in private vs. public provision of services, especially in health care and in prison management.

Pedagogical Aids and Supplements

The glossary, the definitions of terms placed in the text margins, and the captions under each figure help readers to master the basic language and analytical tools of economics. The boxes in the text present additional facts about the economy and raise issues that can be the basis of classroom discussions.

The "Sources of Economic Information" section near the end of the book can help readers locate economic information from official and other sources, both in print and online.

Three-Dimensional Economics and the Neoclassical Paradigm

When the first edition of this book was published in 1985, many thought of it as an "alternative text" and welcomed it as a counterpoint to the neoclassical paradigm that was the dominant approach to economics at the time. Since that time, economics has changed in significant ways. Many of the themes central to this book are now addressed by many economists as well as by other social scientists. In recent years economists have turned their attention to the problem of inequality, the importance of ethical values and unselfish motives in economic behavior, the exercise of power, the way that history shapes economic events, and how the economy shapes who we are as individuals and as people in societies and cultures. The rapid pace of economic, scientific, political, and other developments in today's world has forced economists to face the issue of change.

Since the first edition of *Understanding Capitalism* was published, the Nobel Prize in economics has been awarded to many of the economists and other social scientists who have inspired our own work. Among them are Amartya Sen, Ronald Coase, George Akerlof, Joseph Stiglitz, Robert Fogel, Douglass North, Daniel Kahneman, Vernon Smith, John Nash, and Elinor Ostrom.

Of course, economics remains a controversial topic. There is, however, no longer a single dominant school but rather many distinct approaches, each with its own merits and shortcomings. All of the Nobel Prize winners listed above have been sharp critics of some aspects of the neoclassical approach (while endorsing others). Unfortunately, the teaching of economics to undergraduates has lagged behind what is widely understood by leading economists. The conventional "neoclassical" model is still taught, often as if it were the only approach to the field. For this reason this book may still be thought of as an "alternative text" because it focuses on questions largely ignored in the standard textbooks and develops concepts and ideas that are at variance with—or not even mentioned in—conventional textbooks.

Samuel Bowles, *Santa Fe, New Mexico*

Richard Edwards, *Lincoln, Nebraska*

Frank Roosevelt, *New York, New York*

Mehrene Larudee, *Amherst, Massachusetts*

List of Figures

Figure 1.1 Productivity increases in U.S. agriculture, 1830–1987 5

Figure 1.2 Real wages in London over seven centuries 8

Figure 1.3 Two centuries of world GDP per capita 9

Figure 1.4 Improvements in lighting technology, 1700–1992 10

Figure 1.5 Capitalism and the population explosion 12

Figure 1.6 Cities of the world with more than a million inhabitants in 2014 14

Figure 1.7 Global warming, CO_2 emissions, and CO_2 in the air 18

Figure 1.8 Cumulative CO_2 emissions per capita, 1990–2012, by region, and trends in total emissions by region, 1950–2012 21

Figure 4.1 The production-reproduction cloverleaf 72

Figure 4.2 Total, net, necessary, and surplus product 75

Figure 4.3 Division of the representative farm family's total output 82

Figure 4.4 Conflict over the consumption level of the producers 86

Figure 5.1 Percent of All U.S. Labor Performed that is Unpaid, Paid Commodity-Producing, and Paid Non-Commodity-Producing, 2013 99

Figure 5.2 The concentration of ownership of capital goods 106

Figure 6.1 The American voter: An endangered species? 126

Figure 6.2 Voter turnouts around the world 127

Figure 6.3 Who votes? U.S. voting patterns in the presidential election of 2012 128

Figure 7.1 U.S. wages and union membership, 1930–2015 150

Figure 7.2 Who belongs to unions in the U.S.? 152

Figure 7.3 The growing importance of international profits, 1948–2015 156

Figure 7.4 Percent of U.S population born in other countries, 1850–2010 157

Figure 7.5 Region of origin of U.S population born in other countries 158

Figure 7.6 The changing composition of employment in the U.S., 1939–2015 159

Figure 7.7 The financial sector's share of all reported U.S. domestic corporate profits, 1948–2015 164

Figure 8.1 The demand curve and the supply curve 172

Figure 8.2 A shift of the demand curve 178

Figure 8.3 A shift of the supply curve 179

Figure 9.1 Market messages and market motivation 189

Figure 10.1 Exchange and command in the profit-making process 210

Figure 10.2 The U.S. corporate profit rate, 1948–2014 216

Figure 10.3 Division of a firm's total output into material input costs $P_m M$, wage cost wH, and total profit R 221

Figure 11.1 Finding the selling price at which profit is highest 240

Figure 11.2 Average cost curve for an auto plant with marginal cost $10,000 per car and fixed cost $60 million per year 245

Figure 11.3 Dynamic cost advantage 259

Figure 11.4 Exit and entry of firms from and into markets tend to equalize profits 260

Figure 11.5 Diverse trends in industry concentration, 1997–2012 265

Figure 12.1 The wage and the intensity of labor 291

Figure 12.2 Output per dollar of wage paid 292

Figure 12.3 The labor extraction curve and profit maximization 293

Figure 12.4 The wage, labor extraction, unit labor cost, and profit maximization 295

Figure 12.5 Effects of more unemployment insurance 296

Figure 14.1 The determinants of well-being 329

Figure 14.2 The lucky few: Income share of the top 1 percent of U.S. taxpayers, 1913–2015 331

Figure 14.3 U.S. CEO pay in relation to the average worker's pay, 1965–2015 332

Figure 14.4 Income shares by quintile in the U.S., 1967 and 2014 333

Figure 14.5 Growth rate of family income, by quintile, 1966–2014 334

Figure 14.6 Composition of wealth holdings at different levels of wealth distribution in the U.S. in 2010 335

Figure 14.7 Distribution of median non-home net worth among U.S. households by race/ethnicity, gender, education, and income, 2011 337

Figure 14.8 Increasing concentration of wealth in the U.S. between 1969 and 2010 338

Figure 14.9 Unequal chances: Family background and economic success in the U.S. 339

Figure 14.10 Racism by any other name: Labor market discrimination in the U.S., 2001–2002 343

Figure 14.11 Good resumes pay off, if your name is OK: Labor market discrimination in the U.S., 2001–2002 344

Figure 14.12 Slowed progress in the U.S. toward a color-blind and gender-neutral economy, 1971–2011 345

Figure 14.13 Women's work, women's wages: Gendered occupations and unequal earnings in the U.S., 2015 347

Figure 14.14 Men's and women's median weekly earnings in selected occupations in the U.S., 2015 348

Figure 15.1 Income, life expectancy, child mortality, and schooling in seven countries 356

Figure 15.2 The income distribution of the world, 2011 357

Figure 15.3 Comparative inequality: Incomes by quintile in India, Brazil, and the U.S., 2011 358

Figure 15.4 Progress on a global scale 360

Figure 15.5 Global vehicle production, 1961–2011 362

Figure 15.6 Average annual growth of real GDP per worker, 1960–2011 370

Figure 15.7 Uneven development, 1960–2010 374

Figure 15.8 Hourly wages and benefits in the manufacturing sector in selected countries, 1996 and 2012 376

Figure 15.9 U.S. Direct Investment Abroad, 2014 379

Figure 16.1 Long swings and business cycles 384

Figure 16.2 The U.S. unemployment rate, 1890 to 2014 385

Figure 16.3 Aggregate supply 393

Figure 16.4 Aggregate demand with investment constant 397

Figure 16.5 Macroeconomic equilibrium 398

Figure 16.6 Macroeconomic equilibrium with deficit spending 402

Figure 16.7 The employment multiplier 405

Figure 16.8 How investment responds to changes in employment, the interest rate, and the business climate 414

Figure 16.9 Effect of an increase in investment due either to a decline in the interest rate or an improvement in the business climate 415

Figure 16.10 A wage-led employment situation with investment constant 418

Figure 17.1 Average unemployment rates for six countries, 1971–2012 424

Figure 17.2 Employment and real wages during the 1990s boom in the U.S. economy 429

Figure 17.3 Unit cost, output, and the profit rate 432

Figure 17.4 The employment rate and the profit rate during the 1990s boom in the U.S. 434

Figure 17.5 The determination of the interest rate 444

Figure 18.1 Real quarterly GDP, growth rate 457

Figure 18.2 Real Home Price Index, national, Case-Shiller, 1890–2014 459

Figure 18.3 Interest rates, 1971–2016 460

Figure 18.4 Annual sales of new single-family homes 460

Figure 18.5 Total employment in the construction sector 461

Figure 18.6 Prime and subprime mortgages sold to other financial institutions by twelve months after origination 468

Figure 18.7 Increasing default rates on both prime and subprime mortgages, 2004–2006 469

Figure 18.8 Mortgage delinquency rates on single-family residences, by bank size 470

Figure 18.9 Two official U.S. unemployment rates 474

Figure 18.10 The poverty rate in the United States. Persons in poverty as a percent of all persons 475

Figure 18.11 Financial sector campaign donations to federal candidates and soft/outside money organizations, by two-year election cycle 484

Figure 18.12 Average Wall Street bonus, by year 485

Figure 19.1 Health spending as percent of GDP, six countries 494

Figure 19.2 Expenditure on pharmaceuticals per capita, in U.S. dollars at PPP 495

Figure 19.3 Government outlays as a share of GDP, 2015 497

Figure 19.4 Federal government civilian executive branch employment as percent of total U.S. employment, 1970–2010 498

Figure 19.5 U.S. National defense spending as a percent of federal consumption and investment spending, 1929–2014 499

Figure 19.6 Percent of earnings replaced by pension, OECD countries, 2013 499

Figure 19.7 Annual spending on federal lobbying, defense industry (top) and total (bottom), 1998–2014 502

Figure 19.8 Federal corporate income tax revenue as a share of all federal income tax revenue (individual and corporate), 1934–2015 505

Figure 19.9 Increase in average federal prison time served, 1988–2012 510

Figure 19.10 Transnational corporations' response to the tax holiday in 2005 517

Figure 19.11 Child poverty rate, comparison of selected OECD countries, 2009 525

List of Tables

Table 3.1	Contrasting Perspectives on the Economy 65
Table 5.1	Varieties of Economic Systems 95
Table 5.2	Sources of Income for U.S. Taxpayers, 2011 107
Table 7.1	Percent of the Population Engaged in Different Economic Systems in the U.S. in 1780 137
Table 7.2	The U.S. Class Structure in 1780 139
Table 7.3	The U.S. Class Structure in 1990 140
Table 7.4	The Stages of U.S. Capitalism: Four Social Structures of Accumulation 145
Table 8.1	Determinants of the Positions of Supply and Demand Curves 180
Table 9.1	Prices Are a Decentralized System of Information and Motivation 190
Table 9.2	The Invisible Hand: Assumptions and Conclusions 190
Table 9.3	Positive and Negative Externalities 195
Table 9.4	Prisoner's Dilemma 201
Table 9.5	North's Action/South's Action 202
Table 10.1	Total Profit R and the Profit Rate r for the Good Cod Fishing Company 224
Table 10.2	Profit and its Determinants per Unit of Output, for the Good Cod Fishing Company 225
Table 10.3	Conflicts Over the Profit Rate 230
Table 12.1	The Wage-Output Relationship 290
Table 12.2	The Wage Rate, Output per Dollar of Wage Paid, and Unit Labor Cost 294
Table 13.1	Employers' Systems of Control 307
Table 16.1	The Official U-3 and U-6 Measures of Unemployment, May 2016 388

Acknowledgments

Understanding Capitalism presents an approach to teaching about economics and the economy that we have developed, along with colleagues too numerous to mention, in our introductory economics courses during more than four decades at Harvard University, the University of Massachusetts at Amherst, Sarah Lawrence College, DePaul University, Earlham College, and Hampshire College. To the thousands of students and dozens of teaching assistants whose criticisms, suggestions, enthusiasm, and (sometimes) indifference have guided us, we owe special thanks.

Many colleagues have used previous editions of this book in their classes and given us the kind of feedback and advice only outstanding teachers and researchers can offer. In this regard, we especially thank Arjun Jayadev at the University of Massachusetts, Boston (who also authored the Instructor's Manual of the third edition and contributed greatly to the figures in that edition); Frank Thompson at the University of Michigan; Martin Hart-Landsberg at Lewis and Clark College; Manuel Pastor at the University of Southern California; Bob Pollin and David Kotz at the University of Massachusetts at Amherst; Jim Devine at Loyola Marymount University; Eric Nilsson at California State University at San Bernardino; Fritz Efaw at the University of Tennessee at Chattanooga; Eric Schutz at Rollins College; Jeffrey Carpenter at Middlebury College; Geert Dhondt, Mathieu Dufour, Ian Seda-Irizarry, and Joseph Rebello at John Jay College, CUNY; Teresa Ghilarducci at the New School for Social Research; and Peter Dorman at Evergreen State College. Their detailed comments and suggestions have made this a much better book.

We also thank Ed Ford at the University of South Florida and Rachel Balkom at Santa Fe Prep, who used one or another version of the book in their classes and, citing their students' responses, passed along helpful comments.

We are especially grateful to the following individuals who offered their expertise on particular sections of this or a previous edition: Dominique Alhéritière (Food and Agriculture Organization of the United Nations); Bob Allen (New York University, Abu Dhabi); David Belkin (Independent Budget Office of the City of New York); Lourdes Benería (Cornell University); Josh Bivens, Will Kimball, and Lawrence Mishel (Economic Policy Institute); Mike Buckner (United Mine Workers of America); Stephen V. Burks (University of Minnesota, Morris); James Boyce, Doug Cliggott, James Crotty, Nancy Folbre, Carol Heim, and Gerald Epstein (University of Massachusetts at Amherst); Jonathan Cogliano (Dickinson College); Jane d'Arista (Research Associate, Political Economy Research Institute); Andrew Farrant (Dickinson College); Edward Ford (University of South Florida); Kevin Furey (Chemeketa Community College); Janet Knoedler (Bucknell University); June Lapidus (Roosevelt University); Margaret Levenstein (Inter-University Consortium for Political and Social Research, University of Michigan); Victor Lippit (University of California at Riverside), Jamee Moudud (Sarah Lawrence College); the late Gordon

Pavy (AFL-CIO); James Polito (Byrum Business School); Jared Ragusett (Central Connecticut State University); Michael Reich (University of California at Berkeley); Peter Rosset (El Colegio de la Frontera Sur, México); Juliet Schor (Boston College); John Schmitt (Center for Equitable Growth, Washington, DC); Eric Schultz (Rollins College); Ibrahim Tahri (The New School University); Joe Von Fischer (Princeton University); Jim Weeks (ATL International); Edward Wolff (New York University); Martin Wolfson (University of Notre Dame); and the immensely helpful and patient staff of the U.S. Bureau of Economic Analysis.

Last but not least, we thank the following professionals at Oxford University Press who made possible the publication of this book: Ann West, our highly competent acquisitions editor; Alison Ball, editorial assistant; and Michele Laseau, Holly Haydash, and Patty Donovan, who shepherded the book through the process of its production.

Of course, the authors retain responsibility for any remaining errors.

Samuel Bowles, *Santa Fe, New Mexico*

Richard Edwards, *Lincoln, Nebraska*

Frank Roosevelt, *New York, New York*

Mehrene Larudee, *Amherst, Massachusetts*

Capitalism Shakes the World



For more than four decades following World War II, Germany was divided: East Germany was a dictatorship, while West Germany was a democracy. The economic systems of the two Germanys were as different as their systems of government. In the East, the economy, like just about everything else, was run by the Communist Party. Decisions about who should produce what, how, when, and for whom were made by the government and carried out under orders. Communism was not simply a form of government, it was also an economic system based on centralized direction of economic decisions. By contrast, West Germany had what is termed a capitalist economy. West Germans for the most part made economic decisions independently, guided in most cases by what they needed to do to turn a profit, to get and keep a decent job, or to have a particular kind of lifestyle given their means.

In October 1989 the general secretary of the East German Communist Party, Erich Honecker, grandly celebrated the founding of Communist East Germany forty years earlier. He proclaimed that it had been both a "historical necessity" and a "turning point in the history of the German people." Parades and demonstrations commemorated the anniversary. But twelve days after the celebration, Honecker suddenly stepped down as prodemocracy demonstrations broke out first in the East German city of Leipzig and then spread throughout the country. A million and a half Germans participated in these demonstrations in October, and twice that number attended them in November.

Less than a month after Honecker's resignation, East and West Germans danced together on the Berlin Wall and then dismantled it. Less than a year after the grandiose celebration of its fortieth anniversary, East Germany passed out of existence, its territory joined with that of West Germany, and the combined parts became once again simply Germany. As a result, the citizens of the former Communist nation passed from one economic system to another, from communism to capitalism. At about the same time, prodemocracy demonstrators toppled their Communist rulers in the Soviet Union, Poland, Czechoslovakia, Hungary, and, indeed, in all of the remaining Communist-ruled countries except Cuba, Vietnam,

Laos, North Korea, and China. The demonstrators rejected not only Communist dictatorships but also the centralized organization of their economies. All adopted some form of capitalist economic system.

Like communism, other economic systems had earlier fallen to the capitalist onslaught. An economic system based on slavery in the U.S. South ended with the victory of the Union troops in the Civil War and Lincoln's emancipation of the slaves. As a result the South ceased to be a slave economy and became capitalist. Similarly, the simple economics of hunting and gathering—what most humans did to make a living for most of our time on earth—has been abandoned in most parts of the world, to be replaced by other economic systems, and eventually, in most parts of the world, by capitalism. And the process continues. Capitalism is on a roll and has been since its birth.

Nevertheless, capitalism is new, having been a moving force in world history for only the past five centuries or so—less than 1 percent of the time that humans have inhabited the earth. During this relatively short period, however, the world has changed more quickly, more constantly, and more profoundly than during any earlier period of human history. And now the pace of change appears to be quickening, so even greater transformations will most likely occur in our lifetimes.

Capitalism is an economic system in which employers, using privately owned capital goods, hire wage labor to produce commodities for the purpose of making a profit.

In this chapter we look at the evolution of capitalism in some detail. We define **capitalism** as an economic system in which employers hire workers to produce goods and services that will be marketed with the intention of making a profit. The main idea of this chapter is that wherever capitalism has taken root, it has left no aspect of society unchanged. This main idea is expressed in four key points:

1. Capitalism has brought with it unprecedented advances in scientific and other kinds of knowledge, astonishing developments in technology, previously unimaginable ways of sharing information, and rising standards of consumption, health, and education in most of the world.

2. Capitalism has also led to fundamental realignments of power and redistributions of wealth, the abolition of slavery and other archaic forms of bondage, and radical changes in family life, ideals, and beliefs.

3. Since we have lived with rapid change all of our lives, we tend to think of it as normal, even natural. Yet from a historical perspective, rapid and relentless transformation of the social and physical world is anything but normal. Far from being driven by change, earlier economic systems were bound by inertia. The **capitalist epoch** began in some parts of Europe around AD 1500. The capitalist organization of work—employers hiring people for wages to make a profit—first appeared in parts of England, the Netherlands, Belgium, and Italy. Initially the new way of organizing production affected few people, even in the countries where it first appeared, but as it spread and became stronger, the transformative power of capitalism also grew. It would eventually revolutionize the world.

The **capitalist epoch** began in some parts of Europe around AD 1500, when capitalist organization of labor processes first appeared. It continues to the present in most of the world.

4. Capitalism's development and the social changes accompanying it occurred at different times in different places, and its impact was highly uneven. In some places capitalist development occurred quickly, in other places very slowly, and in some regions of the world, capitalism is only now replacing other economic systems.



Europe at the Dawn of the Second Millennium

A mere handful of folk—unending emptiness stretching so far west, north, and east that it covers everything—fallow land, fens, and wandering rivers, heaths, woods and pastureland, every conceivable type of erstwhile forest leaving behind it brush fires and the woodburners' furtive sowing—clearings here and there, wrested from the forest but still only half-tamed; shallow pitiful furrows that wooden implements drawn by scrawny oxen have scratched in the unyielding soil . . . huts of stone, mud or branches, clustered in hamlets surrounded by thorn hedges and a belt of gardens; sparsely scattered towns, streets in ruins, fortifications haphazardly repaired, stone structures dating back to the Roman Empire that have been turned into churches or strongholds.

Such is the Western world in the year 1000. Compared with Byzantium, compared with Cordoba, it seems rustic, very poor and defenseless. A wild world ringed round by hunger, its meager population is in fact too large. The people struggle almost barehanded, slaves to intractable nature and to a soil that is unproductive because it is poorly worked. No peasant who sows one grain of wheat expects to harvest much more than three—if it is not too bad a year that means bread to eat until Easter time.

Source: From Georges Duby, *The Age of the Cathedrals* (Chicago: University of Chicago Press, 1981), 3.

The Permanent Technological Revolution

To see more clearly the changes that have come with capitalism, consider what life was like in Europe before the dawn of the capitalist age. In the year 1000, people there had short life spans; they had almost no experience with people or places farther away than the nearest town; and they depended on the food and other things they could produce by their own efforts, supplementing their consumption with only a few items available in local markets.

During most of human history, people lived in societies that had not changed much since the time of their parents, grandparents, or earlier ancestors. For generation after generation, sons made their livings in much the same ways as had their fathers before them, and daughters also followed in their mothers' footsteps. Tools and utensils, stories and beliefs were passed on from parents to children, just as they had been a century, or even a millennium, before. Good years and bad alternated with the weather, but continuous, rapid, and systematic change would not become an ordinary fact of life until the emergence of capitalism.

Around the world societies were organized in many different ways, but most people were only dimly aware of this diversity because their horizons did not extend beyond the small communities in which they lived. By the beginning of the fifteenth century, however, Europeans began to explore other continents and "discovered" what they called the "New World." Before long, traders and colonists, often financed by investors seeking fabulous riches, were intruding on indigenous peoples in areas located in what is now Virginia, Peru, Barbados, South Africa, and India. The dynamism unleashed by the advent of capitalism in Europe soon began to impinge on the rest of the world.

It is hard to know which came first, capitalism or the great spurt of technical change that came along with it. Whatever the truth may be, the continuous, rapid, and far-reaching scientific discoveries and technological innovations that are now accepted as a permanent feature of modern life emerged more or less simultaneously with capitalism. And, of course, these discoveries and innovations made possible the remarkable economic advances of the last five centuries.

In 1500 goods were made almost entirely by hand, using simple tools. Power machinery consisted of such devices as the water wheel that turned a miller's grinding stone. People's understanding of the physical world was so rudimentary that births, deaths, and harvests, whether abundant or meager, were frequently interpreted with recourse to magic, superstition, or reference to God's will.

As late as 1800 traditional craft-based techniques, using skills that had been handed down from generation to generation, still prevailed in most production processes. But the new era brought new ideas, new discoveries, new methods, and new machines in every field of endeavor, making old ideas and old tools obsolete. And the new ways were in turn quickly made obsolete by even newer ones. As technical change revolutionized production, it reduced the amount of time required to produce most products.

The most important increases in labor productivity were those that occurred in the agricultural sector. As fewer people were required to produce the same or greater amounts of food, more labor could be devoted to the production of other things, particularly in the manufacturing sector. Thus, increases in agricultural productivity had to be achieved before the Industrial Revolution could take place. To illustrate the rapidity with which farm output has increased during the capitalist era, Figure 1.1 shows the growth of productivity in U.S. agriculture during the past two centuries.

There have also been dramatic improvements in methods of transportation during the past five centuries. In 1500 people either walked or used wagons to get themselves and their possessions or freight from one place to another on land. Wagons were pulled either by people or by animals, and the movement of people or freight overland was arduous, costly, slow, and sometimes dangerous. For all but a very few wealthy people, travel beyond a short distance from home was virtually impossible, and shipping freight was so expensive that it did not pay to send anything but very valuable and lightweight goods such as spices and silks.

Water transport on rivers and along coasts was easier, but ships were small, slow, and unsafe. In 1500 there had been few advances in maritime technology beyond what was available to the Romans 1,000 years earlier. Within a century, however, sea transport was greatly improved. Ships began regularly crossing the Atlantic or rounding the Cape of Good Hope en route to the East Indies. By 1800 clipper ships raced from China to London in eighty days and from New York to San Francisco (around the southern tip of South America) in twenty-two days. At the same time, sailing across the Atlantic became almost routine.

The opening of the Suez Canal by 1870 and the Panama Canal by 1914 greatly shortened world trade routes, while coal- and oil-fueled engines made sailing ships obsolete. Before the end of the twentieth century, oil tankers, each carrying 2,500 times the cargo of Columbus's ships, clogged the shipping arteries of the world and became too large to enter all but the largest and deepest harbors. Until the Concorde was grounded for economic reasons in 2003, those who could pay the $9,000 required for a reservation on this aircraft could travel from London to New York in three hours, moving at twice the speed of sound, overtaking the sun, and arriving "earlier" than when they had left.

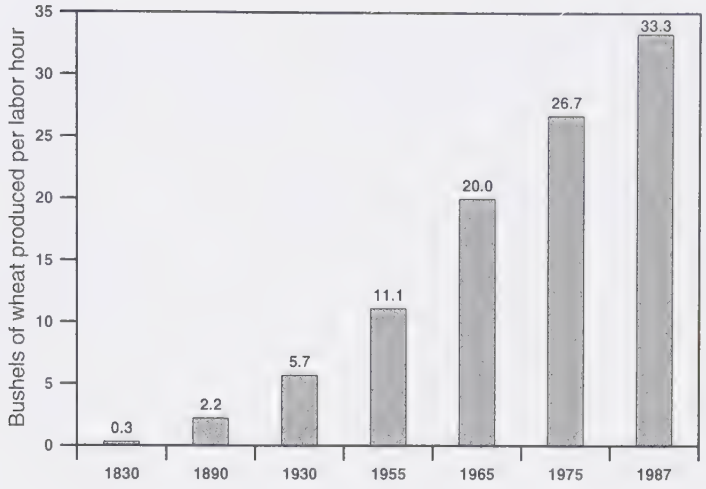


FIGURE 1.1 Productivity increases in U.S. agriculture, 1830–1987.

Focusing specifically on wheat production, this figure shows how labor productivity in U.S. agriculture has risen during the past two centuries. In the eighteenth century American farmers used crude wooden plows, sometimes drawn by horses or oxen. They sowed seeds by hand, cultivated them with hoes, cut the wheat with sickles, and harvested their grain from the wheat using manual threshing devices called flails. Iron plows were introduced early in the nineteenth century, but as late as 1830 it still took about three hundred hours of labor to produce one hundred bushels of wheat on a five-acre farm. In the mid-nineteenth century farmers began to use chemical fertilizers, and soon came to rely more and more on factory-made agricultural tools and later machinery. Over the next century agricultural productivity rose dramatically. By 1987, on a large, highly mechanized American farm, one hundred bushels of wheat could be produced with only three hours of labor on three acres of land, one hundred times more output per hour than could be produced in 1830.

Source: U.S. Department of Agriculture, Economic Research Service, "A History of American Agriculture 1776–1990" (Washington, DC: US Department of Agriculture, Economic Research Service, 1993).

Land transport was revolutionized as well. First, inland canals were dug such as the famous 365-mile Erie Canal in New York State, finished in 1825. Barge traffic through canals greatly reduced the cost of overland haulage. Railroads further increased the speed and cut the cost of moving goods and people. In the United States, the transcontinental railroad was completed in 1867, and by the end of the nineteenth century tracks would crisscross all the world's industrial areas and penetrate the Canadian Rockies, the East African highlands, the Chinese hinterland, the vast Russian steppes, and the plain of northern India as well. Yet all this was but a prelude to the great twentieth-century land transport revolution based on automobiles, trucks, and highways. When air travel and transport, major innovations of the last century, were added to the mix, they partly displaced railroads in global shipping and travel.

The technological advances in transportation were matched by equally significant developments in medicine, agriculture, and communications. Improvements in health care and agricultural productivity made possible the population explosion and urbanization discussed later in this chapter. Moreover, the communications revolution has been central to the process of globalization, also to be discussed later in this chapter.

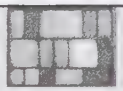
Less beneficial were certain advances in weaponry and the discovery and production of toxic chemicals and biological agents. Today, chemical, biological, and nuclear weapons, sometimes referred to as weapons of mass destruction, are

powerful enough to destroy the entire population of the world. Whether in production techniques, in transportation, in medicine, in agriculture, in communications, or in nuclear, chemical, and biological warfare, technical change has occurred with a speed and pervasiveness that is unprecedented in human history.

The Enrichment of Material Life

The technological changes of the past five centuries have been accompanied by significant increases in people's consumption standards. Before the capitalist epoch, living conditions improved or deteriorated with changes in the weather, epidemics, and other natural phenomena, because most people made their living by farming, herding, or hunting and gathering. But where capitalism took hold, average living standards began to rise in a sustained way, though with variations over space and time. Figure 1.2 shows the sharp upturn in real wages experienced by one group of workers after capitalism emerged as the dominant economic system in Great Britain in the nineteenth century.

While Britain was the first capitalist country, the new economic system soon spread to other countries, producing comparable increases in average living standards wherever it went. In the United States, for example, as best we can measure, the buying power of the average income in 2002 was thirty-two times what it was in 1789 (the year the U.S. Constitution was adopted). This does not mean, of course, that Americans are now thirty-two times happier than they were in 1789, but it does indicate an unprecedented growth in the availability of material goods. With increasing material abundance, diets became more varied (with meat eaten more frequently) and housing larger and more comfortable (at least warmer in the winter).



Why did the English Economy take off? Competition, Coal, or Colonies?

Over the last five hundred years, small regional disparities in living standards grew into enormous gaps, and a large part of the world fell under the domination of Europeans and of Americans of European descent. Why did industrialization take hold so rapidly in Britain and Europe? Was it coal? Colonization? Economic policies?

At first glance England, and Europe generally, had no special advantages that primed these economies to make the leap. Before the "European miracle" Chinese scientific knowledge surpassed that of Europe in many fields and rivaled it in most. And Europe's scientific advances were in fields with few links to the technologies that propelled the industrial revolution there. Nor were the more advanced economic areas of China, like the Yangtze River delta, poorer than the advanced areas of Britain and Europe.

Some economists say it was British economic institutions that made the difference: that the British government kept its hands off the economy, respecting property rights and refraining from setting prices or wages. But recent historical research shows that individual pursuit of economic gain was probably freer from governmental interference in China's Yangtze River delta in the eighteenth century than in England or Europe. Historian Kenneth Pomeranz concludes that neither scientific advances, nor riches, nor a laissez-faire environment explains why capitalism took root in England rather than flowering elsewhere and propelling some other people to world dominance.

What the Yangtze delta lacked was natural resources. Unlike England, it had no rich coal deposits and little water power, and its hinterland was not prepared to supply sufficient raw materials for an expanding economy.

By contrast, British military prowess, honed in centuries of Continental warfare, gave England cheap access to the raw materials of the New World, especially after France lost the Seven Years War (1756–1763). Sugar from Barbados, Jamaica, and the other British Caribbean colonies provided over 10 percent of the calories consumed by British labor. Slave plantations in British colonies fed raw cotton to the booming textile mills of Manchester. Britain's farmland could not have produced enough sugar for British workers or enough fibers for British cloth production. Without the colonies, economic expansion would have pushed up the prices of cotton, sugar, and other raw materials, driving down profits and prematurely grounding the British takeoff.

Pomeranz uses these arguments to dispute other historians' claim that institutions gave Britain the edge: that strong property rights, effective competition, and limited government drove the Industrial Revolution. Instead, he wrote, England and Europe had institutional advantages that seemed "applicable to very few endeavors in the pre-1800 world besides war, armed long-distance trade and colonization."

Source: Kenneth Pomeranz, *The Great Divergence: China, Europe, and the Making of the Modern World Economy* (Princeton: Princeton University Press, 2000). The quote is from p. 166.

Figure 1.3 shows how the rise of productivity in Europe eventually helped to lift output per person in the world as a whole. The main part of the figure shows that the dramatic increase in output per person, averaged for the world as a whole, did not occur until after 1820, while the smaller (inset) chart shows in detail the relative contributions of different regions to the world's total output during the past five centuries. To simplify this chart, the nations of the world have been divided into three groups. The first group, called "Western world," includes western Europe and the largest nations of the former British empire—the United States, Canada, Australia, and New Zealand—where English-speaking settlers rapidly became a majority of the population. The second group includes all of Asia, including China, India, and Japan. The third group consists of all the nations in Africa, Latin America, and the former Soviet Union countries of eastern Europe, including Russia.

It can be seen that the "Western world" was responsible for most of the growth of world output between 1500 and the early 1900s, its share of world output increasing from less than 20 percent in 1500 to more than 55 percent in 1913 and 1950. But, as the chart shows, its relative share has been falling since 1950.

While the West's share of world output was increasing, that of Asia was declining. This was due in part to the fact that the absolute amount of output being produced in the West was growing much faster than the amounts being produced in other regions, so the *percentage* of world output attributable to the other regions had to be falling. Because Asia produced most of the world's output in 1500, it was the region with the most to lose. Most of the decline in Asia's share was due to the spectacular economic success of Europe and North America. But there were other reasons for the precipitous decline in Asia's share of world output between 1820 and 1950. For one thing, British imperialism had a devastating effect on the productivity of India, a major contributor to Asia's output. Whereas India had had a strong and diversified economy in 1800, by the mid-nineteenth century its British rulers had begun to

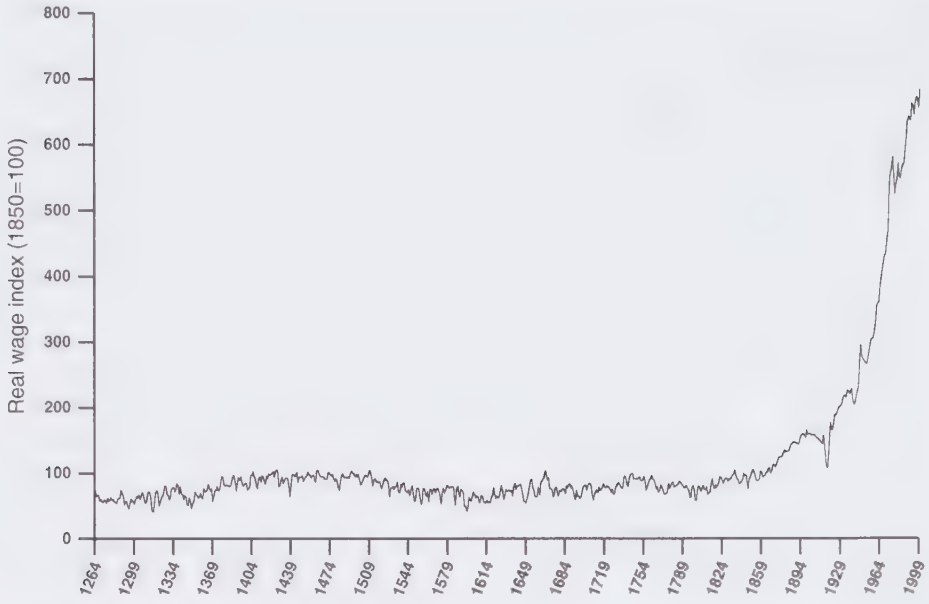


FIGURE 1.2 Real wages in London over seven centuries.

This figure shows the average real daily wage of skilled construction workers in London between 1264 and 1999. The term "real" means that the monetary wage in each year has been corrected for price inflation, so that it measures the buying power of the wage in that year. The data are presented here as an *index*. They show what the average real wage was in each year *relative* to what it was in 1850, the arbitrarily chosen *base year* for which the index is set at 100. For example, if the real wage in the 1930s was twice what it was in 1850, it is shown here as 200; if in the mid-fourteenth century it was half of its 1850 level, it is shown as 50 in that period. While the data are subject to error, they do make clear that real wages did not rise consistently prior to the full development of capitalism. Variations in the real wage before 1800 often resulted from sudden population changes. For example, the Black Death (bubonic plague) killed many in Great Britain and elsewhere in the mid-fourteenth century, creating a labor shortage in Great Britain and driving up wages. Gains made between 1300 and 1500 were eroded after 1500 by rapid price increases caused by the sudden inflow of gold to Europe from the Americas. Inflation often reduces real wages, at least if workers lack organization and bargaining power. But after 1800, and particularly after 1900, increased labor productivity and greater bargaining power of some workers led to dramatic increases in real wages for people like the skilled construction workers represented in this figure.

Source: Robert Allen, "Wages, Prices and Living Standards: The World Historical Perspective," available at https://www.nuffield.ox.ac.uk/People/sites/Allen/SitePages/Biography.

cripple that nation's cotton textile industry in order to gain the entire Indian market for their own cotton textile products. At the same time India was forced to specialize in producing (less valuable) raw cotton for export to English textile factories.

Since 1950, the rising Asian share of world output and the simultaneous decline in the West's share stemmed largely from economic expansion in Japan and China. Measured by total output, the United States has the largest economy in the world, and China (officially the People's Republic of China) is second, with nearly one-fifth of the world's population and with average growth rates of output of 10 percent since 1983. Also important is the contribution of some Asian "tigers"—notably South Korea, Thailand, and Singapore—all of which achieved a 9 percent average growth rate from 1987 through 1996, followed by somewhat slower growth.

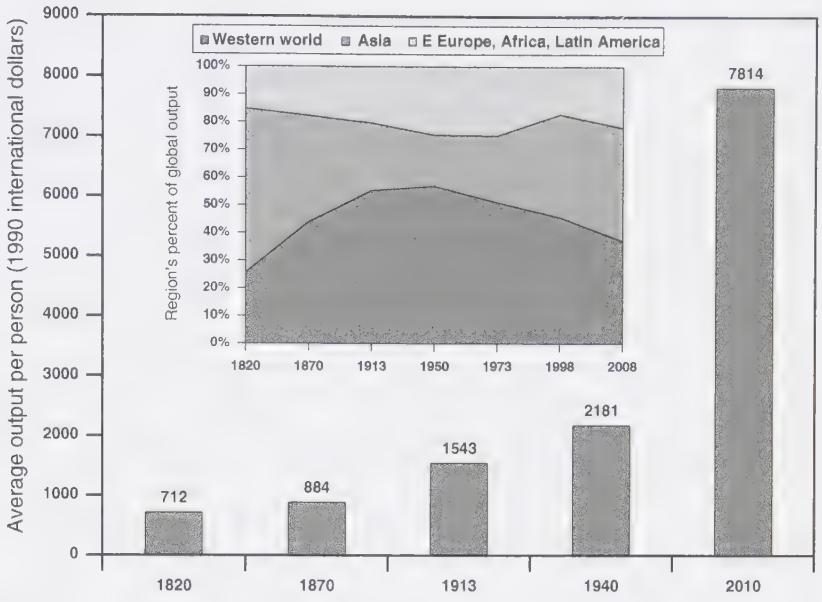


FIGURE 1.3 Two centuries of world GDP per capita.

The larger chart presents inflation-adjusted estimates of output per person (GDP per capita) for the whole world during the last two hundred years. These estimates come from over half a century of research by the late Angus Maddison, an eminent scholar of global population size, technical change, and productivity growth, and his fellow economic historians, who are continuing the project. Maddison estimated that world output per person rose only slightly from the year AD 0 to 1820; his successors agree, estimating it at $600–$700 in AD 0 and $712 in 1820 (in 1990 international G-K dollars). For 1820, among countries for which data exist and have been analyzed, average incomes are estimated from over $300 in Korea to over $2,000 in the UK. For eighteen centuries, global output rose about as fast as global population. But since 1820, technical changes accompanying the rise of capitalism have greatly expanded output per person. The smaller (inset) chart shows how in the past two centuries the share of world output produced by each region has changed. These changes have many causes, including diverse population growth rates, technology, war or peace, and the organization of production and exchange.

Sources: The Maddison-Project, http://www.ggdc.net/maddison/maddison-project/home.htm, 2013 version; Angus Maddison, *Monitoring the World Economy, 1820–1992* (Paris: OECD, 1995), 19, Table 1-1(a); Angus Maddison, *The World Economy: A Millennial Perspective* (Paris: OECD, 2001), 28, Table 1-2, available at http://www.oecd-library.org/development/the-world-economy_9789264104143-en; World Economics: Measuring Global Activity, Maddison Historical GDP Data, Downloadable Data (to 2008); http://www.worldeconomics.com/Data/MadisonHistoricalGDP/Madison%20Historical%20GDP%20Data.efp.

One can also see from the inset chart in Figure 1.3 that the nations of Africa, Latin America, and eastern Europe started out with the smallest share of world output in 1500, and increased their share slightly between 1820 and 1950, but ended up again (in 2010) with the smallest share. The recent decline here is mostly a result of the catastrophic fall in output associated with the difficult transition from central planning to capitalism after the upheavals of 1989 to 1991 in eastern Europe and Russia. It is also partly a result of the very rapid growth in China and other Asian countries, which mean that even the Western world's share of global output shrank.

The data presented in Figures 1.2 and 1.3 showing vast increases in wages and output in much of the world over the past few centuries may actually *understate*

the associated improvements in living standards if the overall *quality* of goods has increased dramatically, an increase that may not be adequately accounted for when we measure prices and output.

Qualitative improvements can be seen most strikingly in the production of light, starting with the campfires of our distant ancestors and then moving on, with the passage of time, to oil lamps, candles, kerosene lamps, and, ultimately, to modern lighting technologies such as filament and fluorescent light bulbs. These changes brought an extraordinary increase in the efficiency of light production. Engineers define lighting efficiency with reference to how much light, measured in units called "lumens," can be produced using a certain amount of energy (measured in watts).

Figure 1.4 charts the advance of lighting technology from 1700 to 1993, showing the lumens per watt of each new lighting source. Not shown in the chart is the fact that the lighting power of a campfire is between .002 and .003 lumens per watt. By 1800 a light source, the tallow candle, had been developed that was more than thirty-two times as efficient as a campfire: it emitted light at nearly .076 lumens per watt. With the coming of electric power in the late nineteenth century, lighting technology began to improve rapidly. By 1900 an advanced carbon filament lamp could light up a room (or a street) at 3.7 lumens per watt. A century later, a 100-watt tungsten

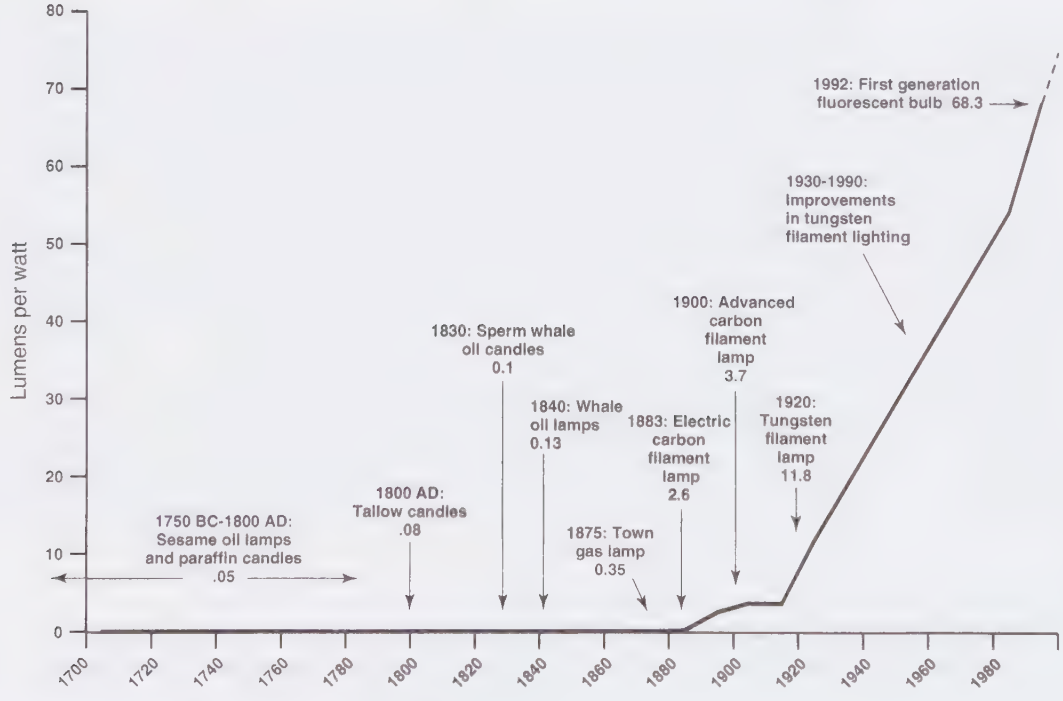


FIGURE 1.4 Improvements in lighting technology, 1700–1992.

This figure shows how the efficiency of lighting technology, measured in lumens per watt on the vertical axis, has increased over the past three centuries. A "lumen" is a unit indicating a certain intensity of light, and a "watt" is a unit measuring the power of a source of energy.

Source: William D. Nordhaus, "Do Real-Output and Real-Wage Measures Capture Reality? The History of Lighting Suggests Not," in Timothy F. Bresnahan and Robert J. Gordon, eds., *The Economics of New Goods*, National Bureau of Economic Research Studies in Income and Wealth, vol. 58 (Chicago: University of Chicago Press, 1996), 29–66.

filament lamp—the standard 100-watt commercial light bulb—could emit light at a rate of more than 14 lumens per watt.

The biggest single advance in lighting technology so far is the "compact" fluorescent light bulb producing 50 to 70 lumens per watt, making it up to 30,000 times as efficient as a campfire. The point is that our typical measures of improvements in standards of living—real wage increases and growth of real gross domestic product (GDP) per capita—do not account for the fact that we often get proportionally more of something, in this case illumination, than increases in our spending on it would indicate. Such measurement issues have grown in importance in the past century and a half as the rate of technical progress—stimulated by capitalism—has accelerated.

Capitalism is not unique in its capacity to promote rapid increases in material well-being. The Soviet Union, with a Communist government and a centrally planned nonmarket economy, achieved very high rates of economic growth between the late 1920s and the mid-1980s. And, as we have just seen, over the past three decades China, also under Communist rule (though with growing market and capitalist institutions), has maintained rates of increase in total output unparalleled by any of the world's capitalist economies, supporting extraordinary improvements in living standards. Indeed, during World War II, when the U.S. economy was in essence centrally planned to ensure that all resources went to supply the war effort, the economy recorded over 15 percent growth in real (inflation-adjusted) output each year during 1941–43, suggesting that under some conditions central planning has the power to foster more rapid expansion in output than does leaving economic growth to market forces alone.

Growing Inequality

The material abundance that has come with capitalism is not evenly distributed. As capitalism has enhanced the generation of wealth in some parts of the world, it has also led to glaring global inequalities. Before the rise of capitalism, most of the world's population lived quite simply, without the material goods that the majority of people in rich countries now take for granted. But as capitalism developed, the gap between rich and poor became a chasm.

One way to measure inequality is by using data on how much wealth people own. A person's *wealth* or *net worth* is defined, at a point in time, as the value of what the person owns (money, house, furniture, vehicles, financial assets, and so on) minus the amount of that person's debt. That is, *wealth = assets – debts*. The distribution of wealth is typically more unequal than the distribution of *income*; income is defined as the flow of money into a person's possession over a given period of time. Someone who says, "I have $200,000 in assets after subtracting my debts," is stating her *wealth* or *net worth*. But if she says, "I received $4,000 last month from various sources," she is stating her flow of *income* for that month.

Several sets of researchers have tried to measure global wealth inequality, using somewhat different methods and data sets. According to one major study, in 2014 the wealthiest 1 percent of the adults in the world held 48 percent of the world's wealth.[1] The *Global Wealth Databook* published by the bank Credit Suisse estimated that in

[1]James B. Davies, Rodrigo Lluberas, and Anthony F. Shorrocks, "Estimating the Level and Distribution of Global Wealth, 2000–14," United Nations University, World Institute for Development Economics Research (UNU-WIDER) Working Paper 2016/3, February 2016.

2015 the richest sixty-two individual multibillionaires alone owned as much wealth ($1.76 trillion) as the poorest half of the world's population, numbering 3.6 billion people. Research on the United States uses better data, and shows clearly that wealth inequality nationally has been on the rise. Such extreme inequality has unacceptable social consequences. Today many people in the world still do not have clean drinking water, and so their children often die of preventable diseases, while the very wealthy own their own jet planes.

The Population Explosion and the Growth of Cities

Along with capitalism's technical progress and rising standards of living, there has been a global population explosion. As Figure 1.5 shows, the population of the world grew very slowly from 10,000 BC to the 1700s. But since then, as the sharp upturn in the curve indicates, the population growth has increased dramatically.

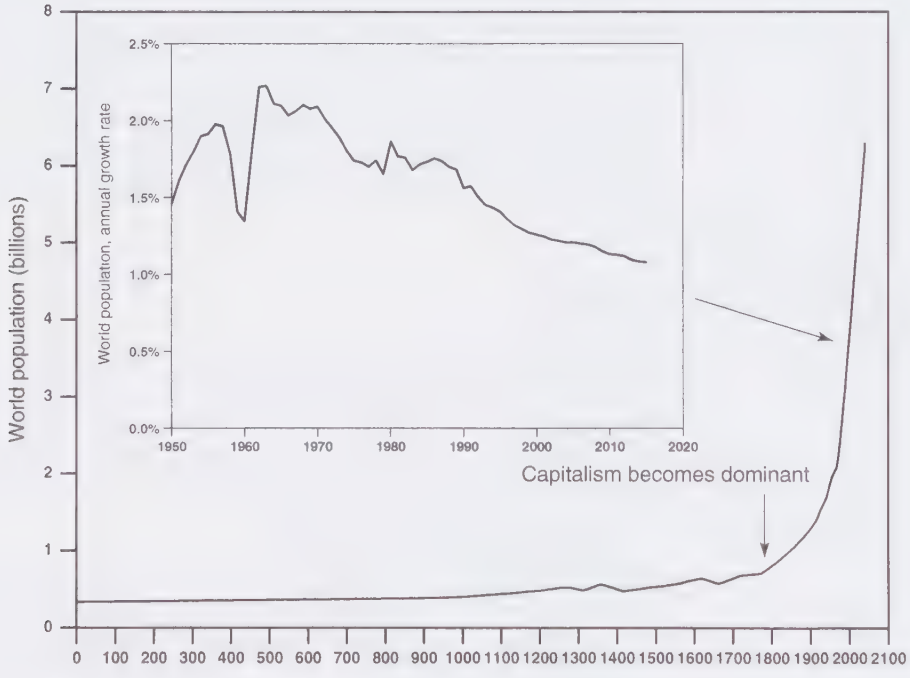


FIGURE 1.5 Capitalism and the population explosion.

This figure charts the population of the world from AD 0 to 2015. For most of the last 12,000 years, including most of the 2,000 years shown in this figure, the total population of the world grew slowly, if at all, with periods of increase in good years followed by intervals of decline in response to bad weather, disease, and other calamities. There are about twenty cities in the world today whose populations exceed the entire population of the world 11,000 years ago (less than 10 million persons), when agriculture began displacing hunting and gathering. Population started growing rapidly in a few countries two centuries ago, but the world's population really exploded in the twentieth century due to the spread of modern medicine, especially public health practices. The world's population is still growing, but at a declining rate (see inset chart).

Sources: U.S. Bureau of the Census, *Historical Estimates of World Population* (updated 2013) available at https://www.census.gov/population/international/data/worldpop/table_history.php; U.S. Bureau of the Census, "Total Midyear Population for the World: 1950–2050" (updated July 9, 2015), available at http://www.census.gov/population/international/data/idb/worldpoptotal.php.

From AD 1 to 1750 the population grew at only 0.56 per thousand annually, so that it took 1,200 years to double in size. Between 1750 and 1950 it grew at a faster rate (about 5.7 per thousand per year), doubling the population every 120 years. In just 38 years from 1950 to 1988, it doubled from 2.4 to 4.8 billion. The population explosion of the past 250 years is a radical departure from the previous trend, a new social phenomenon seen only in the capitalist epoch. However, population growth is slowing. Since 1990, population growth has been less than 0.8 billion per decade, and that number has been drifting downward, so that the next doubling (to 9.6 billion) will likely be completed at the very earliest in 2045 and have taken nearly sixty years.

While world population growth has been dramatic, there have also been important movements of people from one part of the globe to another. In places such as North and South America and Australia, for example, entirely new populations came in, overwhelming and virtually eliminating the indigenous peoples.

Many of the indigenous ("Indian") populations of North and South America were decimated, with their remnants relocated to remote territories. At the same time, millions of Africans—from 10 to 100 million, by various estimates—were forcibly taken from their communities and transported in chains across the Atlantic Ocean to become slaves in the Americas; perhaps as many as half died in the crossing. Numerous Chinese and (Asian) Indians were recruited to work under near-slave conditions and shipped to faraway places—the Chinese to build railroads in North America, the Indians to build them in East and Southern Africa.

Other populations migrated long distances when their traditional livelihoods were destroyed by changes in their homelands. Germans, for example, had to leave their country when they were subjected to political repression. Italians left when declining grain prices made farming no longer a viable way of life. Others emigrated when they found their traditional crafts made obsolete by new capitalist factories. A large number of Irish emigrated in the mid-nineteenth century when the potato crop, their main food source, was wiped out by blight. For similar reasons Poles, Greeks, Jews, Hungarians, and Russians also migrated. Whether pushed by worsening opportunities at home, or pulled by hope for a better life abroad, or both, they experienced fundamental life changes, saw old routines disrupted, and pursued alternative opportunities in new lands.

Along with migration came significant changes in occupations. In 1800 the overwhelming majority of Americans were food producers: independent farmers, food-producing slaves, or fishermen. Today only 2 percent of the U.S. population lives and works on farms, while another 3 percent works in the food processing and food service industries. This 5 percent of the population grows, processes, and serves enough food for the whole country and still produces a surplus that is exported.

As people left farming, another change became apparent, namely, *urbanization* and the growth of cities. Before the emergence of capitalism, most people lived in the countryside, not in urban areas. In the last few centuries, however, people have been drawn, or in some cases pushed, into cities. In 1800 only 6 percent of Americans lived in towns or cities with more than 2,500 people. Today four-fifths of the U.S. population lives in such urban areas.

London's population, only 70,000 in 1500, grew to 600,000 by 1700 and approached three-quarters of a million in 1800, making it then the largest English-speaking city in the world. Today, London has a population of 10 million, the New York metropolitan area 21 million, and Mexico City 20 million. Tokyo, the world's biggest urban area, is home to 38 million people. Seven other urban areas in Asia each have between 20 and 30 million people.



FIGURE 1.6 Cities of the world with more than a million inhabitants in 2014.

Each of the dots in this figure represents a city of more than 1 million people (but not all cities of more than a million people are represented in the figure). Increasing concentration of people in cities– urbanization—has all along been an important consequence of the spread of capitalism. By 1850, three cities had grown to a size of more than 1 million people: London, Paris, and Beijing (the three arrows in this map pinpoint their locations). By 2014, as a result of rapid population growth and the expansion of capitalism throughout the world, there were 498 such cities (only some of which are shown in this figure).

Source: Tertius Chandler, *Four Thousand Years of Urban Growth: An Historical Census* (Lampeter, Wales, UK: St. David's University Press, 1987). Map generated by Deepta Sateesh.

The process of urbanization is not limited to the United States, England, Mexico, or Japan. Urbanization happens wherever industrialization takes place. As the economic historian Eric Hobsbawm argued in *The Age of Capital: 1848–1875*, capitalist economic development propels an exodus from rural areas to cities, pushes people to migrate from one region to another, and drives them to move from one town to another. In 1900, nine of the ten largest cities in the world were in Europe or North America, with Tokyo as the exception. Today, with the global spread of capitalism, nine of the ten largest urban areas are now in Asia or Latin America, with New York as the exception. In 1850 there were only three cities with populations exceeding 1 million people—London, Paris, and Beijing—but as Figure 1.6 shows, by 2017 there were 520 urban areas of this size in all parts of the world.

Urbanization happened without anyone *planning* for it to happen: it occurred because individuals chose, for one reason or another, to leave their homes in the countryside and move to cities. Thus, the urbanization of the past few centuries is a prime example of how free markets can foster fundamental changes in society based on individual decisions.

In the waning years of the twentieth century, Bill Gates, the founder of Microsoft and the richest man in the world (see Chapter 5), embarked on a program of offering free computers to all libraries in poor neighborhoods in the United States, most of which were in rural areas. He imagined—and intended—that his program would halt the process of urbanization, since people in rural areas would now have access through the computers in their local libraries to most of what is available to denizens of urban areas. A story in the *New York Times* reported on this as follows: "Bill Gates predicted in 1995 that the Internet would help rural people stay put, in part because they would have the same advantages as city slickers in the virtual world." Having offered this prophecy in his 1995 book, *The Road Ahead*, Gates later revisited "the land of no stoplights" and had to concede that "the road ahead was full of blind curves." The fact was that the introduction of computers had done nothing to halt, or even to slow down, the exodus of people from rural America. Indeed, many rural Americans used the new computers to help them find jobs in urban areas. Gates concluded: "I thought digital technology would eventually reverse urbanization, and so far that hasn't happened."

The Changing Nature of Work

The way people earn their livelihoods has changed as well. At the dawn of the capitalist era, most families consumed only what they themselves produced or what they could obtain by selling their own products. Except for slaves, most families owned the tools they needed to make a living. With the rise of capitalism, people have become increasingly dependent on employment, that is, on getting a job working for someone else. At the same time, they have become subject to the dangers and hardships of unemployment. When there are not enough jobs to go around, unemployment leaves families and even whole regions destitute and desperate. It becomes a capitalist form of plague, potentially affecting everyone except the independently wealthy and thus threatening almost everyone with insecurity.

Rapid change in the workplace makes even the most skilled workers vulnerable to unemployment as technological change renders their skills obsolete. Before capitalism, the son of a blacksmith could be confident that the skills learned in his father's shop would be a secure source of his livelihood. Under capitalism, a worker may spend arduous years learning specific skills, but new production processes can

make them obsolete almost overnight. Blacksmiths in the nineteenth century, for instance, completed long apprenticeships, learning to work with wrought iron, needing to know exactly how much to heat the iron and how to process it. When steelmaking processes were improved, and steel replaced iron in some uses, blacksmithing began to shrink as an occupation. However, farriers (those who shoe horses) are still needed, and ornamental ironwork—making elegant gates and fences—still supports a few of these skilled workers. Those who mainly produced iron castings, on the other hand, by pouring hot iron into molds, were called foundrymen. But when iron production was partly replaced by steel production, some foundry workers and blacksmiths who were not yet ready to retire had to learn new skills, or else resign themselves to personally pay some of the costs of technological progress.

The changes also transformed working conditions. As time went on workers found employment in huge factories and mills, in circumstances where dangerous machinery, poor lighting, intense heat, long hours, and the pressure to produce quickly combined to make such jobs hazardous and exhausting. In the past century and a half before the passage of protective legislation, the number of people wounded in industrial accidents multiplied like casualties on a battlefield. Even now, occupational safety and health laws have far too few inspectors to enforce them.

Nothing has been left untouched by the expansion of the capitalist economic system. Even people's experience of time itself has changed. Precapitalist lives tended to follow natural time, marked by the passing of the seasons and the movement of the sun and moon across the sky. Individual work tasks were performed irregularly, with periods of high work effort alternating with periods of rest. The work pattern could follow the natural rhythms of the worker, or it might be dictated by the natural rhythms of the weather. But in capitalist employment labor is paid for by the hour, and work tasks are defined with reference to how much time they take to perform. Starting and stopping time, lunch, coffee breaks, and even bathroom breaks are often measured in minutes. Clock time has supplanted natural time. Clocks can now be seen on public buildings, in schoolrooms, at factory gates, on people's wrists, and of course on their cell phones. "Time is money," Benjamin Franklin said more than two hundred years ago, previewing what was to come.

In recent years the trend toward working outside the home has reversed for some, as modern communications technology has made it possible for people who do certain kinds of jobs—mostly well-paid writers, lawyers, professionals in finance, and others working for themselves—to work at home, "meeting" with clients and collaborators and selling their services over the Internet.

The Transformation of the Family

Social and family life has also been transformed in the capitalist era. Although families remain important to our emotional lives and procreation, the household has been entirely reshaped and, as a productive unit, nearly eliminated. Before the advent of capitalism, a family of three generations and several married couples often lived within a stone's throw of one another or even under the same roof, sharing tasks and meals; in some parts of the world this is still common. By the mid-twentieth century in Western societies, the typical household had been reduced to a nuclear family (a family unit consisting of a mother and father and their children), with grandparents living in a nursing or retirement home and aunts and uncles scattered throughout the country. By the end of the twentieth century, divorce, improved birth control,

abortion, greater longevity, and increasing numbers of couples choosing not to get married had taken us even further from the traditional household. Households in the United States are much smaller now, consisting on average of fewer than three people; by 2013 unmarried couples constituted more than one out of eight of the 60 million U.S. households that were headed by couples.

At the same time, many of the customary functions of the family have been removed from the family's domain. Activities such as making clothes and preparing and preserving food that were once carried on at home are now performed mostly in factories or other market-oriented enterprises. The people engaged in such work outside the home today are often the very women (or their grandchildren) who once worked in their own kitchens or at the home loom. Much of the work of rearing children, providing education and medical care to family members, and accomplishing other tasks in the home has also been entrusted to outside professionals.

Finally, today's family, at least in advanced capitalist countries, finds itself in a greatly changed social network. Once, families tended to live in the same community from generation to generation, with skills and occupations being passed from parents to children and each family having an acknowledged and often hierarchically ordered place in that community. Today, few families remain in one spot from generation to generation. Many have to move in search of work. Indeed, it is common now for families to move several times in one generation, making it difficult to sustain their ties to any particular community. As a result, families cannot rely as much as they used to on a local network for support or assistance. This further contributes to people's use of purchased services and government assistance. In many countries the day care center and the babysitter have largely taken over for the grandmother or the older children or relatives. For many people today the idea of a neighborhood as a community of families lives on only as a source of nostalgia.

Threats to the Ecosystem

Scientists have long warned of the damage that human activities can do to the natural environment. In 1962, marine biologist Rachel Carson described the damage pesticides were doing to the water and soil, and the already visible harm to flora and fauna. And over thirty years ago the first edition of this text cited a 1983 U.S. government warning about growing atmospheric concentrations of pollutants that would have a "greenhouse effect," warming the planet. It described the likely consequences: possible melting of polar ice caps, rising sea levels, and changing climates and rain patterns that might destroy prime food-growing areas.

Today we see all these effects. The evidence of global warming is stark: Since the late 1980s the planet has rapidly grown hotter, as the top panel of Figure 1.7 clearly shows. Although there are fluctuations from year to year, this thirty-year trend in the global average temperature has been relentlessly upward. The year 2016 was the hottest on record since 1880, and 2015 the second hottest. In fact, all ten of the hottest years on record have happened since 1998.

The Effects of Climate Change

As climate scientists predicted long ago, heat is not the only problem. In recent years changes in wind currents and rain patterns have brought more frequent extreme weather events. The year 2013 alone brought drought that killed pasture and

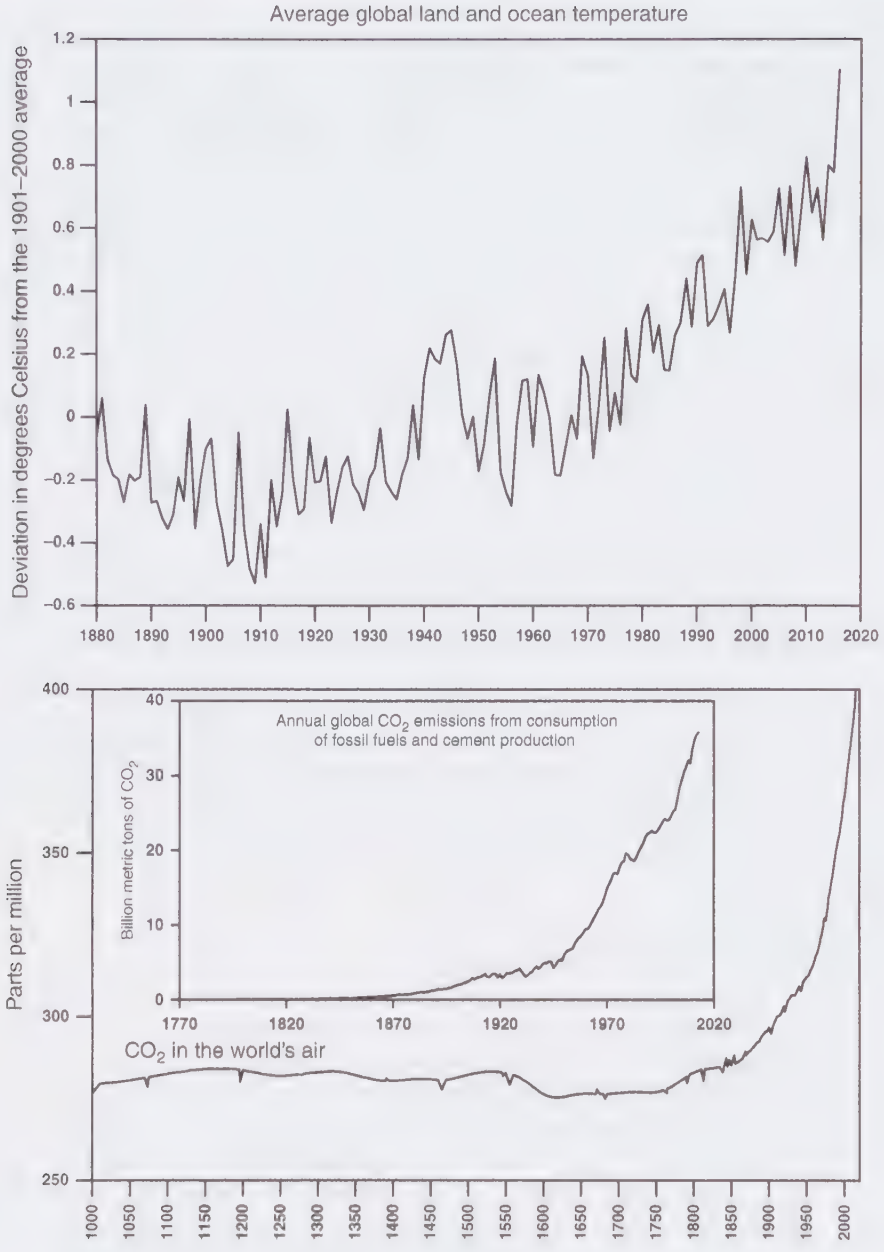


FIGURE 1.7 Global warming, CO_2 emissions, and CO_2 in the air.

For most of the past 1,000 years Northern Hemisphere temperatures and amounts of CO_2 in the world's air remained more or less constant. However, the growing use of fossil fuels such as coal and petroleum products in the twentieth and twenty-first centuries is associated with rising CO_2 emissions and therefore more intense concentrations of CO_2 in the world's air. The smallest (inset) chart shows that CO_2 emissions from fossil fuel consumption and cement production have risen exponentially over at least the last century and a half. Greenhouse gases—gases that trap heat and prevent it from escaping—have also risen, causing global temperatures to rise. The result is global warming. Past temperatures are estimated from studies of tree fossils, centuries-old ice, and other phenomena.

Sources: Top: National Oceanic and Atmospheric Administration (NOAA), National Climate Data Center, Climate at a Glance, Global Land and Ocean Temperature Anomalies, April, Base Period: 1901–2000. Available at http://www.ncdc.noaa.gov/cag/time-series/global/globe/land_ocean/1/4/1880-2016.csv. Bottom, main and inset: T. A. Boden, G. Marland, and R. J. Andres, "Global, Regional, and National Fossil-Fuel CO_2 Emissions." Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, 2016. Available at http://doi.org/10.3334/CDIAC/00001_V2016.

livestock in the African Sahel, causing hunger and famine affecting 23 million people; a vast area of drought-fed wildfires in New South Wales, Australia; the violent super-typhoon Haiyan in the Philippines; and floods covering one-fifth of Pakistan, affecting 20 million. Oxfam International estimated that in the five years between the 2009 and 2014 climate summits, natural disasters caused mainly by climate change cost nearly $500 billion, affected nearly one-tenth of the world's population, and killed 112,000 people.[2]

Why is this happening? After the late nineteenth century rise of refined oil as an energy source, the development of capitalism meant that fossil fuels were put to work powering engines that ran industrial production, public transportation systems, and a growing number of private vehicles. Burning fossil fuels (coal, gasoline, diesel) releases carbon dioxide (CO_2) into the air. The CO_2 and other "greenhouse gases" (GHG) absorb infrared rays from the sun, heating the air but also trapping the heat within the earth's atmosphere. This is called the "greenhouse effect" because it resembles the way the sun's rays enter a glassed-in space and heat the air inside, without allowing the heat any way to escape.

Since GHG persist in the atmosphere for tens to hundreds of years, even cutting back GHG emissions would not immediately solve the problem, though it would help. The consequences are far-reaching: The Arctic ice sheet already is less than half its former size and shrinking fast, and unless temperature increases are soon halted, scientists now estimate that Antarctic ice will likely melt far more rapidly than previously predicted. Rising sea levels already are beginning to encroach on some coastal areas, and many cities and states are completely unprepared. The ocean is also affected as a marine habitat. Part of the carbon dioxide in the air dissolves into its waters, making it currently 30 percent more acidic than three decades earlier. Ocean temperatures are also rising, which like rising acidity is hard on marine life. Coral reefs are being destroyed.

To be sure, not all climate change is caused by greenhouse gases. Average temperatures of the earth fluctuate from decade to decade under the influence of many factors. One is reduced sunlight due to volcanic events such as the massive Tambora eruption in 1815 and the Krakatau explosion in 1883. The Tambora volcano spewed forth so much ash in 1815, blocking the sun's rays, that 1816 became known as the "year without a summer." In that year frost covered the southern United States on the Fourth of July. The 1883 Krakatau eruption produced an ash cloud that circled the world for more than two years, lowering the earth's temperature. In the last century, however, average temperatures have risen due to higher greenhouse gas concentrations, much of which comes from burning fossil fuels. Methane gas produced by livestock, and GHG from cement production, are two of several other significant sources. Cutting down tropical forests worsens the situation because, according to a December 2014 NASA study, they absorb a considerable amount of carbon dioxide.

Is there a point of no return?[3] If so, have we reached it? There are two kinds of lines that scientists say we should not cross. One line is a temperature limit. In 2009, many countries signed the Kyoto Protocol agreeing that the global average

---

[2]"The Summit that Snoozed?" Oxfam Media Briefing, September 19, 1964, available at https://www.oxfam.org/sites/www.oxfam.org/files/file_attachments/bkm_climate_summit_media_brief_sept19.pdf

[3]The following discussion is based largely on the documentary film *Do the Math: Bill McKibben and the Fight over Climate Change* by Tony Hale, Bill McKibben, Mike McSweeney, Kelly Nyks, and Jared P. Scott (PF Pictures Productions; San Francisco, CA: Kanopy Streaming), available from the Media Education Foundation (Northampton, MA, 2013).

temperature should not be allowed to rise beyond 2° C (3.6° F) above the preindustrial global average temperature. The Paris Agreement of 2015 stated an even lower limit of 1.5° C, but with no real plan to achieve it. By 2016 the earth was already 0.9° C above the preindustrial temperature.

The other upper boundary is the level of atmospheric concentration of CO_2. NASA scientist James Hansen and his research team have estimated that the danger zone is above 350 parts per million (ppm) of CO_2.[4] As the bottom panel of Figure 1.7 shows, the world already exceeded this level in 1987, and by 2014 it even passed the 400 ppm level, continuing to rise at about 2 ppm per year. There is no clear cliff-like drop-off, but it would be wise to take steps to stop and then reverse this growth.

One problem with too much warming is that new processes are triggered that could easily accelerate the warming process. In places like Greenland, the frozen earth called the *permafrost* melts as temperatures rise, releasing at least some of the carbon now locked up in it, either as methane gas or carbon dioxide. Likewise, the drought induced by global warming increases the frequency and size of grass fires and forest fires, events which in turn further increase the CO_2 in the atmosphere.

Is there a technological solution? More technical knowledge may make it easier and cheaper to stop global warming, but the main obstacle to doing so is political. Nations that have emitted little carbon, and so bear little responsibility for causing global warming, want to see the major carbon-emitting regions be the ones to spend most heavily to solve the problem. So who are the biggest emitters?

The column graph in Figure 1.8 shows cumulative carbon dioxide emissions per capita for 1990–2012. The United States, Canada, and Europe have been the highest per capita emitters. Over these twenty-three years the United States and Canada emitted a cumulative total 376.6 metric tons of CO_2 per capita, more than four times the world average of 90.7 metric tons per capita. For Europe that number was 172.8 metric tons, just under twice the world average. The United States, Canada, and Europe together emitted about 41 percent of the cumulative world emissions over the period, though they have only about 14 percent of the world's population.

This was much lower than their share during the previous forty years, 1950–1989, however, when they emitted more than three-fourths of the world total; but this was largely because after 1990 other regions, particularly Asia, rapidly developed, and so increased their annual carbon emissions, as shown in the inset to Figure 1.8. Even so, Asia's population is around 4 billion, or about four times that of the United States, Canada, and Europe combined, so its per capita emissions remain well below the world average. Europe's annual emissions leveled off in the 1980s, while U.S. annual emissions only fell beginning with the 2007–2009 Great Recession. There is still much to be done. Among developed countries, Germany has gained recognition as a leader in shifting from fossil fuels to other forms of energy.

Researchers estimate that burning more than 20 percent of the fossil fuels currently reported in world reserves of private corporations and major state-owned oil reserves would lead to temperature increases beyond the acceptable limits described above. For this reason, environmental action groups ask that 80 percent of these deposits simply be left in the ground. The campaign to persuade universities to divest from fossil fuels hopes to make burning fossil fuels an unprofitable business activity, so that fossil fuel companies will shift to less destructive forms of production.

---

[4]James Hansen et al., "Target Atmospheric CO_2: Where Should Humanity Aim?," *The Open Atmospheric Science Journal* 2 (2008): 217–31.

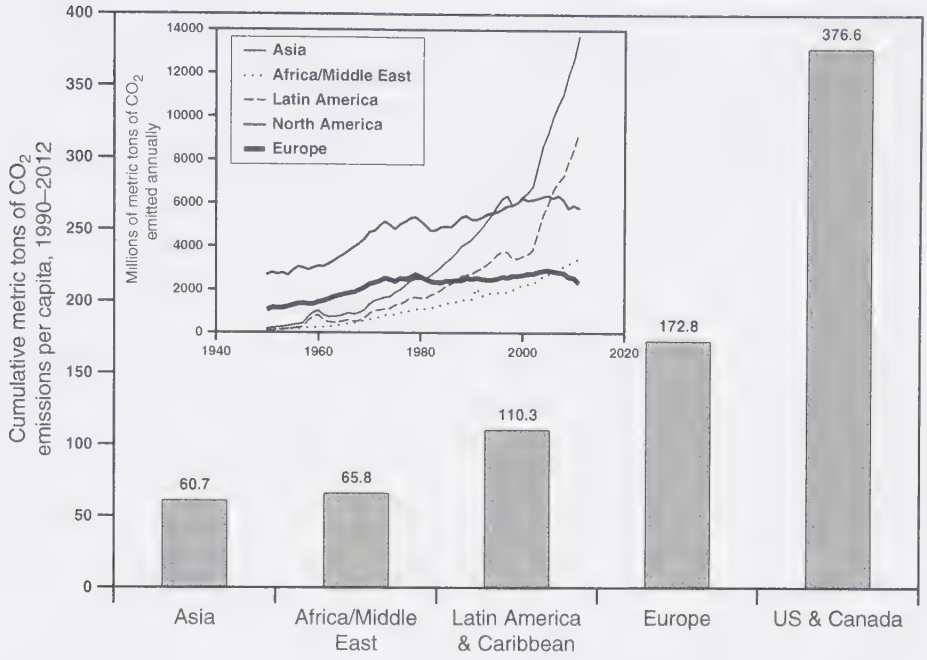


FIGURE 1.8 Cumulative CO_2 emissions per capita, 1990–2012, by region, and trends in total emissions by region, 1950–2012.

Changes in (net) CO_2 emissions result from changes in the amount of fossil fuels burned, but also from changes in cement production, from changes in forests that affect carbon absorption and emission, and other factors.

Sources: CAIT Climate Data Explorer, (Washington, DC: World Resources Institute, 2015), available at http://cait.wri.org and at http://www.wri.org/resources/data-sets/cait-historical-emissions-data-countries-us-states-unfccc. Data incorporate land use and forestry data from FAO 2014, FAOSTAT Emissions Database. Data were calculated and compiled by T. A. Boden, G. Marland, and R. J. Andres, in "Global, Regional, and National Fossil-Fuel CO_2 Emissions," Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, 2016. Available at http://doi.org/10.3334/CDIAC/00001_V2016. Data in metric tons of carbon were converted to metric tons of carbon dioxide by multiplying by 3.667.

Contamination

Global warming is not the only environmental danger. Rachel Carson's acclaimed book *Silent Spring*, published in 1962, warned that we were fouling our nest with pesticides and other toxic chemicals contaminating groundwater, rivers, lakes, oceans, and soil, and so killing off all kinds of creatures—birds, insects, and other—essential to the ecosystem. And in 1984, in the city of Bhopal, India, the discharge of toxic chemicals from a Union Carbide chemical plant there killed at least 3,700 people and permanently damaged the health of many more. Bhopal became known worldwide as a symbol of environmental destruction. In the years since the Bhopal disaster, many more environmental crises have occurred, and people have become increasingly aware of the threats facing the earth's ecosystem. The BP oil spill in 2010 in the Gulf of Mexico caused massive damage to marine life, shrimp farming, and the coastal economy. Such events have raised public awareness about the dangers of environmental destruction.

With oil spills, the dumping of toxic and radioactive chemicals, and routine discharges from tankers and industrial plants, the oceans are being spoiled. The destruction of natural habitats is endangering rising numbers of animal and plant species. As Carson warned, insecticides and herbicides are poisoning prime farming soils. Routine massive use of antibiotics in raising animals for meat, as well as in fish farming, has helped to foster the growth of drug-resistant bacteria ("superbugs"), which soon will be impossible to kill with any known antibiotics. These changes in our physical world, especially when taken all together, threaten destruction on a scale that we cannot now calculate, but must nevertheless do everything possible to prevent. Prevention may well end up being far cheaper, at least from the standpoint of society, than remediation after the damage is done.

The public has increasingly demanded policies to protect and restore the environment. Some of these have been very effective, resulting in cleaner air and water in many parts of the United States and reduced energy use in Europe. The main lesson is that environmental problems often cannot be addressed without cooperation among nations. The successful international effort to reverse the depletion of the ozone layer is an example of the kind of cooperation that is needed. But so far, the United States has been one of the countries least eager to make the changes needed to stop global warming, at least at the national level.

New Roles for Government

Government and people's relations to government have also been altered since the rise of capitalism. In 1500 much of what was to become the capitalist world was ruled by some type of despot, with kings and emperors basing their claims to authority on God's will, hereditary right, or simply brute force. There were few ways in which ordinary people could protect themselves from the arbitrary powers of such rulers. But governments played a minor role in most people's daily lives; tax collection was typically nonexistent or ineffective, while compulsory schooling, permanent police forces, and standing armies did not become common until the nineteenth century.

Hereditary rulers were challenged by the British revolutions of the seventeenth century and by the American and French Revolutions of the eighteenth century. These revolutions established the important principle of governing with the consent of the governed, although only property-owning freeborn (nonslave) males gained the right to vote. But in Europe and the United States these revolutions placed important limitations on absolute rulers, and eventually they were followed by written constitutions, the abolition of slavery, widespread male suffrage, elimination of property qualifications for voting, and the extension of the franchise to women and minorities. In the nineteenth and twentieth centuries, largely as a result of intense and lengthy struggles by workers, antislavery groups, suffragists, and others, all the major capitalist countries participated in the growth of democratic government. As governments became more democratic, they came to play a major role in providing such services as public education for the young and income support and health care for the elderly.

In the late nineteenth century in Europe and later in North America and other parts of the world, governments increasingly also took responsibility for providing assistance to the unemployed and those unable to work. Today, in western Europe and to a lesser degree in the United States, many people expect governments to provide something like a social safety net, especially during periods of economic decline. Moreover, governments have become major employers as well.

But while government in most countries has become more democratically accountable, everywhere it has also become more intrusive. In the past century governments have attained increased powers to invade the privacy and influence the sentiments of the citizenry. Television and other modern communications media give heads of state enormous influence in shaping people's opinions, while modern information technology allows government easy access to our location, private messages, and economic activity.

Globalization

The finale of Warner Brothers' blockbuster Matrix film trilogy, *The Matrix Revolutions*, opened simultaneously on November 5, 2003, on 10,013 screens in more than fifty countries and in forty-three languages. On that day at precisely the same moment—6 a.m. in Los Angeles, 9 a.m. in New York, 5 p.m. in Moscow, 11 p.m. in Tokyo, and so on—moviegoers watched the title shots appear on the screen as this first-ever global debut of a film began.

The president of Warner Brothers Entertainment, Alan Horn, called it "showmanship." "It's theatrical, it's fun, it's exciting," Horn said about the mega-opening. He also explained that by making the film available in theaters around the world he hoped to limit the profits of film pirates, who copy a film and distribute it globally (as they often do if a film opens in one country but cannot be seen in theaters in other countries until much later). Like many others, Horn understood that globalization means that many aspects of our life, from our entertainment to the enforcement of property rights, are no longer local or even national matters; they are global.

Capitalism has accomplished what not even the most powerful rulers in the past were able to do: it has brought the entire world into a single all-encompassing system. Alexander the Great conquered much of the world in the fourth century BC, expanding the reach of Greek civilization as far as India, but he soon retreated, leaving only traces of Greek culture behind. The Roman Empire at its height extended from the British Isles to the Middle East, at most a few thousand miles. In the two centuries following Muhammad's death in AD 632, his influence expanded well beyond Arabia, and Islam became the dominant force in a swath of loosely affiliated states stretching from Spain to what is now Indonesia—a third of the way around the world. But none of these empires touched more than parts of the globe. Only the capitalist economy, with a built-in tendency to expand and with the help of the twentieth-century revolution in communications technology, has reached entirely around the world, obliterating distances, spawning common languages and appetites, and bringing most of the world's peoples into a single interdependent system.

In Bangalore, India, women in a call center respond to customer service inquiries placed via 800 numbers to companies such as General Electric, Dell Computer, America Online, and British Airways. They and others like them are doing back-office work for global corporations that outsource tasks such as customer service to India or other countries where English is spoken and people are willing to work for a fraction of what American workers are paid. To facilitate the growth of call center employment, the Indian government over a decade ago installed reliable high-capacity telephone lines so that a person working in an Indian call center can communicate with customers overseas without sounding any different from a person working in the same country as the corporation being called.

To remove any suspicions that the person responding to a service call from the United States is a foreigner, the young Indian women who answer the phone are trained to sound just like Americans. To practice, they watch old video episodes of shows like *Friends*, and learn that "Bimmer" refers to a BMW. They also take American names like "Naomi Morrison" and "Susan Sanders," according to the *New York Times* story illustrated by the photo of the two women. With accents one might hear at Wrigley Field, the women pretend to be residents of Chicago. Call centers elsewhere in the world similarly train their workers to respond to calls from (and to sound like residents of) some specific U.S. region.

There are now over 60,000 call centers in the United States employing around 5 million people, the majority of them women, most accepting relatively low wages. Call centers outside the United States earn tens of billions of dollars, and for years it appeared that such jobs would continue to go abroad. But for several years, as the wage gap between the United States and the main call-center countries (India and the Philippines) narrowed, these jobs began returning. That trend has slowed down, however, and call centers, or "contact centers" outside are still growing.

One technological change that has smoothed the way to more global trade has been containerized shipping. In any major port, one can see stacks of cargo containers that will be loaded onto container ships, or have just been unloaded. Today, most of non-bulk seaborne shipping is containerized (bulk shipping is of coal, iron ore, petroleum and its products, and the like, and by weight is about five-sixths of all seaborne trade), Containers carried one-sixth of the 10 billion tons of cargo shipped in 2014 by sea, but about half its total value. Container traffic is measured on the basis of "twenty equivalent units," or TEUs, with each 20-foot-long container being counted as one TEU and each 40-foot long container as two TEUs. The average container ship in use by the top twenty operators holds 5,000 to 6,000 TEUs, but new ships being ordered typically will each carry 10,000 to 20,000 TEUs. About 700 million TEUs passed through the world's ports in 2014.[5] Containerization is one technological change that has made world trade cheaper and easier.

The fact that there is now a single world market for goods and services of all kinds means that producers depend on buyers worldwide for their livelihood. When the U.S. economy is booming and American wages are on the rise, automobile workers in South Korea prosper. The globalization of the economy also means that Indian software developers are constantly looking for ways to edge out competitors in Silicon Valley (and vice versa).

Similarly, the clothing worker stitching shirts in the United States (an endangered species, with only about 50-60,000 production workers now engaged in sewing apparel) is looking for ways to stay ahead of the competition in Sri Lanka, where the stitchers are paid less than one-tenth of what a comparable U.S. worker makes. Most large companies now consider the entire world not only as the market for their products but also as offering potential locations for production units. And, as readers of this book may already know, American students seeking admission to colleges or graduate schools now face competition from students from all over the world. The same is true for many jobs in the rapidly expanding information-based sectors of the U.S economy.

---

[5]United Nations Conference on Trade and Development, *Review of Maritime Transport 2015*, New York and Geneva: United Nations.

Trading over long distances has been integral to the functioning of capitalism since its inception. The huge market fairs in Europe from the fifteenth to the seventeenth centuries attracted merchants from hundreds of miles away and even from beyond Europe. They often brought with them lightweight luxuries such as spices and precious metals. However, long-distance trade in other goods was relatively insignificant. As noted earlier, transportation costs for heavier products were very high, and other difficulties blocking the transport of goods over long distances included brigandage, piracy, and the many local transit taxes levied by the rulers of the small kingdoms and principalities that then made up Europe. In the last half-century, however, there has been a dramatic increase in long-distance trade. In 1950 only 8 percent of the world's output was exported to other nations; now the comparable figure is almost 30 percent.

It now makes sense to think, to some extent, of the entire world system of investing, buying, and selling as a single global economy. National boundaries still matter, of course, but less so than in the past: goods, money, and information pass with only minor impediments from country to country. Even movement of people among countries has been somewhat eased with the formation of the European Union and some regional agreements—although for refugees, even transit countries have been quick to put up barriers. The process of globalization has challenged national governments, for it makes them more interdependent. Decisions made by the U.S. government or by the European Central Bank have ripple effects that extend throughout the world. The inability of Mexican or Russian borrowers to service their foreign debt can send shock waves through Wall Street and from there to the U.S. Treasury Department.

Globalization, as we have said, concerns the international movement of money, goods, and even people in search of a better livelihood. But globalization is not simply economic: it concerns languages, political rights and movements, what people value and admire, how people worship, what they eat and dance to, and the arts. The Indian women at the call center in Bangalore are able to imitate Chicago accents and discuss recent Chicago Bulls games because they see the Bulls on TV and hear Chicagoans on the radio.

Globalization, like technical change, is a source of affluence for many, but it can quickly make a once cutting-edge industry or job skill obsolete, causing unemployment and financial hardship for the people involved. The citizens of most democratic nations have pressed their governments for protection against the vagaries of the global marketplace, in some cases seeking to limit the extent of global exchange. In some democracies, especially those in western Europe, governments have implemented extensive unemployment insurance and offered subsidies for retraining and relocation to workers in declining industries. These programs are expensive, of course, and many governments are reluctant to raise the taxes necessary to fund them for fear that higher taxes will induce some businesses to relocate to other (lower-tax) nations, which would only increase economic insecurity. Many fear that globalization increases economic insecurity while at the same time diminishing the power of governments to implement programs that might make people more secure.

Underlying such fears is a simple fact: the economy is now global while government is still local. The global economy is thus interconnected, while the power to coordinate it is dispersed among more than two hundred national governments, a handful of international bodies, and a few thousand large corporations. Power in the

global economy is not distributed among individual citizens and consumers. Rather, it is held by a small number of powerful institutions. For example, the World Trade Organization, which regulates the terms on which nations exchange goods and services, is deliberately structured to limit input from ordinary citizens.

The heads of the world's one thousand largest corporations, producing 80 percent of the globe's industrial output, could all fit into a medium-sized concert hall, for example, Alice Tully Hall at Lincoln Center in New York City. About two thousand corporate leaders, along with some government officials, actually do come together annually at a ski resort in the small town of Davos, Switzerland. Calling themselves the World Economic Forum, this exclusive group has been assembling there for more than forty years. From 2000 to 2015, protesters showed up at Davos to criticize the exclusive meeting, but by 2016 they had decided it was no longer worth it, saying that the Davos Forum was losing credibility.

Some observers have hoped that global leaders such as those who meet in Davos would work together to regulate the world economy in ways that would protect both people and the environment from the social costs generated by multinational capitalist enterprises. However, the fact is that corporations have at least as much interest in competing as in cooperating, while their interest in protecting society and the environment from the consequences of global profit seeking is tempered by their opposition to restrictions on their own profit making. And in the absence of anything resembling a world government, all international cooperation must come about through negotiations among independent governments; but of course any government can refuse to participate in such negotiations.

Conclusion

To lay the foundations for an understanding of capitalism, this chapter has focused on the technological revolution, increasing material well-being, the population explosion, the growth of cities, the transformation of work and family, threats to the ecosystem, new roles for government, and globalization. The list of changes that have come with capitalism could be expanded, but the set of transformations discussed here establishes the basic point of the chapter: capitalism generates perpetual change.

It could be argued that the rise of capitalism was not so much the cause as it was the effect of the changes outlined in this chapter. Might not advances in science and technology have led to the development of capitalism? Or could the population explosion have been the cause and capitalism the consequence?

Science and population growth are undoubtedly important, but before capitalism they did not have cumulative effects. Scientific knowledge and technology were more advanced in the Islamic world and China, for example, than they were in Europe before 1500. But neither Islamic science and mathematics nor the Chinese inventions of gunpowder, magnetic compasses, cast iron, moveable type, canal locks, and machines for keeping time led to sustained technological progress or industrial development. It is also true that periods of rapid population growth have accompanied short periods of economic expansion throughout the 100,000 or so years of human existence, but as Figure 1.5 shows, it was not until the advent of capitalism that rapid population growth became the rule rather than the exception.

In the last five hundred years virtually all traditional patterns of life and livelihood have been disrupted and reconstructed. The world and the world's peoples have been shaken up and remade. In the chapters to follow we discuss the reasons why capitalism is such a powerful source of change and why it affects not only the economy but also politics, beliefs, and many other dimensions of social life.

Suggested Readings

Sven Beckert, *The Empire of Cotton: A Global History* (New York: Vintage, Reprint edition, 2015).

James K. Boyce, *The Political Economy of the Environment* (Cheltenham, UK: Edward Elgar, 2002).

Fernand Braudel, *Capitalism and Material Life 1400–1800* (New York: Harper & Row, 1967).

Ha-Joon Chang, *23 Things They Don't Tell You about Capitalism* (New York: Bloomsbury Press, 2010).

William Cronon, *Changes in the Land: Indians, Colonists, and the Ecology of New England* (New York: Hill & Wang, 1983).

Jared Diamond, *Guns, Germs, and Steel: The Fates of Human Societies* (New York: W.W. Norton, 1999).

Maurice Dobb, *Studies in the Development of Capitalism* (New York: International Publishers, 1947).

Bill Gates, *The Road Ahead* (New York: Viking Penguin, 1995), especially Chapter 8, "Friction-Free Capitalism," 157–83.

Eban Goodstein, *Economics and the Environment*, 7th edition (New York: Wiley, 2014).

John Gray, *False Dawn: The Delusions of Global Capitalism* (New York: New Press, 1999).

Eric Hobsbawm, *The Age of Capital: 1848–1875* (New York: Vintage Books, 1996).

Naomi Klein, *This Changes Everything: Capitalism vs. the Climate* (New York: Simon & Schuster, 2014).

David Landes, *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor* (New York: W. W. Norton, 1999).

Joel Mokyr, *The Lever of Riches: Technological Creativity and Economic Progress* (New York: Oxford University Press, 1990).

Barrington Moore, *The Social Origins of Dictatorship and Democracy* (Boston: Beacon Press, 1966).

Robert Pollin, *Greening the Global Economy* (Cambridge, MA: Boston Review Books, MIT Press, 2015).

Robert Reich, *Saving Capitalism: For the Many, Not the Few* (New York: Alfred A. Knopf, 2015).

George Soros, *The Crisis of Global Capitalism: Open Society Endangered* (New York: PublicAffairs, 1998), especially Chapter 6, "The Global Capitalist System," 101–34.

Eric Wolf, *Europe and the People Without History* (Berkeley: University of California Press, 1982).

People, Preferences, and Society

2

Parents everywhere are sometimes late in picking up their children at day care centers, thereby inconveniencing the staff. An experiment, carried out in Haifa, Israel, was designed to find a solution to the problem of tardy parents. At six randomly chosen centers a fine was imposed for lateness, and a few other centers were selected to serve as a "control group" (nothing was changed at these centers). Staff at the centers with the newly instituted fines expected that punctuality would improve. Contrary to these expectations, however, there was an *increase* in tardiness when the fines were imposed: the number of parents picking up their kids late more than doubled. Even more striking was the fact that when the fines were revoked, the parents' higher rate of tardiness persisted. Meanwhile, the amount of parental lateness at the centers in the control group did not change.

The economists who designed the Haifa experiment were quite surprised by the results. Most economists assume that people seek monetary gain and try to avoid losses. From this perspective, the day care centers' fines should have given the parents an incentive to be more punctual. But the plan backfired. After analyzing the results, the designers of the experiment concluded that the imposition of the fines must have unintentionally suggested to the parents a new way of thinking about their behavior. Whereas before the experiment lateness had been seen as a violation of a *moral obligation* (to pick up the kids on time), after the imposition of the fines being late could be viewed as a *choice* between picking up the kids on time and paying a price (the fine) for being late. And under the new system many parents were apparently willing to pay the price. The designers of the experiment titled their report "A Fine Is a Price." Their main finding was that imposing the fines had signaled to the parents that they were now in a market-like relationship to the day care staff—one in which they could *buy* lateness. Once the fines had been introduced, revoking them could not restore the initial situation; it just lowered the "price of lateness" to zero.[1]

Why were the economists who designed the day care experiment surprised by its results? It was because they took it for granted, as have most economists until very

[1]Uri Gneezy and Aldo Rustichini, "A Fine Is a Price," *Journal of Legal Studies* 29, no. 1 (2000): 1–17.

recently, that people care little about others, act only to promote their self-interest, always seek opportunities for personal gain, even at the expense of others, and abide by the moral standards of their community only when it serves their own purposes.

Economic man (*Homo economicus*) refers to the assumption that human beings are calculating, amoral, and self-interested.

The assumption that people are calculating, amoral, and governed by a self-interested predisposition is referred to as the *Homo economicus,* or "**economic man**," assumption. To say that people are governed by a **self-interested** disposition means that they consider only how their actions will affect themselves, not how their actions will affect others.

Self-interest refers to a disposition to consider only how one's actions will affect oneself, not how they may affect others.

One of the truly radical ideas in economics is the idea that given the right laws and institutions, individual selfishness can be harnessed to serve the public good.

The economic man assumption, of course, leaves out a lot. While it is certainly true that selfish behavior is common, so are acts of compassion, selflessness, and altruism. People show concern for their friends' well-being, volunteer for military duty, care for their infants or aged parents, risk their lives for strangers, and forgo opportunities to steal even when no one is looking. When such acts are motivated by a concern for others—or for what happens to others—they are not self-interested acts. We might better call them *other-regarding* because they are motivated by a regard for others. (Self-interested behaviors are, of course, *self-regarding.*)

The key to whether an act is self-interested is its motivation. The deciding factor is whether the act is motivated by a concern for others, not whether it produces happiness in the actor. For example, many generous people take pleasure in helping others in need. But this pleasure does not make them selfish people: since they act from unselfish motives they are *not* self-interested. Not all other-regarding acts are as admirable as helping others and obeying moral codes, however. Hurting another person out of spite, jealousy, or intolerance of his or her religion or race is also other-regarding. Such an act is intended (based on a motive) to make something bad happen to someone else, just as generous actions seek good outcomes for others.

Also left out of the economic man assumption is the fact that people change. *Homo economicus* is believed to be a "natural" phenomenon, and, accordingly, the type of (self-interested) behavior associated with it is assumed to have been prevalent in every kind of society, unchanging across the entire span of human history, and sure to be characteristic of any future economic system. The great eighteenth-century philosopher-economists Adam Smith and David Hume thought that the key to a well-ruled society was not to deny self-interest (which they thought to be impossible) but to find a way to harness selfish motives to serve socially valued objectives:

> *It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest.*
>
> —Adam Smith, *The Wealth of Nations* (1776), book I, chapter II

> *Political writers have established it as a maxim, that in contriving any system of government . . . every man ought to be supposed to be a knave and to have no other end, in all his actions, than his private interest. By this interest we must govern him, and, by means of it, make him, notwithstanding his insatiable avarice and ambition, cooperate to public good.*
>
> —David Hume, *Essays: Moral, Political and Literary* (1742)

It is well known, however, that people frequently change as a result of experiences they have in the economy. For example, a long but unsuccessful job search

can turn a confident and happy person into a depressed and violent threat to his or her family or community. Even a whole group's culture can change when its way of making a living is altered. For example, when the sons and daughters of farmers become office or factory workers, it is quite likely that they will develop new patterns of behavior, discover new wants, and be guided by different values.

In this chapter we consider various ways in which economists attempt to explain individual behavior. A common starting point is that people make choices and do things for reasons—even if not always for good reasons, and even if their reasoning does not always correctly anticipate the outcomes of their actions. Thus, behavior is seen as intentional, or goal-seeking.

The main idea of this chapter is that *while the intentional view of behavior is essential to understand why people do what they do, economic man is a fictional character. Real humans are* diverse (*some selfish, others generous*) *and* versatile (*sometimes bold, other times cautious*), *and their values, tastes, habits, and beliefs are very much the product of their upbringing, work experience, and national, ethnic, and cultural backgrounds.* This main idea is expressed in five key points.

1. Explaining behavior requires taking into consideration an individual's *constraints* (limits on his or her actions), *preferences* (evaluations of outcomes), and *beliefs* (understandings of how particular actions may bring about specific outcomes).
2. Laboratory experiments as well as ordinary observations of daily life show that selfishness is but one of our motives. We are also generous, even toward those we do not know, and we are willing to reward those who treat us well and to punish those who treat us or others badly, even if such actions are costly to us in terms of lost income or missed opportunities for personal gain.
3. People are similar in many respects the world over, reflecting our common genetic inheritance, and these commonalities are sometimes termed *human nature.* But in other respects our behaviors differ greatly, reflecting differences in the things we have learned from others in our society; such differences are termed *cultural differences.*
4. Families, schools, neighborhoods, religious institutions, and workplaces all play a part in the processes by which we come to have our particular values, desires, and beliefs.
5. All animals *compete* with other members of their species, but humans are unique in the extent to which we can also *cooperate* with those to whom we are not related. We have become the "cooperative species" because throughout history cooperative people prospered, and their cooperative behaviors were copied by others, in part because groups that succeeded in cooperating survived and grew, eclipsing groups that did not.

Aggregate or **population-level outcomes** refer to the economic totals, averages, and relationships that are generally studied by economists.

Economics is about *totals* (the output of an economy or the number of unemployed people, for instance) and *averages* (per capita income, for example). It is also about *relationships*—the power of employers over workers, the price of bread relative to the typical wage, the distribution of income between rich and poor, and so on. These totals, averages, and relationships are sometimes called **aggregate** or **population-level outcomes**.

Economics is not about what particular individuals do, but economists need to know how individuals behave in order to explain totals, averages, and relationships. In most cases it is difficult to understand aggregate, or population-level, outcomes without understanding why people do what they do. Hence, individual behavior comes in as part of an explanation of the larger picture.

Constraints, Preferences, and Beliefs

To explain why people do what they do, economists make use of three terms: *constraints*, *preferences*, and *beliefs*. An example will elucidate the meaning of each of these terms. Imagine that you are planning to drive across the country and are going to purchase a used car for the trip. You will need to pick out the car, decide how long you want to spend on the road, and select the kind of accommodations you will stay in while traveling.

Constraints are the limits on the actions that an individual or a society can take.

Preferences are the relative values one places on the various outcomes that one's actions might bring about.

Beliefs are one's understandings of the actions necessary to bring about particular outcomes.

Constraints put limits on the various actions available for you to take. Such constraints might include physical limitations (one cannot travel in a car from Massachusetts to California in less than two days); your own capacities (you cannot drive for more than twelve hours at a stretch without going to sleep); your social relationships to others (the availability of friends you might ask to go with you); facts about institutions or your own ethical rules (you may not want to buy your car from a dealer you know is corrupt); and your wealth and capacity to borrow (you have a limited bank account and possess little that might serve as collateral in return for a loan). Constraints will limit your actions in such a way that you will face what is called a *trade-off*: achieving more of one thing means getting less of something else. Given your budget—limited by your wealth plus what you can borrow—you can buy a better car if you are willing, say, to give up staying at upscale motels.

Preferences are the relative values you place on various outcomes that your actions might bring about. For instance, as you plan your car trip to California you might assign different degrees of importance to comfortable nights in upscale accommodations, fatigue from all-nighters on the road, ownership of a decent car at the end of the trip, and arrival in California by a certain date.

Your **beliefs** are your understandings of the actions necessary to bring about a particular outcome. (Note that this is a particular usage of the term *belief*, the broader definition of which refers to a conviction regarding the truth of something.) For example, you may believe that getting to California without a breakdown requires buying a better car.

Information about constraints, preferences, and beliefs is generally sufficient to explain why a particular set of actions was taken. Continuing the road trip example: You bought a seven-year-old Honda Civic because you could not afford an Accord and you believed that the Civic would probably get you there. Also, you asked a friend to join you so that you could drive straight through, enjoy his company, and surprise your parents by getting home early.

In sum, individuals make choices to take various actions (within their constraints); they seek to bring about the outcomes they desire (according to their preferences); and they base their choices on their understandings (beliefs) about how certain actions may bring about the desired outcomes. The important thing to note here is that behavior involves choosing. The choices that the driver in the above example made may have been quite limited due to constraints such as a lack of wealth, but that does not mean that the actions taken were not choices. It may also be that

the choices made were bad choices (the seven-year-old Civic might not have been adequate to get her all the way to California), but, in any case, choices were made.

We stress the element of choice because some views of behavior see choice as unimportant or even nonexistent. Such views are based either on the argument that individuals are not free (their constraints dictate a particular action) or that they are creatures of habit or conformity. There certainly are situations in which we are not free: the choice between "your money or your life!" is not much of a choice. It may also be the case that we sometimes repeat our past actions (out of habit) or copy (conform to) the actions of others without considering alternative possibilities. Habit and conformity certainly play a part in our behavior: just think of what you ate for breakfast this morning or what you wore to the last social event you attended. But views of behavior as coerced, habitual, or conformist fail to recognize the important elements of choice in most of what we do.

The constraints, preferences, and beliefs approach to understanding human behavior has therefore been widely accepted, not only in economics but in other social sciences as well. However, no theory can explain anything by itself. To explain behavior we need to know facts about the particular constraints, preferences, and beliefs pertaining to a given situation, and such facts will differ from person to person and among groups. Men and women face different constraints, for example, as do members of different classes, races, ethnic groups, and nationalities.

Moreover, two important aspects of the constraints, preferences, and beliefs approach have not yet been mentioned. The first is that preferences are not necessarily selfish: there is no reason to say that people must always be self-interested. The person driving across the country in our earlier example may have wanted to arrive early not for selfish reasons but to please her parents. Second, we have not said anything about where preferences come from. Was her concern for her parents' happiness an expression of "human nature"—an expression of a genetically transmitted trait? Or was it the result of her happy childhood? Was her lack of status consciousness (indicated by her purchase of the seven-year-old Civic) derived from a considered decision not to throw money away, or was it the result of a vague awareness that her cash-strapped friends would have frowned on the purchase of as nice a car as an Accord?

The view of preferences adopted in this book is fundamentally different from the one that prevails in the neoclassical school of economics, an approach to understanding the economy that appears in many textbooks and is discussed in detail in the next chapter. Neoclassical economists build their theories on the *Homo economicus* assumption that people have entirely self-interested preferences. They assume that people care about outcomes that involve themselves but not those that affect others. They also assume that everyone is an economic man and everyone knows that everyone is an economic man.

In addition, neoclassical economists generally do not ask where preferences come from. Rather, they take preferences as "given," meaning that the preferences that guide economic decision making are thought to be simply there, possibly as an expression of human nature, or possibly due to advertising, socialization, or other factors that are of no concern to economists. Moreover, neoclassical economists usually view preferences as being *exogenously* determined, formed by forces *outside* the economy.

Because the neoclassical approach is out of line with many scientifically determined and widely known facts regarding human behavior, we take a different view. As we have said, in our understanding of the economy preferences are not necessarily selfish. More important, we do not assume that preferences are exogenously determined. Rather, we view them as *endogenously* determined, that is, determined

mostly by processes *internal* to the economy. In the next section we set forth our reasons for doubting the assumption that preferences are entirely selfish. Later in the chapter we explore the question of where preferences come from.

"Economic Man" Reconsidered

Nobody would pick *Homo economicus* for a housemate, a spouse, a friend, or (if we could choose them) a parent or a child. The economics Nobel Laureate Amartya Sen has called economic man a "rational fool." But the implications of the concept are actually worse than Sen suggests: mental health professionals use the term *sociopath* to refer to a person whose behavior is governed entirely by calculation of self-interest. Sociopaths have no sense of right and wrong, and they lack any concern for the well-being or pain of others.

It is not surprising, then, that since the inception of neoclassical economics in the late nineteenth century, even its adherents have sometimes had difficulty with the assumption that human beings are motivated solely by self-interest. A founder of the neoclassical approach, F. Y. Edgeworth, wrote: "The first principle of economics is that every agent is actuated only by self-interest." In his next sentence, however, he cautioned that this "first principle" is strictly applicable only in situations of "contract and war."[2] With regard to war, Edgeworth was not entirely right: bravery under fire often is not based exclusively on self-interest. And with regard to contracts, sometimes a handshake is a handshake, even if one party could benefit by violating the unwritten contract.

Just how mistaken the assumption of universal selfishness is has been revealed in a series of behavioral experiments in recent decades. Subjects, usually students, often economics or business majors, volunteer to play a game in which they can win real money. Each is paired with a person, for example someone in another room, whom they do not meet and do not know, for a single interaction. One (usually chosen randomly) is termed the "proposer," the other, the "responder." The proposer is provisionally given, for example, $10, an amount of money referred to as the "pie." Both participants know the size of the pie. It is explained to both of them in advance that the proposer will decide how much of the pie to offer to the responder; after the proposer makes the offer, the responder decides whether to accept or reject it. If the responder accepts, the responder gets the offered portion, and the proposer keeps the rest. If the responder rejects the offer, both get nothing. The pie is often a small sum, such as the $10 described above, but the game has been played in the United States for $100 and in Indonesia with a pie equal to three months' salary. This experiment is called the "ultimatum game," where the "ultimatum" refers to the proposer's offer, which the responder can either accept or reject.

How would "economic man" play this game? As the proposer, he would reason that the responder (assumed also to be an economic man) would accept *any* offer greater than zero, for rejecting an offer of even one penny would deprive the responder of a penny. For an economic man a penny is better than nothing, no matter how it is acquired. This being the case, the proposer would offer a penny (or the smallest amount possible), anticipating that it would be accepted.

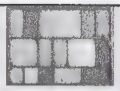
---

[2]F. Y. Edgeworth, *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (London: C. Kegan Paul, 1881), 104.

But this is not what usually happens when the ultimatum game is played with real people for real money. Before we tell you what *does* happen, however, think about what you would offer if you were a proposer with $100 to share, in some proportion, with a responder. Also try to imagine the lowest offer you would accept if you were the responder.

The ultimatum game has been played in hundreds of experiments with university students as subjects in the United States, Japan, Israel, Germany, Russia, Slovakia, Slovenia, Indonesia, and many other countries. Few play the game as economic man would. The vast majority of proposers offer between 40 and 50 percent of the pie, and the most common offers are typically 50 percent. Equally striking is the fact that offers of 25 percent or less are frequently rejected.

These experiments show that neither the proposers nor the responders behave like economic man. A responder is often willing to reject a low offer and end up with nothing, just to punish a proposer who makes an unacceptably low offer. Many have interpreted this behavior as evidence of a *preference for reciprocity*—a tendency to be generous toward another person as long as you are treated well by the other person but a willingness to pay good money to punish someone who has crossed or insulted you, even if you will never see that person again.



Coke Versus The "Just Price"

The Coca-Cola Company has tested a vending machine that automatically raises the price of a soda on hot days. Doing this does not require rocket science, just a thermostat and a computer chip. The company's chairman and chief executive at the time, M. Douglas Ivester, noting that the desire for a cold drink goes up with the temperature, concluded: "So it's fair that it should be more expensive." Airlines charge more when the demand is high, so why should Coke not do the same? "The machine will simply make this process automatic," Ivester explained.

Not everyone agrees. A Pepsi spokesman, no doubt seeking a competitive edge, took the high road: "We believe that machines that raise prices in hot weather exploit consumers who live in warm climates." Another beverage executive wondered: "What's next? A machine that X-rays people's pockets to find out how much change they have and raises the price accordingly? . . . It's another reason to move to Sweden!"

Apparently the price of a Coke—or that of any other commodity—is not, for some people at least, only something to be left to the market (or to the influence of large corporations). Economics Nobel Prize winner Daniel Kahneman (a psychologist by training) and his collaborators asked consumers if they thought it was fair for stores to raise the price of snow shovels during winter storms. The answer: they did not.

The Coca-Cola Company's new machine strikes at least some people as unfair because they think that when two parties engage in a mutually beneficial exchange—one, say, that increases a company's profits while quenching a consumer's thirst—the distribution of the benefits and burdens should not violate ethical norms. This idea can be traced to Thomas Aquinas, the medieval Catholic philosopher, and his concept of a "just price." Most economists think the idea of a just price is a nonsensical expression—like a "yellow logarithm." They would side with Coke. But judging from the reception that Coke received, they have not persuaded everyone yet.

Source: Constance Hayes, "Coke Tests Vending Unit That Can Hike Prices in Hot Weather," *New York Times*, October 28, 1999.

One interpretation is that a proposer could have well-informed beliefs and be motivated by selfish preferences. Suppose the proposer believes that the responder will not play the game like an economic man, one willing to accept a penny. If the proposer believes that the responder will reject low offers, then making a 50–50 offer could be nothing more than self-interest guided by prudence.

While the responders' behavior is understood as reciprocity, the behavior of proposers seems to be more complex than a simple preference for reciprocity would suggest. It is possible that high offers could reflect unconditional generosity toward the responders or a concern for their well-being irrespective of any behavior on the responders' part. If this is the correct interpretation, the proposers can be said to have *altruistic* preferences—preferences that lead them to act to benefit others at some cost to themselves (even with no expectation that reciprocal benefits will be received later).

All we can say for sure in interpreting the most frequent outcomes of the ultimatum game is that neither the proposers nor the responders behave like economic man. Even the self-interested but prudent proposer just described does not believe that the responder is an economic man. And in virtually all cases the proposers assume that the responders will depart from the assumption of perfect selfishness.

Violations of the selfishness assumption are not confined to these experiments, and neither are they limited to such dramatic but exceptional examples as heroism in warfare. Most people do not steal or cheat on their taxes even when they are sure they can get away with it. And in all of the richest nations in the world (Canada, the United States, and the European countries, for example), large majorities vote for income transfers to the poor, knowing that these programs require higher taxes on most income-earners. Even in the United States, where such programs are relatively unpopular, there is substantial support for income transfers to the poor, even among rich and upwardly mobile people who will probably never be able (or have the need) to benefit directly from such transfers.

However, we cannot conclude from our recitation of the above facts that people are not selfish. What is probably true is something like Abraham Lincoln's assertion about people being fooled: being selfish is what some of the people are all of the time and what all of the people are some of the time. But the rest of the people, the rest of the time, are sometimes altruistic, sometimes reciprocal, sometimes spiteful, occasionally vengeful, and so on.

Human Nature and Cultural Differences

These experiments show that people differ within societies. This lesson is important because it reinforces what we know from our own observations, namely, that assuming everyone is selfish (or generous, or spiteful, for that matter) ignores the facts. But people's behavior also differs—sometimes strikingly—from one society to another. To some extent this is due to the particular requirements of making a living in each society—people in Kansas farm, people in Iceland fish. But does the extent to which people resemble economic man vary from one society to another?

One of us (Bowles) and a team of anthropologists and economists designed a set of experiments to explore the connections (if any) between how people make their living and their preferences. We conducted our experiments in fifteen societies in Africa, Asia, and Latin America where people live in sharply contrasting ways. In

some, hunters and gatherers live in ways not very different from the ways in which our early human ancestors lived before the domestication of animals and plants. In others, herders and farmers use technologies that have been in use thousands of years to make a living from their animals and plants. Most of the groups we studied live in inaccessible places such as the New Guinea highlands and the Peruvian part of the Amazon, and they have very limited connections with modern governments or the world of markets. None of the groups we studied was large; most were settlements of less than one hundred people and had had little contact with the outside world. For this reason, and because they have so little else in common, we call them "small-scale societies."[3]

The results of the experiments surprised us. When the ultimatum game was played among the Au and Gnau peoples in Papua, New Guinea, for example, offers of more than half of the pie were common. (Such offers were almost never encountered in experiments conducted with American students.) Even more interesting was the observation that in these societies high and low offers were rejected with equal frequency. (Most fifty-fifty offers were accepted.) This unusual result probably occurred because competitive gift-giving is a means of establishing status in these and many other New Guinea societies. We reasoned that proposers making high offers (offering more than half) may have been seeking to enhance their status, while those rejecting these offers were simply refusing (albeit at a high price) to accept a lower status. The frequent rejection of low offers was probably due to a sense that accepting them would mark the responders as already being of a low status.

In contrast, when the game was played among the Machiguenga forest agriculturalists in Amazonian Peru, the average offer was 27 percent of the pie. Nearly three-fourths of all the Machiguenga offers were less than 25 percent, and only one offer was rejected! This was a pattern strikingly different from the results of the other experiments we conducted. We were left wondering: if the Machiguenga were really so stingy (as suggested by the frequency of low offers and the infrequent rejection of them), why did they offer more than a penny?

Analysis of our experiments in the fifteen small-scale societies led us to the following conclusions. First, and most important, is that typical behaviors vary significantly from group to group. The subjects in some groups were much more generous (and willing to punish stinginess) than were American, European, and other students, and some were much less so. Second, in no group did we find "economic man" behavior to be typical. Third, variations in behavior from one group to another seemed to reflect differences in how people in each group make their living. For example, the Aché people in Paraguay acquire some kinds of food (meat and honey) by hunting and gathering, and these foods are shared equally among all group members. When playing the ultimatum game, almost all Aché proposers offered about half of the pie, and *none* of their offers was rejected. (This behavior differs, of course, from that of the Machiguenga and other groups such as American students.)

Another example of how a particular group's economic circumstances affect its typical behavior comes from Indonesia. There the Lamalera whale hunters need to hunt in large crews, and they divide their catch according to strict sharing rules, with each part customarily going to a particular individual or group—crew, sailmaker, blacksmith and boat builder, for example. When they played the ultimatum game, the average offer was 58 percent of the pie.

---

[3]Joe Henrich, Robert Boyd, Samuel Bowles, Ernst Fehr, and Herbert Gintis, *Foundations of Human Reciprocity: Economic Experiments and Ethnographic Evidence in 15 Small-Scale Societies* (Oxford: Oxford University Press, 2004).

Why do people play the ultimatum game so differently from one society to another? We know that the players in each society, from Orma herders in Kenya to Aché hunters in Paraguay, face the same constraints in the experiments. Therefore, the answer must be either that the players' beliefs vary or that their preferences differ. If it is their preferences, where do the differences come from?

Some of our preferences are influenced by our genes and are hence thought of as reflecting our "nature." Thus, we sometimes say that a certain person is generous or stingy (or something else) "by nature." To take another example, the taste for sweet or fatty foods seems universal and is probably genetically transmitted. But most food tastes vary greatly among countries, and genetic differences between human populations are not great enough to account for such variations. Spain, Italy, and France are famous for their distinctive national cuisines, but the crops grown in each of these countries can be grown in others, and there are no relevant genetic differences between their populations. Why do the Italians eat pasta while the French prefer bread or potatoes? These tastes are not inherited genetically. Rather, they are *learned* from parents, neighbors, and others.

Preferences that are learned from others—passed on from parents, elders, teachers, heroes, competitors, neighbors, or friends rather than being genetically transmitted just from parents—are part of what is called *culture.* We define culture as aspects of behavior that we learn from others.

As noted earlier, beliefs influence our behavior because our choices about what actions to take are based in part on our understandings (beliefs) regarding cause-and-effect relationships. Since our beliefs are either learned from others or gathered from our own experiences, they too (like preferences) are part of our culture. Similarly, learned skills, transmitted to us by parents, schools, friends, neighbors, and others, are part of the culture in which we are situated.

Culture and *human nature* have long been controversial terms: is it "nature" or "nurture" that explains why some people lead and others follow? Is it "genes" or "environment" that make some rich and others poor?

What is not controversial is that people do not differ much genetically from group to group. *Within* any group, of course—whether it is citizens of the United States, the Aché people of Paraguay, or Italians—the genetic differences are very large. Researchers have found, however, that such within-group differences are much greater than the differences between a typical person in a certain group and one in some other group. For example, if you were to pick two Americans at random, even ones who share the same skin color or height, the genetic differences between them would most likely be huge in comparison to the differences, say, between the average American and the average Aché.

Behaviors, however, are another matter. People behave very differently in different societies. (Recall the varying results of the ultimatum game played in different societies.) Behaviors differ mainly because cultures are very different from one country or group to another: what we learn from others as we grow up and even when we are adults varies greatly from place to place. That is why the preferences and beliefs of the peoples studied by the authors differ so much. The culture of Lamalera whale hunters (who on average offered more than half of the pie) is different from that of the Machiguenga forest agriculturalists (who offered barely more than a quarter), which in turn is different from that of the Tanzanian Hadza hunter-gatherers (who rejected almost half of low offers)—and all of these cultures are different from that of the Ecuadorian forest people, the Tsimane (who rejected none). Why do people learn such different things from one culture to another?

The Economy Produces People

As we will discuss more in Chapter 4, the economy produces more than just goods and services; it also produces people. We call the creation of goods and services "production" and the production of people "reproduction" (see Figure 4.1). The term *reproduction* comprises not only biological procreation but also all the processes entailed in the formation of an individual, including what happens in families, schools, and all the other institutions in which parents, teachers, caregivers, spouses, and others combine their labor with other inputs to raise and support each new generation. Societies accomplish the tasks of reproduction in various ways, and some of the differences in methods of reproduction result from differences in the way people make their livings (production). This is what we mean when we say that the economy produces people.

We've seen that ultimatum game experiments can provide information about the relationship between an economy and a culture. In our small-scale society experiments individual choices appeared to reflect everyday life, especially the way people made their living. For example, we saw that the Aché, who acquire much of their food by hunting and gathering and then share it, tended to divide the pie equally, sometimes offering more than half of it to the responders in our experiments.

Similar information can be gathered using other types of experiments. Among the Orma herders and throughout Kenya there is an important cultural institution they call the *harambee system*. With this system it is customary to collect money to build a school or repair a road by assessing each herder a certain amount, expecting him to make a contribution that varies with the size of his herd. We asked the Orma to play a different game called the "public goods game."

A public goods game is also explained to each player beforehand, and, like the ultimatum game, it is played anonymously and for real money. But in this game individuals play in groups rather than with a single partner, and they are asked to contribute to a common pot for the benefit of all. Once all the contributions are made, the amount in the pot is doubled and the total is then distributed in equal amounts to all the players. In this game, each player benefits from *the others* contributing, but each would personally gain the most by contributing nothing.

In the public goods game an Orma herder with a large herd who contributed one Kenyan shilling—quite a lot of money among his people—would see his contribution to the common pot doubled and then divided and distributed equally among the players in the game. Suppose there are five players. Then, the share distributed to the herder as a result of his contribution would be, say, two-fifths of a shilling, less than his original one shilling, so he would have been better off remaining on the sidelines and just holding on to his shilling. Despite the fact that self-interest would prescribe contributing nothing, however, the herders in fact contributed generously—and those with large herds contributed more than those with smaller ones.

We wondered if the similarity between the local customs and the experimental play of groups such as the Aché hunters in Paraguay and the Orma herders in Kenya comes about because preferences are affected by a particular group's social institutions and norms of fairness. The large differences in institutions and norms in our sample allowed us to address this question. Accordingly, we ranked each society with reference to two aspects of its economic institutions and then sought to use the rankings to predict the results we achieved in the ultimatum games.



Fifty-Fifty: The Importance of Norms

On January 11, 1886, Fenner Powell, a former slave in Wade County, North Carolina, placed his X next to the signature of his landlord, W. S. Mial. Powell thereby agreed to "do all manner of work . . . as directed, and to be respectful in manners and comportment to said Mial . . . and to give to said Mial one half of all crops raised." Remarkably, this was exactly the same crop share that free-born white farmers in Wade County and throughout the South agreed to pay their landlords. Why would a former slave—illiterate, excluded from voting, and subject to denigration and lynching—be allowed to keep the same share as free-born farmers with greater social status and bargaining power? The answer lies far from the post–Civil War U.S. South.

In Illinois today, growing corn is big business. Using capital-intensive technologies and computer-generated business plans, some farmers cultivate over 1,000 acres, much of it on plots rented from multiple owners. In the mid-1990s over half the contracts between farmers and owners were sharecropping agreements, and of these, over four-fifths divided the crop fifty-fifty. In southern Illinois, where the soil is less fertile, in some counties most contracts give the tenant two-thirds of the crop and the owner one-third, although land quality varies widely.

Rice cultivation in West Bengal in the mid-1970s seems light years away from Illinois. In West Bengal, poor illiterate farmers eked out a bare living on plots averaging just two acres; their villages lacked modern communication, and were isolated by impassable roads much of the year. Yet, as in Illinois, crops were divided fifty-fifty between sharecroppers and owners in more than two-thirds of the contracts. Ibn Battuta, the famous Arab geographer who visited Bengal, India, in 1347, had recorded exactly the same division of the crop six centuries earlier. (Of course, if each landlord had twenty sharecroppers, fifty-fifty crop sharing would not give owners and farmers equal incomes: each landlord's income would be twenty times that of the typical farmer.)

John Stuart Mill, the nineteenth-century English philosopher and economist, noted both the widespread pattern of equal sharing of crops and local conformity to other proportions where fifty-fifty was not the rule. Mill's explanation: "The custom of the country is the universal rule." But why did the Bengalis and the Americans come up with the same proportion? And, more puzzling: why do owners offer a 1:1 or a 2:1 split, when they could make huge profits if they only offered lower shares on more fertile land? And when the shares do change, as in West Bengal in the 1980s and 1990s, why do they all change at once?

Fifty-fifty crop shares are social *norms*—widespread practices that are followed because violating them would bring criticism, retaliation, or ostracism. Norms play a key role in all economies, limiting the extent to which people can simply pursue their self-interest. Often norms become values in their own right, adhered to not to avoid sanction but because people would not feel right doing otherwise.

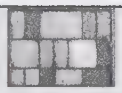
Sources: Samuel Bowles, *Microeconomics: Behavior, Institutions, and Evolution* (Princeton: Princeton University Press, 2004), Ch. 3; Roger L. Ransom and Richard Sutch, *One Kind of Freedom: The Economic Consequences of Emancipation* (Cambridge: Cambridge University Press, 1977); Peyton Young and Mary Burke, "Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture," *American Economic Review* 91, no. 3 (2001): 559–73; Pranab Bardhan, *Land, Labor and Rural Poverty: Essays in Development Economics* (New York: Columbia University Press, 1984); John Stuart Mill, *Principles of Political Economy with Some of Their Applications* (London: Longmass, Green, Reader, and Diver, 1867 [1848]).

The first basis for ranking, *cooperation*, is a measure of the extent to which the local ecology allows for a more productive use of labor when many work together. The Lamalera whale hunters were ranked first because successful whaling requires large numbers of hunters to work together, and the dispersed Machiguenga forest agriculturalists were ranked last because their production is more individualized and they gain little by collective production activities. We speculated that in groups

that cannot benefit much from cooperative production there would be few norms of sharing. We also guessed that in groups such as the Lamalera, whose livelihood depends on large-scale cooperation, ways of sharing would be well developed, and these would affect how the Lamalera played our games.

The second basis for ranking, *market integration*, is a measure of the fraction of a people's livelihood that is acquired through market exchange. The rationale for this measure is as follows: the more frequently people experience market transactions, the more they will also experience beneficial sharing of the gains made possible by trading with strangers. Historically, it is a fact that before markets became widespread, most interactions with strangers were potentially dangerous, often providing occasions for violent confrontation, theft, or worse. As markets developed they habituated us to the benefits of regular interactions with strangers in which both parties can benefit as long as they follow certain rules (you pay at the checkout, you do not take the groceries and run). Our speculation was that such experiences would give rise to societal sharing norms and that these would be reflected in the results of the experimental games.

Using the measures of cooperation and market integration we sought to explain both a group's average ultimatum game offer and its frequency of low offer rejection. We found that the two measures—cooperation and market integration—enabled us to predict the results of ultimatum game play in most of our societies. In societies with more cooperation or greater market integration, proposers made higher offers (on average), and low offers were more likely to be rejected. The great eighteenth- and nineteenth-century thinkers Karl Marx, Edmund Burke, and Alexis de Tocqueville might be surprised to find that exposure to markets leads to higher offers



Does a Constitution for Knaves Make Knaves of Us?

Unlike Adam Smith and David Hume, the English conservative Edmund Burke, the German revolutionary Karl Marx, and the French liberal Alexis de Tocqueville feared that harnessing self-interest, or (to use Hume's phrase) living under a constitution designed for knaves, would turn us into knaves.

> *[T]he age of chivalry is gone. That of sophisters, economists, and calculators has succeeded . . . Nothing is left which engages the affections . . . so as to create in us love, veneration, admiration or attachment.*
>
> —Edmund Burke, *Reflections on the Revolution in France* (1790)

> *Finally, there came a time when everything that men had considered as inalienable became an object of exchange, of traffic, and could be alienated. This is the time when the very things which till then had been communicated, but never exchanged; given, but never sold; acquired, but never bought—virtue, love, conviction, knowledge, conscience, etc.—when everything, in short, passed into commerce. It is the time of general corruption, of universal venality.*
>
> —Karl Marx, *The Poverty of Philosophy* (1847)

> *Each* [person] . . . *is a stranger to the fate of all the rest . . . his children and his private friends constitute to him the whole of mankind; as for the rest of his fellow citizens, he is close to them but he sees them not . . . he touches them but he feels them not; he exists but in himself and for himself alone.*
>
> —Alexis de Tocqueville, *Democracy in America* (1830)

(greater sharing with others) and a greater tendency to reject unfair offers. (See the box, "Does a Constitution for Knaves Make Knaves of Us?")

Our ability to predict behavior in experiments that were entirely novel situations for our subjects on the basis of our two measures of economic structure suggests that economic institutions influence preferences. Our values, likes, dislikes, and morals seem to be affected by living within a particular set of institutions—sharing food like the Aché, cooperating in acquiring food like the Lamalera, pitching in voluntarily to build a school like the Orma, or, for that matter, competing for a job following graduation. How does this come about?

A plausible answer is that people acquire their preferences in part through the way they are brought up, and child-rearing practices stress values and skills that are important in a society's way of life. To test this idea, three anthropologists categorized seventy-nine mostly nonliterate societies (similar to our fifteen "small-scale" societies) according to their prevalent form of livelihood (animal husbandry, agriculture, hunting, and fishing) and their related capacity for food storage or other wealth accumulation. Food storage is common in agricultural societies but not among hunters and gatherers. These researchers also collected evidence on forms of child-rearing, including obedience training ("compliance") and the degree to which self-reliance, independence, and taking responsibility ("assertion") were encouraged. They found significant variations of child-rearing practices, and they also found that these variations were correlated with differences in economic structure. They concluded, "Knowledge of the economy alone would enable one to predict with considerable accuracy whether a society's socialization pressures were primarily toward compliance or assertion."[4]

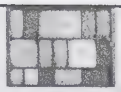
The box "Property Rights among the Arapesh of New Guinea" offers a fine example of how cooperation is embedded in the norms of a culture and particularly the idea of how property rights should be assigned—that is, who "owns" what. A child growing up in such a society learns that producing something does not necessarily imply owning it. It is hard to imagine a society more different from the one we live in.

We do not need to confine our attention to anthropological studies of exotic societies to find evidence that economic institutions influence preferences. Over a period of three decades, a social psychologist, Melvin Kohn, and his collaborators have studied a number of individuals, focusing on the relationship between these individuals' positions in the authority structure of their workplaces and the extent to which they value obedience and discipline or self-direction and independence both in themselves and in their children. The hypothesis was that people who routinely take orders on the job value obedience and discipline, while those who give orders value autonomy. Kohn's collaborative study of Japan, the United States, and Poland (when it was still ruled by a Communist government) found that in all three countries people who exercised self-direction on the job also valued self-direction in other realms of their life (including child-rearing and leisure activities) and were less likely to exhibit fatalism, distrust, and self-deprecation. Kohn and his coauthors argued that "social structure affects individual psychological functioning mainly by affecting the conditions of people's own lives,"[5] and they concluded that "the

---

[4] Herbert Barry III, Irvin L. Child, and Margaret K. Bacon, "Relation of Child Training to Subsistence Economy," *American Anthropologist* 61 (1959): 51–63.

[5] Melvin L. Kohn, *Class and Conformity: A Study in Values* (Homewood, IL: Dorsey Press, 1969), 189.



Property Rights Among the Arapesh of New Guinea

A typical Arapesh man [lives] *for at least part of the time . . . on land which does not belong to him. Around the house door are pigs which his wife is feeding but which belong either to one of her relatives or to one of his. Beside the house are coconut and betel palms which belong to still other people and the fruit of which he will never touch without the permission of the owner. . . . He hunts on bush land belonging to a brother-in-law or a cousin at least part of* [his] *hunting time, and the rest of the time he is joined by others on his bush, if he has some. He works his sago in others' sago clumps as well as in his own.*

Of the personal property in his house, that which is of any permanent value, like large pots, well carved plates, [and] *good spears, has already been assigned to his sons, even though they are only toddling children. His own pig or pigs are far away in other hamlets; his palm trees are scattered three miles in one direction, two in another; his sago palms are still further scattered; and his garden patches lie here and there, mostly on the lands of others.*

If there is meat on his smoking rack over the fire, it is either meat which was killed by another—a brother, a brother-in-law, a sister's son, etc.—and has been given to him, in which case he and his family may eat it; or it is meat which he himself killed and which he is smoking to give away to someone else, for to eat one's own kill, even though it be only a small bird, is a crime to which only the morally—which usually means in Arapesh mentally—deficient would stoop.

If the house in which he is living is nominally his, it will have been constructed in part at least from the posts and planks of other people's houses, which have been dismantled or temporarily deserted, and from which he has borrowed timber. He will not cut his rafters to fit his house, if they are too long, because they may be needed later for someone else's house which is of a different shape or size.

This then is the picture of a man's ordinary economic affiliations, crosscutting every defined line of geography and blood kinship, [and these affiliations are] based on personal ties between individuals which serve to tangle the members of each group into many other groups and to blur every possible distinction between groups.

Source: Margaret Mead, *Cooperation and Competition Among Primitive Peoples* (Boston: Beacon Press, 1961), 31–32. Reprinted with permission from the Institute for Intercultural Studies, Inc., New York, NY.

experience of occupational self-direction has a profound effect on people's values, orientation, and cognitive functioning."[6]

The facts presented above suggest that the way goods are produced and distributed in any society conditions what one must be or do to make a living. Hunters must be independent-minded and physically fit, industrial workers and clerical staff must be willing to take orders, and entrepreneurs must be self-motivated. Economic institutions thus impose characteristic patterns of interaction on the people who make up a society, affecting who meets whom, on what terms, to perform which tasks, and with what expectations of rewards. These patterns, in turn, influence the process by which people mature and change over their lifetimes, forming their personalities, habits, tastes, identities, and values—in short, their preferences.

[6]Melvin L. Kohn and Kazimierz M. Slomcy´nski, *Social Structure and Self-direction: A Comparative Analysis of the United States and Poland* (Cambridge, MA: B. Blackwell, 1990), 967.

Economic institutions shape people's preferences in part because institutions determine what kinds of individuals will be successful, and people try to copy the successful, either in their own likes, dislikes, and values or in raising their children. But in most societies the job of socializing young people is not left entirely to parents. Schools, religious institutions, and other organizations play a major part in bringing up the next generation.

If you look at the curriculum of a school, you might get the impression that its only objective is to teach skills such as reading, writing, math, and the ability to use a computer. But a closer look at what goes on in the classroom and how rewards are distributed among students shows that schools do something else, too: they teach children how to behave. The fact is that getting a good grade requires more than knowledge of the subject, and there is a study that proves this. It shows that to get good grades one also has to develop certain personality traits. (Many students already know this.) More surprising, however, is the study's additional finding that the personality traits rewarded with high grades in the classroom are the same as those rewarded with favorable rankings by supervisors in the workplace.

Here is the study. One of us (Edwards) used a peer-rated set of personality measures to predict supervisors' ratings of workers in both private and public employment. Peer-rated personality measures are based on how individuals are seen by people similar to, or in the same situation as, themselves. Such measures are expressed in words such as *tactful*, *creative*, and *punctual*. One of Edwards's collaborators, Peter B. Meyer, used the same peer-rated personality variables to predict differences in high school students' grade point averages from what would be predicted on the basis of their SAT and IQ scores.

Edwards found that certain peer-defined personality traits—perseverance, dependability, consistency, punctuality, tactfulness, and being able to "identify with work" and "empathize with others"—were highly correlated with positive supervisors' ratings, whereas to be judged by one's peers as being creative or independent meant receiving poor ratings from supervisors.[7] Meyer found virtually identical results for the high school students in his grading study: the correlations between their grade point averages and twelve personality traits are nearly the same as the correlations observed in Edwards's study of employees.[8] Thus, both teachers and employers reward the same personality traits. The conclusion: schools teach more than skills, and they also cultivate (or at least they reward) the kinds of personality traits that employers prefer.

All human societies have developed elaborate ways of teaching the preferences and beliefs required for normal functioning as an adult. In many hunter-gatherer societies children accompanied their parents as they stalked game and searched for fruits and nuts, learning the skills necessary to live by these means. Before the emergence of capitalism, most production took place within families—in small workshops, on farms, and the like—and a person could learn most of what was necessary to function in the economy from parents and relatives. Also, the skills required did not change much from generation to generation.

Capitalism changed all this. It created huge workplaces in which thousands of strangers come into contact with one another, and rapid technological change may

---

[7] Richard C. Edwards, "Personal Traits and 'Success' in Schooling and Work," *Educational and Psychological Measurement* (Spring 1977) and "Individual Traits and Organizational Incentives: What Makes a 'Good' Worker?" *Journal of Human Resources* (Winter 1976).

[8] Samuel Bowles, Herbert Gintis, and Peter Meyer, "The Long Shadow of Work: Education, the Family, and the Reproduction of the Social Division of Labor," *The Insurgent Sociologist* (Summer 1975).

now render the skills of one's parents obsolete even before retirement becomes an option for them. As capitalism has become the preeminent economic system, schools have come to play an essential role in the socialization process. Moreover, the personality traits that schools foster—dependability, consistency, punctuality—now make it possible for large numbers of strangers to work together even if the bonds of kinship, loyalty, and affection are absent.

The Cooperative Species

Humans are unique among animals in that large numbers of unrelated people cooperate to produce the goods and services we require. We also cooperate in pursuing other projects such as raising the next generation and engaging in warfare.

All animals compete: for food, for survival, for reproductive success. Some animals exchange goods and services. For example, fish called "cleaner fish" remove parasites from the skin and mouths of larger fish, providing health services in return for a good meal. The Greek scholar Herodotus described a similar exchange more than two and a half millennia ago:

> Because [the crocodile] spends its life in water, its mouth is filled with leeches. With the exception of the sandpiper, all other birds and animals run away from it. The sandpiper, however, is on good terms with it because it [the sandpiper] is of use to the crocodile. When the crocodile climbs out of the water and onto land, it yawns widely . . . and then the sandpiper slips into its mouth and swallows the leeches. This does the crocodile good and gives it pleasure, so it does not harm the sandpiper.[9]

Some animals even respect property rights. Spiders do not intrude onto the webs occupied by other spiders (unless the intruder is much bigger). A male Hamadryas baboon does not attempt to steal food that is in the possession of another one.

But in no other species but *Homo sapiens* do thousands of unrelated individuals work together to accomplish a common project, whether it be building cars, providing medical insurance for citizens, or making war. (Ants, bees, and some other so-called eusocial insects cooperate on a grand scale, but it is all in the family: the members of a hive or nest, even if they number in the thousands, are mostly relatives.) How do we do it?

In part, these feats of cooperation are accomplished because, distinct from other animals, we are able to devise laws and organizations that go beyond the family, such as governments and firms. These often provide the incentives and constraints that induce people to work together effectively, even if they are entirely self-interested.

Self-interested behavior is not characteristic of successful organizations, however. The soldier who goes to war may do it for the money or because he was drafted and had no choice. But as any officer knows, such motives do not inspire people to become good soldiers.

All forms of human cooperation, including those capable of winning in warfare, are best understood by considering motives other than self-interest. These include

---

[9]Herodotus, *The Histories* (New York: Oxford University Press, 1998), 122.

envy or spite toward others as well as concern for others and the aspiration to see certain principles upheld. The success of humans as cooperators is much less puzzling once one realizes that economic man is just one kind of human, and not a very common one at that. Far more common are people who, at least some of the time, are reciprocators or altruists and, for better or worse, care about others. Humans are cooperative on a scale unmatched by any other animal because we have preferences that lead us to act in cooperative ways.

Conclusion

Our last question: how did we get this way? Part of the answer concerns human nature. We have the intellectual capacity to devise general ethical rules to live by, and we also have the linguistic capacity to communicate these rules among ourselves, to report violations of the rules, and to coordinate the punishment of those who break them. In addition, we are acutely sensitive to praise and blame, experiencing such *moral emotions* as shame, which serve as powerful incentives to avoid wrongdoing. It is worth noting that Adam Smith titled his first book *The Theory of Moral Sentiments* and devoted it to the analysis of exactly this aspect of human life. The moral emotions as well as the intellectual and linguistic capacity to devise and enforce social norms are part of human nature. They are not part of cat nature, or spider nature, or baboon nature.

The content of ethical rules—just what it is that they bid us to do and not do—is also, to some extent, a matter of human nature. Incest evokes disgust and shame among most humans, irrespective of the culture they grow up in, as do a number of unhealthy practices such as living with personal uncleanliness or ingesting unhealthy substances. But most of the content of our ethical norms is learned: it comes from culture, not nature.

People in most societies actively teach the value of curbing our selfish desires and behaving in altruistic or reciprocal ways under appropriate circumstances. For most people (but not, of course, for sociopaths), acting in accordance with such teachings becomes an objective that is embedded in our preferences, becoming thus more than just an external constraint. That is why most of us, most of the time, do not steal even when we could get away with it.

But what about those who, like *Homo economicus*, are clever enough and immoral enough to steal when they *can* get away with it? Why do they not succeed in taking advantage of and eventually outcompeting their more ethical neighbors for the goods necessary for survival? If this happened, would not the ethical ones have to respond by becoming like the immoral ones? We hear of cases of unethical behavior being rewarded and going unpunished all the time.

The answer is that a group made up of economic man types would not function successfully as a unit. Who would come to the defense of the group in an attack by an external enemy or help out during a drought or other ecological crisis? Not economic man. In his second great book, *The Descent of Man* (the first was his better-known *The Origin of Species*), Charles Darwin, the founder of the modern theory of biological evolution, came to the following conclusion:

> When two tribes of primeval man, living in the same country, came into competition, if . . . one tribe included a great number of courageous, sympathetic and

> faithful members, who were always ready to warn each other of danger [and] to aid and defend each other, this tribe would succeed better and conquer the other. . . . Selfish and contentious people will not cohere, and without coherence nothing can be effected.[10]

The point of Darwin's statement is clear: in competitions among groups, those whose members have learned how to cooperate—that is, *not to compete* with one another—often win. Think of team sports. Darwin spoke of tribes as groups that would benefit from having a preponderance of cooperative members. The same reasoning applies to firms, neighborhoods, ethnic groups, and nations.

Thus, it is not that our good cultures beat out our bad nature. Rather, our cultures and our nature work together to make *Homo sapiens* the uniquely cooperative species that we are. The fact that we are cooperative means that nice guys do not always finish last. The reasoning that explains why nice guys do not always finish last also makes clear that neoclassical economists sometimes overrate the value of competition as a source of progress. Cooperation is also necessary.

Suggested Readings

Samuel Bowles and Herbert Gintis, *A Cooperative Species: Human Reciprocity and Its Evolution* (Princeton, NJ: Princeton University Press, Reprint edition, paperback, 2013); http://www.jstor.org/stable/j.ctt7s72v.

Samuel Bowles, *Microeconomics: Behavior, Institutions, and Evolution* (Princeton: Princeton University Press, 2004).

Samuel Bowles and Herbert Gintis, *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life* (New York: Basic Books, 1976).

Ananish Chaudhuri, *Experiments in Economics: Playing Fair with Money* (New York: Routledge, 2009).

Robert H. Frank, *Passions within Reason: The Strategic Role of the Emotions* (New York: Norton, 1988).

Bruno S. Frey, *Not Just for the Money: An Economic Theory of Personal Motivation* (Cheltenham, UK: Edward Elgar Publishing, 1997).

Herbert Gintis, Samuel Bowles, Robert Boyd, and Ernst Fehr, eds., *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (Cambridge, MA: Massachusetts Institute of Technology Press, 2004).

Uri Gneezy and Aldo Rustichini, "A Fine Is a Price," *Journal of Legal Studies* 29, no. 1 (2000) 1–17.

Joseph Henrich, Robert Boyd, Samuel Bowles, Ernst Fehr, and Herbert Gintis, *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford: Oxford University Press, 2004).

---

[10] Charles Darwin, *The Descent of Man* (Amherst, NY: Prometheus Books, 1997 [1871]), Ch. 5, "On the Development of the Intellectual and Moral Faculties During Primeval and Civilized Times."

Joseph Henrich, *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter* (Princeton: Princeton University Press, 2015).

Albert O. Hirschman, *The Passions and the Interests: Political Arguments for Capitalism Before Its Triumph* (Princeton: Princeton University Press, 1977).

Daniel Kahneman and Amos Tversky, *Choices, Values and Frames* (Princeton: Princeton University Press, 2000).

Robert E. Lane, *The Market Experience* (Cambridge: Cambridge University Press, 1991).

Peter Richerson and Robert Boyd, *Not By Genes Alone* (Chicago: University of Chicago Press, 2004).

William F. Whyte, *Money and Motivation* (New York: Harper & Row, 1955).

A Three-Dimensional Approach to Economics



Understanding capitalism has become essential for anyone who reads the newspaper, watches the nightly news, listens to political candidates, or simply wonders why it is so hard to find a good job or get enough free time. Since capitalism is an economic system, understanding capitalism requires some knowledge of economics. But what *kind* of economics? We call the approach presented in this book *three-dimensional economics*, and we often refer to it simply as *political economy*.

Until the beginning of the twentieth century the term *political economy* was used to refer to all of economics, and the field itself encompassed most of what is now divided up among the social sciences: anthropology, sociology, psychology, and political science, as well as economics. But around 1900 the term **political economy** was replaced by *economics*, and the boundaries of the discipline were narrowed: the study of markets became its primary focus. Inquiries into politics, psychology, history, and other aspects of society were then left to the other social sciences. Thus, it is no accident that anthropology, sociology, psychology, and political science came into existence as separate fields at about this time.

Political economy is a term we use for a theory that analyzes capitalism in terms of the three dimensions of competition, command, and change.

We prefer to use the older term *political economy* (rather than *economics*) to describe our approach because one cannot understand contemporary societies very well unless politics, economics, psychology, and the other social science disciplines are all brought together to study the complexities of modern life. Another way of describing the *political economy* approach, then, is to say that it is *interdisciplinary*.

Many people believe that the approach presented here makes more sense and is a more useful way of understanding our economy than what is sometimes called *neoclassical economics*—the "conventional" approach that is set forth in most economics textbooks. Ultimately, however, everyone who seeks to understand capitalism needs to consider a number of approaches to economics and decide which one, or which combination, makes the most sense and is the most useful.

What is certain is that no one should accept a particular approach to economics just because it is what some experts believe. Each person must make a choice for herself or himself, asking: "Does this make sense to me?" "Does this help me understand things that I have experienced and believe to be true?" Of course, not all opinions about economics are equally valid. A useful approach to economics,

whether it be political economy or some other approach, must be logically sound, internally consistent, and helpful in explaining what is known about economic reality ("the facts").

The main idea of this chapter is that *political economy considers all three dimensions of economic life: competition, command, and change.* The main points of the chapter are:

1. Capitalism is an *economic* system.
2. Every economic system can be analyzed in terms of three dimensions: a horizontal dimension (*competition*), a vertical dimension (*command*), and a time dimension (*change*).
3. Economics is about *values* (what ought to be) as well as *facts* (what is). It is useful to make values—and their role in any particular economic analysis—explicit. The values adopted in this book are simple: an economy should provide all members of a society with an equal opportunity to lead a flourishing life, and this objective is more likely to be achieved if the economy is *efficient*, *fair*, and *democratic*.

The fundamental concepts introduced in this chapter will be used throughout the book and will be further developed in later chapters.

Economic Systems and Capitalism

Over the span of human history people have organized their economic activities in many different ways. The variety of economic systems runs the gamut from tribal commonwealths, slavery, and feudalism to self-sufficient households, capitalism, and state socialism.

An **economic system** is a set of relationships among people that organizes the labor processes all societies need to sustain life.

What all these systems have in common—what makes them **economic systems**—is that they are all *ways of organizing the human labor needed in every society to produce the goods and services that support life.* No matter what type of society it is situated in, an economic system will determine *what* work is done, *how* it is done and *by whom*, and *to whom the resulting products are distributed.*

Economic systems are *relationships among people.* These relationships may be direct, face-to-face relationships, such as the interaction that occurs when you buy an orange from a grocer. At the other end of the spectrum, they may be relationships between people in different parts of the globe, for example the Iowa farmer producing grain that will later appear as bread on a table in Egypt. The relationships that make up economic systems may also be embodied in customs, laws, constitutions, political parties, or business corporations.

Economic relationships are shaped in important ways by the physical things and technologies used in production as well as by other factors such as geography, customs, religion, and whether production is primarily agricultural, industrial, or postindustrial (knowledge-based). Nevertheless, the *social relationships* among the various economic actors—producers and consumers, slave owners and slaves, feudal lords and serfs, employers and employees, borrowers and lenders—are the defining qualities of an economic system. Thus, the distinguishing features of any economic system may be seen in the social interactions among the economic players These

interactions may, in varying degrees, be cooperative or competitive, altruistic or avaricious, equal or unequal, democratic or authoritarian.

How human work is organized differs from one economic system to another. To understand how work is organized in any particular society one must examine its economic system. The economic system that is the focus of this book, capitalism, is the one that prevails, in one way or another, throughout most of the world today.

Capitalism is quite familiar to most of us. In various forms it is the economic system not only of the United States but also of Great Britain, Japan, Chile, France, Russia, Mexico, Brazil, Germany, South Africa, and indeed, more than one hundred other countries. Thus, we can study capitalism directly because we experience it every day of our lives.

In a capitalist economic system most goods and services are produced at the direction of employers (businesspeople, entrepreneurs, capitalists, or managers of firms) who seek to make profits by selling the produced goods and services in markets. Most people in capitalist economies work for someone else (their employers) and receive a wage or salary in return. So work is organized for the purpose of making a profit; the employer, or his or her appointed manager, is the boss at the workplace; and goods, services, and people's capacity to work—their labor time—are all exchanged through markets.

To understand capitalism one must find answers to a number of questions. How is work organized? How do markets operate? How much of the output when sold will go to profits, how much to wages, and what will determine the relative magnitudes of these two types of income? Why do some workers get paid more than others? Who decides what technologies will be used, and on what basis will such decisions be made? How does our capitalist economy affect the way we develop as human beings? How does it condition our culture, influence our political system, and alter our natural environment? And, in turn, what reverberations will all these effects have on the capitalist economy itself?

Three-Dimensional Economics

In this book the complex relationships of a capitalist economy are examined, taking into account all three dimensions of an economic system: competition, command, and change.

Competition

Competition, or the horizontal dimension in economics, refers to aspects of economic relationships in which voluntary exchange and choice among a large number of possible buyers and sellers play the predominant role.

The first dimension is called **competition**, and it refers to that aspect of an economic system in which exchanges of one sort or another play the most important part. In capitalism, of course, competition and exchange occur primarily in markets. For example, when a motorist chooses to buy gasoline at a particular gas station, it is obvious that he or she is making a choice between competing suppliers.

The competition dimension of the economy is a *horizontal* one: it can be thought of as involving a relative equality of power among those offering the choices, engaging in exchanges, and competing with one another. In the gas station example, for instance, the sellers of gasoline must compete with one another to entice the motorist to come to their pumps. In the contest for customers, the gas suppliers are equal in the following sense: none can dictate to any other—or to the motorist—where the motorist will buy gasoline.

Political economy shares with conventional economics the view that an analysis of how competition works is essential to any attempt at understanding the economy. (As we point out in Chapter 11, however, political economy differs from the conventional approach on how competition actually takes place in a capitalist economy.) When, as in capitalism, much of economic life is organized in markets—not based, say, on ancient customs or on decisions imposed by central planners—markets are the terrain on which most of the competition is carried on and where most of the choices are made by individuals and firms.

Command

Command, or the vertical dimension in economics, refers to aspects of economic relationships in which power plays the predominant role.

The second dimension is called **command**, and it refers to those aspects of economic relationships that involve power, coercion, hierarchy, subordination, or authority. In capitalist (and many other) societies, command is a central aspect of the workplace, the household, and the government. It concerns relations among nations, classes, races, men, women, and other groups in society as well.

The command dimension is regarded as *vertical* because it necessarily involves people or groups who are unequal, some being "higher up" in a hierarchy than others. One person or group is "dominant," while the others are "subordinate."

It is not always easy to distinguish command from choice. Suppose a thief points a gun at someone and demands, "Your money or your life!" This is *literally* a choice. The victim could presumably choose to surrender either the money or his or her life, yet the thief's threat is easily recognizable as *in fact* a command to hand over the money.

A less extreme example would be a situation in which a boss asks a worker to do something and the alternative to carrying out the boss's request is to get fired. It may seem here that the worker has the choice of doing or not doing what the boss has asked, but, in reality, the boss's request is a command.

Although the prospect of getting fired is not comparable to the possibility of losing one's life, for many people the loss of a job will cause financial disaster. This, in turn, will make it difficult to put food on the table, pay the rent or mortgage payment, or land another job.

A has **power** over B if by imposing costs on B (or threatening to do so) A can cause B to act in a way that is to A's advantage.

One form of command is what we call **power**. We define power as follows. A has power over B if by imposing costs on B (or threatening to do so) A can cause B to act in a way that is to A's advantage. The employer has power over the employee in this sense.

Command may be exercised, however, without threats and costs being imposed. It may just be that one party influences or shapes the conditions under which another party will be making a choice. Thus, corporations often use their financial resources to alter the conditions under which consumers make choices. If, for example, an aspirin producer's television commercials can convince consumers that its product is "stronger," works "faster," or is "recommended by most doctors," people going to the store in search of relief for headaches will tend to choose this product more frequently.

The point of the aspirin commercial example is that command is not only the ability to impose costs on others. It is also the ability of one person or group to control others' information, playing upon their fears, hopes, insecurities, or other emotions and thereby influencing their actions in order to promote the interests of the powerful person or group. Thus, in the case of advertising, command is often used in subtle—and sometimes not so subtle—ways to shape or condition choice.

Command is used in many other ways to influence outcomes. One example of such influence is corporations making campaign contributions to sway the voting patterns of legislators or other politicians. Another is companies hiring recent immigrants to work long hours for low wages in unsafe or unhealthy workplaces because such people have few other employment opportunities. A third is a large meatpacking firm that dominates sales of meat to supermarkets and other retail outlets, giving it the power to write contracts with small chicken farmers that leave many of these farmers in poverty. In short, command refers to circumstances in which money and power give one person or organization the ability to influence the actions of other organizations or people. As we observe in later chapters, many economic relationships involve both choice and command, with neither one operating exclusively.

Change

Change, or the time dimension in economics, refers to the historical evolution of people and economic systems.

The third dimension of economic systems is called **change**. It concerns the passage of time and the ways in which, over time, the operation of an economic system will change the system itself. In capitalism, change occurs because big profits can be made by changing the existing conditions—by building new and better machines, by designing novel products to meet previously unknown needs, by expanding production, building factories in distant corners of the world, or succeeding in changing the rules of the game.

Central to the changefulness of capitalism is the system of investment for profit that creates an inexorable tendency for the economy to expand. With its expansionary drive, capitalism alters the conditions in which it operates—and within which people live and die. Capitalism's continuous expansion also transforms the ways in which the system itself actually works.

Change is called the "time" dimension because change always occurs through time. Thinking about change necessarily involves using concepts such as "before" and "after," "old" and "new," or "early" and "late."

Emphasizing the change dimension of economics reminds us that each economic system works differently at different points in time. It also brings out the fact that people participating in an economic system change over time.

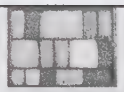
Each economic system has a history, and the way each system functions at any particular point in time will depend, in part, on its history. For example, American capitalism works differently in the twenty-first century from the way it worked in the nineteenth and twentieth centuries, and although it is still capitalism, an analysis of how it works in the present must be informed by a consideration of the ways in which it has changed. Similarly, American capitalism today differs from, say, German capitalism, in part because the nineteenth- and twentieth-century histories of these two countries differed.

Any economic system will undergo further change in the future. The present and even the future will become the past: what *is* today will *not be* tomorrow. Indeed, as noted in Chapter 1, capitalism is the most dynamic—or "changeful"—economic system yet to be observed in history.

Of course, many factors other than the normal, everyday functioning of the economic system may cause economic and social changes. A list of other kinds of change-inducing events might include wars, plagues, new discoveries, climate change, religious upheavals, or new scientific breakthroughs. In some cases the

economic analysis presented in this book may help to explain why these events occur. However, this analysis is mainly concerned with change as it emerges from the routine and persistent operation of the economic system. Similarly, with regard to human development the concern here is with how people adopt new tastes, values, ways of life, and even religions at least partially in response to their changing experiences of working and making a living.

The emphasis it gives to change is one of the qualities that distinguish political economy from neoclassical economics. This and other differences are summarized in the table at the end of this chapter.



Economics, Politics, and History

Economics is the study of *how people interact with one another, with nature, and with the other things they require in order to produce their livelihoods.* The three-dimensional approach to economics takes the view that, as important as they are, the processes of competition and market exchange on which conventional economics focuses are only part of the story. Processes of competition and market exchange are important. Indeed, they are the first dimension referred to in the triad of "competition, command, and change" that defines three-dimensional economics. However, the other two dimensions, command and change, are just as important.

The vertical dimension of political economy, command, takes the existence of power relationships into account. Conventional economists do not include these relationships in their analyses; instead, they leave power to be studied by political scientists. Abba Lerner, a prominent conventional economist, once commented that economics had become the "queen" of the social sciences by focusing only on those political problems that had already been solved. When political problems—such as a society's choice of a particular framework of laws or a system for the administration of justice—have already been solved, the ensuing relations can be conducted solely on the basis of contracts and market exchange.

Three-dimensional economics does not limit itself to the study of solved political problems. Focusing on command as one of the most important aspects of human social life, it sees the economy as a place where power plays a decisive role and where there have been—and continue to be—endless and often bitter struggles between workers and their employers, between buyers and sellers, among giant corporations, and between people and corporations to determine whose interests will be most benefited by public policy decisions. Thus, one reason for the presence of the word *political* in *political economy* is the recognition of the fact that power relationships are an important aspect of any economy.

The third dimension of three-dimensional economics, change, suggests that studying economics also means studying history. The process of change in society cannot be understood without considering the past and how it changed, eventually becoming the present. Change in political economy may be contrasted with the static approach of conventional economics that freezes time at a moment. Economic reality, according to political economy, is better represented dynamically—as a process of change rather than a frozen state of affairs. It is a movie compared to the neoclassical snapshot.

From the standpoint of political economy, the usual distinctions among the disciplines in the social sciences—history, political science, economics, sociology, anthropology, and psychology—are quite arbitrary. These distinctions divide social reality into parcels that reflect the traditional boundaries among university departments, but they obscure our understanding of how the economy works.

Neoclassical Economics

Neoclassical or **conventional economics** is an economic theory emphasizing the horizontal dimension of markets and voluntary exchange.

The **neoclassical** (or conventional) approach to economics, mentioned in both this chapter and the previous one, sees capitalism as a system of markets. The label *neoclassical* is given to conventional economics because this approach presents an updated version of some of the ideas of eighteenth- and nineteenth-century "classical" economics, articulated by Adam Smith in his influential book, *The Wealth of Nations*. Neoclassical economics is thus primarily an explanation of how markets and market systems work.

Neoclassical economics is mainly about competitive markets, that is, markets with many buyers and sellers, and it offers explanations of how economic systems made up of many competitive markets function. While understanding markets is essential to understanding capitalism, the neoclassical approach is founded on three very restrictive assumptions. The first, "economic man" (*Homo economicus*), was discussed in the previous chapter.

The second assumption underpinning the neoclassical approach is that when studying a market transaction, all essential aspects of it are covered in a contract. Accordingly, all other aspects and consequences of the exchange are treated as being of secondary importance.

A **contract** is an agreement, either written (explicit) or unwritten (implicit), that commits two or more parties to taking certain actions, such as making payments and delivering goods or services.

The concept of a **contract** is significant not only in economic analysis but also in many other areas, for example, the law. It is defined as an agreement, either written (explicit) or unwritten (implicit), that commits two or more parties to taking certain actions, such as making payments and delivering goods or services.

Neoclassical economists assume that contracts are *complete* in the sense that the *prices* resulting from them take into account everything that is important about a particular transaction. This is sometimes referred to as the *complete contracting assumption*. A **complete contract** is one that fully specifies, in ways that the courts will enforce, everything that each party to the contract is to do as a result of the contract.

A **complete contract** is one that fully specifies—in ways that the courts will enforce—everything that each party to the contract is to do as a result of the contract.

In making the assumption that all market transactions are based on complete contracts, neoclassical economists take for granted that the contracts into which we enter, explicitly or implicitly, whenever we buy or sell something are "complete" in the sense that (a) they cover everything of interest to both parties to an exchange, and (b) they can be enforced at no significant cost to either party. In contrast, an **incomplete contract** between two parties is one that leaves out certain aspects of an exchange and requires or imposes, upon one party or the other, significant enforcement costs.

An **incomplete contract** is an agreement between two or more parties that leaves out certain aspects of an exchange and requires or imposes significant enforcement costs upon one party or another.

When a new car is purchased the contract is quite complete: the specifications of the car are described, the price is given, the payment plan is made clear, the warranty is spelled out, the limits on the liabilities of the producer are stated, and so on. When an employer hires a worker, however, the contract does not even mention some of the most important aspects of the bargain, such as the exact task (or tasks) the employee may be assigned to do or how hard he or she will be expected to work. The incompleteness of employment contracts is one of the most important issues given attention in the political economy approach. (Credit contracts are incomplete, too, but for a different reason: the exact amount the borrower is to repay is clearly specified, but the contract may be unenforceable if the borrower is broke when the time comes to repay.)

In making the complete contracting assumption, conventional economists picture a world in which exchanges are *voluntary* (the very idea of a contract implies that both parties have voluntarily agreed to it). Coercive—"command"—relationships are not in the picture because if everything that matters in an exchange has already been settled by contract, there is nothing for the exercise of power to be *about*. As the example of the employment contract makes clear, however, what is left out of an incomplete contract may have to be resolved by command. Moreover, when an employer issues an order to an employee, it may be necessary to hire a supervisor to make sure that the order is obeyed. In this case, the salary paid to the supervisor is a cost to the employer of having the contract enforced.

If the price paid for something in a market exchange *did* reflect all relevant aspects of the exchange, one could say "You get what you pay for," and vice versa. Thus, in the neoclassical economists' world, the things you might enjoy but do not have to pay for (such as the love of a friend or your neighbor's beautiful garden) are assumed to be either unimportant or the subject matter of some other discipline. Similarly, the things we get but would rather not get and are not paid to accept (such as a difficult supervisor or environmental pollution) are not given much attention.

Externalities occur when some of the effects of a market exchange are not reflected in the price and are thus "external" to the participants in the exchange.

Side effects on people other than those directly involved in a transaction are termed "external effects," or **externalities**. They are called this because they are said to be "external" to the transaction itself. For example, the price one pays for gasoline does not reflect the costs imposed on others as a result of its consumption—carbon monoxide emissions, smog, health costs, traffic congestion—so all such effects are externalities of the purchase and consumption of the gasoline. Economists generally agree that externalities are a fact of life in any modern economy. But in practice they are taken to be the exception, not the rule. The issues are: How widespread are they? And what should be done about them? Externalities are discussed at length in Chapter 9 of this book.

The third important assumption of neoclassical economics is that *economies of scale* generally do not occur. The term **economies of scale** refers to a situation in which expanding the quantity of output produced in an enterprise per unit of time—or the *scale* of a productive activity—makes possible a reduction in the average cost of producing a unit of output.

If by expanding the quantity of output an enterprise produces, the firm is able to reduce the average cost of producing a unit of output, then the production process has **economies of scale**.

Neoclassical economists assume that, beyond a modest scale of production, economies of scale are rare enough to ignore for most purposes. This in turn allows them to say that, in general, the average cost of producing a unit of a good will rise (or at least will not fall) as the rate of output is increased. This assumption flies in the face of the facts, namely, the prevalence in modern economies of large-scale production that allows many goods and services to be turned out at lower and lower costs per unit. (Think of music, pharmaceutical products, and this book.) The significance of the widespread presence—and growing importance—of economies of scale in modern economies is discussed briefly in Chapter 9 and at length in Chapter 11.

Why is it essential to neoclassical theory to assume that economies of scale do not exist or are of little importance? If economies of scale exist and are important, then the normal state of the economy cannot be competition among many small- or medium-sized firms. Smaller firms will find it impossible to thrive because larger firms will be able to produce at lower cost. When this is the case, the larger firms will be able to drive the smaller firms out of the market or to buy them out. There will then tend to be more monopoly than competition—and in that case the economics of competitive markets will be of little interest.

There is an additional reason why economies of scale may make the competitive markets of neoclassical economics the exception rather than the rule. The outcomes

of competition will often depend not just on which firm delivers a better product at lower cost. Market success may also depend on a firm's political influence, its ability to get a head start and enter a market first, or just the luck of being in the right place at the right time. Whatever the reason, if a firm gets to be a certain size before others do, it will have the advantage of being able to produce at a lower cost than its (actual or potential) competitors. Its greater size—and the cost advantage flowing from economies of scale—will enable it to leave the competition behind.

The three assumptions that underpin and define neoclassical economics can best be understood as part of a worldview dating from the seventeenth-century physics of Isaac Newton. From the standpoint of this worldview, all social and physical phenomena involve knowable and predictable motions of atomlike particles. Thus, the complete contracting assumption is a way of limiting an observer's view of the interactions among the particles so that these interactions may be seen as obeying a few simple laws. Similarly, the economic man assumption establishes the principle of motion of each particle. Finally, the assumption that economies of scale are negligible eliminates the advantages of head starts and accidents, and so implies that the past history of the interacting particles does not influence their current relationship. The end result, in neoclassical theory, is that the economy is viewed as a smoothly running machine, not one with the sometimes harmonious but more often conflict-ridden and sporadically chaotic human interactions that actually occur in a capitalist economy.

The neoclassical approach may be summarized in three interrelated points. First, the economic machine, as seen by a neoclassical economist, operates continuously and indefinitely into the future without any change in its basic design. It may need repair or replacement parts from time to time, but the machine itself—the economy—is relatively trouble-free and not very accident prone. Its few problems, such as recessions or technological unemployment, can easily be taken care of.

Second, change does not occur as a result of the workings of the economic system itself. If something in the economy should happen to change, it will do so only as a result of an *external* influence, such as a major technological innovation or the spontaneous emergence of a new fad in consumer tastes. The development of the Internet, with its corresponding expansion of electronic communication or the sudden proliferation of peoples' desires for running shoes and designer clothing, might be seen as examples of such "external" influences. In fact, however, such innovations and changes in consumer tastes do not occur entirely as a result of forces external to the economy. Rather, they are most often brought about by capitalism's drive to increase profits, expand markets, and sell more products.

Finally, since the economy remains unaffected by its own operation, it does not have a history. It does not, in itself, have a past, present, or future. Thus neoclassical economics presents only a "static" analysis of the economy—static (or "stationary") in the sense that it is not "dynamic." When this shortcoming is considered along with its (previously discussed) lack of interest in power (command) relationships, one may conclude that the conventional approach offers only a one-dimensional analysis of capitalism, focusing exclusively on competition and exchange on the supposedly level playing field of the market.

Values in Political Economy

Most people with an interest in economics care not only about how our economy works (or does not work), but also about what is good or bad about it and how it might be made to work better. The economy is the subject of much controversy and debate.

It used to be said that if you wanted to avoid an argument, stay away from politics or religion. Today, better advice might be: don't mention economics.

Debate in economics is not only about "what is"; it is also about "what should be." This is sometimes referred to as the difference between "positive" (or "scientific") economics and "normative" (or "policy-oriented") economics, but, in fact, the boundary is not always clear. The "what is" question has to do not only with facts but also with their *interpretation*. Thus, differences arise when people disagree about what the facts are—*and what they mean*. The question "what should be," on the other hand, cannot even be addressed without explicit reference to *values*. In this case differences arise when people disagree about which situations are better and which are worse.

Getting the facts and their interpretation right is an essential task for *any* approach to economics. The facts regarding "what is" must be determined with enough accuracy that people with varying points of view can agree on them. Statements about how the economy works are either true or false, however difficult it may be to determine which, and one's judgment of their truth or falsity should not depend on one's values.

When one is choosing to examine a particular aspect of "what is," however, the choice will usually be strongly influenced by one's view of "what should be." If one places a high value on individual freedom of choice and less on fairness, one might be more interested in studying the way that markets and governments may affect one's freedom than in figuring out why women are generally paid less than men. This last topic would be of greater interest to someone who attaches a higher value to fairness. Nobody can be equally interested in all aspects of the economy, and your values will help you decide which economic questions you would most like your economic investigations to illuminate. You need to have some idea about where your lost car keys are in order to know where to point your flashlight.

Values also provide a basis for judging whether we think an economic system is good or bad, or, more specifically, what processes or outcomes of an economic system are better or worse. If you value democracy highly you probably have a very negative opinion of dictatorial regimes such as the ones that prevailed in the former Communist countries. But if you care a lot about increasing the amount of material goods available in a society, you have to be impressed by the record-setting increases in output achieved by Communist-ruled China in the last quarter of the twentieth century. There are, of course, a number of different values, or criteria, on the basis of which one may, explicitly or implicitly, evaluate an economic system.

We evaluate economic systems on the basis of how well or how poorly they organize economic activities so as to provide opportunities for all their participants to lead flourishing lives. Our use of the biological term *flourish* is deliberate: plants flourish—as long as they have enough water, sunlight, and nutrients. In Chapter 14 we discuss some of the elements that people need in order to flourish. But certainly they cannot flourish when they lack adequate food and health care, when they are not free or are denied opportunities for learning, or when they are not respected both as individuals and as members of groups.

The most basic question regarding an economy is: How does it affect people? And while each individual may have his or her own definition of "the good life," an economy that provides more people with more opportunities to lead flourishing lives is judged to be better than an economy that provides fewer people with fewer such opportunities. An economy is thus evaluated on more than just "economic" outcomes—individuals' incomes, for example. It is assessed on *all* of the ways it affects its participants.

How, then, are we to determine what is needed for a flourishing life? Can we say that one person's passion for opera is such a need, while another's craving for a BMW is not? In fact, though, judgments are not hard to make about many of the important issues to which our evaluation criteria may be applied. Imagine a child starving in a country where large quantities of grain are fed to cattle on their way to becoming steaks. Most people would not regard this as a good allocation of grain, even though this conclusion requires making a judgment about the value of a steak to one person relative to the value of enough nutrition to produce a healthy child to another. The world as we know it—sadly—poses many more problems similar to the steak versus the hungry child than to the BMW versus the opera.

An economy can provide the means for a flourishing life, or it can hinder the achievement of one. Of course, whether individuals actually do lead happy and free lives will be influenced by many more variables than how the economy is organized. To a significant degree, the achievement of a flourishing life will depend on the choices one makes. But economies can create conditions favorable to the achievement of a flourishing life, or they can generate conditions that make it difficult to live with dignity, freedom, and happiness.

An economy will impede the achievement of flourishing lives if jobs are mind-numbing or unsafe, if children die young of easily preventable diseases, if people remain illiterate or are not free to speak their minds or practice the religion of their choice, if malnutrition is common, if people are stigmatized because of their race, sex, or sexual orientation, or if other conditions that limit opportunities are generated. An economy is better to the extent that it reduces or does not create such conditions—not just for a few people but for everyone and to the greatest possible extent.

Whether an economy contributes to a flourishing life for all the members of a society will depend on a number of aspects of the way the economy works. Among these are its efficiency, its fairness, and the degree to which it is democratic.

Efficiency

The term **efficient** is applied to a labor process if the effort, time, intelligence, creativity, raw materials, natural environment, information, and machinery used in it are applied in a way that enhances people's well-being by equipping them with the things and the free time needed to lead a flourishing life.

One criterion for evaluating an economic system is that it should be **efficient**. There are many definitions of efficiency in economics. We use the term to mean that labor and inputs are used well rather than wasted. Inputs include effort, time, intelligence, creativity, raw materials, the natural environment, and machinery. Using these inputs well (rather than wasting them) means using them to enhance people's well-being by equipping them with the things and the free time needed to lead a full life. An economic system that uses its resources more efficiently than another is, according to this criterion, a better economic system. It is better because it increases people's freedom to use their time and energy for noneconomic ends such as leisure, play, and learning.

INPUTS AND OUTPUTS Engineers think of efficiency as a relationship between physical inputs and output. A production process is said to be *technically efficient* if, given the existing technology, the output of it cannot be increased without using more of at least one of the inputs.

In evaluating efficiency, we need to look at both the inputs into production and the outputs from production in a particular way. Efficiency is not to be confused with *profitability*, which occurs when a firm's sales revenues exceed its costs, taking account of all the inputs paid for by the firm. Chapter 13 gives examples of why profitability and efficiency are not the same thing.

Efficiency refers to the production of *useful* goods and services. The production of goods and services that are not useful should be left out of the efficiency calculation. For instance, the production of advertising (beyond the extent to which it simply informs the consumer) is highly profitable—this is why it is produced—but it is not useful. The same may be said of some forms of military goods production during peacetime.

All inputs used in the production of useful goods and services, whether they are paid for or not, must be entered into the efficiency calculation. One input that is often not taken into account is the natural environment. A profitable factory that uses up clean water by adding toxic wastes to it may, in fact, be inefficient when all of its inputs are considered. This is because in addition to the inputs it pays for, the factory is also using up (consuming or destroying) a part of the natural environment.

Still another input that is frequently ignored is household labor. When we consider all the useful inputs and outputs of an economic system, we include work in the home as well as the labor that is organized in factories and elsewhere.

A further input that is often overlooked is labor effort. This is not the same as the *time* that a worker spends "at work." An assembly line that speeds up production may increase a company's profits, but it may not be efficient if the increased output is made possible only by an even larger increase in the employees' work effort, leaving them with greater fatigue at the end of the day and increased susceptibility to health problems. One way of thinking about this aspect of efficiency is to consider people and their health not only as inputs in the production process but as outputs of it as well.

PARETO EFFICIENCY The definition of efficiency preferred by many economists is *Pareto optimality*, named after the economist Vilfredo Pareto (Pa-RAY-toh), who first thought it up about a century ago. It is the definition of efficiency preferred by many economists. An outcome is a Pareto optimum (sometimes called *Pareto efficient*) if there exists no *other* outcome (using available resources and technologies) that would make at least one person better off without making anyone worse off. If there is some alternative use of inputs and outputs that has the win-win quality that some could be made better off and none worse off, the outcome is said to be *Pareto inefficient*.

Notice two things, however. First, there are a very large number of Pareto (efficient) optima, each with a different distribution of goods among the members of society. Second, even if some people are starving while others feed caviar to their cats, the result will be Pareto efficient if there is no way of redistributing the goods that makes the poor better off *without making the rich worse off* (the cats do not count). There is nothing about Pareto efficiency that implies fairness. Fairness is a separate criterion.

Fairness

Fairness means that people in an economic system suffer the burdens and enjoy the benefits of that economic system equitably.

A second criterion for evaluating an economic system is **fairness**, which involves the distribution of the system's burdens and benefits. Burden refers to sacrifices such as the work necessary to produce goods and services; benefit refers primarily to the use of what is produced. Who does how much work? And who consumes the products of the work that is done? Just as in evaluating the efficiency of an economy, all the costs of—and gains from—production need to be included.

A fair economic system would be one in which its burdens and benefits were distributed *equitably*. When judged according to this criterion of fairness, an economic

system that distributes its burdens and benefits more equitably than another does is a better economic system. It is better because it recognizes the equal worth of all human beings. When each person is valued equally, the pleasures of each are equally worthy of being promoted, and the pains of each are equally worthy of being avoided. Thus, though each of us is different, we are the same with regard to our right to enjoy pleasures and avoid pain.

The belief in the equality of human worth as an ethical benchmark comes from a variety of sources. Virtually every religion considers each human being to be equal in the eye of whatever god is held to be supreme in the particular religious faith. The authors of the U.S. Declaration of Independence proclaimed in 1776: "We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty, and the pursuit of Happiness." (But the right to vote was not one of these "Rights." Voting rights went only to free white male property owners at first in most states. Many Native Americans and Asians were excluded from citizenship and voting for decades, and only got the full legal right to vote in the mid-twentieth century; ex-slaves got citizenship in 1878 but were often prevented from voting until the Voting Rights Act of 1965. Women got voting rights in 1920.)

Many, possibly most, people would agree with the framers of the U.S. Constitution: people should have equal rights to participate in governing their societies and to be respected as individuals. But *what else* should be equal? Hardly anyone would argue that every individual should have exactly the same things, for this would not respect differences among people with regard to what they like and dislike. Some people work hard and long because they value material things, while others value leisure or nonmaterial pleasures more and therefore work less; it does not seem fair that both sets of people should receive the same quantity of material goods. Thus, while equality is a value endorsed in this book, it is important to be able to answer the question: *Equality of what?*

The answer offered here—and one that is widely shared not only in the U.S. but also throughout the world—is that people should have an *equal opportunity* to live a good life, however they may conceive of this objective. Equal opportunity requires that, insofar as it is possible, people should be equally free of impediments to fashioning for themselves the kind of life they want, as long as the exercise of their freedom does not limit or otherwise impinge on the pursuit of a good life by others.

Of course, there will always be impediments to doing what we would like to do. For example, it sometimes happens, as it did to one of the authors of this book, that a person dreams of becoming a musician, but falling short on talent, has to settle for going through life as an economist! Equal opportunity simply means that people should be equally free of those impediments that are within our power to remove.

Some violations of equality of opportunity are obvious, as when employers or landlords discriminate against people on the basis of their race, sex, or age. But there are many less blatant forms of inequality, and they are to be found in most societies. When some children attend well-equipped schools and are taught by outstanding teachers while other children do not have comparable facilities or teachers, opportunity is unequal. When poor people must pay more for groceries because, for example, lacking a car they must shop at small neighborhood stores, or when people must pay higher interest rates to borrow money because of their race, sex, or where they live, opportunity is unequal. If some young people start off life with an expensive college education and a trust fund while others have only debts that have been passed on to them by their parents, opportunity is unequal. When, as a result

of different starting points, some hard-working people earn only the minimum wage while other people, working no harder, take home hundreds or even thousands of dollars per hour of labor, opportunity is unequal.

Unlike the idea of *equal outcomes*, the objective of *equal opportunity* often requires people to have access to different things or to be treated in different ways. For example, people with health problems need more medical attention than do those without them. Children with dyslexia or other learning disabilities need more attention at school to have the same learning opportunities as do others. Children of parents who cannot or will not help them with their homework may also need more help at school if equality of opportunity is to be achieved.

The criterion of fairness will always be controversial. Should it apply equally to everyone in the world? Or should it be applied only within a single nation? If the answer is that it should be applied worldwide, then we say that it is unfair that a child growing up in a particular part of India will, as a result of poor living conditions, live twenty years less than a child of the same age growing up in Norway.

A final question: to what extent is it fair to give people second or third chances? If certain people drink their way through college and, as a result, end up poor and unemployed, does fairness require that they be temporarily supported while they undergo job training?

Democracy

The third criterion for assessing an economic system is *democracy*. One important part of this criterion questions the extent to which the economic system promotes (or hinders) the democratic functioning of the government. Another part asks whether the system allows for the accountability of power when it is exercised in the economy. An economic system that promotes democracy in both areas is better than one that does not. It is better because the ability to influence a decision that has an effect on you gives you greater control over your life.

Democracy is a process with three characteristics: the exercise of power is accountable to those affected by it, civil rights and personal liberties are guaranteed, and citizens have relatively equal access to political resources and influence.

Democracy is a process that has three characteristics: accountability of power, respect for civil liberties and other guarantees of individual choice, and equal opportunity for effective political participation. First, decision makers (whether in the government or elsewhere) must be held accountable to the people affected by their decisions. This requires periodic review and possible replacement of public officials by means of democratic elections. Those who wield power in the economy—owners of firms, for example—can be held accountable by similar means. Democratic governmental bodies can regulate the actions of firms. If there is competition among many firms, consumers can hold sellers accountable for bad products by switching to other sellers—"voting with their pocketbooks"—and firms failing to serve consumers well will be eliminated.

Second, there must exist guarantees for the exercise of the civil rights and personal liberties that are commonly associated with democratic citizenship. The rights to freedom of speech and assembly, for instance, are essential for democratic decision-making.

Third, the citizens in a democratic polity must have approximately the same amount of resources with which to participate in the democratic process; this is necessary if citizens are to have roughly equal opportunity to influence how decisions are made. A system in which everyone can vote but a few people—campaign contributors, for example—have more political influence than everyone else is not democratic.

Democracy does not mean that all decisions have to be made through voting. Individuals should be free independently to make any decisions that will have consequences entirely or mainly felt only by the individuals themselves. For example, the choice of which food to consume for dinner is almost always a decision that affects only the eater or the eater's family.

When a decision imposes unavoidable effects on many people, however, democracy requires that individual choice give way to collective democratic decision making. For example, the choice between closing down or modernizing an old factory will affect many people: the investors, the workers, the consumers of the product, the people who live near the plant and may be bothered by its noise or pollution, the community that depends on the property taxes on the plant, and so forth. In this case, the democracy criterion holds that individual choice—for instance, the plant owner's right to choose whether to close the plant—is undemocratic. If the democracy criterion is applied, all those affected by the decision must be able to participate in the decision-making process.

Of course, to call the owner's decision undemocratic does not mean that the value of democracy should trump the values of efficiency or fairness. Either or both of these may recommend against a democratic decision. In the plant-closing example, for instance, it is not easy to see how the competing claims of the consumers, neighbors, workers, owners, and others could all be accommodated. Should each be given one vote? Or should plant closings be regulated by democratic national governments to minimize their adverse effects?

Some economists would think it odd to suggest that a private firm should be run democratically. Conventional textbooks ignore the exercise of power in the economy and treat organizations such as corporations simply as pass-through structures: inputs go in one side and products come out the other. Texts that ignore the power that is exercised in corporations will find nothing in them to democratize. Moreover, the right of people to participate in a decision-making process, the outcome of which affects their entire community, is not considered to be a problem that falls within the purview of economics.

Democracy and command are not necessarily inconsistent, although a command can certainly be undemocratic. For instance, orders issued by dictators or rules imposed by employers will be undemocratic if the people affected by the orders or rules have had no influence in the processes that led to their issuance or imposition. But a command may also be a means of carrying out a democratic decision. Take, for example, a national environmental protection law that was formulated with popular participation and then voted on and passed by a democratically elected legislature. To enforce this law, a governmental agency, backed by a federal court, may have to order polluters to stop polluting. In a different kind of situation, the command given to a worker by a democratically chosen manager in a worker-owned and worker-run cooperative would be another example of a democratic command. Implementing and enforcing democratic decisions requires commands.

Balancing Efficiency, Fairness, and Democracy

Economic systems may be judged according to how well they meet the criteria of efficiency, fairness, and democracy. Some systems may be evaluated more favorably according to one or two of these standards, while other systems perform well in

relation to one or more of the other criteria. For example, slavery was efficient—at least in the production of some crops—but it was unfair and undemocratic. In contrast, production by independent producers (for instance, independent farmers who owned their own land in colonial New England) was probably less efficient than slavery, but it was fairer and might be thought of as more democratic. It is also possible that the ability of an economic system to perform well in relation to the standards of efficiency, fairness, and democracy will change over time as the economic system itself changes.

Moreover, it may be difficult for any economic system to make consistent progress toward meeting all three criteria simultaneously or with the same speed. For example, the efficiency criterion may conflict with the democracy standard in the following way: the achievement of efficiency—say, production of the greatest possible amount of useful goods and services with limited quantities of inputs—may require intense competition and a high degree of mobility of labor and capital, while these very same elements may make it difficult for worker-owned (or otherwise democratically controlled) enterprises or stable, democratically governed communities to survive.

Necessarily, then, our evaluations of economic systems will result in more complex judgments than simply "good" or "bad," "better" or "worse." Also, it is highly unlikely that everyone will agree on all of the issues involved. Each person's conclusions will depend on his or her own values, and as long as we remain autonomous individuals our values and the conclusions we reach on the basis of them will inevitably differ. To the degree that diversity and debate are good indicators of a healthy democratic society, such differences are to be welcomed.

In this book the values of efficiency, fairness, and democracy are not explicitly brought into the description and analysis of how capitalism works. Indeed, our analysis of capitalism should stand or fall without regard to a particular reader's values. The important question is whether the political economy approach offers a more fruitful way of understanding capitalism than that provided by neoclassical economics. We conclude this chapter with Table 3.1 summarizing the differences between the two approaches.

The contrast in Table 3.1 between neoclassical economics and political economy is of course not the only way one can subdivide economics. Indeed the subject is usually broken down into **microeconomics**, which is concerned with what individuals, families, and firms do (and why), and **macroeconomics**, which is about how decisions of the same individuals, families, and firms, together with government policies, determine outcomes for society as a whole. In Part 2 of this book—"Microeconomics"—we look at how buyers and sellers interact in markets, how firms seek to increase their profits, and how conflicts between employers and their employees regarding wages and work are resolved. In Part 3—"Macroeconomics"—we examine the interactions of individuals, firms, and governments, asking how they result in economy-wide outcomes affecting such things as incomes and opportunities, wealth and poverty, growth and stagnation, employment, unemployment, and inflation.

Microeconomics deals with decisions that individuals, families, and firms make (and why they make them).

Macroeconomics is about how the decisions of individuals, families, firms, and governments produce outcomes—such as economic progress or stagnation, inflation or unemployment—for society as a whole.

Having introduced the basic concepts of three-dimensional economics in this chapter, we proceed in the next chapter to explain how political economists analyze production in a capitalist economy, and how different relationships to the production processes define different groups with different economic interests.

TABLE 3.1 CONTRASTING PERSPECTIVES ON THE ECONOMY

NEOCLASSICAL ECONOMICS	POLITICAL ECONOMY
The main social relationships studied involve competition among self-interested people or between the firms in which they work.	The social relationships studied are cooperative as well as competitive, and generosity and reciprocity are considered along with self-interest.
Most economic interactions take the form of complete contracts.	Many economic interactions are not governed, or are not governed completely, by contracts.
Economic outcomes are determined by market forces. Power is exercised only by monopolies and governments.	The exercise of power is an important determinant of economic outcomes, even in competitive markets. Many economic outcomes are determined through bargaining between the parties or agents involved.
Constancy is the rule; change occurs only in response to forces outside the economic system.	Change is the rule, constancy the exception. Change, both in economic systems and in people, takes place through the workings of the economic system itself.
People's tastes and needs are determined largely by human nature or by other influences outside the economic system.	People's tastes and needs change and are strongly influenced by the economic system.
Knowledge and science evolve outside the economic system, governed by noneconomic forces.	Knowledge and science are strongly influenced by the economic system and by the exercise of power within it.
Economic inequality is given little attention and is measured by a single scale: income inequality.	Economic inequality is many-sided, encompassing differences of race, gender, status, property ownership, authority, income, political rights, and citizenship.
Economies are evaluated according to how well they do in relation to a limited view of efficiency.	Economies are evaluated according to how well they foster everyone's chance to lead a flourishing life; economic efficiency, fairness, and democracy can support the achievement of this goal.
Economies of scale (costs declining as output expands) are absent or may be ignored.	Economies of scale are common in modern economies and therefore must be taken seriously.

Suggested Readings

Ari Berman, *Give Us the Ballot: A Modern History of the Struggle for Voting Rights in America* (New York: Farrar, Straus, and Giroux, 2015).

Robert A. Dahl, *Democracy and Its Critics* (New Haven, CT: Yale University Press, 1989).

Nancy Folbre, *The Invisible Heart: Economics and Family Values* (New York: New Press, 2001).

Robert Heilbroner, *Behind the Veil of Economics: Essays in the Worldly Philosophy* (New York: W.W. Norton, 1988).

Robert Nozick, *Anarchy, State, and Utopia* (New York: Free Press, 1974).

Louis Putterman, *Dollars and Change: Economics in Context* (New Haven, CT: Yale University Press, 2001).

John Rawls, *Justice as Fairness: A Restatement* (Cambridge, MA: Harvard University Press, 2001).

John Roemer, *Free to Lose: An Introduction to Marxist Economic Philosophy* (Cambridge, MA: Harvard University Press, 1988).

Charles Sackray, Geoffrey Schneider, and Janet Knoedler, *Introduction to Political Economy*, 8th edition (Boston, MA: Economic Affairs Bureau, Inc., 2016).

Michael Sandel, *What Money Can't Buy* (New York: Farrar, Strauss, and Giroux, 2012).

Amartya Sen, *Development as Freedom* (New York: Knopf, 1999).

Amartya Sen, *The Idea of Justice* (Cambridge, MA: Belknap Press of Harvard University Press, 2011).

Philippe van Parijs, *Real Freedom for All: What (If Anything) Can Justify Capitalism?* (New York: Oxford University Press, 1998).

The Surplus Product: Conflict and Change

4

Walk into almost any travel agency and you will see posters that tell a very interesting economic story. You may see a poster of India showing the Taj Mahal. Beside it may be a picture of the Egyptian pyramids. Another poster invites you to visit Rome and see the ancient Coliseum or the fabulous treasures of St. Peter's. Then on to France to view the delicate opulence of the chateaux and the sheer splendor of the palace of the Sun King (Louis XIV to his close friends) at Versailles. Far-off China beckons with the attraction of the Great Wall. One may also take a tour of the southern United States, visiting the great plantation houses of the former slaveholders.

What story do these posters tell us? What do the Taj Mahal, the pyramids, the Great Wall, and the slaveholders' mansions have in common? Simply this: even today, these great edifices inspire awe (and attract tourists) not only because of their great beauty but also because of the almost inconceivable amount of labor that was required to build them. How was it possible in poor societies such as ancient Egypt and medieval France for so many resources to be devoted to such construction projects? Where did the rulers of these societies—kings, bishops, emperors, or slaveholders—obtain the means to realize such magnificent plans?

The answers to these questions may be found in the concept of *surplus product*, another of Adam Smith's important ideas. The surplus product is *that part of the total output of an economy that is in excess of what is needed for reproducing and replenishing the labor, tools, materials, and other inputs used or used up in production*. There is no reason why a surplus *must* exist, but it *does* exist and *has existed* in all but a few human societies. The surplus product may be used in a variety of ways. It can take the form of cathedrals, palaces, luxury goods, military spending, more or better productive equipment, higher levels of education, improved health, and many other things.

The surplus product concept is a powerful lens that can help us understand how a society works. Why, for example, did the economy of China grow rapidly over the last two decades of the twentieth century, greatly improving the living standards of most Chinese, while at the same time the economy of the Philippines stagnated, leaving most Filipinos desperately poor? The simple answer is that in China the surplus

was invested in new factories, new equipment, and increased education, while in the Philippines the rich used much of the surplus in luxury consumption.

How the surplus comes to exist is also a key to understanding how a society works. As in other economic systems, the existence of a surplus in a capitalist economy depends on the power exercised by the dominant class over the producers. In a capitalist economy, however, the surplus results not only from the direct exercise of power by an employer over his or her employees; it also arises indirectly from the way markets work, particularly from the prices at which goods and services, including labor, are exchanged.

In the Soviet Union under communism, a large surplus came about due to the central administration of the economy. In yet other societies, for example the bands of hunting and gathering ancestors of modern humans who made their living before the dawn of agriculture, there was no surplus at all. The absence of a surplus in such societies helps to explain why they left behind no monuments for archaeologists to study—or for travel agencies to put on posters.

A focus on the surplus product is of primary importance in political economy. This focus is one of the main ways in which political economy differs from neoclassical economics. Looking at society through the lens of surplus product enables the political economist to see economic systems as having historically specific qualities, with one economic system being differentiated from others by the ways in which the surplus product is generated and controlled. How do we know what "capitalism" is? The answer offered in this book is that it is an economic system with a distinctive way of generating and disposing of its surplus product.

Understanding the surplus product requires the use of all three dimensions of the three-dimensional approach to economics. First, the command (or vertical) dimension must be called upon to understand how a society's producers are induced to settle for less than the whole of what they have produced, for only then can there be a surplus product. The horizontal dimension helps us understand how the surplus is affected by relationships among people at the same level in society. Are relationships among the producers competitive or solidaristic? Do members of the ruling group just compete with one another? When individuals act in concert with others at the same level of society, how do they suppress their competition? Finally, the historical, or change, dimension comes into play because the disposition of the surplus—be it luxuriously consumed, used to build monuments, or invested to create new productive capacity—will determine whether a particular society is stationary or experiencing change.

The main idea of this chapter is that *how the surplus arises, the size of the surplus, who controls it, and how it is used are the most important issues to be considered when analyzing the structure of any society, tracing a society's evolution over time, or determining the extent to which its economy allows for and supports a flourishing life for all of its people.* In this chapter the general concepts needed for understanding the surplus in any type of society are introduced; in the next chapter these concepts are applied specifically to the analysis of capitalism as an economic system.

This chapter's central idea is expressed in three main points:

1. An economy is a collection of *labor processes.* Each labor process is made up of a *technology* (a relation between inputs and outputs) and a *social organization of production* (the relations of people to the production process and to one another in the workplace). Each labor process produces one or more outputs.

2. A *surplus* is produced whenever the labor processes in an economy produce more than is needed to maintain the producers at the standard of living to which they are accustomed and to replace the materials used up and restore the machines used in production.
3. The size of the surplus in any economy depends on relationships that are often highly conflictual.

Economic Interdependence, Production, and Reproduction

The **division of labor** (or **economic specialization**) exists when people are not economically self-sufficient but instead produce things used by others and use things produced by others.

Horizontal economic interdependence is based on specialization and is not necessarily based on unequal advantage or command.

Vertical economic interdependence exists when one person controls the labor and the products of another; it is based on a relationship of unequal advantage and command.

In all societies people are **economically interdependent**. The story of Robinson Crusoe is of such great interest because, among other things, it is a story about someone who is entirely self-sufficient: he grows his own food and makes all the things he needs to sustain himself. This provokes fantasies exactly because it is so different from our own experience. In all real societies everyone depends on the products of other people's labor, and when we work together we generally do better than we would if we just subsisted independently. This point is illustrated in Daniel Defoe's novel itself: Crusoe does better when he happens to acquire a slave (Friday).

If people worked only in isolation, we would not need economics; Daniel Defoe would be good enough. The most important thing that requires us to engage in economic inquiry is the fact that people generally do depend on other people and are thus interdependent with them. Economic interdependence is virtually universal because it allows for greater productivity. The superiority of working in interdependent ways can also be derived from the observation that many of us work in firms with thousands of coworkers, thereby producing particular goods or services more cheaply than they could be produced in small enterprises (recall the point about economies of scale in Chapter 3).

Every society, no matter how it is organized, must solve two economic problems: first, how to organize the interdependent economic activities of people and, second, how to distribute the resulting products. Addressing the first of these issues—how to organize our interdependence—we find that economic interdependence itself occurs in two ways. The first involves *horizontal* relationships that may be explored with reference to **economic specialization**, or **division of labor**. The people who make up an economic system produce different things: some make shoes, others generate electricity, others cook meals. No one produces all the things that he or she needs, and, at the same time, people usually produce (or help to produce) more of something than they can consume. The excess of each person's output beyond what he or she needs is then distributed through some kind of process to others in return for the others' excess output.

The second form of economic interdependence involves *vertical* relationships. A vertical relationship occurs when one person *controls* the labor and the products produced by another. The slave owner, for example, was economically dependent on the slaves' labor, for it was this labor that kept the plantation going and enriched its owner. The slave, in turn, was dependent on the owner for a livelihood, however meager it was. The modern day employer in a capitalist economy is likewise dependent on his or her employees: without their labor there can be no production or profit. And the employee, in turn, is dependent on the employer for a job and a wage.

With a relationship of vertical interdependence there is a top and a bottom, and the relationship is one of *command*, for example between the slave owner and the slave or between the employer and the employee. Although both parties in these relationships are mutually dependent, the slave owner and the employer are bosses: they command the work activities of the slave or the employee and benefit from the fruits of their subordinates' labor. Vertical interdependence thus differs from horizontal interdependence since the horizontal kind does not involve a ranking of superiority and inferiority between, say, the buyer and the seller of something.

In the eighteenth century Adam Smith was the first major economist to illuminate horizontal economic interdependence, while in the nineteenth century Karl Marx advanced our understanding of vertical interdependence. It is worth noting, however, that Smith was not oblivious to vertical interdependence. Indeed, in his analysis of medieval European agrarian life in *The Wealth of Nations*, he used the term *surplus produce*, defining it with reference to "a great proprietor" who lives off "[that part of] the produce of his lands which is over and above the maintenance of the cultivators."[1]

Production

A **labor process** is any activity performed with the intention of producing something, which may be either goods or services.

Technology is the relationship between inputs and outputs in a labor process.

Technical progress is a change in the relationship between inputs and outputs that permits the same output to be produced with less of one or more of the inputs.

An **economy** is a collection of labor processes.

In order to analyze vertical economic interdependence we must look at production. All human production must involve at least one **labor process**, defined as any activity of people performing purposeful work with the intention of producing something. Thus, any labor process will require *inputs* (such as human work and means of production such as machines and raw materials), and it will also produce *outputs* (such as a ton of steel, a computer program, or a haircut).

Think about making something . . . say, pancakes. If you made a list of all the ingredients needed (stating how much of each) and you included all the various activities required to mix the batter, turn on the stove, flip the pancakes, and so on, you would have a very complete recipe for pancakes. This recipe would describe the labor process for making pancakes. Similar lists of inputs and activities could describe the labor processes for making the flour used in the pancakes, for producing the gas that is burned in the stove, for making the stove itself, and for supplying the other inputs as well.

A recipe such as the one for making pancakes—or any other set of relationships between inputs and outputs in a labor process—is called a **technology**, and *technical change* refers to a change in any of these input-output relations. An example of a technical change would be a way to make pancakes with a different kind of flour, say, one that does not need to be sifted. **Technical progress** is a technical change that allows the same amount of some output to be produced with less labor or less of another input.

In its most fundamental sense, an **economy** is *a collection of labor processes.* Of course, countless numbers of different things are produced in an economy—it has many types of outputs—and an economy also uses a great variety of inputs, including many different kinds of labor. In order to discuss these inputs and outputs and their corresponding labor processes, we have to group them in some way, and how we group them depends on what we want to find out.

---

[1]Adam Smith, *The Wealth of Nations* (New York: Random House, Modern Library, 1937 [1776]), Book 3, Ch. 4, fifth paragraph.

For our purposes here we divide the economy into two sectors, each producing a certain kind of output. The first sector produces as its output all the goods and services that we consume or that businesses use as inputs. We are quite familiar with outputs of this kind, for they correspond to the normal everyday meaning of the term **production**. The second sector is not usually thought of as being part of the economy at all: it produces people. (Recall "The Economy Produces People" section of Chapter 2.)

Production is a labor process whose output is a good or a service.

Reproduction is a labor process whose output is people; it includes not only biological reproduction but also such activities as child rearing, training, feeding, and caregiving.

The fact that people are produced by labor is suggested by the expression that a mother giving birth is "in labor." But these people-producing labor processes involve not just biological reproduction but also feeding, caring for others, training, imparting skills and values, and all the other tasks of caring for and rearing children that are associated with family life. Not only children but adults as well need to be fed and cared for. When we prepare a meal, eat, and relax after a day's work we are, among other things, replenishing our energy and maintaining our ability to continue functioning in the economy.

To distinguish the two sectors (with reference to the types of outputs that define them) we call the first one *production* (because it produces goods and services) and the second **reproduction** (because it reproduces people). This book focuses primarily on the production sector, especially within capitalism, but we will also take into account the vitally important reproduction sector since production and reproduction are clearly and indispensably interwoven with each other.

Links Between Production and Reproduction

Figure 4.1 is a picture of the economy that shows the ways that the production and reproduction sectors are linked together through their inputs and outputs. Labor processes producing goods and services (production) are at the top; labor processes producing people (reproduction) are at the bottom.

We first consider the links between the production and reproduction sectors by looking at the *outputs* of each sector. The outputs of labor processes in the production sector may circle back into this sector as materials, machines, or services for use in another production process. Alternatively, they may be consumer goods or services to be used by people in the reproduction sector.

Looking at the outputs—people—of the reproduction sector, we can see that some of them find employment in the production sector (producing shoes, steel, haircuts, etc.), whereas others (such as parents who care for children) will continue to work in the reproduction sector. Still others will not be engaged in productive labor at all (we will return to them in a moment).

Of course, many people work in both sectors, contributing both to production and to reproduction. For example, many of those employed outside the home have two "jobs," one that pays, and the other consisting of the housework, child care, and other tasks that throughout much of history have been performed mainly by women in the home.

The other way of identifying links connecting the production and reproduction sectors is to look at the *inputs* to each type of labor process. Starting with the production sector, we can see that some of its inputs (for example, materials or machines) come from this sector itself, while others (the people who do the labor) come from the reproduction sector, and still others (such as inputs from nature) come from outside the economy altogether.

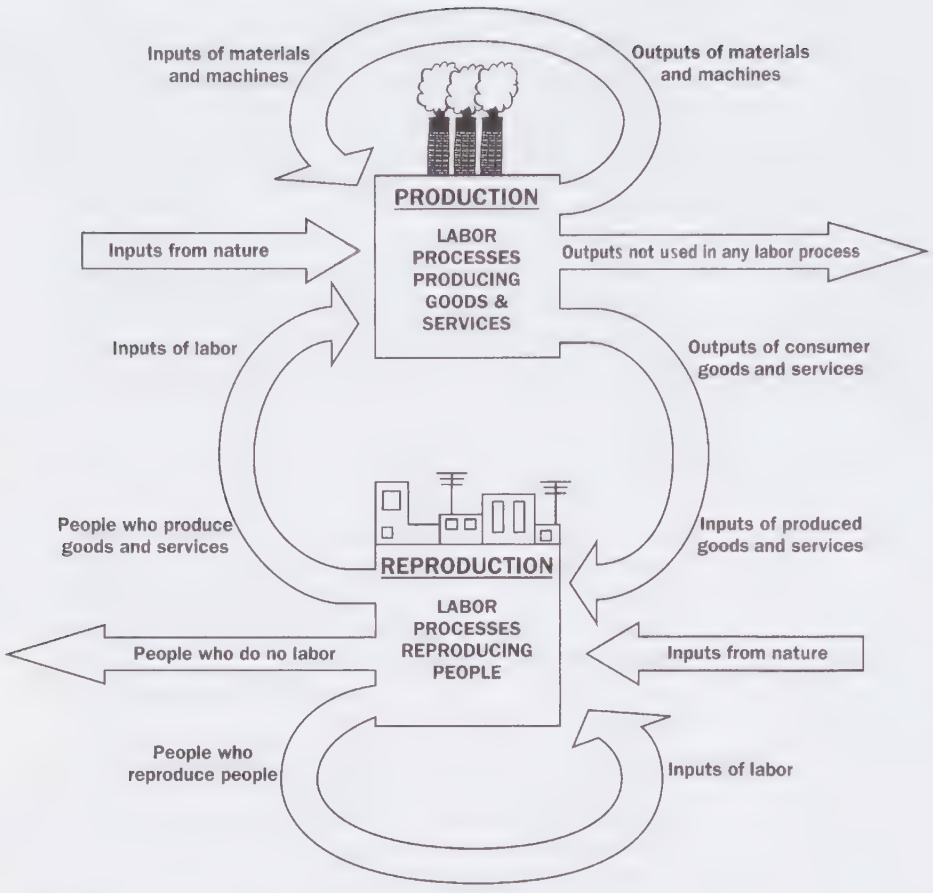


FIGURE 4.1 The production-reproduction cloverleaf.

The economy is composed of two interdependent sectors, one producing goods and services and the other reproducing people. The arrows represent the movement of goods or people from one sector to the other or back into the same sector. Each sector uses three types of inputs: inputs produced elsewhere in the economy, inputs produced in the same sector, and inputs from nature. Each sector also produces three different types of outputs: outputs used in the same sector, outputs used as inputs in the other sector, and outputs rarely used as inputs anywhere in the economy.

Capital goods are goods used in production—machines, buildings, and the like—that are durable and will be used up only over the course of years.

Materials are goods that are used up during the process of production.

If an output of the production sector is a physical object that circles back and becomes an input into a labor process in the same sector, it is classified either as a *material* or as a **capital good**. The term *capital good* refers to durable production equipment and structures, including machines, tools, buildings, and other long-lasting things needed in production.

Materials, on the other hand, are different from capital goods in that they are entirely *used up* in the course of a production cycle, whereas capital goods are merely *used*. While they are being used, capital goods are subject to a certain amount of wear and tear—called "depreciation"—but they are only used up over a period of at least a year. The cotton that ends up in a shirt and the fuel that is burned in the process of making the cotton are examples of materials, whereas the sewing machine that is used to sew shirts to be sold is a capital good.

Labor is any activity performed by people that contributes to production or reproduction.

With regard to the *people* going from the reproduction sector to the production sector, they are put to work as the labor that sets labor processes in motion. We define **labor** as any activity performed by a person that contributes to production or reproduction.

A Labor Process Example: Making Pancakes

We can illustrate the concepts just introduced by going back to our example of pancake making, this time looking at a labor process producing pancakes at a fast-food restaurant:

Inputs:	
Labor	Mixing the batter Pouring the batter into the frying pan, turning the pancakes over, and so on Waiting on customers
Materials used up	Flour Eggs, etc. Gas or electricity
Capital goods used	Wear and tear on the stove Wear and tear on the frying pans, bowls, mixing spoons, and other utensils Depreciation of the building (physical facilities)
Output:	
Consumer good	Pancakes

This is an example of a labor process in the production sector that is turning out a consumer good directly consumed by people. Pancakes are something that could have been produced at home, not for sale, in which case the process of making them would be considered part of the reproduction process. In this example, however, they are an output of the production sector.

Returning to Figure 4.1, we can now summarize how the economy's labor processes are linked. First, some outputs of the production sector are used as inputs into labor processes in the same sector. For example, coal is used to make steel, and machinery used in coal mining is itself made out of steel. Another type of output of the production sector, consumer goods of all kinds, goes to the reproduction sector and is used in the "production" and maintenance of people.

The output of materials and machines may be sufficient not only to *replace* but also to *expand* the supply of machinery, raw materials, buildings, and the other things used or used up in production processes. Similarly, the supply of consumer goods and the reproduction of people may support an increasing number of producers. As we saw in Chapter 1, one of the defining characteristics of capitalism is its tendency to expand both production and reproduction.

There are two kinds of outputs in Figure 4.1 that are special in that they might be thought of as being used little or not at all in any labor process. Certain physical objects might appear to be in this category. Luxury goods might not be used in a labor process—except when a yacht is repaired or refurbished by a paid worker, or a fur coat is cleaned by a commercial cleaning service. A cathedral or a monument

might seem unconnected with any labor process after it is built—but it may be a crucial input into tourist services, and if it is later maintained or reconstructed by paid craft workers, it is an input into that labor process of maintenance or reconstruction. When a cathedral is used for worship services, it also may be thought of as an input into a reproductive labor process by which people strengthen values and have their spirits lifted so that, among other things, they become more effective in both productive and reproductive labor processes. Apart from physical objects, there may also be people who are created and maintained in the reproduction sector—from panhandlers to playboys—who may rarely participate in any labor process, yet might do so from time to time.

There are also inputs not produced by a labor process. We call these "inputs from nature," using "nature" to refer to our physical environment—air, water, arable land, the minerals in the earth's substratum, the sun, and so on. These inputs are presented in Figure 4.1 as arrows entering the two types of labor processes from outside the system. Saying that they are provided by the environment does not mean they cannot be used up. Although there are a few inputs from nature (such as the sun) that can be used indefinitely (to grow crops, to warm houses) without being exhausted—these are "renewable resources"—many natural inputs are nonrenewable and can be used up. Indeed, we now know that human production and consumption are using up so many things in our environment—such as clean air, clean water, and the viability of the world's oceans as a habitat for marine life—that the future of the ecosystem itself is in doubt.

Summing up what we have said thus far, an economy is a collection of labor processes that are specialized in the production of distinct outputs and that are connected by relationships of both horizontal and vertical economic interdependence. To pursue our investigation of the vertical dimension we now turn to the surplus product.

The Surplus Product

What did the great kings and queens of the past live on? And the legendary warriors and other worldly saints? Who produced what they ate, and why? What made possible the glittering Versailles Palace, the Great Wall of China, the imposing Jefferson Memorial, and the war machine of Adolf Hitler? And how was metalworking transformed from the humble blacksmith's shop to the great mills of Pittsburgh, Kobe, and Dortmund? These questions all refer to the tangible, physical manifestations—impressive in their sheer magnitude—of the surplus product.

No less impressive, however, are the intangible forms of the surplus product, such as the vast accumulated knowledge, the result of untold years of work by scholars, now available on the smartphone in your pocket); the awesome beauty of the ceiling of the Sistine Chapel in Rome, the result of many hours of creative work by Michelangelo; and the wisdom in the sacred texts of the Muslim, Hindu, Jewish, and Christian religions, the Koran, the Bhagavad Gita, the Talmud, and the Bible. To understand both the tangible and the intangible phenomena just described, we need to define the concept of the surplus product.

In defining the surplus product, the first step is to note that a society produces a certain quantity of goods and services during a year. This is the *total product*. The

total product is in turn divisible into two parts, the *necessary product* and the *surplus product*:

total product = necessary product + surplus product

The **total product** is the total amount of goods and services produced in an economy during a given period of time.

The **surplus product** is what remains of the total product after the necessary product has been deducted.

Depreciation is the cost (due to wear and tear) of restoring the capital goods that suffered wear and tear in producing last year's output.

The **surplus product** is what is left of the total product after the necessary product has been deducted. But what is the necessary product? It consists of three parts.

First, part of the total product must be used for *replacement of materials used up in production*. This allows the next production cycle to begin with the same conditions that existed at the beginning of the present cycle, with the capacity to produce exactly the same level of output as before.

Second, because capital goods suffer wear and tear during the production process, another part of the total product is used for *maintenance and repair of capital goods used in production*. It is clearest to think of this maintenance and repair process as happening fully each year. However, in reality this may be postponed until particular machines or tools are ready to be replaced, so if the capital goods are not fully maintained each year, the annual reduction in the value of the capital goods is called **depreciation**.

These first and second parts of the total product are combined in Figure 4.2 and labeled *replacement of capital goods and materials*.

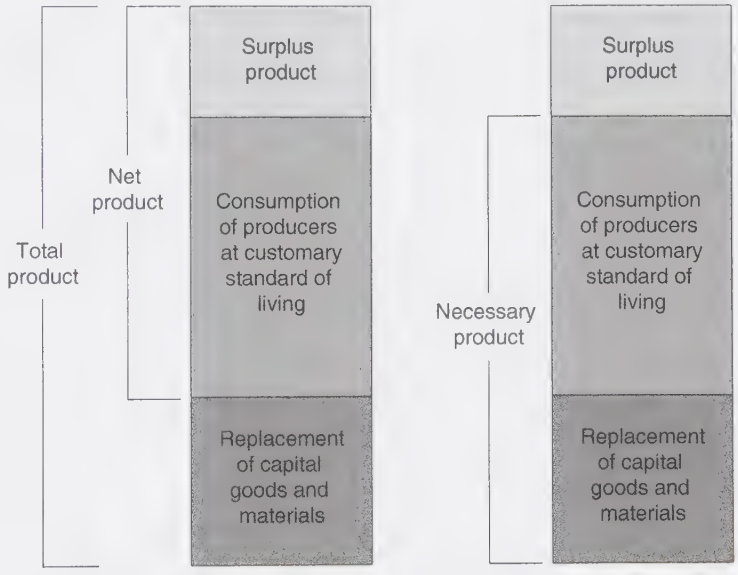


FIGURE 4.2 Total, net, necessary, and surplus product.

This figure shows how the total product of an economy (whatever its specific form) is allocated among various uses, and illustrates how key concepts are defined in the text. The components of the total product can be combined in two different ways. The bottom two segments combined are the *necessary product* (right hand side), while the top two segments combined are the *net product* (left hand side). We talk about the *necessary product* when we are interested in how much of the total product needs to be used just to maintain the previous level of production, by maintaining the workforce, the material inputs, and the equipment at the same level or in the same condition. The *surplus product* is then the remainder of the total product, and is available for investment in expanding the level of production, or for capitalist consumption or other purposes. On the other hand, if we are mainly interested in how much the value of the total product exceeds the value of the material inputs into it during a given year we use the *net product* (or *value added*) to measure this concept.

The third component of the necessary product is the *consumption of the producers*. This refers to that part of the total output—food, clothing, and other items—that must be distributed, through some mechanism, to those who produce the total product so that they may maintain themselves at their "customary" standard of living. It is shown as the middle segment in Figure 4.2. In a capitalist system, conflict between workers and capitalists (those who own the capital goods and control the work process) often plays a part in determining workers' customary standard of consumption.

The **necessary product** is the portion of the total product that is needed to maintain the inputs in the labor process—workers, materials, and tools—at their current level or in their current condition for the next round of the production process.

Together, the bottom two segments in Figure 4.2 are called the **necessary product**, the part of the total product required to sustain production at its previous level. The remainder of the total product, the top segment in Figure 4.2, is the *surplus product*, which in capitalism is owned and controlled by the capitalists. They can do as they choose with the surplus, such as spend it on luxury goods for themselves or reinvest it to expand productive capacity.

Putting together the three parts of the necessary product, it can be defined as:

necessary product	= **replacement of materials used up**	+ **depreciation of capital goods used**	+ **consumption of the producers at their customary standard of living**

One can see, then, that the word *necessary* in the term **necessary product** refers to what is required to maintain the inputs into the labor process—producers, tools, and materials—so that production may be carried on in the same way, at the same level, in the next period. The continuance of an economy at its current level is the basic idea of the necessary product. It does not refer to the adequacy of a society's living standards, the level of its technology, or the diligence of its workers.

Although we began this discussion referring to the total product of a society, total product is, in fact, not a very useful measure of economic output. This is because it includes that part of the total output that must be set aside to replace the machinery and materials used or used up in the production process. A more useful measure of output is **net product**, defined as the total product minus ("net of") the amount of goods and services that must be used for replacement of materials and depreciation of capital goods. Net product is sometimes called **value added**. Figure 4.2 summarizes the relationships among the concepts introduced here.

Net product (or **value added**, or **net output**) is the total product minus materials and capital goods used up in the course of producing the total product.

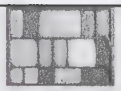
How, then, does the concept of the surplus product help us understand who produces what, and why? Will it help us ascertain why there is a surplus product? And will it help us answer a fundamental question about a capitalist economy, namely, *how is the rate of profit determined?* Chapter 10 will address this question in detail; here we just touch on several basic points.

If an economy produced only enough to support its producers and replace the machines and materials used up in production, it would not produce any surplus at all. In this case, the total product would be just equal to the necessary product, and there would be no surplus product. People lived this way a long time ago and, in fact, did so for most of the 100,000 or so years of human existence before the domestication of plants and animals. This ancient but long-lasting way of life preceded the emergence of hierarchies, the division of societies into classes, and the appearance of distinctions between political elites and those they govern.

Although many early human societies were far from affluent, a society without a surplus is not inevitably a poor one, nor must it be a "subsistence" society. As long

as the "necessary product" is being produced, a society can maintain its members at their "customary standard of living." Every society's customary standard of living is defined by its particular culture: it may be high or it may be low, as long as it allows the producers to meet their culturally defined needs. Moreover, the normal workday in any given society might be longer or shorter than eight hours; many hunters and gatherers work for shorter days than office and factory workers in the United States, for example. Whatever the length of the normal workday, the producers in a no-surplus economy are not supporting anyone but themselves.

Historically, the appearance of a surplus can be attributed to the emergence of some kind of power relationship. Depending on the type of economic system, the surplus product may take the form of the rents and feudal dues paid by serfs to feudal lords, it may constitute the income of slave plantation owners, it may be that portion



Superyachts and Bull Market Boats

Want freedom?

A superyacht, one advertiser says, offers "the freedom to explore the world in absolute privacy and luxurious comfort." A superyacht is defined as a yacht longer than about 80 feet (or 27 meters), and the standard version features sleek design, palatial bedrooms, and exquisite dining and lounging areas for large parties. Many have at least one large hot tub (and some have a second, perhaps smaller one). Each has its unique features: a grand piano in one, an in-water swimming pool enclosed by anti-jellyfish nets in another. Some are designed for extended journeys and are called *expedition yachts*, a term apparently first applied in 1996 to a yacht used for cruising in the Arctic. There is generally a *tender* (a service boat) on the yacht.

The magazine *Superyacht World* now recommends a range of optional accessories, many of them other vehicles: an amphibious Jeep-ish device in which one can get from the yacht to the beach and then travel on land; a sort of smallish submarine; a more or less motorized surfboard.

But how much does freedom (in this form) cost? Good question. Recent models are priced very roughly between half a million and a million dollars per meter of length: a 66-meter luxury yacht for $58 million, for example. The world's largest superyachts, about 200m in length, cost more—up to $600 million—to build, and tend to be owned by monarchs, princes, major film directors, famous actors, and successful businesspeople.

The population of *megayachts* (the old term) or *superyachts* (currently the favored term) ballooned from several hundred in the mid-1980s to over five thousand by the end of the 1990s, a fact that is attributed to the long "bull" market (the stock market boom). The 1 percent got richer. But it also helped that in 1993 the 10 percent tax on new boats costing over $100,000 was repealed, inspiring many wealthy boaters to trade up to bigger and more luxurious yachts.

One thing to remember: If you're going for the yacht, do budget for paying the crew (of as many as 150 on some of the largest superyachts), for dockage fees, for fuel (possibly half a million dollars a year). And if that puts the whole thing beyond your budget, relax: You can charter one of these floating paradises for a week for less than a million dollars.

Sources: "The Latest and Greatest Superyacht Accessories in 2016," July 20, 2016, https://www.fraseryachts.com/en/news-detail/the-latest-and-greatest-superyacht-accessories-in-2016/; listings of yachts for sale, https://www.fraseryachts.com/en/yachts-for-sale/; Alicia Adamczyk, "How Much Does a Superyacht Really Cost?," April 8, 2015, available at www.forbes.com/sites/aliciaadamczyk/2015/04/08/how-much-does-a-superyacht-really-cost/#418fe99536bc; Doug Gollan, "Sneak Preview: Best New Superyachts You Can Charter," *Forbes*, April 4, 2015; Doug Sease, "Bull-Market Boats: Size Does Matter," *Wall Street Journal*, April 23, 1999, p. W1; Penelope Green, "Go-Anywhere Luxury Yachts Bull Out to Sea," *New York Times*, July 22, 2001.

of the harvest that is turned over by sharecroppers to their landlords (less any inputs the landlord has supplied), or it may appear as profits for the owners and top executives of today's business corporations.

The surplus product is important because it allows for economic growth and change. All or part of the surplus may be used to improve or expand the stock of capital goods, making possible either increased output or shorter working hours or both in the future. Devoting the surplus to research activities may expand the stock of knowledge relevant to production. When the surplus is used in these ways to increase productivity, it is called **investment**. When the surplus is invested, it can make it possible for people to work less hard or for fewer hours without experiencing a reduction in their living standards. Indeed, such productivity-enhancing investment can even support increases in the living standards of the producers without reducing those of the class in power. An enlargement of the surplus may also allow for population growth: instead of the system simply reproducing the current number of producers, it could support a larger number of people at the same standard of living.

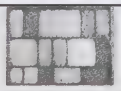
Investment means spending money to add to, or improve, a firm's productive equipment, software, facilities, or workforce skills in order to increase productive capacity and productivity.

Of course, it is also possible that the surplus will be used in ways that do not increase productivity. For example, all or part of it may be used by those who control it to engage in displays of opulence. In the world today, an important example of the use of the surplus in a way that does not increase productivity is *capitalist consumption*. (See box, "Superyachts and Bull Market Boats.")

A Model of Production and Reproduction

The idea of the surplus product is so important that we need to make it more precise. We will do this by means of a simple example—a model—designed to bring out the most significant points while ignoring what is not essential to our present inquiry.

The point of our model is to clarify the relationships among the concepts introduced in this chapter: specialization, technology, technical change, replacement of inputs (materials and capital goods), reproduction of producers, and the total, net, and surplus products.



Economic Models

Economic models, like all scientific models, are attempts to simplify a problem so that we can understand it better. The purpose of a model is to represent what is important for the task at hand, not to represent every detail of the problem. For example, an airplane model used in a wind tunnel does not have seats in it, but great care is taken with every detail concerning the shape of its fuselage, wings, and tail. Similarly, an architectural model focuses on the visual and spatial aspects of a proposed building; the fact that the model is made of wood and plastic, not the concrete and steel that will go into the building itself, is unimportant.

Models allow us to think about complex problems in simple ways. Economic models that successfully highlight what is important without ignoring essential aspects of a problem are tools as necessary to economic reasoning as the telescope is to astronomy. However, models can be very misleading when they are based on unrealistic assumptions or when they oversimplify things to such an extent that essential aspects are left out. Economists would do well to heed Albert Einstein's advice: "Make things as simple as possible, but no simpler."

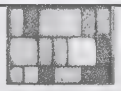
To simplify things we will focus on the labor processes of a single family that is part of a society that includes many similar families. Our representative family engages in two kinds of labor processes, child-rearing and farming, but produces only one good, grain. Another simplifying assumption in our model is that grain is the only input used, other than labor and nature, in the labor processes. These assumptions will allow us to highlight various economic relationships without introducing more complicated elements, such as money.

In our model, farm work produces grain, and once it is produced the grain serves three different functions. First, it is a *consumer good* used in the reproduction of people: people eat the grain (as bread) and they drink the grain (as beer). Second, it is an input used to feed the draft animals employed in the fields so they will be able not only to aid in the current production of grain but also to reproduce themselves, ensuring that there will be draft animals to help with future crops. In this situation the draft animals may be thought of as *capital goods*. Third, grain is a *material input* because grain itself must be used as seed for the next crop. Summarizing the different roles that grain plays in the production sector of our simple model, we can represent the various inputs and outputs as constituting a technology.

The way the grain technology is presented here makes it easy to distinguish between total product and net product: *total product* is all the grain that is produced, while *net product* is all the grain produced minus the amount of it that must be put aside to replace whatever is used or used up in production (in this case the feed for the draft animals and the seed for next year's crop).

Is our model family producing a surplus? In order to answer this we will have to determine how productive the technology is, how hard and how long the family members work at grain growing, and what their customary standard of living is.

Suppose the family comprises two adults who together spend a total of 1,000 hours of labor on child rearing and 1,000 hours of labor growing grain each year. Also suppose that it takes 10 hours of labor to produce a bushel of grain. (This is unrealistically high for a modern economy, as a glance at Figure 1.1 will show.) Under these assumptions, the family will harvest 100 bushels of grain each year.



The Grain Technology

Inputs:	
Labor	Planting, harvesting
Materials used up	Grain used as seed
Capital goods used	Grain for maintaining and reproducing draft animals
Inputs from nature	Land, rainfall, sunshine
Outputs:	
Consumer good	Grain to be consumed
Materials	Grain to be used as seed the following year
Capital goods	Grain to be used as food for draft animals

Now imagine that the harvest is in, and the 100 bushels make up a huge pile of grain. The family divides this large pile into three smaller piles: the first pile must be large enough to replace the materials used up and maintain the capital goods used in production (seeds and draft animal feed); the second pile is the amount set aside for the family's own consumption (at the customary level) until the next harvest; and the third pile, whatever is left over, is the surplus. The first and second piles together are the necessary product, while the second and third piles together are the net product. (The different piles in this example correspond to the three segments of the bar in Figure 4.2.)

One possibility is that the model family owns its farm and is therefore in control of the surplus it has produced. In this case the family may choose to do whatever it wants with the surplus, which may be anything from improving the quality of the technology it uses to building a swimming pool in the backyard. In Chapter 7 we examine a time in American history when independent or self-employed producers of commodities—farmers, craftsmen, self-employed artisans, and other independent producers—accounted for approximately 60 percent of all people participating in the production sector.

But the other possibility, the one that corresponds more to realities in the world today, is that the family is working someone else's land and has to get by with the portion of the total output that is allocated for consumption by the producers. In this case, who gets the surplus? This is one of the most important questions addressed by political economy.

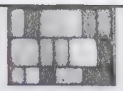
Who Gets the Surplus?

Much of human history has been about battles over the surplus. In some economies the surplus product is claimed as rent by a landlord, and so in many farming communities the third pile is literally trucked off to the landowner's warehouse. Or perhaps the family in our model is employed by a large agribusiness corporation. In this case the family gets only the second pile, its wages; the corporation sets aside the first pile for replacement and claims the third pile, the surplus, as its profits. Alternatively, the family in our model may have taken out a loan from a bank, perhaps in the form of a mortgage, to purchase its land and draft animals. In this case the surplus may go entirely to the bank. The important point in each of these cases is that the producers themselves do not get the surplus. We will return to this point in greater detail when we discuss classes and economic systems in the next chapter.

To see how big the surplus pile will be, imagine, in the first instance, that 30 bushels must be set aside for replacement of materials and capital goods, specifically to provide seed for the next year and to feed the draft animals used in tilling the fields. (The allocation here is 10 bushels for the seed and 20 bushels for the animals.) Subtracting the 30 bushels needed for replacement from the total product of 100 bushels, the net product is 70 bushels. If the customary consumption level for the family is 50 bushels per year, the surplus product will be 20 bushels.

Enlarging the Surplus

How might those who control the surplus make it larger? There are a number of possible answers to this question. First, the standard of living of the farming family might be cut, say, from 50 to 40 bushels per year. In this case the surplus would rise from 20 to 30 bushels. This illustrates why, in any society in which some people control and benefit from the labor of others, there will be an incentive for those who



The Surplus Product: An Example

(1) Total hours of labor devoted to grain growing per year:	1,000 hours
(2) Bushels of grain produced per hour worked:	1/10 bushel/hr.
(3) Total product of grain per year [(1) × (2) = (3)]:	100 bushels
(4) Replacing materials used up + maintaining capital goods used in producing the total product (seed + draft animal feed) = (Pile #1):	30 bushels
(5) Net product per year [(5) = (3) – (4)]:	70 bushels
(6) Customary consumption level of the family = (Pile #2):	50 bushels
(7) Necessary product [(7) = (4) + (6)]:	80 bushels
(8) Surplus product [(8) = (3) – (7) = (5) – (6)] = (Pile #3):	20 bushels

control the surplus to try to reduce the standard of living of those who produce it, setting up a situation that is inherently conflictual.

The consumption of the farming family in our model can be reduced from 50 to 40 bushels without changing the total amount of grain produced if a **labor-saving technical change** is introduced. A labor-saving technical change is defined as a change in the technology that increases the amount of output that can be produced with a given amount of labor. But the benefits of a labor-saving technical change can be taken either as an increase in total product using the same amount of labor or as production of the same total output with less labor. Here we assume that a labor-saving technical change increases the productivity of labor by 25 percent and that those who control the labor process choose to maintain total output at 100 bushels while reducing the amount of labor employed from 1,000 hours to 800 hours. This allows reducing the amount of grain used for producer consumption from 50 to 40 bushels, since only 800 labor hours are now needed to produce the total output of 100 bushels. (We need not worry that the family will go hungry; the saved labor could be used to fish in nearby streams.)

A **labor-saving technical change** is a new technology that increases the total output produced with a given amount of labor.

A second way of enlarging the surplus would involve changing the technology in such a way that fewer seeds or less animal feed would have to be set aside for the next year. This technical change could take the form of better seeds, or it might be a new way of storing the animal feed so that less is lost to spoilage. In our example, if only 20 bushels of grain (instead of 30) had to be set aside to replace the capital goods and materials used up in production, the surplus product would increase from 20 to 30 bushels (assuming no increase in the standard of living of the producers). What we have in this case is a **capital goods–saving technical change**, defined as one that reduces the amount of capital goods or the quantity of materials required to produce the total product.

A **capital goods–saving technical change** is a new technology that reduces the capital goods and materials necessary to produce a given amount of total output.

The two ways of enlarging the surplus product by means of technical change are shown, along with our example's original allocation, in Figure 4.3. This figure uses bars divided in the same way as the bar in Figure 4.2. The left-hand bar represents the division of the total product in our original example. The middle bar shows how, with

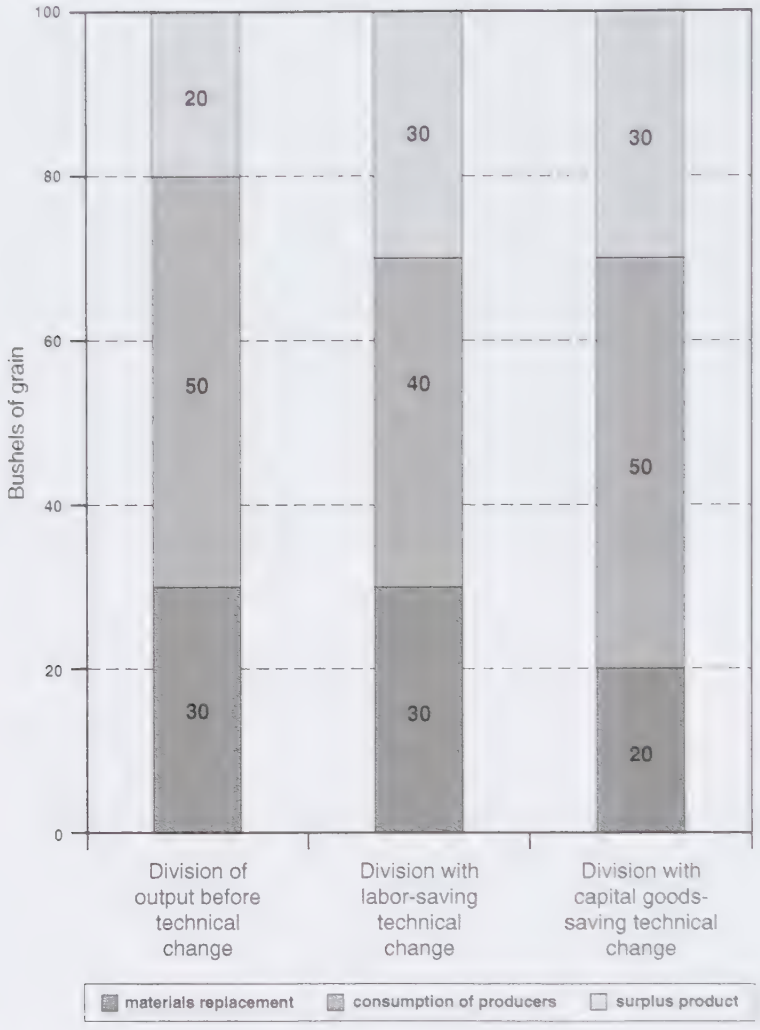


FIGURE 4.3 Division of the representative farm family's total output.

The left-hand bar represents the division of total output as set forth in the original example in the text: total output is 100 bushels, with 30 used for replacing the material inputs to the production process (capital goods depreciation as well as materials used up), 50 allocated for the consumption of the producers, and 20 remaining as the surplus product. The middle bar shows the division of total output after there has been a *labor-saving technical change*. Since there has been an increase in labor productivity, the same total output (100 bushels) can now be produced with less labor (800 rather than 1,000 hours) so only 40 bushels (rather than 50) need be allocated for the consumption of the producers. The right-hand bar shows what happens when there is a *capital goods-saving technical change*. In this case, both the total output (100 bushels) and the consumption of the producers (50 bushels) are unchanged from the original example, but the reduction in replacement costs from 30 to 20 bushels allows the surplus to be enlarged to 30 bushels.

a labor-saving technical change, the surplus product can be enlarged to 30 bushels while the consumption of the producers is reduced to 40 bushels. The right-hand bar shows the surplus product expanded by means of a capital goods–saving technical change. An important point here is that the surplus can be increased with *either* a labor-saving technical change *or* a capital goods–saving technical change.

Just as a labor-saving technical change does, a capital goods-saving technical change results in more net product produced for any given amount of labor performed. Accordingly, these two types of innovations make it possible for both the surplus product and the standard of living of the producers to be increased. Many technical changes are both capital saving and labor saving, meaning that they reduce the amounts of all of the inputs needed to produce a given quantity of total product.

Still another way of increasing the surplus product is to (somehow) induce the producers to work harder without increasing the total number of hours that they work or the portion of the total product allocated for their consumption. (If we assume that the producers are paid according to the number of hours they work, keeping their hours constant will mean that they will continue to receive the same amount of grain for their consumption, even if they are working harder.) This way of enlarging the surplus product is called an increase in the **intensity of labor**.

The **intensity of labor** is how much work effort producers must expend per hour of work, or more simply, how hard they work.

Referring to our grain model once again, if the intensity of labor were to be increased so that the 1,000 hours of grain-growing labor produced 130 rather than the previous 100 bushels, the surplus would increase from 20 to 50 bushels (assuming that the proportions of the total product allocated to producer consumption and replacement of nonhuman inputs remains the same).

A final—and perhaps the most obvious—way in which the surplus might be increased is one that is not based on any increase in the productivity or living standards of the producers. They can simply be induced to spend more hours growing grain. This could be achieved in two ways. First, the length of the working day (or the total number of hours worked in a year) could be increased, in which case the amount of free time (time available for leisure activities, eating, sleeping, etc.) would be reduced. The second possibility, which might well happen as a result of the first, is that the number of hours spent by the family on child-rearing activities (or reproduction in general) could be reduced, with the hours saved being reallocated to grain growing activity.

Application of Model to Labor

Much of the analysis we have applied to the production sector may be used just as effectively to analyze the 1,000 hours spent on child-rearing and other reproduction activities by the family in our model. For example, we have noted that a reduction in the number of hours worked in the reproduction sector allows for an increase in the surplus product because those who are in control of production may then get the producers, who are mostly the same people as those involved in the reproduction activities, to work longer hours producing grain.

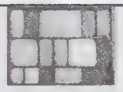
Child-rearing and other reproduction activities are obviously necessary to enable families to continue supplying labor to the production sector. By using analyses like the ones we have applied to activities in the production sector we can speculate about the consequences of introducing an innovation such as a labor-saving technical change in child rearing and other reproduction processes. For example, a TV set might allow parents to spend fewer hours with their children, a vacuum cleaner might enable the family to keep the house clean with fewer hours of labor, the acquisition of frozen foods or an automatic dishwasher might reduce the amount of time that has to be spent in the kitchen, and so on. We can also imagine several families getting together to share day care responsibilities, thereby reducing the number of hours each individual family needs to devote to child-rearing activities. All of these examples

illustrate the point that production and reproduction activities are interconnected in many ways. Accordingly, changes in the reproduction sector may affect—or be required for—changes in the size of the surplus product.

The Surplus Product and Conflict

In a society that produces a surplus product, two things are very likely: conflict and change. Conflict within nations arises when the producers are not allowed to keep the surplus that they have produced. Under such circumstances the producers' interests are opposed to the interests of those who own and control the surplus. The latter—whether they be slave owners, feudal landlords, the owners of capitalist corporations, or dictatorial governing elites—will probably seek to enlarge the surplus. To do this they will need to get the producers to work harder, work longer, or consume less. The producers, however, would like to enjoy less arduous work, have more free time, and benefit from a higher standard of living. Thus, what they want is in direct opposition to what those who control the surplus want. The producers and those who control the surplus constitute different social classes, and the conflict between them can be understood as a class conflict. (We discuss classes and class conflict in subsequent chapters.)

The conflict between the interests of the owners of a surplus and the interests of the producers of that surplus can be better understood if we list the determinants of the size of the surplus product and contrast these with the determinants of the economic well-being of the producers. (See box "The Surplus Product and the Producers' Well-Being.")



The Surplus Product and The Producers' Well-Being: A Recipe for Conflict

How to Increase the Surplus Product	How to Increase the Producers' Well-Being
Reduce the level of consumption of producers.	*Raise* the level of consumption of the producers.
Implement a *labor-saving* technical change.	*Share* the resulting increase in the surplus product with the producers.
Implement a *capital goods-saving* technical change.	*Share* the resulting increase in the surplus product with the producers.
Increase the intensity of labor.	*Reduce* the intensity of labor.
Lengthen the labor time of the producers (without changing their customary consumption level).	*Shorten* the labor time of the producers (without changing their customary consumption level).
Reduce the real price of an imported input (by assertion of power over the exporter).	*Share* the resulting increase in the surplus product with the producers.

As we can see from the accompanying box, the producers will be better off if their customary consumption level is higher, if their work is less intense, and if they have more hours for leisure and for child-rearing. These are the determinants most relevant to the economic well-being of the producers. However, just the opposite movements in these same determinants are required if the surplus product is to be increased. Here, the conflicts of interest between those who produce and those who control the surplus product are evident.

The box also shows that there are some determinants that, if increased, would benefit the surplus owners without necessarily hurting the producers (possibly even allowing an improvement in the standard of living of the producers while at the same time expanding the surplus product). Labor- or capital-saving technical changes would fall into this category. So would a decline in the price of an imported input (oil or computers, for example), which would reduce the portion of the total product needed to replace the capital goods or materials used in production. This could well have knock-on effects on the living standards of those in the importing and the exporting nations; but that is beyond the scope of this chapter.

If the customary consumption level of the producers is very low, of course, the workers may be so poorly nourished, weak, or unhealthy that they cannot produce very much. Likewise, if they work so many hours in grain growing that they have little time to take care of themselves and raise their children, they may not be very productive, and the next generation may be even less so. In these situations those who control the surplus may see that it is not in their interest to impoverish or overwork the producers, for poverty and overwork can serve no group's interests. If consumption is barely above, or actually below, the subsistence level, it is likely that an increase in consumption would *both* improve the living standards of the producers *and* enlarge the surplus product. From the point of view of the surplus owners in this set of circumstances, then, an increase in the producers' consumption would more than pay for itself, and for this reason an increase in their consumption is likely to occur.

Using the numbers from our grain model (when the net product was 70 bushels), Figure 4.4 illustrates the conflict that exists between the producers and those who control the surplus. Reading the graph from right to left, the point at 70 on the horizontal axis shows what would happen if the entire net product were to be distributed to the producers for their consumption. In this case the surplus product would be zero. The part of the line that slopes upward to the left from this point shows that as the consumption of the producers is reduced from 70 to 25 bushels, taking them well below their customary consumption level of 50 bushels, the surplus rises from zero to a maximum of 45 bushels. (The net product remains at 70 bushels; each bushel cut from the producers' consumption is simply transferred to the surplus product.) Beyond this point, however, any further reduction in the producers' consumption level will diminish their ability to produce grain, and the surplus product will fall, eventually reaching zero when the producers' consumption is down to 15 bushels. At this point the net product itself is only 15 bushels, and in order for the society to continue to exist, all of it has to be devoted to sustaining the producers, however inadequately.

If Figure 4.4 is read in the other direction (from left to right), it illustrates how, at very low levels of consumption, the interests of the surplus owners and the producers may coincide. The part of the curve that slopes upward to the right shows that any increase in the consumption of the producers between 15 and 25 bushels will benefit both groups. At 15 bushels of consumption there is zero surplus product, whereas at 25 the maximum of 45 bushels of surplus is produced. Beyond this point, of course,

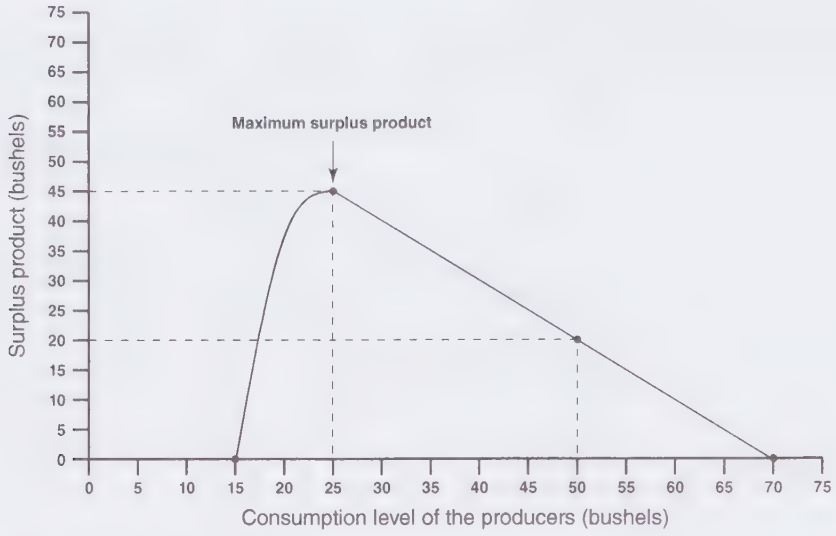


FIGURE 4.4 Conflict over the consumption level of the producers.

This figure shows a hypothetical relationship between the level of consumption of the producers and the size of the surplus product. Each point on the line in the graph is an answer to the hypothetical question: for each level of consumption by the producers, how large will the surplus product be? If the producers' consumption level is 50 bushels, for example, the surplus product will be 20 bushels (as it was in our presentation of the grain model). If the producers' consumption increased to a level of 70 bushels, the surplus product would be zero. Reducing the level of consumption will increase the surplus, but only to a point. The surplus reaches a maximum of 45 bushels when the consumption level is 25 bushels. Consumption levels below 25 bushels actually reduce the surplus product. If the consumption level were to fall to 15 bushels, the workers would be so impoverished that they would not be able to produce more than 15 bushels of net product a year, leaving no surplus product.

further increases in the producers' consumption will reduce the surplus product. This is because once the producers are able to consume 25 bushels of grain they are fully productive, the net product is at its maximum of 70 bushels, and each bushel added to the producers' consumption beyond this level means that there is one less bushel left for the surplus product.

The Surplus Product and Change

The necessary product (as defined above) is dedicated to uses that are conservative in the literal sense of the word: it is used to reproduce existing conditions. The surplus product, on the other hand, may be used to alter the status quo. Such change can take various forms. In capitalist societies a significant portion of the surplus is *invested*. This means that the surplus is used to increase the productive capacity of the economy—by paying for education and training, or for more and better capital goods, or for invention of new technologies. When it is used in any of these ways, the surplus benefits not only those who control it but others as well, and conflicts over its production are therefore diminished.

Other uses of the surplus are conservative in a different sense. Not only do they not produce change, they work against it and may even involve using force to maintain the power of the dominant class. When, for example, the surplus is used merely

to support extravagant consumption by the wealthy, it is not promoting change; it is benefiting only the elite (and luxury good producers). When the surplus does not go into supporting the private consumption of the wealthy but is used to build monuments, temples, or churches, it may enrich the culture or serve the spiritual needs of the population, but it also may have the effect of reinforcing the status quo.

In a capitalist economy the way the surplus is typically used produces rapid change. The reason is that those who control the surplus (mostly the owners of large corporations and the people they hire to manage their affairs) are engaged in a relentless competition for survival as capitalists, and the key to survival is innovation—the introduction of new products, new technologies, new information, or new ways of organizing businesses. The elites of other economic systems—slavery, feudalism, or communism, for example—had it easier: competition was less intense and more easily controlled, and what these elites needed to do in order to win was not to innovate but rather to maintain their position in society.

Those who control the surplus product usually have a considerable amount of political power. Because they have money to spend, they wield a lot of influence over the actions and policies of the government. They also use the surplus to maintain the upper hand in the international and class conflicts that arise from the process of producing and claiming the surplus. Thus, to preserve the existing structure of society—and the corresponding distribution of wealth and income—part of the surplus is used to pay for the services of police, armed forces, security personnel, work supervisors, and even teachers and religious leaders who enforce and maintain the prevailing power relationships or propagate the ideology that explains and supports these relationships.

Our model of the surplus product, technology, labor, and consumption illustrates the relationships among these concepts almost as an architect's drawing illustrates the spatial relationships among the parts of a building, but it does not do much more than that. It does not tell us why the technology is what it is, why the members of the family work a total of 2,000 hours annually, why they work as hard as they do, or how the "customary" level of consumption is established. The model does not even determine who gets the surplus (if there is one) or what is done with it. The job of the model is to help us *ask* these questions in a precise way, not to answer them. To answer these questions we need to take up the idea of class and ask how class relationships define economic systems. We do this in the next chapter.

Suggested Readings

Perry Anderson, *Lineages of the Absolutist State* (London: Verso Books, 1976).

Christopher Boehm, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Cambridge, MA: Harvard University Press, 2000).

Nancy Folbre, *Who Pays for the Kids? Gender and the Structures of Constraint* (New York: Routledge, 1994).

E. J. Hobsbawm, *Industry and Empire: The Making of Modern English Society* (London: Penguin Books, 1968).

Margaret Levi, *Of Rule and Revenue* (Berkeley: University of California Press, 1988).

Karl Polanyi, Conrad Arensberg, and Harry Pearson, eds., *Trade and Market in the Early Empires* (New York: Free Press, 1957).

Marshall Sahlins, *Stone Age Economics* (Chicago: Aldine Publishing, 1972).

Paul Sweezy, *The Theory of Capitalist Development* (New York: Monthly Review Press, 1942).

Charles Tilly, *Coercion, Capital, and European States, AD 990–1990* (Cambridge, MA: B. Blackwell, 1990).

Capitalism as an Economic System



Cambridge, England—home to one of the world's great universities—sits at the edge of a low plain stretching north and east to the sea. Here the earth itself reveals how capitalism has transformed society. Before the advent of capitalism this marshy, swampy ground (called the "fens") had poor drainage, and in the rainy season much of it lay flooded and unusable. A town called Ely grew up on a small hill—the "isle of Ely"—that usually escaped the flooding.

In the Middle Ages a part of the surplus product of feudal England was used at Ely to build the first of several churches. In the ninth century an abbey was established, burned down, and then was rebuilt. Soon after the Normans conquered England in the eleventh century, workers under their command began building the great cathedral that stands on the site today. Massive and elegant, the cathedral consumed the labors of thousands of artisans and required 110 years to build. But it was never really finished: new chapels were built, more elaborate furnishings were added, and the tower was heightened. Its west tower now soars 200 feet above the surrounding floodplain (and so would its east tower if it had not collapsed in the fourteenth century). For centuries it stood as a dry monument to the religious use of the surplus, while all about it the waterlogged fens frustrated cultivation.

Then, in the seventeenth century the fens themselves were transformed. The Duke of Bedford, owner of some 20,000 acres in the area, sensed the possibility for profits and organized an investment company to drain the fens. Engineers were hired, and workers dug two huge ditches, 100 feet wide and 30 miles long, to carry the excess water to the sea. Many additional miles of feeder ditches and channels carried water to the main ditches. This colossal construction project transformed the marsh into well-drained and fertile agricultural land and, as a result, made the duke a very rich man.

The modern visitor to Ely can see both marvels, the cathedral and the ditches. The cathedral, magnificent and dominating, represents feudalism, absorbing the region's surplus product but leaving the economy unchanged and stagnant. The less glamorous ditches represent the capitalist enterprise that invested the surplus product and drained the fens, repaying the investors many times over and supplying food to the region. The important point is that the fen ditches, humble and easy to overlook,

revolutionized the entire system of production in the area. They are an example of the productive use of the surplus product characteristic of the capitalist system.

The story of the fens has a distinctly twenty-first-century (and noncapitalist) twist: in 2003 the Royal Society for the Protection of Birds announced plans to acquire most of the area from the "barley barons" who now farm it and to restore the fens' rivers, pools, and marshland, making it a homeland to the marsh harriers, black tailed godwits, and other bird species that disappeared when the Duke of Bedford drained the area four centuries earlier.

In this chapter we investigate how different societies determine the uses of their surplus products and who benefits from the ways they are used. Often people who do not produce the surplus product nonetheless control it. This was the case, for instance, in ancient Athens as well as in the U.S. South before the Civil War, when slaveholders lived off the products of their slaves' labor. The relationship between those who produce the product, including the surplus product, and those who control the use of the surplus product is called a *class relationship*. Class relationships exist between different *classes*, defined as groups of people who share a common position in the economy. The producers, excluded from control, are one class; the controllers of the surplus product are another class.

Of course, the distinction between producers and controllers is not always as clear-cut as it is on a slave plantation. The owners of a large capitalist corporation, for example, may participate in the production of the surplus in so far as they perform socially necessary tasks such as coordinating the work of many producers. At the same time, because of their position of power in the economy, they also have control over the disposition of the surplus product. Complications such as these mean that it is not always easy to draw clear lines between one class and another.

In any case, a society that is based on an economic system in which most producers exercise little or no control over the use of the surplus product is a *class society*. Capitalism is a class society. Though different from U.S. capitalism, the economic system of China today makes it a class society, too. In both China and the United States the surplus product is produced by the work of millions of workers of all descriptions. But in both countries, even though their economic systems are very different, the disposition of the surplus product is controlled by a relatively small group of people.

The main idea of this chapter is that *economic classes are defined by how the surplus is controlled and used, and classes and the surplus product are the keys to understanding how different economic systems work and how they change.*

This main idea is expressed in four primary points.

1. A *class society* is one divided into two groups of people: those who *produce* the product but do not control the use of the surplus product, and those who *control* the use of the surplus product. In such a society class is the principal aspect of the vertical, or command, dimension of the economy. (Other vertical relationships, such as racial or sexual domination, are also important, and we will discuss them in later chapters.)

2. *Economic systems* represent different ways of organizing and controlling labor processes based on various systems of ownership, and these property systems in turn determine how the surplus product is used. The way in which the surplus product is controlled, and by whom, are what distinguish one economic system from another. Thus, each economic system will have a distinct set of classes—and a dominant *class relationship*—associated with it.

3. *Capitalism* is a particular economic system in which the private owners of *capital goods*, employers, hire *wage labor* to produce *commodities* for the purpose of making a *profit*. In this system there are two main classes, workers and capitalists, although other classes (the "middle" classes) are also important.
4. In capitalism the surplus appears as *profits*. Profits may be either retained within a firm and reinvested to expand production or paid out as rent, interest, dividends, or other payments to the owners of the capital goods used in production. Both those who receive such payments and those who control any retained profits may either save their money or spend it to buy the commodities, including capital goods, that make up the surplus product.

At Ely the various uses of the surplus product left visible reminders of the differences between feudalism and capitalism. The cathedral is hard to miss; the ditches, by contrast, are not so obvious unless we know where to look and what we are looking for. Similarly, while it is reasonably easy to see how the feudal economic system worked, understanding capitalism is more difficult unless we know where to look and what to look for—in this case classes and the surplus product.

Class and Class Relationships

The word *class* is used in many different ways. If a celebrity is said to have class, it may mean that the person dresses stylishly or is simply cool-headed and respectful of others. A family that calls itself middle class may mean that it maintains a certain living standard and perhaps is buying a home. In political economy, however, the word has a very precise meaning: a **class** is a group of people who share a common position in the economy with respect to the production and control of the surplus product.

A **class** is a group of people who share a common position in the economy with respect to the production and control of the surplus product.

Thus, a **class relationship** exists when one group of people—one class—produces the total product, including the surplus product, while another group—another class—controls the use of the surplus product. Both groups, the *producers* of the surplus product and those who *control* its use, are classes. Different classes have existed throughout history. During the Middle Ages, the serfs who worked the land in medieval Europe constituted one class, their feudal lords a different one. In slave-holding societies (both in ancient Rome and in the United States before the Civil War), slaves were one class, slaveholders another.

A **class relationship** exists between the producers of the total product, including the surplus product, and those who command the use of the surplus product.

Four aspects of the concept of class are important. First, every class is defined in terms of a relationship. It is like the word *cousin*, which refers to someone, but which is meaningless unless there is another person who is also a cousin. You cannot be your own cousin; you cannot be a cousin by yourself; you can be only your cousin's cousin. A class cannot exist by itself; it can exist only in relationship to some other class.

Second, a class relationship refers to a labor process. Classes are defined by the particular positions that they occupy with respect to the labor process. A class is not defined either by its status or by the social esteem that others may attribute to its members, nor is a class defined by the income of its members: the rich and the poor are not classes, though it is generally the case that most of the rich are in one class and most of the poor in another. A status group such as the "well educated" or an income group—say, those receiving more than $50,000 a year—is a collection

of people defined by some characteristic. However, they are not defined by their relationship to another group in the labor process. For this reason they are not classes.

Members of a class may be very different from one another. They may live in different parts of the country or even in different areas of the world, they may be more or less well off, they may have different lifestyles, and if they are workers they may have very different jobs. They may even be in competition with one another—to find jobs, to sell products, or in other ways. Thus, it is possible that the only thing two members of a class will have in common with each other is their relationship to the labor process and to the other class (or classes).

A number of social groups very important to economic analysis are not classes. Nations are not classes, races are not classes, and sexes (genders) are not classes: the inclusion (or exclusion) of people in (or from) such groups is not based on their particular position in relation to a labor process. Females and males, for example, are defined biologically and culturally, and though men and women often do different kinds of work, it is not their position in a labor process that identifies them as men or women.

Third, class relationships are hierarchical, or vertical: they refer to a group on top and a group below. What is vertical about the relationship? Why do we call it hierarchical? The group on top is not necessarily happier, nor are they better people or more productive. Consider the slaveholders, for example. A group is said to be on top because it controls the labor and the products of those below. This is why class relationships are central to what we call the "command," or vertical, dimension of an economy.

The slaveholder employed the overseer who controlled the work of the slaves. What the slaves produced, in turn, belonged to the slaveholder. In the medieval era serfs worked about half of each week in the lord's fields, where the lord's bailiff directed their work. What the serfs grew on the lord's land belonged to the lord. Today, the owner (or owners) of a company, usually working through a board of directors, selects the management that hires and directs the work of the company's employees. The commodities produced by the employees are then owned by the company's owner (or owners).

In each economic system, of course, the producers must consume some of the output. This may be either a part of what they themselves have produced or goods and services produced by other workers. Slaves and serfs were allowed to till small plots of land on the plantations or estates on which they lived. In this way they were able to produce for themselves most of what they ate and wore. In contrast, workers today must buy most of what they consume in markets using the wages or salaries they are paid in return for their work. Note that in each case the producers—slaves, serfs, or workers—produce more than they consume. Correspondingly, the class on top receives something that it did not produce.

Those on top may also participate in production. Capitalist employers, for example, may put in some hours on the shop floor. As noted earlier, top executives perform work when they function as coordinators of the production process. But the income of employers does not flow primarily from productive work. Rather, it derives mainly from their ownership of the capital goods used in production and their control over the labor of others.

Fourth, the interests of producing and controlling classes are usually, but not always, in conflict. In most cases a gain for workers will show up as a loss for employers, but there are exceptions to this general rule. In certain situations both classes will have a common interest in relation to classes in other countries. Also, as we have seen in the previous chapter, both classes may benefit from some forms of

technical change. Finally, almost everyone will be better off if a country's government is honest and efficient.

There are often more than two classes in a society, so usually the class relationships are more complex and involve more than just the relationship between the producers and those who control production. For instance, in the nineteenth century in the United States, an important relationship existed between the (northern) working class and the (southern) slaves. The relationship between these two classes was not vertical—the slaves did not live on the surplus product of the workers, or vice versa. Their relationship was partly one of class interdependence, since slaves grew the cotton needed in the mills where workers were employed. Also, the mills' demand for cotton provided the revenues that kept the slave plantations going, and the textiles produced in the mills became the slaves' clothing. Political relations also existed between these two classes, as when workers volunteered in the northern armies to fight for the emancipation of the slaves. However complex the relationship between slaves and workers may have been, both of these classes were defined with respect to a particular vertical relationship: slaves to slaveholders and workers to employers.

Classes and Economic Systems

What is an economic system? We have already named several different economic systems: slavery, feudalism, self-sufficient households, and capitalism. We also know that what they have in *common*—what makes them economic systems—is that they organize labor processes. But what makes them *different* from one another?

Property rights establish the owner's right to control the property, to decide who uses it for what purpose, and to benefit from its use or sale.

Each economic system has a distinct set of class relationships, and each set of class relationships is identified with a specific way of organizing and controlling the system's labor processes. This combination of class relationships and associated methods for controlling labor processes is usually expressed legally in a set of **property rights**. Property rights, in turn, are what establish an owner's right to control property, to decide who can and cannot use it, to determine the purpose for which it is used, and to benefit from its use, improvement, or sale.

Slavery

Slavery was the dominant economic system in the U.S. South before the Civil War; slaveholders obtained the surplus product by owning all of the inputs (including slaves) and the output of slave production.

For instance, **slavery** is an economic system defined by the class relationship between slaves and slaveholders. Inherent in the very definitions of *slaves* and *slaveholders* is a particular method for organizing labor processes based on the fact that the human input to production is itself owned as a piece of property. Thus, the slaveholder was a slaveholder because of his or her ownership of a particular piece of property (a slave), and this ownership conferred certain rights on the owner: the right to direct the slave's labor, the right to own the product of the slave's labor, the right to own any children the slave might bear, and the right to sell the slave.

In slavery as it existed in the American South before the Civil War, there were not only the two principal classes, slaves and slaveholders. There were also other classes such as non-slave-owning white farmers, overseers, urban craftspeople, free blacks, and merchants. However, the slaveholders were the dominant class, and, as noted, they owned all of the inputs to the slave labor process—including the laborers themselves. The slaves produced the total product, all of which was also owned by the slaveholders, and this consisted mainly of tobacco, rice, and raw cotton. Food was also grown on the plantations, and some of this food, together with other supplies purchased with revenues

derived from the sale of the crops, was used to feed, clothe, and house the slaves. This portion of the total product, then, provided for the customary level of consumption for the slaves. Another part of the output had to be set aside to feed mules, replace worn out tools, or otherwise replace capital goods that had been used up in producing the output. What remained—the surplus product—was left for the slaveholder to use for personal consumption, for the purchase of more slaves, for the construction of beautiful mansions, for education of the slaveholder's children, for foreign travel, and so on.

Feudalism

Feudalism was the dominant economic system in Europe in the Middle Ages; lords obtained the surplus product through rents and other customary obligations owed by the serfs.

Another type of economic system, **feudalism**, was how most economic activities were organized in Europe during the Middle Ages. As in slavery, there were two principal classes in feudalism, the lords (either church officials or members of the "noble" class) and the serfs (or peasants). Other classes in feudal Europe included artisans, bailiffs and other people in a lord's retinue, and merchants. The serfs produced the total product (except for goods made by artisans), some part of which, the customary amount of consumption, they retained for their own use. But the lords garnered a significant part of the total output, the surplus product, to support themselves, to build castles and cathedrals, to finance their Crusades, and, more generally, to maintain their noble lifestyles.

How did the lords obtain the surplus product? Why did the serfs turn over part of the total output to them? To answer the question of *how* first: serfs usually owed rents to their lords, rents that were to be paid *in kind* (for example, two sheep and eight chickens) rather than in money. In addition, the serfs on an estate generally owed labor services to their lord: they were obligated to work on the lord's land for a certain number of days per week, and the lord owned what was produced on this land. (This system made it easy for anyone to see that during part of the week the producers were working on their own land to support themselves while during the other part of the week they were producing a surplus product for the lord on the lord's land.) Sometimes serfs were obligated to pay other fees, such as the common one for milling grain at the lord's gristmill or, in a less common instance, to support the lord's dancing bear. In this case the serfs had to continue paying the fee even after the bear died!

In feudalism, then, the surplus product was paid by the serfs in rents, labor duties, and fees owed to the lords. However, the lords did not own the serfs in the same way that slaveholders owned slaves. For example, a lord could not sell a serf nor even, in most instances, displace a serf family from its customary hut and plot. So why did the serfs pay? The answer is that they were legally obligated under feudal law to make these payments, and local feudal courts (usually controlled by the lords) enforced the lords' property rights. As a last resort the lords also had the military power—in the form of knights on horses—to punish any serf who refused to pay. In these ways, then, the lords controlled the production and use of the surplus product in the feudal system. (It should be noted here that sometimes a serf, when driven to desperation, would manage to escape to a nearby town, to move to open land, or to transfer to the service of another lord.)

Distinctions Among Economic Systems

These examples illustrate the underlying point that whether one is considering slavery, feudalism, capitalism, or something else, each economic system is characterized by a distinct class relationship. As we have said, each system also has corresponding

ways of organizing labor processes and establishing property rights that, together with the class relationships, establish who controls labor processes and disposes of the surplus product. In Table 5.1 we summarize the main features of a wide variety of economic systems, offering suggestions for further reading about each type of system (see the suggested readings at the end of this chapter for full citations).

Economic systems are distinguished by *how decisions are made* concerning such things as who does what kind of work; who determines what is to be produced, how, and when; and who controls the resulting goods and services.

TABLE 5.1. VARIETIES OF ECONOMIC SYSTEMS

ECONOMIC SYSTEM	CHARACTERISTIC SOCIAL RELATIONSHIPS	EXAMPLES	SUGGESTED READINGS
Hunting and gathering (foraging bands)	Cooperatively acquiring and sharing food; consensus decision making; no government; little private property; little inequality; no surplus product.	Most of the 100,000 years of human history until agriculture developed 11,000 years ago; some tribal groups today.	Sahlins (1974) Diamond (1997) Boehm (2000)
Slavery	Slaves do most of the productive work; their owners own the equipment and animals used in production, own what the slaves produce, and control the government; the surplus product consists of the slave owners' gain.	Ancient Rome; U.S. in the years before the Emancipation Proclamation of 1863; pockets of slavery exist in certain regions (such as the Brazilian Amazon and the Thai fishing industry).	Finley (1973) Genovese (1965) Fogel and Engerman (1974)
Feudalism	Serfs work both for themselves and for hereditary landowners (lords), who rule locally; serfs pay rent in kind and work a set number of days per year on the lord's land; the surplus product takes the form of rents and the output of the serfs' forced labor.	Much of Europe from 1000 to 1500.	Bloch (1961) Anderson (1974) Aston and Philpin (1985)
Independent production	Families work on land they own, using capital goods and animals they own; they own what they produce and make decisions themselves; government is minimal.	Much of England prior to the development of capitalism; nonslave states of the U.S. during the 19th century.	Allen (1992)
Agrarian despotism	A political elite governs, claiming the surplus product by taxing both landowning farmers and tenants.	Pre-20th century China; India's Mughul Empire (1200–1700); much of Europe 1500-1800.	Spence (1990) Richards (1993) Anderson (1972)
Central planning	Most economic decisions are made in accord with a plan drawn up and implemented by a powerful central government; capital goods are state owned; economic inequality is in theory limited; government controls the surplus product.	Soviet Union, 1921–1990; China, Cuba and other Communist-ruled countries during the 20th and early 21st century.	Dobb (1966) Allen (2003)
Capitalism	Owners of capital goods hire wage labor to produce commodities to sell for a profit; employers make most economic decisions; substantial economic inequality exists; capital owners reap the surplus product as profit and as other property income.	Parts of Europe since 1600 and most of the world today, except for a very few nations like China that are governed by a Communist party and retain some elements of socialism.	Dobb (1964) Wallerstein (1974) Gray (1998) Schumpeter (1942) Gordon, Edwards, and Reich (1982)

Taking the analysis a step further, we can think of an economic system as a set of *rules of the game* governing economic life. From this point of view, economic systems may be understood using an analogy to games. What defines any game is the rules by which it is played. For example, what makes baseball different from basketball is not who plays it, or where it is played, or when, or anything else other than the rules by which each game is played.

The same is true of economic systems. Each economic system can be thought of as having a distinct set of rules of the game, that is, a particular set of rules governing the way in which classes interact with respect to the production and use of the surplus product. Economic systems are not defined according to *what* is produced (grain, steel, haircuts, or computer programs); *where* it is produced (in urban areas or in the countryside, in the United States or in Botswana); or *who* makes up the economic system. As in sports, it is the rules of the game that distinguish one economic system from another.

Class relationships, along with their particular rules of the game, were easy to see in precapitalist societies. In slave and feudal societies, for example, the slaveholder-slave and lord-serf relationships were quite visible. In contrast, the class relationship central to a capitalist economy is not so easy to see, especially when the capitalist economy in question is coupled with a democratic political system. In such a society everyone is said to be "equal under the law," and the rules of the game *do* involve a certain kind of parity among individuals that was not present in either slave or feudal societies. But legal equality is not social or economic equality. As Anatole France quipped, "the law, in its majestic equality, forbids the rich as well as the poor to sleep under bridges, to beg in the streets, and to steal bread."[1]

Regardless of the fact that people have equal legal rights in most capitalist societies, there is nevertheless a distinct—and defining—class relationship at the core of every capitalist economy. We turn now to an examination of the specific qualities that set capitalism apart from other economic systems.

Capitalism

Capitalism is an economic system in which employers, using privately owned capital goods, hire wage labor to produce commodities for the purpose of making a profit.

Capitalism is a particular economic system in which employers, using *privately owned capital goods*, hire *wage labor* to produce *commodities* for the purpose of making a *profit*. Thus, the primary class relationship of capitalism is the one between employers (capitalists) and workers—*the capital-labor relationship*. Other classes—the "middle" classes—are also important, and they are discussed in Chapter 7.

Capitalism is defined by three characteristics of its labor processes. First, most labor processes in the production sector produce commodities. Second, the capital goods used in production are privately owned. And third, the labor time devoted to producing commodities is purchased in labor markets. With each labor market transaction, a person agrees to provide his or her labor services to an employer for a certain period of time in return for a wage (or salary). Accordingly, the capitalist form of work is referred to as "wage labor."

We now discuss, in turn, each of the three aspects of a capitalist labor process: commodity production, private ownership of the capital goods used in production, and wage labor.

---

[1]Anatole France, *Le Lys Rouge* (Paris, 1894), Ch. 7.

Commodities

A **commodity** is any good or service that is produced with the intention of selling it in order to make a profit.

A **commodity** is any good or service that is produced, whether for a person to consume or for a business to use as an input, *with the intention of selling it* to make money. *Commodity production* therefore requires the presence of *markets* in which things are bought and sold.

To clarify the idea of a commodity, think about pancake making. In the previous chapter we focused on the recipe—or technology—for making pancakes. Now we turn to the social organization of pancake making. To do this we will have to say where the pancakes are being made and why they are made. The labor process for making pancakes at home on a Sunday morning is organized in a very different way from the labor process for making pancakes at the House of Pancakes or some other restaurant.

Let us start with the House of Pancakes case. What explains who is doing what tasks in the restaurant? What explains how the pancake eater gets the pancake? The answer to both these questions is markets. The labor market determines the answer to the first question, and the pancake market, along with other markets that allow the pancake consumer to earn an income, determines the second.

The pancake eater gets the pancake by paying for it. He or she gives up something—some money—to acquire ownership of the pancake. In Chapter 8 we note that a market exchange is the (legal) way in which the ownership of things is transferred in a capitalist society. How much money is required? The price is determined by the competitive interaction of buyers and sellers, a process that we analyze in Chapter 8 with the help of supply and demand curves.

How the pancake eater gets the pancake also explains why the pancake is made. It is made to be sold in exchange for money. (This is what is meant by "commodity production.") The pancakes are not made at the House of Pancakes because the workers particularly like making pancakes or because they want the customers to be well fed. They make the pancakes because they are paid to do it. If they did not get paid they would quit. If they could earn more money making hamburgers, they would probably be seeking employment at McDonald's.

Also, the pancake is probably not made because the owner of the House of Pancakes enjoys presiding over pancake making. He or she may live in another town or even another country. He or she may never have seen this particular House of Pancakes. In any case, if the owner could make more money by investing in steel making or sheep raising, he or she probably would not be the owner of the House of Pancakes.

The goal is making money, not making pancakes, and that is why a pancake produced at the House of Pancakes is called a commodity. The purpose of producing a commodity is, as we have said, to sell it. The likes and dislikes of producers or owners—their concern (or lack of it) for the nutrition or enjoyment of buyers—may be very real, but these factors (in most cases) will not determine what commodities are produced or who will get them. The outcomes regarding the *what* and the *for whom* questions will be determined by the *prices* at which the goods can be sold and the *costs* of producing them. Moreover, in a capitalist society commodity exchanges are generally anonymous: anyone who comes into the House of Pancakes and offers to pay the listed price for them will be provided with a stack of pancakes. It does not matter who the buyer is; only the price matters. In this respect capitalism undermines attitudes such as racial and other forms of discrimination.

Around the Sunday morning breakfast table at home, however, the situation is very different. If a stranger walked in, he might be asked to come back later—or never. It is also conceivable that he might be invited to sit down and have some pancakes. But it is not likely that he would be given some pancakes for no other reason than because he offered to pay for them.

Pancakes made in the kitchen of an army base are not commodities, either. They are made because someone in authority ordered them to be made, and those making them do so because they are following orders. No price is charged for the pancakes; to get a pancake one must simply belong to one of the units at the base.

Although the spread of capitalism in the modern world has brought with it the commodification of more and more aspects of social life—turning more and more things, previously not for sale, into commodities—there are still many products and services that are not produced as commodities. Take, for example, elementary and secondary education provided directly by government. Although there are more than 30,900 private schools in the United States, the large majority of American children go to public schools that are financed not by tuition payments (which buy education as if it were a commodity) but rather by communitywide tax dollars distributed through local governments. Activities financed through the government and not sold to the government by profit-seeking enterprises are not thought of as producing commodities, because they are not producing things that are *for sale* in a market. In our example, children have the right to go to a public school in their community because they live in the community, not because they can buy the right to go to a particular school.

With few exceptions, *people* are not produced as commodities. Although the work that a mature person does will almost certainly be influenced by market forces, the birth, care, nurturance, and upbringing of that person—all involving labor processes in the reproduction sector—will not usually be undertaken with the aim of producing an object for sale. Exceptions would include the incubation of an embryo by a surrogate mother for a fee (a service that might be advertised as "womb for rent") and the breeding of slave children on American plantations in the antebellum South for the purpose of selling the children in slave markets for profit.

In any case, Figure 5.1 shows that commodity-producing labor amounts to less than half of all labor performed in the United States today. Surveys of time use (surveys that collect data on how people divide their time among various activities) for two-thirds of the world's population have shown that 41 percent of all work takes place outside the market. This work is not paid, most of it is household labor, three-quarters of it is performed by women, and much of it involves caring for family members, especially young children. Women also work (at paid or unpaid work) an average of 52 hours per week, while men work 48 hours, and have more leisure time. The surveys also found that of men's total work time, nearly four-fifths is spent in paid activities and one-fifth in unpaid activities. But for women the proportions are far different, with about two-fifths of their work time devoted to paid activities and three-fifths to unpaid activities.[2]

Among Americans who sold their working capacities in labor markets in 2013, 77 percent were engaged in the production of commodities (including services) while the other 23 percent were non-commodity-producing employees of government agencies or nonprofit organizations. Thus, the relative percentages

[2]United Nations, *Human Development Report 2015* (New York: United Nations, 2015), 107.

of nonmarket, commodity-producing, nonprofit, and government labor—shown in Figure 5.1 as percentages of the total number of hours worked in the United States in an average week in 2013—were 51 percent, 38 percent, 7 percent, and 4 percent, respectively.

Much household work is devoted to tasks such as child care, grocery shopping, cooking, cleaning, folding laundry, gardening, and mowing the lawn, most of which require large amounts of time. In a society such as the United States, however, where two-earner couples have become the norm, parents of young children (particularly mothers) enjoy very little leisure time. Many such parents, if they have sufficient income, purchase substitutes for goods and services they once provided for themselves, paying for child care and buying more restaurant and take-out meals. There are few paid family leaves from work in the United States, and there is considerably less public support for child care than there is in other rich countries.

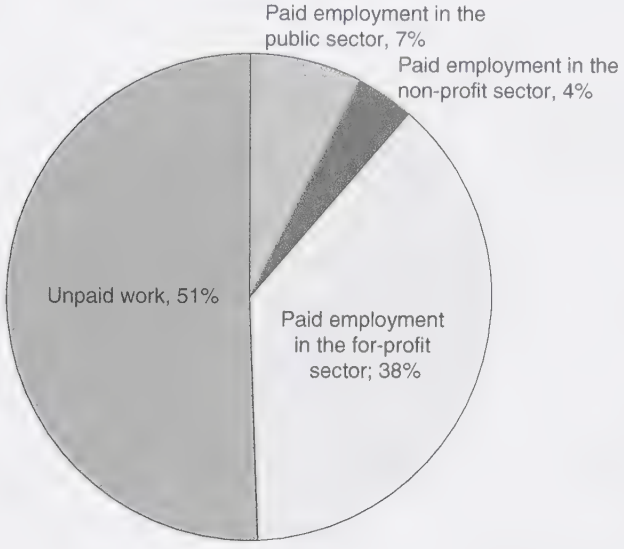


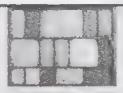
FIGURE 5.1 Percent of All U.S. Labor Performed that is Unpaid, Paid Commodity-Producing, and Paid Non-Commodity-Producing, 2013.

This pie chart shows how much of all labor is household labor and other nonmarket work, how much is paid employment that produces commodities (in for-profit enterprises), and how much is paid employment that does not produce commodities (divided into government work at all levels and work for nonprofit organizations). Approximately 12.5 billion hours of labor were performed in a typical week in the U.S. in 2013, and the percentages in this figure represent shares of this total number of hours. The data are presented in terms of hours of labor rather than numbers of workers because most people participate in both household labor and paid employment. Based on the American Time Use Survey and other government data, about half the work performed in the U.S. is nonmarket (unpaid) labor.

Sources: U.S. Department of Labor, Bureau of Labor Statistics, "American Time Use Survey, Table A-1 for 2013," available at https://www.bls.gov/tus/tables/a1_all_years.xlsx; BLS, "Total civilian labor force in the U.S. in 2013," news release, available at http://www.bls.gov/news.release/archives/empsit_01102014.pdf; U.S. Census Bureau, "Annual Survey of Public Employment and Payroll Summary Report, 2013," available at http://www2.census.gov/govs/apes/2013_summary_report.pdf; BLS, "Nonprofits account for 11.4 million jobs, 10.3 percent of all private sector employment," *The Economic Daily* (TED), October 21, 2014, available at http://www.bls.gov/opub/ted/2014/ted_20141021.htm.

Another important form of nonmarket work is the volunteer labor that people provide to community organizations, schools, and other institutions in their neighborhoods. Such activities sustain our civic life, but people who have to work two or more jobs to make ends meet can hardly spare the time for volunteer work.

Although commodity production accounts for less than half of the labor performed in a modern capitalist society such as the United States, it was even less important in precapitalist societies in the past, and it is also less important in noncapitalist societies today. For example, market exchanges in food were forbidden in some societies because their people regarded nourishment as too important a function to be left to the buying and selling activities of individuals. Even today in the United States, as well as in many other countries, government price supports and other kinds of state intervention prevent the price of food from being determined entirely by unregulated commodity exchange. In the global economy today, however, food *is* a commodity—and this has a lot to do with the persistence of hunger around the world. (See box "Money Talks (And Eats).")



Money Talks (And Eats)

About one-ninth of the world's population—795 million people—were undernourished in 2015, meaning that they got too little food to meet their daily energy needs, even for moderate activity. In the whole world, about 3 million children under the age of five died of poor nutrition in that year.

The overwhelming majority of the world's hungry people, 780 of those 795 million in 2015, live in the poor countries of Africa, Asia, and Latin America; two-thirds live in Asia. Yet the total has fallen from about 1 billion, or almost 19 percent of the world's population, in 1998 to just under 800 million, or 11 percent, today.

Despite gains on a world scale, in the U.S. many more people and households were food-insecure in 2014 than before the Great Recession of 2007–2009. According to the UN's Food and Agriculture Organization, people are *food-insecure* when they lack secure access to sufficient amounts of safe and nutritious food for normal growth and development and an active and healthy life. (Undernourishment is a more serious problem.) In 1998, 36 million people in the United States lived in food-insecure households; in 2014, 48 million did. Among households made up of a single woman with children, over one-third were food-insecure. We live in a world in which money not only talks but also determines whether or not a person can eat.

Why are there so many hungry people in the world, many of them children who die before reaching five years of age? The main reason is that land and income, both within and among nations, are very unequally distributed—and those who do not have much land or income cannot produce or buy enough food to nourish themselves adequately. Paradoxically, many of the world's hungry people are farm workers or farmers who depend on part-time employment in nonagricultural jobs to make ends meet. Unemployment is a far more common cause of hunger than crop failure.

Much of food is produced as a commodity, and this means, first, that it is distributed though markets and, second, that market prices determine how much is produced, and for whom. In a highly integrated global economy, food will be purchased and consumed by those, wherever they may live, who can afford to pay the highest prices for it.

In many places, small farmers, once able to grow their own food, have been displaced by large commercial farms that produce cash crops for export. Food is sold to richer countries where

continues

already-well-fed consumers pay relatively high prices for it. In a world of persistent hunger, food flows through global markets from areas of need to areas of wealth.

Sources: Alisha Coleman-Jensen, Matthew P. Rabbitt, Christian Gregory, and Anita Singh, *Household Food Security in the United States in 2014* (U.S. Department of Agriculture, Economic Research Service, September 2015), available at www.ers.usda.gov/publications/err-economic-research-report/err194.aspx; FAO, IFAD, and WFP, 2015, *The State of Food Insecurity in the World: Meeting the International Hunger Targets: Taking Stock of Uneven Progress*, Rome: Food and Agriculture Organization of the United Nations, http://www.fao.org/3/a-i4646e/index.html; U.S. Department of Agriculture, Economic Research Service, 2015, "Food Security in the U.S: Key Statistics and Graphics," http://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/key-statistics-graphics.aspx#foodsecure; World Food Programme, c2016, http://www.wfp.org/hunger/stats.

Commodity production, then, is very different from other ways of organizing production and distributing its results. Other ways of coordinating economic activities include relations within a family, military commands at an army base, and the democratic processes of deliberating and voting. However, understanding commodity production is one of the keys to understanding the economies of the United States and the world today.

By itself, commodity production does not define capitalism, nor is capitalism synonymous with commodity production (or "market economy"). Historically, there have been economic systems that were oriented to commodity production but in which there were no capitalists. One such example, already discussed, is the slave economy of the pre–Civil War American South, in which slaves produced tobacco, rice, and cotton on plantations and the owners sold the slaves' products as commodities in the world market.

At roughly the same time, another way of organizing economic life coexisted with slavery in the United States. This economic system—to be discussed at greater length in the next chapter—consisted of home production of items for direct use and the production of commodities by independent self-employed producers such as farmers, craftsmen, and artisans. This economic system, in which production was carried on by households that owned their own capital goods and frequently exchanged their products—but not their labor—in markets accounted for approximately 60 percent of all production in the United States at that time. It was something like this, a property-owning democracy, that was advocated by Thomas Jefferson.

During the twentieth century, proponents of *market socialism* argued that markets might be used to coordinate an economy that would form the basis of a socialist rather than a capitalist society. What they envisioned is a system in which worker-owned firms would sell their products competitively in markets (thus avoiding the need for central planning) but in which there would be neither capitalists (since capital goods would be owned by worker-controlled firms) nor wage labor (since workers would have access to their own capital goods and would, in effect, be hiring themselves). Whatever one may think of such proposals, the fact is that commodity production has occurred historically in various noncapitalist economic systems and could occur in future noncapitalist systems.

For commodity production to be termed capitalist production, there must also be employers and workers. The relationship between employers and workers

(the capital-labor relationship) is, in turn, defined by two other characteristics of the labor process: private ownership of the capital goods—the tools, factories, offices, and other (durable) goods used in production—and labor employed as wage labor.

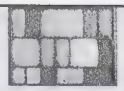
Privately Owned Capital Goods

Private property is a social institution (or rule) that gives individuals or firms the right to use, lend, or sell things such as land, buildings, and artistic or intellectual creations of any kind; it means that one can have or use such things only if one has made, rented, purchased, or been given them.

The second defining feature of capitalism (after commodity production) is that the capital goods used in production are the *private property* of capitalists. **Private property** means, first, that what is owned may be used or sold at the discretion of the owner or people he or she selects; second, that the owner has a claim on any income or other benefits that may result from the use of the object owned; and third, that the owner is entitled to exclude others from using whatever it is that is owned, for example, land or factories. A sign that reads "Private Property, No Trespassing" conveys the third meaning of private property.

Private ownership of capital goods, while not unique to capitalism, is nonetheless far from universal. For instance, many Native American peoples did not have a concept of private property in land, and this led to misunderstandings between them and the European settlers who thought they were buying things from the Native Americans in accordance with Western notions of private property. The derogatory expression "Indian giver," referring to someone who gives something and then wants it back, illustrates this point: the phrase arises from the Europeans' (and, later, some Americans') inability to understand that among Native Americans some things such as land were neither mine nor yours but ours or no one's. Other societies have also had very complex notions of property, as illustrated by the description in the box "Property Rights among the Arapesh of New Guinea" in Chapter 2.)

Privately owned capital goods include machines, buildings, offices, tools, and other durable things needed in production and whose owner, because of a property right, determines how the property will be used. How could capital goods *not* be privately owned? In some cases a government body, such as a municipal electric company, may own them. Even in the private sector it is possible for the capital goods in a particular company to be owned by its workers. Indeed, it occasionally happens that when a company is on the verge of bankruptcy and needs to be "bailed out" to prevent it from "going under," its managers and workers will get together, raise the capital necessary to buy a majority of the company's shares in the stock market, and—often with wage reductions and other measures—put the company back on a firm financial footing. There have been a number of enterprises in the United States that have been owned and controlled by their workers, the most well known of these being a group of plywood manufacturing companies in the Pacific Northwest until about the 1990s (see box, "Own Your Job. Elect Your Boss"). In the Mondragon region in Spain for many years, more than one hundred workers' cooperatives have manufactured and sold everything from home appliances to machine tools. Similarly, Germany has a system—called "codetermination"—that requires that boards of directors of companies include (as directors) a certain number of labor representatives and that democratically elected workers' councils be allowed to participate in the decision-making processes of German corporations. In Italy there are tens of thousands of worker cooperatives.



Own Your Job. Elect Your Boss

In 1921 a group of loggers, carpenters, and mechanics in Olympia, Washington, formed the Olympia Veneer Plywood Cooperative. For an investment of $1,000, a member gained the right to work in the plant and share equally in any profit. A member who left was to sell his shares, and someone approved by the membership to join was required to buy shares, which by 1923 were selling for $2,550 per share. In 1939, 250 workers in nearby Anacortes each invested $2,000 to set up a second cooperative plywood mill. Strong wartime demand for plywood boosted the value of a share to $28,000 in 1951, and members paid themselves wages at rates double the union wage in nearby conventionally organized plywood mills.

Stimulated by the success of both co-ops, between 1949 and 1956, twenty-one more co-ops entered the plywood industry in Washington and Oregon states, nine of them by buying out conventional firms. (Some co-ops also turned back into more or less conventional firms.) Until the entire industry moved from the Northwest to the Southeast in the 1980s and 1990s, about half the plywood firms were co-ops, and the rest conventional, some unionized and some not. Though the co-ops and conventional firms used similar machinery, the co-ops specialized in the more labor-intensive "sanded" plywood because, as one analyst of the co-ops commented, it "puts a premium on worker effort."

Most worker-owners of the plywood cooperatives in the Northwest received equal pay, and their jobs were often rotated. The body of worker members elected management. Some nonmembers were hired under conventional wage contracts, their numbers making up an average of a quarter of the total workforce. High productivity was maintained through a strong work ethic among members, enforced by peer pressure. The resulting saving in supervision costs was substantial: when one conventional firm converted to a co-op, the number of supervisors was reduced to a quarter of its previous level. Retiring or departing members advertised their shares in local newspapers, so we know that they sold on average for about one to three years' annual earnings.

Conventional firms and cooperatives alike were able to attract both labor and capital over their seventy-five years of coexistence in the Northwest, but the co-ops' productivity (output per unit of all inputs) was substantially higher than that of the conventional firms. Another difference: when demand for plywood was slack, the co-ops reduced the pay of all their workers, thereby avoiding the pain of job loss for any.

Today in the United States more than one-fifth of employees own stock in the companies where they work, though few own substantial amounts, and almost none can elect their boss. While there have been some well-publicized failures of worker ownership, such as United Airlines, firms with a significant amount of ownership by employees generally have higher labor productivity than conventional firms, particularly if employees also take part in decision-making about the work process.

Sources: Adapted from Samuel Bowles, *Microeconomics: Behavior, Institutions, and Evolution* (Princeton, NJ: Princeton University Press, 2004). See also: Douglas Kruse, "Research Evidence on the Prevalence and Effects of Employee Ownership," *Journal of Employee Ownership Law and Finance* 14, no. 4 (2003): 65–90; John Pencavel, *Worker Participation: Lessons from the Worker Co-ops of the Pacific North-West* (New York: Russell Sage Foundation, 2002).

As noted, private ownership of capital goods not only grants their owners the right to use or sell them; it also gives employers the right to exclude others from using them. The right to exclude others implies the right to fire workers, since that is just a way of excluding workers from the workplace. The related right, the right

to determine who is included, is the right to hire. So the power to hire and fire is an aspect of the property rights of capitalist employers.

The right to use one's property at one's own discretion is considerably limited by zoning laws, union contracts, public health codes, occupational health and safety regulations, pollution controls, and the like. Even with such limitations, however, private ownership of capital goods is one of the legal conditions that makes capitalist production possible.

When discussing ownership of capital goods it is useful to distinguish between *wealth*, *property income*, and *earned income* (or earnings). **Wealth** refers to what a person or a firm owns at any particular moment in time, minus what the person or firm owes. Economists think of it as a "stock" because it is measured at a point in time. On the other hand, **income**—whether from property or from labor—is regarded as a "flow," since it is measured over a period of time: an hour, a week, a month, or a year.

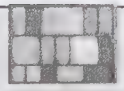
Wealth (or **net worth**, or **net assets**) of a person or firm refers to the total value of assets owned by the person or firm, minus the total value of the outstanding debts, or **liabilities,** owed by the person or firm.

Income can be earned from property or labor and is considered a "flow."

Property income is income that is received in the form of profit, rent, interest, or dividends as a result of owning an asset such as a business, a piece of land, an existing structure, a bond, or a share of corporate stock.

Wealth involves the ownership of tangible or intangible things (assets), such as real estate, household goods, vehicles, money or other financial assets, or copyrights or patents, that may yield income or other benefits to their owner over time. A person's wealth at a given moment may consist of such tangible items as houses, cars, factories, land, horses, or jewelry; it may also include intangibles like money, and financial assets that provide income such as securities or shares of stock. The terms *net assets* and *net worth* refer to everything owned minus liabilities or debt owed to other people or institutions (such as banks and credit card companies). *Wealth*, *net worth*, and *net assets* all mean the same thing. A person is "wealthy" if he or she owns a significant quantity of assets over and above any liabilities or debts. Bill Gates is an example of a wealthy individual: his net worth in January 2016 was $77 billion. (See box "Brother, Can You Spare a Billion?")

Income, on the other hand, is a different matter, though not entirely different. We can calculate from the information presented in the box "Brother, Can You Spare a Billion?" that each member of the "Forbes 400" had, on average, a net worth of $5.85 billion in 2014. If these four hundred individuals invested—or hired someone to invest for them—their assets in such a way as to produce even a 5 percent return each year (not an unreasonable assumption to make for the "Forbes 400"), each of them would enjoy, on average, an annual income of over $290 million. Moreover, such an income would consist entirely of property income: generating it would require only as much effort as it takes to select a skilled investment manager to handle one's account.



Brother, Can You Spare a Billion?

Every year since 1982, *Forbes* magazine has published the Forbes 400, a list of the four hundred richest people in the United States. More recently it has also published a list of the world's billionaires, numbering 1,826 in 2016, with just over $7 trillion in total wealth. In 2016, altogether the four hundred richest Americans had a total of $2.34 trillion in net assets which, if divided up among all U.S. households, would give each household a present of nearly $20,000. Just the twenty wealthiest in the United States—listed below, along with the seven among the world's wealthiest with incomes in the same range—had the same wealth as the entire poorest half of the U.S. population. Many are on the Forbes list because they inherited a substantial amount of wealth from their parents or grandparents, though Gates did not.

continues

THE WORLD'S TOP BILLIONAIRES

NAME	NET WORTH	COUNTRY	SOURCE OF WEALTH
Bill Gates	$77 billion	U.S.	Microsoft
Amancio Ortega	$70 billion	Spain	Zara (retailer)
Warren Buffett	$61 billion	U.S.	Berkshire Hathaway
Jeff Bezos	$52 billion	U.S.	Amazon
Carlos Slim Helu	$52 billion	Mexico	telecom (America Movil)
Mark Zuckerberg	$49 billion	U.S.	Facebook
Larry Ellison	$45 billion	U.S.	Oracle
Charles & David Koch	$40 billion each	U.S.	Koch Industries
Larry Page	$38 billion	U.S.	Google
Sergey Brin	$37 billion	U.S.	Google
Liliane Bettencourt	$37 billion	France	L'Oreal
Michael Bloomberg	$36 billion	U.S.	Bloomberg LP
Bernard Arnault	$34 billion	France	LVMH
Jim, Alice, S. Robson ("Rob") & Christy Walton	$31–34 billion each	U.S.	Wal-Mart heir
Li Ka-shing	$29 billion	China	diversified
Sheldon Adelson	$25 billion	U.S.	Sands Casino
George Soros	$25 billion	U.S.	Hedge funds
David Thomson & family	$25 billion	Canada	Thomson Reuters
Jorge Paulo Lemann	$25 billion	Brazil	Beer
Phil Knight	$26 billion	U.S.	Nike
Forrest Mars Jr., Jacqueline Mars, & John Mars	$23 billion each	U.S.	Mars Candy heirs

Sources: Chuck Collins and Josh Hoxie, December 1, 2015, "Billionaire Bonanza: The Forbes 400 and the Rest of Us: Wealthiest 20 People Own More Wealth Than Half the American Population," Institute for Policy Studies, http://www.ips-dc.org/billionaire-bonanza/; *Forbes*, 2016, "The World's Billionaires," available at http://www.forbes.com/billionaires; Kerry A. Dolan and Luisa Kroll, "Inside the Forbes 2015 Billionaires' List: Facts and Figures," available at https://www.forbes.com/sites/kerryadolan/2015/03/02/inside-the-2015-forbes-billionaires-list-facts-and-figures/#1df828c830ca.

High net worth is usually based on ownership of a significant amount of corporate stock (private ownership of the capital goods used in production). Figure 5.2 shows that if the 117.6 million households in the United States in 2014 are ranked by the amount of corporate stock they owned in that year, the top 1 percent owned 35.0 percent of the corporate stock, the top 10 percent owned 80.8 percent of such stock, and the bottom 90 percent of American households owned 19.2 percent. (See sources cited below Figure 5.2.) This uneven distribution of the ownership of corporate assets helps to explain why some people are employers while others (the majority of the population) are workers.

The last two decades of the twentieth century witnessed a dramatic increase in stock ownership among American households. If *indirect* ownership—through mutual funds, trusts, IRAs, Keogh plans, 401(k) plans, and other retirement accounts—is included, the percentage of households owning at least some stock went from

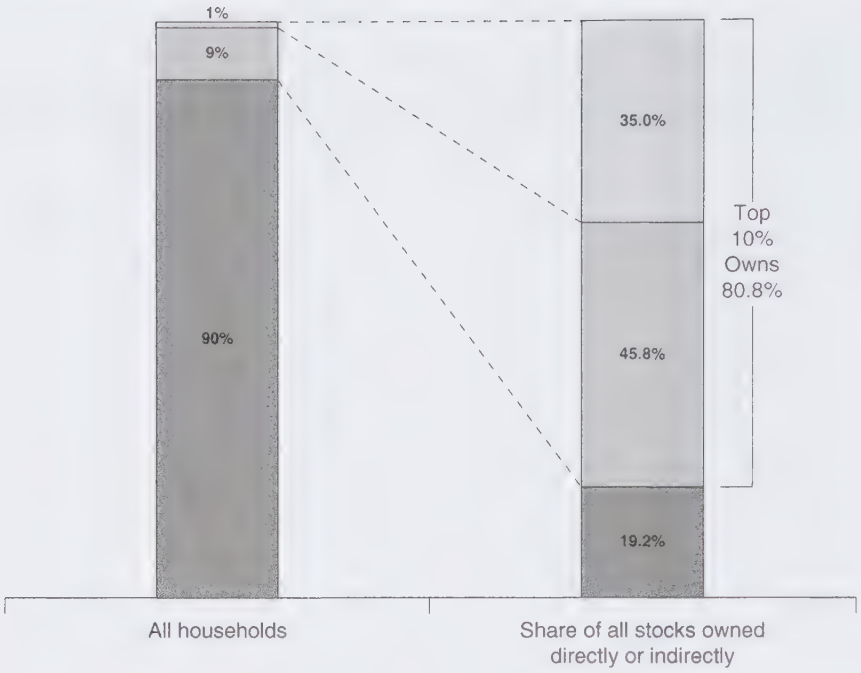


FIGURE 5.2 The concentration of ownership of capital goods.

The bar on the left represents the 117.6 million households (families and individuals) that made up the U.S. population in 2010. The bar on the right represents the value of all the publicly traded stocks of corporations in the same year, and it shows the proportions of stocks that were owned—whether directly or indirectly through mutual funds, trusts, and retirement accounts—by the various wealth groups. The figure shows that the wealthiest 1 percent of households owned 35.0 percent of the corporate stock, that the top 10 percent (which includes the top 1 percent) owned 80.8 percent of such wealth, and that the bottom 90 percent of American households owned 19.2 percent.

Source: Data from Edward N. Wolff, "The Asset Price Meltdown and the Wealth of the Middle Class," NBER Working Paper 18559, 2012, 57, Table 9, available at http://www.nber.org/papers/w18559.

24.4 percent in 1983 to a peak of 51.9 percent in 2001, and then fell by 2010 to 46.9 percent. These numbers also show, of course, that in 2010, and in almost every year since at least 1983, more than half of all American households owned no stock at all.[3]

The significance of privately owned wealth is that the more of it one has, the less one needs to work. Table 5.2, "Sources of Income for U.S. Taxpayers, 2011," presents IRS data for 2011 (the latest year for which data are available). It shows that the higher a taxpayer's income, the less of it comes in the form of wages or salaries. The logic of it is as follows: Wages and salaries are earned income, while unearned income comes in the form of interest, dividends, rent, royalties, and realized capital gains, all of which are different kinds of property income. Other kinds of payments received, such as social security benefits or unemployment compensation, do require previous employment to make a person eligible to receive them, but are not currently "earned." Hence,

[3]Edward N. Wolff, "The Asset Price Meltdown and the Wealth of the Middle Class," NBER Working Paper 18559, 2012, p. 65, Table 17b: Stock Ownership, 1989–2010.

TABLE 5.2. SOURCES OF INCOME FOR U.S. TAXPAYERS, 2011

INCOME LEVEL*		% OF TAXPAYERS	% OF TOTAL INCOME	% OF TOTAL WAGES AND SALARIES	WAGES AND SALARIES AS % OF TOTAL INCOME***
Zero to $20,000		34.6%	3.6%	6.3%	123.2%
$20,000 to $100,000		52.0%	43.7%	47.4%	77.2%
$100,000 to $1 million		13.2%	41.4%	41.3%	71.1%
$1 million or more		0.2%	11.3%	5.0%	31.9%
	Total:	100.0%	100.0%	100.0%	
		Total no. of taxpayers** (millions)	Total income (billions)	Total wages and salaries (billions)	
		145.3	$8,498	$6,055	

Note: Wages and salaries as a percent of total income can exceed 100% because deductions on tax returns can make Adjusted Gross Income lower than wages and salaries. The data on individual tax returns also include S-corporation returns, which are filed by individuals whose businesses are not incorporated.

* "Income level" refers to the "adjusted gross income" reported on the tax returns filed with the IRS in 2011.

** No. of "taxpayers" is the number of *tax returns* filed with the IRS in 2011. Since many multiple-earner families file "joint" tax returns, the number of returns is less than the number of income-earners. The data in this table are based on tax returns.

*** Total income includes, in addition to wage and salary income: interest, dividends, rent, royalties, and capital gains, all of which are forms of property income; and many types of nonwage income (generally smaller amounts than those derived from property), including business and professional income, social security benefits, alimony, and unemployment insurance.

Source: U.S. Treasury Department, Internal Revenue Service, Statistics of Income: Individual Income Tax, All Returns: Sources of Income, Adjustments, Deductions, and Exemptions, 2011, Table 1.4. Available at https://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Returns.

from the columns on the left and right sides of the table we can draw the following conclusion: the higher a taxpayer's income, the lower is his or her wage or salary as a percent of total income. As we follow the income brackets (in the left-most column of the table) from less than $20,000 to more than $1 million, we can see (in the right-most column) that the corresponding percent of total income that is wages or salaries, including employer-paid benefits, falls from 123.2 percent to 31.9 percent.

Although many very-high-income taxpayers—for example, chief executive officers (CEOs) of large corporations, as well as others—work very hard, the magnitude of their property income means that they generally do not *have* to work. We present more complex definitions of the capitalist class and working class—and also the "middle" classes—in the next chapter. However, one important conclusion that can be drawn from the preceding set of facts is that if a person's property income is substantial, he or she *can* choose not to work while still maintaining a decent standard of living, whereas a person without much property income *cannot* make such a choice. It is this distinction (among others) that separates the capitalist class from the working class.

The **capitalist class**, or **capitalists**, are those who own capital goods used in production and exercise control over the labor of others; they receive their income in the form of profits or other payments (such as interest and rent) for the use of their capital goods.

The **working class**, or **workers**, are those who perform wage labor; they neither own the capital goods used in their labor processes nor command control over the labor of others.

Thus, we can say that the **capitalist class** consists of that group of people—a minority of the population—who own the capital goods used in the production of commodities. They control not only their own work (if they work) but also (and more importantly) the labor of others, and most of their income is property income. The **working class**, on the other hand, consists of people—the majority of the population—who do not own sufficient capital goods to live on and must therefore work under the control of others. Workers receive their income in the form of wages or salaries. People in the working class typically do own cars, homes, and other personal property, and some even own corporate stocks or shares in mutual funds, but such possessions do not give workers control of the capital goods used in production.

Wage Labor

The third defining characteristic of a capitalist economic system (after commodity production and privately owned capital goods) is *wage labor.* Wage workers are people who must live on the wages or salaries they receive in return for working for an employer—whether their wage is figured based on how much time they work, or on how much they produce ("piecework" or "commission"); their work is called **wage labor**. Everyone who earns a living as an employee is a worker.

Wage labor is work performed under the direction of an employer in return for a wage or salary.

What other kind of work is there? As we explain in the next chapter, most people in the United States two hundred years ago did not work as wage workers. They were independent farmers or other self-employed persons, or they were slaves. Some producers in the United States today are still not wage workers. These are the people who work for themselves, or in partnerships, or in cooperatively owned establishments. Some work at their own family farms or shops.

People who work for themselves make their living either by producing what they themselves consume or by selling a commodity they have produced (which could be a service rather than a physical product). Wage labor is fundamentally different. People who work for wages or salaries make their living by renting themselves out, or more precisely, by selling their time. In exchange for the wage, they work under someone else's direction, producing things that they will not own.

Concentration of ownership of capital goods in one part of the population implies the need for wage work among most of the rest. Only people who own a substantial amount of property can live on their property income, while those who own few or no capital goods usually must have jobs to survive. Thus, the concentration of ownership of capital goods in the hands of relatively few people and the existence of wage labor are closely related.

An example will make this point clear. In 2014, private sector workers in the United States worked with an average of more than $179,000 worth of capital goods (including software as well as equipment and buildings).[4] Most U.S. workers, of course, do not own the capital goods with which they work. American workers usually own cars, and most own their own homes but little else. Thus a typical worker is not wealthy enough to acquire the capital goods he or she would need in order to be an independent producer. If workers each had an extra $179,000 of wealth, they might be able go into business for themselves, perhaps by pooling their wealth with friends or family members. Since most workers cannot raise enough money to start their own businesses, the tools they work with and the place they work in are generally owned by somebody else. That "somebody" is their employer, on whom they depend for their livelihoods.

Capitalism, the Surplus Product, and Profits

We can now see the link between the basic class relationship of capitalism and the surplus product. Employers (capitalists) hire workers to produce commodities (the total product), which the employers then own. The employers then sell the commodities

[4]U.S. Department of Commerce, Bureau of Economic Analysis, Table 5.10. Changes in Net Stock of Produced Assets (Fixed Assets and Inventories), "Produced Assets, Private, Nonresidential," available at http://bea.gov/iTable/index_FA.cfm; and BLS, series CES0500000001, "All Employees, Total Private, Seasonally Adjusted," available at http://data.bls.gov/timeseries/CES0500000001

Profit is the form of the surplus product in a capitalist economic system; it is what is left over, out of sales revenues, after wages, the costs of materials used up, and wear and tear on machines have been paid.

The contractual payment made by the government to owners of U.S. Treasury bonds, bills, or notes is called **interest**. The contractual payment made by a corporation to owners of its bonds or to banks that have provided loans to the corporation is also called **interest**.

A share of stock represents a share of ownership of a corporation and entitles its holder to receive a **dividend**—a payment per share decided by the board of directors—whenever a corporation decides to pay out a portion of its total profits as dividends to its stockholders. The amount of the dividend varies.

that the workers have produced. Assume that, out of their sales revenues, the employers pay wages sufficient to provide each worker with the customary amount of consumption. Assume also that the employers set aside from their revenues whatever is needed to replace the materials used up and to provide in advance for the depreciation (wearing out) of the tools, machines, and other capital goods used in producing the total product. As explained in the previous chapter, the remaining revenue represents the surplus product. In a capitalist economic system the surplus product takes the form of profit.

In the simple grain model of the last chapter the surplus product was a quantity of grain representing a certain amount of labor time. In capitalist economies, however, the surplus is not only a quantity of goods (or labor time); it is also a quantity of money, called **profit**. Profit provides the basis for the capitalist's income—whether such income comes in the form of interest, dividends, rent, or capital gains—and it flows, as we have explained, from ownership of capital goods and control over labor processes.

An employer may not be able to keep all the profit resulting from a capitalist labor process. For one thing, the government takes part of the surplus product in taxes to pay for public services, public goods, and amenities of various kinds. (It is also true that to finance its activities, the government also imposes taxes on income that is not part of the surplus product, for example, the income of wage earners.) We discuss the government's role in the economy in Chapters 6 and 19.

Another part of the surplus may be used to pay *interest* to a bank or to an individual who has lent money to a firm. Interest payments are often regarded as costs, and they *are* costs to a business. However, they are also just a different form of profit, since they are paid to those who *indirectly* own the capital goods used in production. If a bank has lent money to a firm, it will have received in return a commitment on the part of the firm to pay interest and ultimately to repay the money that was loaned. Although the bank does not directly own any of the capital goods, it has acquired a claim on part of the profit, or surplus product, generated by the use of the capital goods.

Similarly, a firm may pay out part of its profit in the form of **dividends** to holders of stock in the company. A share of stock signifies that its holder is a part-owner of the company and is therefore entitled to a dividend if the company decides to pay dividends. Although they are not contractually obligated to do so, companies usually do pay dividends every quarter in order to maintain the attractiveness of their stock to current or potential investors, thereby helping to keep up the price of the stock in any stock market in which it is traded. Any profits that are not paid out either as interest or as dividends are called "retained earnings." These are part of the profits that companies set aside to be used to pay for new investments. Dividends are different from interest in that companies are not obliged to pay dividends to shareholders, but they are obliged to pay interest to bondholders or lenders. However, dividends are similar to interest in that they are one of the forms that the surplus product takes in a capitalist society. Accordingly, they are counted as part of total profits, or property income.

Most of the profits received by capitalists (including such property income as interest, rent, and dividend payments as well as the "retained earnings" of corporations) are spent to buy commodities of one kind or another. When capitalists' profits are spent on buying new machinery for a factory (above and beyond what is needed to replace the capital goods worn out during production), or on research and development activities, or on other goods and services intended to raise productivity in the future, such

spending is called *net investment*. Profits may also be used to pay for *consumption* of necessities and of such items as mansions, yachts, luxury cars, private jets, and so on. (Recall the box on "Superyachts and Bull Market Boats" in Chapter 4.) The surplus product may be thought of as all the things on which profits are spent, and accordingly it consists of a bewildering variety of goods and services. What all these goods and services have in common—what makes them part of the surplus product—is that none of them goes toward maintaining the producers at their customary standard of living or toward replacing the materials and machines used or used up in production.

As we have said, it is ownership of capital goods that gives control over the disposition of the surplus product to the capitalist class. This control may, in turn, be seen in two ways. On the one hand it may be seen as granting *income rights* to capitalists. These are the rights that establish claims on the interest, rent, dividends, and other payments resulting from ownership of the capital goods used in production. On the other hand, ownership of capital goods gives employers *control rights* that establish their right to control what goes on in the workplace, including the way in which it is organized. The significance of this dimension of capitalist rights is that it gives the owners of capital goods *power* over their employees, giving the owners the right to make the decisions (or select the managers who will make the decisions) that affect the jobs, incomes, and safety of their workers. We discuss at some length in Chapters 12 and 13 the various ways in which this control may be exercised.

An additional dimension of the power of capitalists has to do with their participation in politics. Owners of substantial wealth are not only rich; they are also highly influential. Their high incomes give them the ability to hire lawyers and lobbyists, to make major political campaign contributions—sometimes even financing their own campaigns—and to engage in other activities that may enable them to influence the way in which legislation is written or policies selected. Such legislation or policies will sometimes make it easier for employers to maintain the control that they have over the labor process and over their employees, thus ensuring that they will be able to continue making profits in the future. This is discussed in Chapter 6. Our point here, however, is that even if they have nothing to do with the government, substantial owners of capital goods have a lot of power—over those they employ and others affected by their decisions about the technologies used, the products to produce, and where to locate production.

Conclusion

To conclude this chapter, we return to our basic characterization of capitalism as a way of organizing labor processes involving commodity production, private ownership of capital goods, and wage labor. Our definition of capitalism stresses the social organization of the labor process; nothing is said about its technology or about what is produced. A capitalist economy may turn out factory goods or farm products, tanks or dolls. What is capitalist about it is how it is organized.

This definition of capitalism also says nothing about the form of government. Capitalism may coexist with democratic governmental institutions, as it does in the United States, with fascism as it did in Hitler's Germany, or with other forms of dictatorship. Moreover, democratic government may exist alongside other economic systems besides capitalism. For instance, as we explain in the next chapter, during the early days of American democracy capitalist organization hardly existed at all.

Capitalism operates in accordance with one set of "rules of the game," one set of procedures and class relationships that determine how social decisions are made. Slavery and feudalism went by other sets of rules and other methods for making social decisions. In families the authority of adults, particularly male household heads, as well as custom and tradition, provide still a different way of determining how things are done. This book is mainly about the capitalist rules of the game.

Suggested Readings

Robert Allen, *Enclosure and the Yeoman* (Oxford: Clarendon Press, 1992).

Robert Allen, *From Farm to Factory* (Princeton: Princeton University Press, 2003).

Perry Anderson, *Lineages of the Absolutist State* (London: Verso, 1972).

Perry Anderson, *Passages from Antiquity to Feudalism* (London: NLB, 1974).

T. H. Aston and C. H. E. Philpin, eds., *The Brenner Debate: Agrarian Class Structure and Economic Development in Pre-Industrial Europe* (Cambridge: Cambridge University Press, 1985).

Marc Bloch, *Feudal Society, Volume I: The Growth of Ties of Dependence* (London: Routledge, 1961).

Christopher Boehm, *Hierarchy in the Forest* (Cambridge, MA: Harvard University Press, 2000).

Jared Diamond, *Guns, Germs, and Steel* (New York: Norton, 1997).

Maurice Dobb, *Studies in the Development of Capitalism* (New York: International Publishers, 1964).

Maurice Dobb, *Soviet Economic Development Since 1917* (New York: International Publishers, 1966).

M. I. Finley, *The Ancient Economy* (Berkeley: University of California Press, 1973).

Robert Fogel and S. Engerman, *Time on the Cross: The Economics of American Negro Slavery* (Boston: Little, Brown, 1974).

Eugene Genovese, *The Political Economy of Slavery: Studies in the Economy & Society of the Slave South* (New York: Pantheon, 1965).

David M. Gordon, Richard Edwards, and Michael Reich, *Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States* (Cambridge: Cambridge University Press, 1982).

John Gray, *False Dawn: The Delusions of Global Capitalism* (London: Granta, 1998).

Eileen Power, *Medieval People* (New York: Barnes & Noble, 1963), especially Chapter 2, "The Peasant Bodo: Life on a Country Estate in the Time of Charlemagne."

John F. Richards, *The Mughal Empire* (Cambridge: Cambridge University Press, 1993).

Frank Roosevelt and David Belkin, eds., *Why Market Socialism?* (Armonk, NY: M. E. Sharpe, 1994), especially Chapter 2, "Thinking about Socialism: Achievements, Failures, and Possibilities" by Irving Howe, 51–82.

Marshall Sahlins, *Stone Age Economics* (Chicago: Aldine Publishing Company, 1974).

Joseph Schumpeter, *Capitalism, Socialism, and Democracy* (New York: Harper & Row, 1942).

Jonathan D. Spence, *The Search for Modern China* (New York: Norton, 1990).

Immanuel Wallerstein, *The Modern World-System: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century* (New York: Academic Press, 1974).

Erik Olin Wright, *Class Counts, Student Edition: Studies in Marxism and Social Theory* (Cambridge: Cambridge University Press, 2000).

Government and the Economy



In August 1963 more than 200,000 people came to Washington, DC, and heard Dr. Martin Luther King Jr. give a speech that marked the peak of the civil rights movement and is today as memorable as Lincoln's Gettysburg Address. "I have a dream that my four little children will one day live in a nation where they will not be judged by the color of their skin but by the content of their character," King told the crowd.

By mobilizing huge marches, challenging racial discrimination in the courts, provoking mass arrests by peacefully violating discriminatory laws and practices, and lobbying Congress for policies to rectify racial inequality and poverty, the civil rights movement transformed America. During the 1960s, new laws made the denial of equal voting rights and the barring of equal access to public accommodations and employment illegal. Deliberate policies to segregate schools by race came under attack by the U.S. Justice Department. Partly in response to the civil rights movement, President Lyndon Johnson inaugurated the "War on Poverty" in 1965.

Did the new policies make a difference? During the 1960s the pay gap between African Americans and others narrowed. But it had narrowed in the 1950s, too, before the civil rights movement. And it is difficult to say how much of the improvement in the 1960s was due to new policies and how much resulted from the fact that the demand for labor was rapidly expanding during that decade (partly due to government spending on the war in Vietnam). The less well-off generally do better when the unemployment rate is very low.

In at least one area, however, the causal connection between civil rights activism and improvement of the lives of African Americans is clear. In 1964 African-American babies were twice as likely to die in infancy as were European-American babies. During the next seven years the infant mortality rate among African Americans fell by more than a quarter (from 41 to 29 deaths per 1,000 live births). During the same period, the ratio of African-American to European-American infant mortality fell by a third, having risen during the fifteen years before 1964.

The improvement in the health of black infants was especially great in the poorest parts of the South, where, prior to the civil rights movement, access to

medical care had been segregated and the facilities open to blacks had been inferior. After passage of the 1964 Civil Rights Act, medical facilities were integrated, and this brought about a significant improvement in the quality of medical care for black citizens. Especially large improvements for black infants were made in reducing deaths from easily treatable causes such as diarrhea and pneumonia.

The connection between the civil rights movement and infant mortality makes it clear that government policies matter. Government can affect the way the economy works, and it is no less true that the economy has a big impact on how the government works. This chapter explores the relationship between the government and the economy.

As we have argued in previous chapters, a capitalist economy operates on the basis of a set of principles, or "rules of the game," designed to organize commodity production for profit using wage labor and privately owned capital goods. Governments, on the other hand, are organized according to different principles, a different set of rules. These rules make possible collective action. Governments—or government leaders—act on behalf of the entire population of a nation, and their actions can be enforced on all those under its jurisdiction.

The main idea of this chapter is that *government, especially democratic government, operates on different principles from those of a capitalist economy, and each shapes how the other functions.* This central idea is expressed in four key points:

1. *The rules of the game are crucial, and are of two main kinds: those that determine how the government functions* (in a democratic government, how that democracy works) *and those that determine how the economy functions. The government affects the economy; but the economy also affects the government.*

2. *A democratic government can ideally be thought of as a process in which the people collectively provide for their own welfare.* Ideally, policies are decided by all. And whether or not everyone participates in decision-making, all are equally subject to those policies, which mandate both rights and obligations backed up by the power of the government to compel compliance. This differs from interactions in the capitalist economy.

3. *Government activities (making and implementing laws, regulations, policies, and programs) have consequences for the distribution of income, wealth, and welfare.* These activities include collecting taxes from people and organizations, as well as paying them benefits or subsidies as provided by law; providing services, or contracting with firms or other organizations to provide goods or services; making many kinds of rules regarding rights, democratic processes, and other matters, and enforcing and implementing them; and managing government-owned resources like land, mineral deposits, timber and water resources, and so forth.

4. *Capital owners and labor (or the public) each try to influence government policy to shift the distribution of income, wealth, and welfare in their favor.* Because workers are in the majority, and in a well-functioning democracy should therefore be able to defend their welfare, capital owners sometimes adopt extraordinary strategies to assert their interests. Therefore, *to ensure that the public interest is served requires an active democratic process with widespread popular participation.*

Rules of the Game: Government and the Capitalist Economy

The government of a country is in the unique position of being able to make rules and then exercise the power to compel both individuals and organizations to follow those rules. No other agent can legally force a person or organization to comply with its wishes unless the agent has already signed an agreement with that individual person or organization. This government monopoly on the use of the power of police and the courts to ensure that people comply with the law is what motivates both owners of capital on the one hand, and workers organized on behalf of their economic interests on the other, to advocate strongly for laws in their favor. Each wishes to have on its side the power of government to compel compliance with the law; neither wishes to be compelled to obey laws that appear contrary to its interest.

For example, most employers want to be able to hire or fire workers at will. But workers want protection against being fired (or not hired) because of pro-union sentiments and activity, or because of characteristics irrelevant to the qualifications needed to do the job, such as race, religion, nationality, sex, sexual preference, and so on. Rules often redistribute income and wealth by serving the interest of one class over the interest of another, or over the public interest. So proposed changes in rules are likely to be fiercely contested. A good example is changes during the nineteenth century in the way corporations could legally function.

The Emergence of Modern Legal Forms of Corporate Business

As capitalism was developing into its modern forms throughout the nineteenth century, owners of capital from time to time saw opportunities to increase their power and profit if they could succeed in changing the laws in their favor. Sometimes this meant seeking new legislation; sometimes it meant winning court decisions overturning previous rules and replacing them with rules more favorable to business ambitions; and sometimes it meant securing action by the executive branch in favor of capital owners. Any of the three main branches of government—legislative, executive, and judicial—might be involved. One such change was intended to reduce the risk of investing in a business, and thereby to encourage potential investors to put more money into business enterprises.

Initially, in the early nineteenth century, many businesses were partnerships. Putting money into a business as a partner meant taking responsibility for the business's debts, however large they might grow. So becoming a partner in a business was taking considerable risk; you could lose your shirt. An alternative to forming a partnership was to incorporate a business—that is, to establish a corporation and issue stock, with approval of the corporation's charter by the legislature. This could spread ownership of the assets of the business more widely. However, in this period buying stock in a corporation did not necessarily alter the amount of risk an investor faced. The prevailing legal philosophy was that stockholders could be held liable for at least some of the debts of a business in which they invested, beyond the value of the shares of its stock that they purchased.

To say that an investor in a corporation has **limited liability** means that the investor is not legally responsible for the debts of the corporation. The most the investor can lose if the corporation's stock value falls to zero is the money the investor paid to buy shares of stock.

Limited Liability

The key new legal concept that changed this situation in the nineteenth century was the principle of **limited liability** for investors in corporations. This idea was that those who provided investment funds to a corporation by buying shares of stock in it

(and so becoming stockholders) would thereby risk only the dollar amount of value of the stock, but would not incur any legal responsibility beyond that if, for example, the business failed. If the corporation went bankrupt, the value of the shares of stock would typically fall to zero, so stockholders would lose the money they had invested; but they would lose nothing else, and in particular could not be made to help pay the debts of the corporation.

In the early nineteenth century the principle of limited liability was not yet widely accepted in the law. However, by about 1860 in the United States, laws were on the books in most states establishing that corporate stockholders did enjoy limited liability. This naturally made it easier to convince potential investors to invest in a corporation chartered in one of these states, and so attracted new funds to the corporation that it could use for investment. This helped capitalism to flourish in the last half of the nineteenth century. In the United States, over a century later, limited liability was extended by law even to partnerships that were not corporations, in the form of the "limited liability company" or LLC, which is now quite widely used.

The main point of this story is twofold. First, government has the power to alter the conditions under which businesses operate—including the degree of risk faced by those who get involved in business enterprises. This is a strong incentive for capital owners to press, in all branches of government, for laws and regulations favorable to their activities. Second, historically, once capital owners envisioned new ways to increase their profits and power that required only a change in laws or regulations (the rules of the game), they did persistently press to get such changes. In other words, the development of the capitalist economy itself, opening up new possibilities for profit, inspired capital owners to press for changes in laws, regulations, and court interpretations of the laws.

Corporate Personhood

The next major change was to define a corporation as a legal person. After the Civil War the Fourteenth Amendment, ratified in 1868, said:

> Section 1. All persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the State wherein they reside. No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; **nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.** (http://www.archives.gov/exhibits/charters/constitution_amendments_11-27.html; emphasis added)

For two decades from about 1868 to 1886, lawyers for corporations sought to get the courts to agree that a corporation was a *legal person*, and therefore entitled to share equally in these already established legal rights of *natural persons* (human beings). These included, for example, the right for a corporation to buy stock in another corporation (something only human beings were permitted to do up to that point), the right to privacy, and the right to equal protection under the law (that is, equal with the protections accorded to human beings). Their arguments relied in part on the fact that corporations had already been frequently described as "artificial" persons in legal contexts. They tried to argue that corporations were entitled to the protections

accorded to all other persons, namely human beings. But the courts repeatedly rejected such arguments, so that for two decades no court decision recognized corporations as legal persons.

Then a sequence of events took place which Thom Hartmann recounts in his book *Unequal Protection* (2004): The idea that a corporation was indeed a legal person appeared in the unofficial notes accompanying the Supreme Court's 1886 judicial decision in *Santa Clara County v. Southern Pacific Railroad Company*, even though the official decision did not give the idea of legal personhood of corporations any credence. From there this unofficial language somehow leaked into legal doctrine and began to be treated as a precedent, and the principle was then established in official Supreme Court decisions in the following years. (Hartmann discusses whether the unofficial notes in the 1886 decision were a simple clerical error, or may possibly have been intentional.) By 1889 the Supreme Court had officially ruled that a corporation is a legal person entitled to due process of law and to equal protection under the law. Corporations then quickly took advantage of their newly acquired legal status to secure legal decisions that (1) struck down prohibitions against corporate political campaign donations and lobbying, relying on the legal right of "persons" to free speech; (2) struck down laws requiring a corporation to open its books as a condition of getting a legislative charter, on the grounds that a "person" has a right to privacy; and, (3) using the idea that "persons" were protected against discrimination, successfully overturned a law in Florida that charged a lower business license fee to small businesses than to chain stores.[1]

From Competition to Monopoly

One impact of this change in the rules was on the ability of businesses to form monopolies, or at least large firms that combined many of the firms previously competing within an industry; such a move typically allowed the combined entity to charge higher prices and reap higher profits. However, until the late 1880s definition of a corporation as a legal person, the best available procedure for doing this was cumbersome. To be sure, an individual human being could own several businesses, or buy controlling shares of stock in several businesses. But almost the only way in which businesses as such could join together was the one initiated by oil magnate John D. Rockefeller in the form of the first *trust*, the Standard Oil Trust, in 1882. The way it worked was that corporations that wanted to join the trust would sell their stock to a board of trustees that headed the trust, and would receive in exchange an equal value of certificates in the trust itself, along with rights to receive any dividend payments that the trust might decide to pay out of the profits it earned. In that way all the businesses participating in the trust were dependent on it, and it was the board that ran the trust. In all forty-one firms were part of the Standard Oil Trust, under a board of trustees of nine, including Rockefeller. The oil trust consolidated, and in a sense legitimized, the immense power that to some degree Rockefeller had already concentrated in his hands. Now one enterprise could smooth out the boom-and-bust cycles in oil prices and profits, making production in some sense more efficient—but at the same time extraordinarily profitable. Monopoly took the risk out of the oil

[1]Thom Hartmann, *Unequal Protection* (Rodale, 2004), 120–21; Chs. 6–7 of the book are the source for much of this account of changes in nineteenth century legal principles regarding the evolution of the forms and status of corporations.

business by replacing *competition* with *command*. This benefited owners of capital in the oil industry, but hurt both the public and other industries that bought oil as an input into their own production.

Up until 1886 or a little later, another option for combining the efforts of two corporations was for corporation A to become a holding company, owning corporation B, by buying B's stock. The snag was that at that time only natural persons—human beings—were permitted by law to buy stock. True, a state legislature could in principle make an individual exception and pass a bill expressly approving the holding company arrangement; but public sentiment was running strongly against monopolies, and this was generally not a viable option.

Some states responded by passing laws against the formation of trusts ("antitrust laws"). Trusts, which to the public (and to small companies not included in the trusts) meant monopolies with far too much power, also provoked a response at the federal level, and the Sherman Anti-Trust Act was passed in 1890. By 1892, successful legal action by the state of Ohio forced the Standard Oil Trust to dissolve. But Standard Oil of New Jersey (SONJ) was by then able to become a holding company, and after 1899, when state law was changed to allow a corporation to buy stock in a corporation established in a different state, SONJ expanded its holdings, moving in the direction of the kind of sprawling corporate structure we know today. A few years later, though, a federal antitrust case was brought against Standard Oil, leading to a 1911 court decision breaking it up. At this stage in history, some combination of the public interest, the interests of smaller competitors not included in the monopoly, and the interests of businesses that were buyers of oil, prevailed.

From National to Transnational

A **transnational corporation** (TNC), or multinational corporation (MNC), is a firm earning profits on its global operations by locating its corporate entities, called **affiliates**, in multiple countries, in whatever combination of sites yields the highest profit for the global corporate structure as a whole, based on the wages, materials costs, markets, government policies, taxes and subsidies, and local markets for relevant outputs that exist in various locations.

The legalization of holding companies also had an enormous long-term impact in allowing the growth of transnational corporations. A transnational corporation is one whose corporate structure includes entities, or affiliates, in multiple countries. The global management maximizes overall profits by deciding where to locate its activities based on wages, materials costs, markets, government policies, taxes and subsidies, and local markets for relevant outputs.

A transnational corporation typically consists of a headquarters firm (for example, in the United States) that owns a portion, often a majority, of the stock of each of various **affiliates** (also corporations) in other countries. Although corporations were already creating holding companies by the turn of the twentieth century, it took some time for the full possibilities of transnational corporations to be realized. This was in part because even the basic technology of international communications was still in its infancy; telegrams could be sent across the ocean via cables laid in the nineteenth century, but trans-oceanic cables for telephone communications were first laid in the 1950s. Only well after World War II did the majority of large capitalist enterprises in the United States become transnational.

So it was that the early nineteenth century principle that corporations should only be granted existence if they served the public interest became gradually transformed into the principle that corporations had their own independent rights, with little or no general obligation to take account of the public interest, as long as they did not explicitly violate the law. And if corporate activity conflicted with the law, in some cases it was eventually not the corporations, but the law itself, that ended up changing. As we will see in Chapter 7 as well as Chapter 19, the role of corporations in the economy continued to shift.

An **affiliate** is a corporate entity in which the parent corporation holds enough shares to give it control, which usually means it holds more than 50 percent of the affiliate's voting stock (some shares of stock carry with them the right to vote on company policies; others do not).

As recently as the 1980s, when some corporate executives violated laws during the savings and loan debacle (discussed later in this chapter), many of them ended up serving prison time for fraud and other crimes. In contrast, in the last decade, following the subprime crisis and Great Recession, even when corporations had violated the law, they paid fines often in the billions of dollars, but the executives who carried out or oversaw the illegal activity generally did not suffer penalties. This is due not mainly to changes in the law, but to changes in the ways in which the laws were and are enforced (which is also part of the rules of the game).

History thus confirms that rules and rule changes matter. Rules that allowed a capitalist to attract investment made a big difference in how fast a corporation could grow. Later rules that evolved from corporate personhood to make holding companies legal allowed controlling a large linked group of enterprises through a legal ownership network, and this magnified what one entrepreneur, or a group of entrepreneurs, could do. The world was thereby transformed—and the sheer size of such firms increased the challenge for democratic institutions to regulate such firms in the public interest.

Democratic Government as Collective Provision for the General Welfare

At its best, democratic government can be thought of as a process by which the people collectively provide for their own welfare. Ideally, rules and policies are deliberated and decided by all, and all are equally subject to those rules and policies, which define general rights and obligations, as well as rights and obligations within specific government programs like Social Security and Medicare. The two aspects of this democratic process, namely, the *collective* nature of the decisions and the *compulsory* nature of compliance with those decisions, are both necessary for the whole positive result: that the people together indeed, as the preamble to the U.S. Constitution says, "promote the general welfare, and secure the blessings of liberty" to this generation and the next.

Interactions involving government differ in this respect from interactions in the capitalist economy. In a capitalist economy, at least in theory, interactions are meant to be voluntary exchanges. A person or firm makes a decision to buy or not to buy, to sell or not to sell, at a particular price. An employer decides whether or not to hire a job applicant, and an applicant decides whether or not to accept a job offer. Nothing about this is compulsory within the bounds of the capitalist economy, although there are usually clear differences in how much power each participant in the exchange has—a topic to be discussed at length in Chapter 12.

The fact that in and of themselves, exchanges in a capitalist economy are voluntary creates a need and a role for government, to bring into play both the public interest, and the interest of all business enterprises taken together. Even advocates of small government tend to agree that government should protect property rights by preventing and penalizing theft. Most people would agree that fraud should also be illegal and should be prosecuted. This is because, under capitalism, nothing prevents sellers from overstating the value of what they offer for sale, or failing to tell the whole truth about the qualities of an item or a service for sale. Nothing, that is, except the danger that the seller will gain a bad reputation, and lose future sales—a possibility that was perhaps a greater danger in small pre-capitalist communities

than in the big cities of today. Moreover, the majority of business enterprises may well have an interest in preventing fraud, for two reasons. One is that households are more likely to spend money freely if they feel sure that any defective product or service will be made good—replaced, or the purchase price refunded. The second is that businesses themselves may sometimes be misled or defrauded by suppliers, and they, too, want the assurance that they will be compensated for flaws in products or services they buy.

Similarly, in the absence of government rules against discrimination on the basis of race, sex, or other characteristics of the applicant, firms are free to hire whomever they please. In a capitalist economy, in a certain sense everyone is on his or her own. What democratic government potentially brings to this scenario is greater fairness and equality, and lower risk of personal or business disaster, than would be produced by the capitalist economy in the absence of government.

Government also is a vehicle for promoting the general welfare. As early as the late nineteenth century in some European countries, notably Germany, governments initiated old-age pensions and some forms of publicly supported health care. Before the United States legislated Social Security in 1935, pension programs had already existed in many European countries for decades. Similarly, collective provision for health care is now very widespread among countries at the U.S. level of average income and even far below—because it reduces the cost of medical care to everyone by largely eliminating costly paperwork; the population pays taxes, and the government simply provides health care free, or largely free, to patients. The United States stands almost alone among countries at its income level in lacking universal health care.

Universal pension and health programs, except for their enormous size, bear a strong resemblance to ways in which people have gathered together locally in developing countries to provide for their own welfare. Everyone in a local mutual assistance group, for example, regularly contributes a small fixed sum of money weekly, and when someone needs the money for an emergency (injury, illness, funeral, or other need), that person is allocated some of the money by the group. Some labor unions also originated in such mutual aid societies. The advantage of such arrangements is that, by spreading the risk, it greatly eases the worries of those who otherwise would have nobody to rely on but themselves. Insurance is less risky when the risk is pooled among a large number of people. A considerable share of the growth of government among the highest-income countries is accounted for by growth in programs that "promote the general welfare" through supporting the elderly—who are now more numerous relative to the nonelderly population, because advances in medicine and public health allow people to live longer. Of course, other major expenditures by governments, including programs that "provide for the common defense" (another quote from the Preamble to the U.S. Constitution), are also a significant factor in government spending, a topic that will be taken up in Chapter 19.

Government and the Distribution of Income, Wealth, and Welfare in the Nineteenth Century

Every American high school student knows the 1800s as the era of building canals and railroads. And the 1860s and early 1870s brought the transcontinental railroads that unleashed economic growth by connecting distant reaches of the country with one another. Now agricultural implement factories could send their

products nationwide; in this period, farming using animal-drawn plows became widespread. In turn, the expanded markets for agricultural equipment stimulated further innovations that drove productivity increases such as those shown in Figure 1.1 for 1830–1890.

Government played an enormous role in the building of both canals and railroads.[2] In the first half of the nineteenth century, state and local governments donated much land, and also invested funds, first in canal construction and then in local and regional railroads. An estimated 70 percent of the money that went into funding canals came from public sources, while up to 1861 about 25 to 30 percent of the entire investment in building railroads also came from public funds. Mostly the source was state and local governments in places where businesses expected to benefit from new transport facilities. State and local governments also gave about 300 million acres of land to railroads.

Although available data show federal government revenue and spending at the end of the nineteenth century at around 3 percent of GDP or less (although in the Civil War it is estimated to have risen to around 7 percent), the federal government gained control of vast tracts of land in areas where railroads were projected to be built. It did this in part through negotiated treaties with Plains Indian peoples (until 1871, when Congress forbade future treaties)—and despite the Indians' view that the land belonged to no one. The tens of millions of bison who were central to Plains Indians' way of life were decimated in a few decades by the railroads and settlement of the West, while the Indians resisted, but were defeated.

Until 1862 the federal government's contribution was mainly to give relatively small amounts of land to the states, which the states often granted to railroad-building companies as part of a package of subsidies linked to the obligation to actually complete the rail lines. In the 1860s, federal government subsidies came to the fore with ambitious plans for a transcontinental railroad. Congress judged that a massive multi-state project requiring building rail lines through remote areas and over mountain ranges was too risky to attract sufficient private capital. So beginning in 1862 it gave the railroad companies hefty financial subsidies, plus 10 square miles of land for every mile of track that a railroad building company committed to construct, in exchange for a legal commitment to finish the project within a specified time period. Some builders failed to meet their commitments, and about one-sixth of the 175 million acres that the federal government granted to the railroad companies in this period was forfeited back to the government as a result. Still, the majority of the railroads were built as planned.

Railroad Financing: Public and Private Gains

What lessons can we learn from this episode about the relationship of the government to the economy? First, some kinds of government intervention can benefit both private business and the public. If government subsidies to railroads had appeared to be simply gifts to private business, benefiting the railroads but not the public, they would likely have garnered less legislative support. Railroads allowed small towns with new access to a transport network to flourish, and allowed new inventions (like agricultural implements) to spread quickly to all parts of the country, thus rewarding

[2]Much of the following discussion is based on the detailed account in Carter Goodrich, *Government Promotion of American Canals and Railroads 1800–1890* (Westport, CT: Greenwood Press, 1960).

innovation. Private business unquestionably benefited, but there were also benefits to the public. Economic growth cheapened consumer goods, both directly by reducing the cost of transporting them, and indirectly by making available better tools to produce them. This stimulated production and employment.

Of course, in assessing the overall benefit to the general public, we need to take into account not only the benefits the public received, but also the costs the public indirectly paid when governments, supported by taxpayers, gave land and loans to the railroad companies. If loans were repaid, or the government purchased railroad stock that later paid dividends over and above the cost of buying the stock, then the government's decisions did not entail a loss to the public. In fact, though, where state and local governments subsidized railroads by buying shares of railroad enterprises, they more often than not ended up breaking even or losing money. The federal government also was not a big winner, and quite possibly not a winner at all, given that the valuable land it donated multiplied in value due to the building of the railroads, and the proceeds mainly went to railroad stockholders and owners rather than being returned to the public coffers.

At the same time, many of those who ran railroad companies—and in some cases did so without putting in much of their own money—became stupendously rich, and public subsidies certainly helped them do this (especially in this era before income taxation). Those who made a considerable part of their fortune from railroads included Leland Stanford, who founded Stanford University, and Cornelius Vanderbilt, who founded Vanderbilt University. Vanderbilt is said to have been the second wealthiest man in history after John D. Rockefeller, with Vanderbilt's fortune at his death variously estimated at $140 to $200 billion in today's dollars, more than the wealth of the richest person in the world today. And there were many more.

The second lesson, then, is that while government spending can create public benefits, it can also have effects—sometimes small, sometimes large—on the distribution of income, wealth, and welfare. Extreme inequality of wealth and income arose in the last half of the nineteenth century (The Gilded Age). It was driven in part by extraordinary government generosity to private railroad companies, in part by the absence of income and inheritance taxes, and in part by lobbying and bribes that persuaded legislators to act in the interests of owners of capital. This polarization of income and wealth had far-reaching effects on the future development of capitalism.

The third lesson is that, in light of the lucrative gains to be made from business enterprises when laws, policies, and programs are designed in favor of business, owners of capital have a powerful incentive to spend funds to shape legislative, judicial, and executive decisions in their favor. Stories of bribery and business influence on politics in the era of railroad-building are rife, including at the state and local levels. But none is more spectacular than the story of the Union Pacific Railroad, which oversaw the building of the eastern portion of the transcontinental railroad in the 1860s.

While the details can be read in other sources, in brief the story is this: The directors of Union Pacific set up a firm called Credit Mobilier of America to carry out the construction of the railroad and to finance itself through issuing stock. The Union Pacific directors completely controlled Credit Mobilier, but led everyone to believe that it was an entirely independent entity with which Union Pacific had an arm's-length contractual relationship. The two enterprises got both loans and land grants from Congress, together amounting to over $100 million in value. Yet

the entire cost of construction was just over half that amount, so that the profit made on the whole project amounted to $44 million in dollars of that time (about $60 *billion* in today's dollars). A news exposé in 1872, followed by an investigation, revealed that in 1867 a congressman whom Credit Mobilier had made its head distributed shares of Credit Mobilier stock to various other members of Congress at below its then-current market price, in exchange for their votes favoring the company's fortunes. Ultimately as many as thirty members of Congress were found to have received shares of stock, which ended up paying lavish dividends. Yet much of what Union Pacific directors did was technically legal, and virtually no one received a penalty: some congressmen were censured, but apparently no one was jailed.

In this instance, what owners of capital sought to influence was government decisions about a specific project, rather than permanent rules about government or the economy. Yet these government decisions, which we include here as changes in the rule of the game, sharply altered the distribution of income and helped lead into a stage of capitalism in which large corporations and large fortunes held sway.

Other Rules: Winners and Losers

Though the land grants and generous loans to railroads are a particularly striking example, in fact nearly everything the government does has some kind of distributional effect: someone gains or someone loses, or both. Laws and programs may specifically increase the profits of some business sector, or may specifically help workers or poor people. Even those designed to benefit the general population (such as legislation in favor of clean air, or in favor of safety in trucking) may benefit some more than others, and impose costs on some but not others.

For example, capital owners or the wealthy (two groups that overlap considerably) typically gain disproportionately when the government reduces or eliminates inheritance taxes (estate taxes), because inheritance taxes are typically imposed only on large estates. The wealthy also tend to be the ones who gain when government reduces personal income taxes on high incomes, and when the government lowers capital gains taxes (taxes on gains from buying assets cheap and selling them dear), because the incomes of the rich consist to a much greater extent of gains from trading assets than is true for those with lower incomes. All these types of changes have been implemented in the United States in recent decades, a topic that will be explored further in Chapter 19.

On the other hand, workers benefit from a higher minimum wage, as well as from low interest rates on credit cards, home mortgage loans, and other lending. They also benefit from strong labor rights, including overtime pay for long workdays or work weeks, generous maternity and parental leave policies, holiday pay, sick pay, and employer health benefits (at least in the United States, one of the only high-income countries that does not provide health care to the whole population with low or zero fees). There have been some gains for workers in recent decades, but at the same time many losses. The purchasing power of the minimum wage has fallen since the 1970s (see Chapter 7), and many kinds of labor rights have been rolled back, or remain far more limited than in European countries with similar levels of income.

In fact, it is striking that, although workers, not owners of capital, are the large majority of the population, nevertheless the rule changes during about the last four decades have tended to be more in favor of capital owners and the wealthy than

in favor of the bottom 90 percent of income earners. Taxes fall less heavily on the rich than in the past; real hourly wages stopped growing in the 1970s in the United States. We have seen in previous chapters that inequality of wealth has been growing. This is in part because government has changed its rules and practices in such a way as to shift income, wealth, and welfare toward the "haves" and away from the "have-nots," as well as toward corporations and away from labor. But is it inevitable that "them that has, gets"? Of course not; it depends on how strong a democratic movement the majority are able to mobilize. When capital owners want to shift the rules governing the economy in their own favor, part of their strategy is typically also to shift the rules that govern the democratic process in their favor. Therefore, to avoid a continuing shift in the rules of the game in favor of rich and the owners of capital, and against the interest of the rest, requires an active democratic process and a vigilant public.

Consider an example that touches on both rules that benefit capital owners and rules that benefit the rest of the population. In the United States in 2003, Congress passed a new feature of Medicare (the program that provides subsidized medical care mainly for those sixty-five and over): Medicare, Part D, an insurance program for purchases of pharmaceutical products. The plan then went into effect in 2006. A startling provision of the bill was that it prohibited the federal government from negotiating with drug companies for discounted prices, although in the past it had been customary (and sensible) for bulk purchasers to expect to bargain successfully for price discounts. It was startling because tying the federal government's hands and forbidding bargaining for lower prices was clearly not in the taxpayers' interest; instead, it was a gift to the pharmaceutical companies. In late 2004 a *New York Times* editorial[3] publicized the fact that the congressman whom it identified as the "chief architect" of the bill had taken a position as the principal lobbyist for the pharmaceutical industry. Such a move was legal for a member of Congress as long as the mover did not lobby congressional colleagues for a period of one year; even so, he was still permitted to socialize with members of Congress and to make campaign donations. But why does the rule only prescribe a one-year moratorium on lobbying for such a person, and not, for example, five years? Where there are those who seek to influence, there are also those willing to be influenced, for a reward. But the rules of democratic government could and should be structured to guard against such influence. Making sure that happens is the job of a vigilant public actively involved in democratic decision-making.

Contention Between Capital and Labor over Rules

Because the rules are so important, as the story of the legal emergence of corporate capitalism shows, capitalist enterprises (and sometimes their industry associations) often strategize to change the rules in their favor over a period of years or decades. They may either seek to change the rules governing the economy, or to change the rules governing the workings of democracy. Changing the rules of democracy may be indirect, but in the long run may reap more gains for owners of capital.

There are numerous recent examples. One is changes in the rules about campaign financing. If campaigns are publicly financed, as some advocate, there is less opportunity for the wealthy to control the democratic process. If campaign

[3]"The Drug Lobby Scores Again," *New York Times*, editorial, December 17, 2004. http://www.nytimes.com/2004/12/17/opinion/the-drug-lobby-scores-again.html?_r=0

ads, even issue-oriented ads, are outlawed unless publicly financed, then this limits the weight of wealth in democracy further. But the 2010 Supreme Court ruling in the case *Citizens United v. Federal Election Commission*, a 5–4 decision, overturned certain campaign finance rules passed by Congress in 2002, and allowed independent nonprofit corporations to have an enormous impact on election campaigns, particularly through television advertising. The decision now also applies to for-profit corporations, associations, and labor unions. It paved the way for a manyfold multiplication of the role of the wealthy in political campaigns, with dramatic results.

Another example is efforts at the state level to redraw legislative districts to favor one party. If one party manages to get control of a legislature and redraw congressional districts, it can in principle semi-permanently alter the political landscape. As a hypothetical example, imagine a state with ten congressional districts and a population of 5 million voting age citizens, of whom 3 million (or 60 percent) are registered in the Green Party, and 2 million (or 40 percent) are registered in the Purple Party. In such a state, if the Purples temporarily get a legislative majority (perhaps due to low voter turnout in an off-year election), they might well prefer to redraw districts, putting half the Greens (1.5 million of them) into three districts of half a million Greens each, in each of which they would be 100 percent of the voters. They would then distribute the other 1.5 million Greens evenly over the remaining eight districts, where with about 214,000 voters per district (= 1,500,000/8) they would make up 43 percent of the 500,000 voters in each district. All 2 million Purples would be distributed evenly over those seven districts, so that with about 286,000 voters in each district (= 2,000,000/7) they would form 57 percent of the 500,000 voters in each district. (Of course this redistribution would happen through redrawing the district boundaries rather than by actually moving people physically from one district to another.) The three 100 percent Green districts would, of course, elect Greens. But the remaining seven districts would usually elect Purples. In this way a minority party with 40 percent of registered voters could win 70 percent of the seats in Congress (seven out of ten districts). If they did this in enough states, they might not only have enough seats to pass whatever legislation they wanted in Congress, but also have a two-thirds majority sufficient to override any presidential vetoes. Of course, legislative redistricting can only be done every decade or so, and can be challenged as unfair in the courts, sometimes with success, so it is not a sure thing. But it is clearly in use at present as one of a number of tools by which a minority party can gain long-term control of a legislature. The point, once again, is that the rules matter. A method of voting that ruled out such schemes would be more likely to preserve democracy than one vulnerable to them. Such concerted efforts, especially if well funded, have a good chance of succeeding, in part because the general population may not be aware of contention over the rules, as these efforts are often carried out away from the public eye.

Of course, workers and their organizations also can, and sometimes do, seek favorable legislation, and there have been numerous examples. One was the Fair Labor Standards Act of 1938 that limited work hours and mandated additional pay for hours worked above the standard work day or week. Still another was the Occupational Safety and Health Act of 1974, which established standards and inspection procedures to try to assure worker health and safety on the job. A highly organized and well-informed populace, relentlessly advocating for a fair set of rules that fosters democracy, can win rules that favor both democratic government and a fair economy. But it is always a struggle.

Voter Turnout in the United States and Around the World

A key problem for the democratic process, and therefore for the welfare of the large majority of the population, has been the fact that, particularly in the United States, many citizens who have the right to vote have not been exercising this right. Figure 6.1 shows that in the United States the share of eligible voting-age citizens actually voting in presidential elections fell from an average of 78 percent from 1872 to 1896 to an average of just 58 percent from 1900 to the present. For example, in the 1872 presidential election, 77 percent voted. In the 1896 election, when the agrarian populist William Jennings Bryan challenged the eastern establishment, 80 percent voted (Bryan lost). As for midterm elections, voter turnouts were typically lower than in presidential elections, but also fell from an average of 65 percent in 1870–1898 to an average of 43 percent during 1900–2014.

Voter turnout fell sharply in the twentieth century. One cause was that southern states imposed poll taxes and absurdly difficult "literacy tests," to prevent African Americans from voting. In the 1920 and 1924 presidential elections, a low of 49 percent voted. Even in the 1936 election, seen as a referendum on FDR's New Deal, only 61 percent of eligible voters cast ballots (FDR won by a landslide). In the 1980 election, when the conservative counterattack against New Deal policies came to fruition, only 54 percent of those eligible voted for Ronald Reagan or Jimmy Carter. In 1996 and 2000, barely half of those eligible to vote went to the polls; so in 2000 George W. Bush won with the votes of less than 26 percent of the voting-age population.

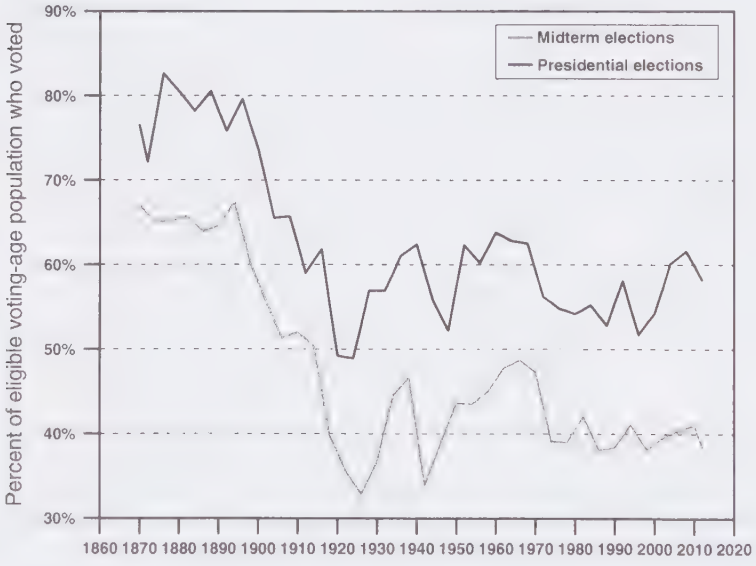


FIGURE 6.1 The American voter: An endangered species?

The data are for the percentage of all eligible voters who actually voted. Note that these figures may be inconsistent with other reported voter turnout data because of differing definitions of voters (did they cast a ballot at all? or: did they cast a vote for president?) and of the group of which they are a share (the whole voting age population? only voting age citizens? or only those voting age citizens who have registered to vote?).

Sources: Michael P. McDonald, "National General Election VEP Turnout Rates, 1789–Present," United States Elections Project. Accessed June 24, 2017, available at http://www.electproject.org/national-1789-present.

Many other countries have higher—in many cases consistently higher—average voter turnouts, however. As Figure 6.2 shows, during 2002–2014 nations as different as Bolivia, Costa Rica, Iceland, and South Africa had higher voter participation rates than the United States. About a dozen countries both make voting compulsory and enforce that law, usually with exceptions such as for elderly persons. Of the countries shown in the figure, Argentina and Australia have, and enforce, compulsory voting.

Of course, low voter turnout in the United States is not evenly distributed among citizens. As Figure 6.3 shows, the lower the family income, the less likely it is that a citizen of voting age in that family actually went to the polls in November 2012; and this pattern has been true for decades. In families with incomes above $100,000, 77 percent of their members who were citizens of voting age reported voting in the 2000 presidential election; but among the poorest of the poor (families with incomes below $10,000), only 47 percent of eligible citizens actually voted. The inset chart within Figure 6.3 shows that participation in the 2012 presidential election was higher for employed persons than for the unemployed, and higher for whites and blacks than for Hispanics or Asians.

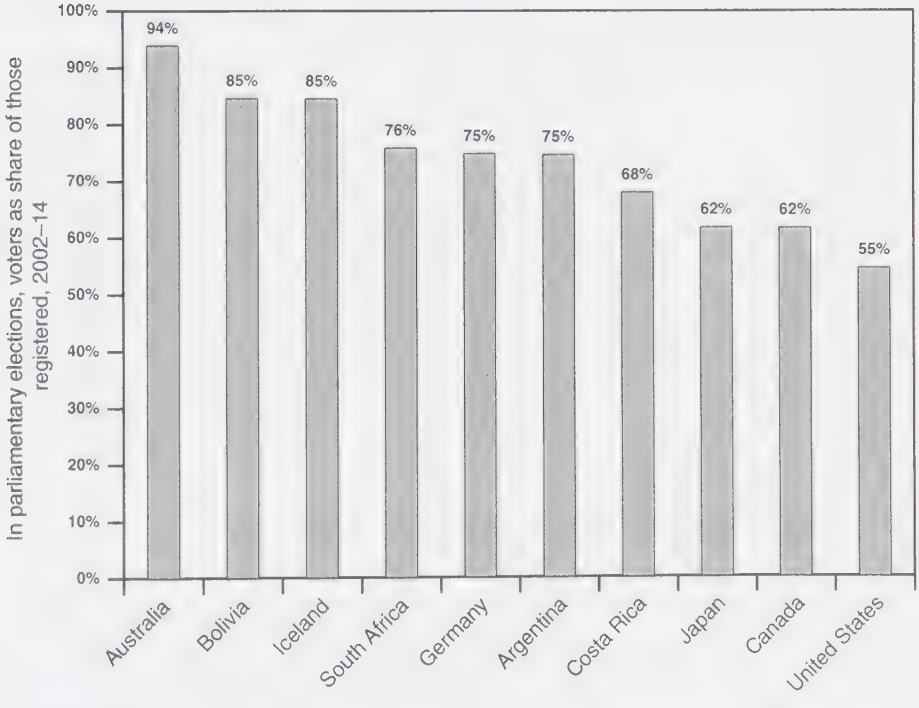


FIGURE 6.2 Voter turnouts around the world.

For selected countries, voter turnouts for parliamentary elections are averaged over 2002–2014, including all parliamentary elections, whether or not they coincided with election of a president or other chief executive. Voters are counted as a percent of all those registered to vote. Both Argentina and Australia have compulsory voting laws, and enforce them; the other countries shown do not.

Source: Data from International Institute for Democracy and Electoral Assistance, "Voter Turnout," available at http://www.idea.int/data-tools/data/voter-turnout.

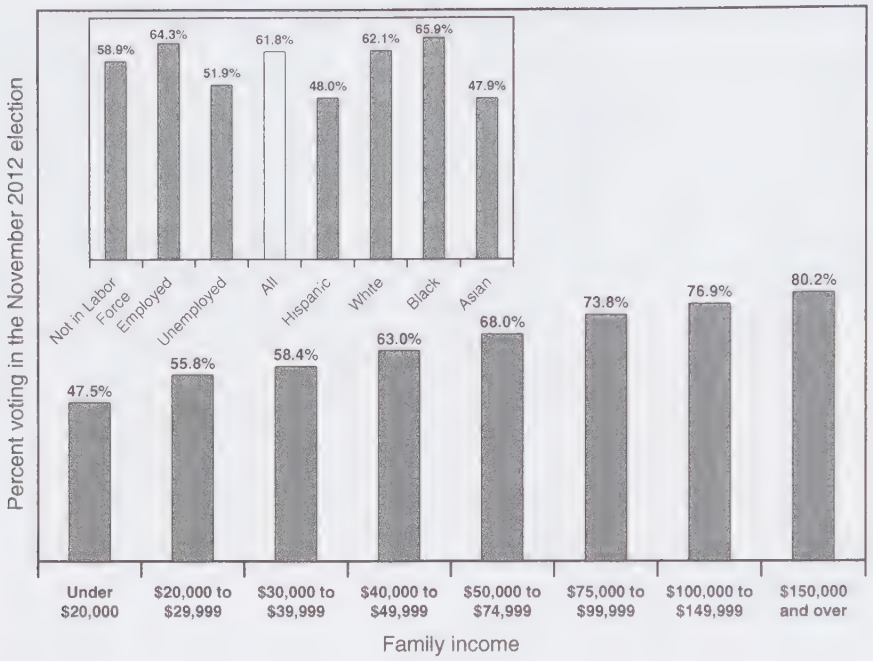


FIGURE 6.3 Who votes? U.S. voting patterns in the presidential election of 2012.

This figure shows the fraction of citizens in various segments of the population, by family income, employment status, race, and ethnicity, reported to have voted in the U.S. presidential election of November 2012. Voter turnout can be measured in various ways: by the number voting as a percent of the voting age population (some of whom are noncitizens, or not registered to vote), or as a percent of the number of citizens of voting age who actually voted (although those citizens who have not registered could not actually vote), or as a percent of the number who are registered to vote (a smaller group still). The numbers in this figure are based on the second method: for each category, the number in that category who self-reported having voted in the November 2012 election, divided by the number of citizens of voting age in that category. Some of those surveyed did not answer the question either yes or no, and so they were not counted.

Source: U.S. Bureau of the Census, Current Population Survey, November 2012 and November 2014, Tables 4b, 6, and 7; available at https://www.census.gov/data/tables/2012/demo/voting-and-registration/p20-568.html

Democracy in the Twentieth Century

Historical changes in the relationship between the government and the economy illustrate the principles of this chapter in the United States in the twentieth and early twenty-first centuries just as well as does U.S. history of the nineteenth century. Rules and rule changes had profound effects. Democracy was used as a way collectively to promote the general welfare. Changes in the rules affected the distribution of income, wealth, and welfare among people and between the owners of capital and workers. And contention over the rules often fell along class lines, with the large majority, the workers, having to use the democratic process actively just to defend their incomes and welfare against proposals to reduce them. As income and wealth inequality grew from about the 1970s on, the contention over the rules of the game also grew, and the rules changed in ways that predominantly disadvantaged workers and advantaged the owners of capital.

The Great Depression of the 1930s was a personal disaster for many. With a quarter of the workforce unemployed at the height of the Depression, many families were reduced to poverty. The banking crisis of 1933 pushed the economy deeper into depression, and led almost at once to new regulations prohibiting commercial banks (those that took deposits from the public and made loans to the public) from engaging in the risky business of trading in securities. Likewise, the desperation of poverty generated both social unrest and widespread support for new laws that protected the rights, wages, and working conditions of wage workers, along with new social programs like Social Security that provided old-age pensions. These new rules reflected the sentiment that if people became unemployed and poor because of economic forces larger than themselves which they were largely powerless to reverse, then the government should take care of them and help them get back on their feet. The idea of using democracy to provide for the general welfare through the minimum wage, social security, workers' compensation (insuring workers against on-the-job injuries) and job-creating programs like the Civilian Conservation Corps was key to helping people survive the Depression. Such programs were also important in helping people through the recent Great Recession.

And just as businesses in the nineteenth century were able to secure the principle of limited liability for corporate stockholders to reduce the risks they faced, so too, for workers, did the New Deal reduce the risks associated with unemployment. It did this with the passage of state unemployment insurance programs by half a dozen states, and then a national unemployment insurance program in 1935. The program provided, and still provides, income support in the form of unemployment benefits to persons who qualify under state rules.

To help support child nutrition, the program Aid to Families with Dependent Children (AFDC) was also created in the 1930s, helping both children and their families. (It was replaced by the Temporary Assistance for Needy Families (TANF) program in the mid-1990s.) To reduce the risk of losses to depositors when a bank failed, a new law created the Federal Deposit Insurance Corporation, which insured that deposits up to a specified amount would be replaced if the bank in which they were held collapsed. Likewise, the social security system was created in 1935, with workers and employers paying into it during a person's working years, entitling them to draw a monthly check after they retired. Social Security is credited with reducing poverty among seniors from 45 percent to 9 percent by lifting about 13 million seniors out of poverty.[4] These new social insurance and income support programs were created between one and four decades later than similar programs in European countries at comparable levels of income.

Just as nineteenth century events were used to explain the principles of this chapter, so too can events in recent decades illustrate the same concepts. In the 1980s, under the presidency of Ronald Reagan (1981–1989), Congress deregulated a particular category of financial institutions, namely savings and loan institutions (S&Ls, or "thrift" institutions). These alternatives to ordinary commercial banks existed mainly to collect people's savings, put them into savings deposits that paid a reasonable interest rate, and lend the money out for home mortgages and other purposes.

Other features of the financial sector were deregulated as well. Regulation Q, which had limited the interest rates that banks and S&Ls could pay on various kinds

---

[4]Paul N. van de Vater and Arloc Sherman, "Social Security Lifts 13 Million Elderly Americans out of Poverty," Center on Budget and Policy Priorities, August 11, 2010, available at http://www.cbpp.org/research/social-security-keeps-20-million-americans-out-of-poverty?fa=view&id=3260.

of deposits, was lifted on S&Ls, and interest rates were allowed to rise. This and other changes in laws and regulations had an enormous impact. Within a very few years many S&Ls went bankrupt, and government regulators then discovered that widespread fraud was a central causal factor. Although many thousands of those who ran the thrifts were investigated, and around a thousand of them were prosecuted, convicted, and sentenced, some also accumulated considerable wealth.

Reagan also led the way in getting Congress to reduce the top income tax rate on the highest incomes from 70 percent to 28 percent over the course of his administration (later the 28 percent rate was raised again, but only slightly). This and other measures—along with the deep 1982–1983 recession, with over 10 percent unemployment at its peak—contributed to a worsening of inequality of both income and wealth. The rich began getting much richer, while the incomes of the rest stopped growing.

The government continued to deregulate the financial sector in the 1990s, culminating in 1999 with the repeal of the main Depression-era law, the Glass-Steagall Act, that had separated commercial from investment banking and thus insulated the public from the risky activities in which investment banks often engaged. (Investment banks manage the sale of new stock issues for corporations, as well as the sale of various kinds of securities such as corporate bonds.) Chapter 18, on crises and the 2007–2009 Great Recession in particular, will explain in detail the implications of this major change in the rules of the game. Changes in regulations, along with dangerous innovations in unregulated financial markets, also contributed to swelling financial sector profits to over one-third of all reported profits in the U.S. economy. Briefly, along with other developments such as the increasing creation of new kinds of financial assets, and the refusal to regulate them, all these changes ended with house prices rising to levels that were far too high, causing a bone-jarring crash in the prices of both houses and financial assets, and triggering a steep economic downturn from which the majority of the population suffered far more than the wealthy.

From 1980 on, further changes in rules made inequality still worse. Estate taxes (inheritance taxes) had in the past reduced inequality. The estate tax has been gradually reduced: in 2001, estates of over $675,000 were subject to estate taxes at a rate of about 50 percent, while in 2015 only estates of over $5 million were subject to estate taxes of 40 percent (after deductions and so forth). There also exist schemes for circumventing such taxes, although not all the wealthy take advantage of them.

Another feature of the last few decades that has also involved important rule changes has been major agreements on trade, investment, and other issues. Of these, the best known are the North American Free Trade Agreement (NAFTA) of 1994 and the (1994) agreements that created the World Trade Organization in 1995. Both of these agreements granted new powers and protections to transnational corporations, a topic that will be discussed in Chapter 15, and in Chapter 19, which continues this discussion of government and the economy.

Conclusion

Capitalist economic activity operates under principles that will be spelled out further in the coming chapters, especially Chapter 10. Yet it does not operate independent of government. Government defines what it can or cannot do, and because of this, capital owners seek to shape what rules the government imposes—what bills are legislated, what orders are issued by the executive, and what court decisions are made.

All these affect profitability of enterprises by affecting how enterprises compete with one another, how capital owners are permitted to treat workers, what rights people have and what additional rights citizens have. These all affect what is produced, how it is produced, and how the product is distributed among the population.

We have focused mainly on nineteenth century examples of these principles in order to lay the groundwork for Chapter 7's discussion of the stages of capitalism and varieties of "social structures of accumulation." Three additional very important functions of government will be discussed in Chapter 19: taxation and benefits, government contracting, and certain new supranational rules that have increasingly been embedded in agreements on international trade, investments, and other issues, in some cases bypassing or overriding decisions made by national governments.

Suggested Readings

Daniel Altman, *Neoconomy* (New York: Perseus Books, 2004).

Samuel Bowles and Herbert Gintis, *Democracy and Capitalism: Property, Community, and the Contradictions of Modern Social Thought* (New York: Basic Books, 1987).

James K. Galbraith, *The Predator State* (New York: Free Press, an imprint of Simon & Schuster, 2008).

Carter Goodrich, *Government Promotion of American Canals and Railroads 1800–1890.* (Westport, CT: Greenwood Press, 1960).

Thom Hartmann, *Unequal Protection: How Corporations Became "People"—and How You Can Fight Back* (San Francisco, CA: Berrett-Koehler, 2nd edition, 2010).

Godfrey Hodgson, *More Equal Than Others: America from Nixon to the New Century* (Princeton: Princeton University Press, 2004).

Peter Lindert, *Growing Public: Social Spending and Economic Growth Since the 18th Century* (Cambridge, UK: Cambridge University Press, 2004).

Adam Przeworski, *The State and the Economy under Capitalism* (New York: Harwood, 1990).

Adam Przeworski, Michael E. Alvarez, Jose Antonio Cheibub, and Fernando Limongi, *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990* (Cambridge, UK: Cambridge University Press, 2000).

7 U.S. Capitalism: Accumulation and Change

In 1774 Joseph Hubbard invited his few neighbors to a meeting in his recently built barn, where they founded the town of Leverett, Massachusetts. Like his neighbors, Hubbard and his family grew crops, made tools, processed food, slaughtered livestock, wove cloth, made clothing, tanned leather, made shoes, and pressed apples for cider. The Hubbards also tended their own sick, educated the young Hubbards, and entertained themselves with stories and song. Of the things they needed, there was little that they did not produce.

When Joseph passed away his will listed (in beautiful script) his entire belongings, mostly tools for types of production that today would be carried on not under a single roof but in dozens of sectors of the economy. The Hubbards bought only a few things from their neighbors or others outside their household: salt, some iron goods, Joseph's gun, and not much else. The Hubbard family was a mini-economy in its own right. Though items of personal clothing were virtually absent from Hubbard's will—only "the coat" was mentioned, suggesting that he did not have a closet full of them—they were far from poor, and in the years following Joseph's death, the next generation of Hubbards added larger rooms with plastered walls. We know all of this because one of the authors of this book lives in the house that Joseph Hubbard and his family built.

The Hubbard family was typical of most families of European descent then living in North America: they were independent producers. The Hubbards probably never even considered *hiring* someone to operate their cider press or harvest their fields. They certainly never hired themselves out to anyone else. *Slaves* worked for someone else, but not by choice, and while there were a few slaves in Boston and the other seaports, they were unknown in the rural western hinterland of the Massachusetts Bay Colony. In the homes of the rich, *servants* worked under the direction of the madam of the house, but no one in Leverett or anyplace nearby was that rich. The single word that best describes the Hubbard family economy is *independence*: they neither worked for anyone else, nor did they depend on others for their livelihood.

In the United States today, most families make their living by working for someone else, employing others to work for them, or both. Independent family economies

like that of the eighteenth-century Hubbards are rare. The home is still an important work site, but gaining a livelihood generally involves working for some entity outside the home, either for a wage or salary, or as a contractor. Most people work in for-profit businesses or for a government.

One essential part of employment, whether in business or in government, is taking orders or giving orders. Whether subtly or harshly, employment in the modern economy involves what we term *command*. (Command was not absent in Joseph Hubbard's world, but he exercised it over his wife and children, at least until his failing years, not over his neighbors.)

A second essential aspect of employment is *dependence* on others: one needs a job to live, and someone else ultimately controls whether one has a job. The Hubbard family was buffeted by the occasional dry summer, cold winter, or other vagary of nature, but unlike employees today, their livelihood did not depend on the decisions of other people. As the social structure of economic life has changed, material abundance has grown, health has improved, and communication with the rest of the world has multiplied. The Hubbard family would have been dazzled.

This chapter is about how fundamental *change* in economic life comes about. It brings the time dimension of our "three-dimensional" approach to the fore, focusing on the forces for change inherent in the capitalist organization of production, and their effects. In this chapter we examine U.S. capitalism and its historical development as a case study of the process by which the forces inherent in a capitalist economy can bring about wide-ranging social change. In Chapter 1 we saw that some enormous social, political, and economic changes have taken place since the capitalist era began. Now we ask: How does a capitalist economy drive change, and what kind of change does it produce? Further, how has U.S. capitalism been changed by the process of its own historical development? Chapter 6 introduced the way in which capital owners often seek to change the rules of the game in their favor. This chapter explores the dynamics of change internal to the capitalist economy itself, and the kinds of transformations that has brought about.

Our answers to these questions form the main idea of this chapter: *Normally, the competition that is built into a capitalist economy compels business owners to innovate, and this has often meant producing better products and services more cheaply. Throughout most of capitalist history this has required them to invest—rather than consume—much of the surplus product. This investment, in turn, has not only raised productivity and living standards, but has also generated continuous technological, economic, social, and political change, thereby transforming both the class structure and the institutional framework of the economy. Yet while many kinds of change both increase profit and improve life, other changes have more mixed effects, and some may reveal a need for corrective regulation.*

This main idea is expressed in eight primary points:

1. *Competition for profits* arises because the only way a firm can stay in business is to make profits. Each business owner is caught in a never-ending race to avoid falling behind. An important way to stay ahead is to produce better goods or services at lower cost. To keep up each firm must not only maintain its current production process, but must also expand and improve its own product line, enter new markets, introduce new technology, and find lower-cost ways of getting the necessary work done.

2. Competition thus typically compels business owners to *invest* (rather than consume) most of the profits they make: to install better equipment, build or buy production units in new locations, conduct research to develop new products or processes, buy useful patents or trademarks, and more. Investing to compete for greater profits is called *accumulation*.
3. As a result, the capitalist class is the first economic elite in world history whose members have usually been compelled by competition to revolutionize the way goods and services are produced and distributed. It is this race that has made capitalism so productive and dynamic.
4. The competition for profits during the eighteenth and nineteenth centuries made *capitalism the dominant economic system* in the United States. As in many other countries, capitalism replaced independent production, slavery, and other economic systems that prevailed in the past.
5. U.S. capitalism has, over time, been characterized by a series of distinct institutional settings that have facilitated the accumulation process, and we call these institutional settings *social structures of accumulation* (SSAs).
6. The social structures of accumulation that have defined successive *stages of U.S. capitalism* since the mid-nineteenth century are *competitive* capitalism, *corporate* capitalism, *regulated* capitalism, and contemporary *transnational* capitalism. The power of organized labor in the United States waxed for a time under regulated capitalism, but has waned since about 1980.
7. In the last several decades U.S. capitalism has become highly transnational. Many large corporations have established affiliates around the world, and cross-border investment has led to growing trade. The rules laid down in new international agreements and institutions like the World Trade Organization have powerfully influenced national economies. Globalized production has also undermined labor's bargaining power.
8. *Since 2000, the link between profits and investment in the United States has been partly severed*: corporate executives, instead of using profits for productive investment, have often used them to buy back a company's own stock or pay dividends. This trend undermines investment and growth, and tends to endanger the viability of the current SSA.

Accumulation as a Source of Change

Accumulation is profit-driven investment, including the process of mobilizing, transforming, and exploiting the inputs required in capitalist production and in selling the output.

Profit-driven investment is normally the motor of change in a capitalist economy. Both investment and profit-making are shaped by competition as owners of businesses vie with one another for profit opportunities. In turn, the activities of profit-seeking and investment usually transform labor processes, draw in new supplies of labor, increase the stock of capital goods, create new knowledge, and mobilize or transform other resources to make them available for use in commodity production. The word **accumulation** refers to this whole interrelated set of activities.

The Accumulation Process

In the accumulation process, profit is both a lure and a lever for change. It is a lure because the possibility of making a profit is what usually motivates business owners to reinvest their profits; it is a lever because profits, when invested, create change.

The nineteenth century political economist Karl Marx (1818–1883) believed this process to be the central feature of capitalism: "Accumulate, accumulate!" he said. "That is Moses and the prophets!"[1] Just as Moses and the biblical prophets have inspired people in the Judeo-Christian tradition, so does accumulation drive the capitalist economic system.

The idea here is that given the dynamism of capitalism, *the present is different from the past*, and *what has happened in the past affects both the present and the future.* When there is accumulation and change in year 1, the beginning of year 2 may well be different from year 1: technologies, equipment, knowledge, products, and the relative size and power of firms may all have changed. Likewise, change in the earlier part of a longer historical period may in the same ways transform the later part. Thus, the introduction of political economy's third dimension, change, can shed light on how the process of profit making evolves. To see the link between accumulation and change in a capitalist economic system, consider the following example.

Suppose an entrepreneur decides to start a T-shirt manufacturing business. She uses her funds to buy the necessary inputs (machines, materials, and labor) and then during time period 1 supervises the production of T-shirts. When the period ends, the T-shirts are sold for more than the cost of the inputs, yielding a profit. After paying all the production costs—maintaining the machinery, buying the materials used, and compensating the labor employed—the entrepreneur reaps the profit. What then will she do with it?

If she personally owns the whole business, she may choose to use most of the profit to buy goods for her own consumption, allocating just the same amount of money as before to continue in the T-shirt business. The production equipment, materials, and workers may well be just as physically productive as they were at the beginning of period 1. But the business environment—factors outside the manufacturing process—may have changed. So if the T-shirt business owner does in period 2 what she did in period 1, she cannot necessarily expect to earn the same profit. For example, if the price of some input has increased, the same amount of money will buy less of this input in period 2 than in period 1. Or T-shirt customers' preferences may have changed, so that they now want to buy fewer of this factory's T-shirts. Possibly some other businessperson, eyeing the profit made by our T-shirt manufacturer, may start a competing T-shirt business. If any of these things happen, our T-shirt producer will be forced to change her business strategy in period 2 if she hopes to make the same amount of profit as in period 1.

These cautions do not apply to a noncapitalist economic system, where productive activity is much more stable and predictable because there is little competition. The self-sufficient household provided most of its own inputs, and its output was mostly consumed by its members, not sold. The medieval manor was also largely self-sufficient. Thus there were few reasons to worry about input prices or the preferences of customers. Production could often be carried on in the same way year after year without difficulty.

---

[1]Karl Marx, *Capital*, vol. 1 (Harmondsworth, UK: Penguin Books, 1976), Chapter 24, Section 3, 742.

But in a capitalist economy, businesses producing commodities for profit normally cannot stand still without falling behind. Not only do they face uncertainty with regard to unpredictable events, they also face the certainty of facing market competition. They know that other businesses, also fearing uncertainty and seeking higher profits, will change their production processes and introduce new products in order to keep their profit rates high. Thus, all business owners are forced to make changes as well to defend their own profits. Necessarily, change becomes constant.

Competition for Profits

Competition for profits is the struggle for survival and expansion among firms seeking new ways of doing business, new markets, new products, and other possibilities for profitable investment.

What normally drives all businesses in a capitalist economy constantly to change their operations and seek improvements is the **competition for profits**. Those who do not engage in this competition will find themselves left with higher-priced inputs, more costly methods of production, or obsolete and unsold output. Businesses constantly revolutionize their operations to have a chance of surviving, and of those, only those who make the right changes will succeed.

Those who succeed in one period are like runners who are ahead halfway through a race: the leaders maintain their lead only by continuing to run fast, and those who slow down or stop will see the other runners go by. This capitalist "race" has two peculiar features. First, new competitors keep jumping into the contest from the sidelines as the race progresses, intensifying competition. Second, there is no end to the capitalist race: it just keeps going. Those who fall behind have no chance to start again at a common starting line in a new race. Recalling our discussion of the "rules of the game" in Chapter 5, capitalist competition works like this: if you finish at the end of one game 15 points behind, you start the next game with a score of minus 15.

One main way a firm grows is by investing, that is, by using its profits to increase the quantity or the productivity of its capital goods and other inputs. Employers can do this by buying more and better machinery, building new factories and offices, acquiring ones that already exist, supporting in-house research activities, or enabling their workers to gain additional skills. If a firm is not making a profit, it cannot grow by itself. And if a firm does not grow, it will soon be outpaced (and perhaps acquired) by others that do grow. In a capitalist economy, survival requires growth, and growth requires profits. This is capitalism's law of the survival of the fittest, analogous to Charles Darwin's notion of the evolution of species through natural selection.

Capitalism is differentiated from other economic systems by its drive to accumulate, its predisposition toward change, and its built-in tendency to expand. In earlier economic systems—feudalism and slavery, for instance—the surplus product was mostly *consumed* by social elites in economically unproductive ways such as consuming luxuries and building monuments. The presence of a capitalist elite with even more lavish consumption habits than those of feudal lords or slaveholders is a highly visible aspect of our economic system. But it is not the superyachts that make capitalism distinct (see box Superyachts and Bull Market Boats," this chapter); the difference is that, to thrive, this economy requires the *investment* of a major portion of its surplus product.

In feudal society there was relatively little market activity, hardly any economic competition, and even less long-distance trade, so the lords had little incentive and few opportunities to invest the surplus to increase productivity. Indeed, their positions as lords or church officials were strengthened by their consumption of the surplus product. Secular nobles spent their portion of the surplus product on things

that demonstrated and enhanced their power and prestige: building castles, going on crusades, and maintaining a large retinue of subordinate knights and other retainers. The high clergy, including those at Ely, spent their portion of the surplus product in similar ways, erecting cathedrals and supporting cardinals, bishops, priests, and other functionaries in the church hierarchy. These uses of the surplus may well have reinforced the ideology, class structure, and institutions of feudal society—but they left the productive system mostly unchanged. Old ways of producing things persisted, or were modified only gradually.

Capitalism is the first system to have an elite whose members, throughout most of capitalist history, have been driven to invest the surplus and so revolutionize production. A feudal lord whose fields had low yields was just a shabby noble. But a capitalist whose factories make high-cost or low-quality products may soon be an ex-capitalist.

In sum, capitalism's normal functioning produces tremendous pressures for change. It creates not only the incentive (competition for profits) for individual capitalists to introduce change but also the means (those very profits) to implement change.

Capitalism Comes to the United States

The U.S. economy did not begin as a capitalist economy. Rather, in the first of its two centuries, capitalism grew from being a relatively small part of the economy to being its dominant economic system.

In 1780—after the Declaration of Independence (1776) and before the U.S. Constitution (1787)—few participants in the economy were engaged in capitalist production, as shown in Table 7.1. People involved in capitalist class relationships

TABLE 7.1. PERCENT OF THE POPULATION ENGAGED IN DIFFERENT ECONOMIC SYSTEMS IN THE U.S. IN 1780

QUESTION:		ARE THE CAPITAL GOODS USED IN PRODUCTION PRIVATELY OWNED?	
		Yes	No
Is wage labor the main form of productive work?	Yes	Capitalism (6%)	Government activities (less than 1%)
	No	Slavery (32%) Home production and commodity production by independent producers (60%)	Native American communal/kinship production (2%)

Economic systems can be sorted into categories based on the answers—yes or no—to two questions: (1) Is wage labor the main form of productive work, and (2) are the capital goods used in production privately owned? Each of the boxes in the table represents one of the four possible combinations of answers to the two questions, with the answers to the first question given in the *rows* and the answers to the second question given in the *columns*. The percentages in each box indicate, for all of the persons engaged in production, the approximate percentage involved in each economic system in 1780. Because of the absence of available data on them, people who were engaged in reproduction—child rearing and related tasks—are not included in this table, and the percentages do not add up exactly to 100 percent because of rounding of the particular percentages.

Sources: Jackson T. Main, *The Social Structure of Revolutionary America* (Princeton, NJ: Princeton University Press, 1965); and Gary B. Nash, *Class and Society in Early America* (Englewood Cliffs, NJ: Prentice-Hall, 1970).

(capitalists and workers) were no more than 6 percent of the population. Their activities coexisted with two other economic systems—slavery and independent commodity production—that together accounted for over 90 percent of those engaged in economic activity. Slaves and slave owners together made up about a third of the population.

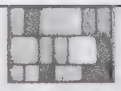
Yet the biggest economic system in the early American economy was neither capitalism nor slavery, but consisted of people like the Hubbard family engaged in **independent production of commodities**. Independent, or self-employed, producers of commodities—farmers, craftspeople, self-employed artisans, and other independent producers—made up about 60 percent of those participating in production. Independent producers of commodities owned the capital goods needed for their own labor processes. For instance, most small farmers owned the land they farmed as well as the tools, livestock, barns, and other capital goods needed in farming. They used mainly their own labor and that of their family members, and only rarely—or to a small degree—depended on the labor of slaves or wage workers.

Independent production of commodities is an economic system in which the producers own the capital goods needed in production and use (primarily) their own labor.

Independent producers of commodities often produced some items for their own consumption, like food and clothing, as well as goods to be marketed. Who did what work in independent commodity production was determined partly by markets and partly by the assignment of tasks within the family, decisions often dominated by the father. The hallmark of independent commodity producers was that they were self-employed: they were their own bosses.

Early capitalism emerged in the United States, then, within an economy that consisted mainly of independent commodity production and slavery. Soon, however, competition intensified and expanded, especially between independent producers and capitalists. (Slave-based production was concentrated mainly in tobacco, rice, and cotton in the South, and it was only displaced by capitalist production after the Civil War ended in 1865.)

Independent shoemakers, for example, increasingly had to compete against shoemaking wage laborers working under the direction of capitalists. The capitalist producers had the advantage that they could organize production on a larger scale and so achieve higher productivity (output per worker)—gains that, as Adam Smith



How the Accumulation Process Transformed Workers' Lives: A Nineteenth-Century View

The village blacksmith shop was abandoned, the roadside shoe shop was deserted, the tailor left his bench, and all together these mechanics [workers] *turned away from their country homes and wended their way to the cities wherein the large factories had been erected. The gates were unlocked in the morning to allow them to enter, and after their daily task was done the gates were closed after them in the evening.*

Silently and thoughtfully, these men went to their homes. They no longer carried the keys of the workshop, for the workshop, tools, and keys belonged not to them, but to their master. Thrown together in this way, in these large hives of industry, men became acquainted with each other, and frequently discussed the question of labor's rights and wrongs.

—Terrance Powderly, Grand Master Workman, Knights of Labor (1889)

Source: Thirty Years of Labor 1859–1889 (Columbus, OH: Excelsior Publishing House, 1889).

noted in 1776, resulted from greater division of labor in the workplace. The capitalist firms could produce and sell shoes more cheaply, and still earn a profit. This enabled them to drive the independent shoemakers out of business. Similarly, self-employed blacksmiths, weavers, spinners, harness makers, and other independent producers were increasingly drawn into competition with the new capitalist firms that invaded their markets with cheaper products.

Soon the independent producers had to choose between surviving or falling behind. To survive, some reinvested their profits, expanded their operations, and sought cheaper methods of production—in short, they tried to become capitalists themselves. Others fell behind, suffered declining incomes, and eventually lost their independence and took jobs in capitalist firms. At the same time, a significant influx of immigrants, many of whom possessed little wealth, out of necessity joined the growing supply of wage workers, feeding the capitalist firms' need for labor.

Tables 7.2 and 7.3 compare the class structure in 1780 with that in 1990 in the United States, dividing all persons engaged in productive activity into four classes. A person's class membership depends on the answers to the two questions listed in each table: (1) Does the person own the capital goods used in production? (2) Does the person control the labor of others? A "yes" to both questions (cell (1)) means that a person both owns the equipment used in production and has control over the labor of others, and hence is a capitalist or a slave owner. A "no" to both questions (cell (4)) means that a person has neither ownership of productive equipment nor control of others' labor, and so is a worker or a slave.

Those with ownership of capital goods, but not control of labor (cell (2)), or control of labor but not ownership of capital goods (cell (3)), are both in the middle classes, to be discussed below. If we compare the U.S. class structure in 1780 in Table 7.2 with the class structure two centuries later in 1990, in Table 7.3, we see what changes took place over that period.

TABLE 7.2. THE U.S. CLASS STRUCTURE IN 1780

QUESTION:		OWNS THE CAPITAL GOODS USED IN PRODUCTION?	
		Yes	No
Controls the labor of others?	Yes	(1) Capitalist class (1%) Slave-owning class (3%)	(3) New middle class (1%)
	No	(2) Old middle class (60%)	(4) Working class (5%) Slave class (30%)

Classes can be categorized based on answers to two questions: (a) Do the members of the class control the labor of others and (b) do the members of the class own the capital goods used in production? (To be consistent with the available data, we say that the answer to the first question is negative for entrepreneurs who employ fewer than ten workers.) Each of the boxes represents one of the four possible combinations of answers to the two questions. The two most important class relationships in 1780 were between box (1) and box (4)—between the capitalist class (answers yes and yes) and the working class (answers no and no) and between the slave-owning class (yes/yes) and the slave class (no/no). The numbers in parentheses indicate, for each class, what percentage of the labor force its members were. The labor of those involved in child-rearing and other aspects of reproduction does not appear in this table or in Table 7.3 because such labor is not directly part of the class structure.

Sources: Estimates based on data in Jackson T. Main, *The Social Structure of Revolutionary America* (Princeton, NJ: Princeton University Press, 1965); and Gary B. Nash, *Class and Society in Early America* (Englewood Cliffs, NJ: Prentice-Hall, 1970).

TABLE 7.3. THE U.S. CLASS STRUCTURE IN 1990

QUESTION:		OWNS THE CAPITAL GOODS USED IN PRODUCTION?	
		Yes	No
Controls the labor of others?	Yes	(1) Capitalist class (5%)	(3) New middle class (29%)
	No	(2) Old middle class (12%)	(4) Working class (54%)

This table presents recent answers to the same two questions that were posed in Table 7.2 for Americans in 1780. As in that table, each of the boxes presents the numbers (as percentages of the labor force) for one of the four possible combinations of answers to the two questions. Between 1780 and 1990, the slave-owning class disappeared from box (1), while the capitalist class grew from 1 percent to 5 percent of the labor force by the definition used here. The old middle class declined from 60 to 12 percent, while the new middle class grew from 1 to 29 percent. Perhaps most significantly, slaves were freed and the working class grew from 5 percent of the labor force in 1780 to 54 percent in 1990. Thus, the working class of 1990, having grown by a factor of 10, is not only a majority of the working population but is a larger percentage of it than were the slave and working classes combined in 1780. While the data are not available to make exactly the same calculation for the year 2016, not much has changed. The capitalist class in the U.S. is still small but powerful, and the working class is still the majority of the population, and probably larger than in 1990. The middle classes shrank in the Great Recession, and appear to have only partly recovered since that time.

Source: Estimates based on data for 1990 in Erik Olin Wright, *Class Counts: Comparative Studies in Class Analysis* (Cambridge, UK: Cambridge University Press, 1997), 99, Table 3.2.

In 1780 capitalists and slave owners together constituted about 4 percent of everyone who was involved in labor processes in the production sector. By 1990 the slave-owning class had, of course, long since disappeared, but, as Table 7.3 shows, the dominant class in production—the capitalists—was approximately the same percentage of the labor force (5 as opposed to 4 percent) as the slave-owning and capitalist classes together had been in 1780. In the study used for this table, the category "capitalist" was defined to include those who employed at least ten workers. (If the class were redefined to include only employers of at least five hundred workers, the percent of the population in the capitalist class would be well under 1 percent. The U.S. Small Business Administration defines a "small business" as one that employs fewer than five hundred workers.)

Cell (4) in Table 7.2 counts slaves and wage workers in 1780, while in Table 7.3 the same cell, of course, counts just wage workers in 1990. In 1780 slaves were about 30 percent of those participating in production. Today they are nearly zero percent; however, the U.S. State Department, in efforts to combat human trafficking, has found thousands of undocumented foreign workers trafficked into the United States and forced into various kinds of labor under conditions of slavery. *Wage workers grew from only 5 percent in 1780 to more than half the labor force in 1990.*

Middle classes in capitalist society possess one but not both of the attributes of capitalists; they therefore stand between capitalists and workers.

The **old middle class** consists of those who do own the capital goods used in their own labor processes but who do not regularly control the labor of others; they are self-employed or are small business employers.

As for cells (2) and (3) in the two tables, they represent the **middle classes** who possess either *one*, but *not both*, of the defining characteristics of the capitalist class. There are two quite distinct middle classes.

The **old middle class** (cell (2) in the tables) consists of people who *do* own the tools of their trade or whatever capital goods they need in order to carry on their work, but *do not* regularly employ or control the labor of others. Such persons *may* employ a few (say, fewer than ten) workers. (As noted in the caption under Table 7.2, we set the limit at ten workers to be consistent with the available data.) They differ from the capitalist class in that capitalists are involved in a power

relationship with their employees, while members of the old middle class generally are not.

The old middle class includes self-employed persons and independent commodity producers such as self-employed doctors, carpenters, artists, or family farmers, and small retailers such as "Mom and Pop" storekeepers, as well as other small business owners who employ few or no workers. Members of the old middle class do not sell their time as wage workers. Rather, they sell a commodity, whether it be their crops or services such as medical treatment. These independent producers do not perform capitalist labor processes, but they will most likely have to relate to capitalist enterprises through buying and selling. The old middle class is called "old" because it existed as a category before capitalism developed, even though some of the kinds of labor performed by it are quite new, such as independent contracting for computer programming. An independent computer programmer who works on contract is, in fact, a classic example of the "old" middle class—but may at some point in her career become part of the "new" middle class if she takes a salaried managerial position in a firm.

The **new middle class** consists of those who do not own the capital goods used in their own labor processes but who do regularly control the labor of others; it includes managers and supervisors.

The **new middle class**, on the other hand, is composed of those who *do* control the labor of others but who generally *do not* own the capital goods used in production; they appear in cell (3) in the tables. The new middle class is composed of managers, supervisors, and people who are now called "expert-managers." The people in this new middle class usually receive a salary (rather than a wage), are paid for their labor time and expertise (rather than for their services or products), and can always be fired (because they are under the authority of others). Although controlling other people's labor is a large part of what they do, their own work is controlled by others such as owners, top executives, or other supervisors in the organizations in which they are employed. The reason they are said to be in the middle class is not that they have middle-level incomes; in fact, their incomes range widely—from that of a top manager to that of a supervisor of a maintenance crew, for example. They are "middle class" because they are *in the middle* between capitalists and workers. Notice that we have defined *middle class* differently from the way it seems to be defined by politicians and the media. They appear to include in that category anyone who is neither particularly rich nor destitute, which likely includes well over half of the U.S. population.

Table 7.2 tells us that the old middle class was the largest category in 1780, with 60 percent of the population. But Table 7.3 shows that as capitalism has grown, the old middle class has withered, and in 1990 it made up only 12 percent of the labor force. Independent commodity production, the economic activity of the old middle class, is simply much less prevalent today than it was in 1780—yet remarkably, the ways of life of the early American economy live on in people's memories and hopes. Indeed, most Americans are very likely, at some point in their lives, to dream of working for themselves, perhaps as the owner of a small business, rather than for a boss. Some succeed.

Social Structures of Accumulation

The accumulation process has not only destroyed earlier economic systems—feudalism, slavery, independent production—transforming the class structure of these economic systems and unleashing a whole series of social changes; it has also

A **social structure of accumulation (SSA)** is the institutional setting within which accumulation occurs; it structures relations among capitalists, between capitalists and workers, among workers, and between government and the economy. Two phases of an SSA can be distinguished: first, its "consolidation" and, second, its "decay."

restructured capitalism itself. This has been true of capitalism in the United States as much as in other countries. The basic "rules of the game" of capitalism in the United States (and elsewhere) remain largely unchanged, however. Private owners of capital continue to employ wage labor with the intention of making a profit. Thus, the U.S. economy continues to be a capitalist economy.

Yet the capitalist economy itself changes. The types of products made, the technologies used, the geographical location of production, and the kinds of work that people do are constantly in flux. The most important changes are the changes in the institutional setting within which accumulation takes place.

We refer to this institutional setting as a **social structure of accumulation (SSA)**, meaning the laws, institutions, and social customs that provide the basic rules of the game governing accumulation. Accumulation, the making and reinvesting of profits by individual capitalists or firms, occurs within this "social structure." The rules of the game, of course, influence how the game is played.

Changing Strategies for Profit-Making

While the goal—profit—remains the same, the conditions under which employers try to make profits change. Thus, *the strategies for profit-making that are open to the firm change as the social structure of accumulation changes, and the profits that can be made with each strategy also vary.*

For example, until the end of the nineteenth century there were few large firms, whereas today there are many. A firm today, no less than its nineteenth-century predecessor, seeks to make profits, but the strategy it adopts will be different, partly because it must consider how to compete with large corporations. Before 1900, relocating some of the firm's operations to another country was not an option except for a few mining firms; today it is easier. Around the mid-twentieth century many large private employers had to negotiate with a labor union in the hiring and management of labor; this is less common today. Before 1970 a firm faced few if any restrictions on the effect of its operations on the natural environment; today, firms must comply with environmental regulations or calculate the likelihood of paying penalties if they do not. All of these developments reflect changes in the social structure of accumulation.

As Chapters 5 and 6 explained, we can think of the capitalist economy as a game. In most games the rules say what the members of each team may do with the members of their own team, with (or to) the members of the other team, and so on. The social structure of accumulation (SSA) is no different. Each SSA can be described by the way its rules influence, and are influenced by, several important relationships, namely,

- relationships among employers,
- relationships between employers and workers,
- relationships among workers, and
- relationships between the government and the main private economic actors

While all capitalist economies have fundamental similarities, particular countries have distinctive SSAs: U.S. capitalism, Swedish capitalism, Japanese capitalism, and Brazilian capitalism are distinctive in important ways. And each of these, including U.S. capitalism, changes as time passes.

Each SSA is long-lived, but no particular SSA is a permanent feature of capitalism. Typically, each SSA lasts for several decades or more. As explained below, the U.S. economy has seen four social structures of accumulation in more than a century and a half (before that the United States was not a capitalist economy). Each SSA lasts a long time because it consists of highly intertwined social relationships: sets of laws, institutions, class relationships, organization of political parties, beliefs, expectations, and customary ways of producing and consuming. It is difficult to change such a structure piecemeal, for it is only interactively that the parts work at all. Moreover, each of these arrangements, once established, works to the advantage of some individuals or groups in the society, who become the existing SSA's natural defenders. Until these defenders lose their influence—or until their interests change—the SSA will persist.

Consolidation and Decay of an SSA

Though long-lived, an SSA is not immortal, and like other "mortals," there are different phases in its life. If the existing structure is successful, the relationships among the key players will be such that gains (mostly profit rates) are high and owners of firms and other wealthy individuals will have confident expectations for the future. As a result, they will normally invest substantial sums in expanding the productive capacity of the economy, and the economy may work well not only for investors but for most people in the society. This type of situation is called the *consolidation* phase of the SSA.

Eventually the social structure becomes less and less capable of providing a favorable environment for accumulation. The consolidation phase is thus followed by a period of *decay* of the SSA. At this point economic problems multiply and people's expectations—especially investors' expectations—for the future become more modest or even pessimistic. As a result investment lags, and the growth of productive capacity stagnates.

A **long swing** occurs over a period of thirty to fifty years. The first part generally coincides with the "consolidation" phase of a particular social structure of accumulation (SSA) and is characterized by relatively high rates of investment and economic growth and relatively little unemployment. The remainder of a long swing is associated with the "decay" phase of an SSA; it is a period of stagnation of economic growth with relatively little investment and high unemployment.

The alternation between consolidation and decay can be seen in what are called **long swings** in economic performance. Long swings are different from the short-term ups and downs of the economy (sometimes called expansions and recessions) that are discussed in Chapter 16. Short term fluctuations, usually referred to as *business cycles*, generally last three to ten years from peak to peak. A number of these shorter cycles may occur during a long swing, but long swings commonly run thirty to fifty years between the end of one period of decay and the end of the next. Consolidation is reflected in a lengthy period of good economic times, or boom, whereas decay is indicated by a long period of hard times. (Near the beginning of Chapter 16 we present data on the long swings that have occurred in the U.S. economy between the Civil War and the second decade of the twenty-first century.)

Why does a boom not last forever? What tips an SSA from consolidation into decay? The answer is that the (successful) accumulation process itself may erode the favorable conditions associated with the consolidation phase of an SSA; this seems to have been the case in the 1920s and 1960s. The requirements for continued accumulation may become more inconsistent or self-contradictory as accumulation proceeds; this, too, occurred in the 1920s and 1960s. And there may be other causes as well.

But the question—why does decay set in?—cannot be answered in general terms. The most that can be said is that whereas an SSA changes slowly because of its interlocking relationships and the defense of it by particular interest groups, the accumulation process organized within it changes rapidly as new technologies are developed and

new kinds of opportunities for making profits emerge. When a tension develops between the (slowly changing, or "sticky") institutions and the (rapidly changing) profit-driven investment process, the old social structure no longer provides a favorable climate for accumulation. The particular causes that force consolidation to give way to decay are the result of specific historical forces; they are likely to be different for each SSA.

The period of decay may create a "crisis," during which large numbers of people become disaffected and social conflict is heightened. This occurred, for example, in the United States during the Great Depression, when more than a quarter of those wanting work were jobless and many turned against capitalism as a system. When this happens, new ways of organizing the economy are proposed, employers are forced to speed up their efforts at innovation, especially organizational innovation, and workers are likely to demand a "new deal." At the same time some of the erstwhile defenders of the old SSA find that they are getting less and less out of it. In the United States many Depression-era businessmen also came around to supporting the idea of a "New Deal" as a way of "saving capitalism."

As the consolidation phase of an SSA turns into decay, the old alignments may change, and new electoral coalitions may emerge, sometimes reforming the dominant political party from within, sometimes establishing a new dominant party. Electoral realignments brought on by the decay of an SSA often produce what political scientists call "critical elections." Examples include the 1896 defeat of the agrarian populism of William Jennings Bryan, the 1932 election of Franklin Delano Roosevelt that led to the adoption of "New Deal" programs, and the 1980 Reagan revolution that ushered in conservative populism and made the South Republican for the first time since before the New Deal.

Out of a period of crisis, decay, conflict, and change may come the construction of a new set of social relationships—a new SSA. Such a reconstruction is far from inevitable; it depends on the outcome of many specific conflicts and negotiations. Yet the crisis creates a situation in which the inadequacies of the old SSA become apparent, the economic problems become more serious, and the demands for action become increasingly urgent. It may then be possible to overcome the inertia of the old institutions and develop a new SSA.

The Stages of Capitalism in the United States

Each social structure of accumulation defines a stage of U.S. capitalism. Table 7.4 illustrates these stages and the distinctive aspects of the key social structure relationships discussed above.

Competitive Capitalism (1860s–1898)

The **stages of capitalism** in the United States are distinct phases in the development of U.S. capitalism, with each being defined by a particular social structure of accumulation.

At the beginning of the first stage, competitive capitalism, there was a small capitalist sector overshadowed by both slavery and independent commodity production. Capitalist activities grew, establishing the economic system that would clearly become dominant in the future. Slavery, a competing economic system, was outlawed, and the competition between capitalism and independent producers of commodities turned decisively in favor of capitalism.

U.S. capitalism in this stage mainly consisted of small businesses employing fewer than twenty people and competing with one another in the widening markets mainly by price cutting. The prices of most goods fell over the latter part of the

TABLE 7.4. THE STAGES OF U.S. CAPITALISM: FOUR SOCIAL STRUCTURES OF ACCUMULATION

KEY RELATIONSHIP	COMPETITIVE CAPITALISM (1860s–1898)	CORPORATE CAPITALISM (1898–1939)	REGULATED CAPITALISM (1939–1991)	TRANSNATIONAL CAPITALISM (1991–)
Capital-capital	Small business; competition in local and regional markets; late in this phase, corporations become legal persons and may own other corporations	National-level competition among large corporations (trusts); oil and mining go partly global; Sherman Anti-Trust Act restrains concentration	World War II leaves U.S. corporations dominant over Europe; most become transnational; mergers and acquisitions produce giant firms	Large U.S. firms go global in production and sourcing; financial sector is deregulated, so its profit share grows; CEO pay grows huge
Capital-labor	Strong craft-based unions in some occupations; much conflict; at most 10–20% of workers join broad-based labor organizations	Employers are dominant, labor unions weak and struggling; corporate paternalism and company towns in some sectors, open conflict in others	National Labor Relations Act has legalized collective bargaining; unions gain members, and strength in wage-setting and politics; "labor accord"—real wages rise with productivity	No more labor accord; global mobility of capital pushes labor into backing down; union membership falls; political and economic polarization of the nation
Labor-labor	Craft-based distinctions between skilled and unskilled workers; immigration increases labor supply	Homogenized labor; non-unionized semi-skilled factory workers grow in numbers; immigration is limited	Segmented labor markets: unions are initially strong in "primary" markets and mass production; their strength declines as production is moved abroad	Inequality among workers grows; "good" industrial jobs decline and both "high-end" and "bad" jobs grow; U.S. workers compete with workers abroad; immigration grows
Government-economy	Limited government: military and police functions (land policy, tariffs, canal building, subsidies to railroads)	Federal Reserve System is created to regulate money and banking	Macroeconomic stabilization through deficit spending, expansion of social security, medical, unemployment, and other insurance; U.S. military supports U.S. corporate interests worldwide	Deregulation of financial sector, environmental protections, and other rules; government and Fed spend trillions on crisis bailout; global governance institutions like World Trade Organization expand corporate rights

U.S. capitalism has passed through three stages and is now in its fourth (though some economists who study SSAs think the fourth SSA is already in decay or has fully decayed). Historically, there have been the stages of *competitive* capitalism, *corporate* capitalism, and *regulated* capitalism. We call the current stage *transnational* capitalism. Each stage is defined by a distinct social structure of accumulation. The key aspects of the social structures of accumulation are listed in the first column. They include relationships (a) among the owners of the capital goods used in production, (b) between the owners and the workers, (c) among workers, and (d) between all the economic actors and the government. The other four columns summarize the particular characteristics of the U.S. economy in each of its four stages.

nineteenth century. Workers, many of them with skills they had honed in independent commodity production, now found themselves working for a wage under capitalist supervision. In many cases they also retained their associations ("brotherhoods" or trade unions) to bargain over wages and working conditions. The government played a crucial role in enforcing the rules of the game (such as enforcing contracts and breaking strikes), but otherwise it played a minor part in the economy, except—as described in Chapter 6—in distributing lands as homesteads, giving land grants to builders of railroads, and approving important changes in the rules governing how businesses could legally be organized.

Corporate Capitalism (1898–1939)

The second stage of U.S. capitalism was inaugurated with the rise of large corporations with multiple affiliates (see Chapter 6) around the turn of the twentieth century. This stage lasted until the end of the 1930s. Corporate capitalism in the leading industries featured competition among large firms with substantial market power. Employers opposed labor unions, and workers who had tried to organize and defend unions suffered such serious defeats that in many industries unions enjoyed limited influence until near the end of the period. Workers had few rights, and until near the end of this period there was no unemployment insurance except in a few states, so the loss of a job meant the loss of a livelihood. The government, although slightly more involved in the economy than it had been during competitive capitalism, continued to play only a limited role in the economy until the late 1930s. This SSA decayed under the onslaught of the Great Depression and the mass unemployment of workers and low profits for employers. Congress established labor's rights to organize, and new industry-wide unions were created in the late 1930s. New social programs and financial regulations were also passed, setting the stage for the next SSA.

Regulated Capitalism (1939–1991)

The next stage of U.S. capitalism brought an increase both in direct regulation of economic relationships and in indirect regulation of the economy as a whole, especially of employment and growth. New regulatory institutions had been set up during the 1930s, including the Securities and Exchange Commission (SEC), created in 1934 to regulate financial markets and the financial practices of corporations. Congress had passed the National Labor Relations Act (Wagner Act) in 1935, creating the National Labor Relations Board (NLRB), which established procedures and protections for workers seeking to form labor unions, and regulated the relationships between employers and unions.

During World War II the economy was massively transformed as the government took over the war effort, forming committees to manage the economy, control prices, and direct resources and labor to maximize wartime production.

After World War II the ideas of British economist John Maynard Keynes influenced macroeconomic policies designed to influence job creation and economic growth, and so avoid another depression. These ideas will be explained beginning in Chapter 16.

The size of the federal government grew in relation to the total economy, as the new Social Security system collected taxes and made payments to retirees. The new system of unemployment insurance did the same, and later a new program of medical insurance for the elderly (Medicare) was established in 1965 and extended to the disabled in 1972.

After World War II workers in labor unions were powerful enough to force employers to recognize and bargain with them, producing a compromise arrangement called the *labor accord*. The Civil Rights Act of 1964 outlawed race and sex discrimination of many kinds, including in employment. In 1970 Congress passed new laws to protect workers' health and safety on the job and to address environmental problems.

In addition to regulating employment and growth and providing new social programs, the government maintained a military force that extended over much of the globe. The inability of that force to win the Vietnam War of the 1960s and 1970s

was one of several factors that contributed to the decay of this SSA—including the rise of newly independent nations and of the Organization of Petroleum Exporting Countries (OPEC). During the late 1960s the reported corporate rate of profit began to fall, although probably this was at least in part because firms that established new foreign affiliates were using them to shift reported profits overseas in order to reduce the U.S. income taxes they paid.

Transnational Capitalism (1991–)

We describe the form of capitalism under the most recent SSA as *transnational*. This name captures the global fragmentation of production by transnational corporations (TNCs): They now produce components in, or buy components from, many different countries, and the fragmentation of production has been a major cause of increased world trade. These components are often shipped across borders to be subassembled, and then again across other borders to be assembled into the finished product. International migration of labor is another feature of the transnational capitalism SSA. Some of these changes began in the 1980s, a period when the old SSA was decaying and at the same time changes in laws and institutions were giving shape to a new SSA in formation.

Calling the current SSA *transnational* also reflects the rising importance and power of international organizations like the World Trade Organization, established in 1995, with new rules magnifying the power of TNCs. This parallels the rise of numerous other bilateral and multilateral agreements about investment and trade—another part of the shift of rule-making to the international sphere. This period also has seen a further rise in the use of tax havens and global accounting practices by U.S.-based (and other) TNCs to reduce the amount of corporate income tax they pay.

Although we call the current SSA *transnational capitalism*, this does not capture all of its key features. Capitalism has simultaneously become more deregulated, particularly in the financial sector; government deregulation has particularly freed the financial sector to develop new financial assets with potential to increase financial sector profits, but ominously also to risk major financial and economic collapse, as happened in 2008. *Deregulation* and *financialization* are also key features of the current SSA.

We now look at labor organizing, its strengths, successes, and weaknesses, in each of the four SSAs, and then at the way labor markets work. Later in the chapter we further explore the nature of the current SSA.

U.S. Capitalism: Labor Organizing and Labor Markets

In Chapter 6 we saw how in the nineteenth and twentieth centuries the environment for the development of capitalist enterprises was shaped by changes in laws, regulations, and court decisions. But how did this same period look from the standpoint of labor? How did labor unions arise? And why has union membership been declining for over half a century?

The Rise and Fall of Labor Unions

During the late nineteenth century, as many in the old middle class had to give up independent production and become wage workers subject to the command of a boss, employees often sought the protection of a union. Workers in a particular

trade or craft formed their own unions, which were sometimes able to become quite powerful. Furriers, railway engineers, iron molders, carpenters, and plumbers each belonged to a separate *craft union*. Then, stitching together their specialized unions at the national level, they established the American Federation of Labor (AFL).

But each craft union enrolled only skilled workers in the particular occupation it was organized to represent. In the same industry or even the same workplace, each type of skilled worker would belong to a different union, and less skilled workers might belong to none. At the same time, there was a series of other labor organizations (the National Labor Union, the Knights of Labor, and others) that were open to unskilled white male workers, and some women and African-American workers (though the latter sometimes were relegated to segregated locals). These broader organizations at times had 10 to 20 percent of all nonagricultural workers as members. In contrast, craft unions, although they unified workers within a craft, perpetuated division and disunity among the workers of different crafts.

The natural strategy was to seek to form *industrial unions*, including all types of workers in an industry or workplace. When workers did so, however, they were frequently attacked by private police hired by employers, and sometimes by local or state police as well. Labor organizers were often jailed; some were beaten or even killed. Attempts to form industrial unions, such as by the Industrial Workers of the World (Wobblies) in the lumber industry and elsewhere, had some short-term successes, but virtually no long-term success in establishing or maintaining unions from the 1890s through the mid-1930s. The power of employers, backed up by government and privately hired police power, was buoyed by the ample supply of labor resulting from the rapid displacement of workers by labor-saving technical change, the decline of employment in agriculture, the wave of immigration from Europe, and then the high unemployment of the Great Depression that began in 1929; at the height of the depression a quarter of the labor force was out of work.

Efforts to organize workers continued, and conflict between capital and labor led labor organizers to use whatever strategies and tactics seemed appropriate. Despite continuing high unemployment that weakened workers' bargaining position, the longshore and maritime workers in San Francisco embarked on a strike in 1934 lasting several months. Following brutal police violence against the strike, it erupted into a *general strike*: Instead of workers striking in just one workplace or industry, workers in a much wider range of industries walked off the job to support the longshore strike, an event reminiscent in some ways of an earlier general strike in Seattle in 1919.

Organizing to build industrial unions gained strength in the mid-1930s. By 1934–35, after five or so years of disastrous depression, few workers believed that capitalism was functioning well, and having a union to defend their interests seemed more important than ever. In the mass-production industries—steel, auto, textiles, electrical products, rubber tires, and others—many workers were not eligible to join craft unions, and yet a strike could only succeed if it included a large majority of workers in a workplace. In a number of industries, organizers offered to workers—whether skilled or unskilled, and without regard to craft—the chance to join a single industry-wide union.

An important victory for workers was the passage of the Wagner Act in 1935. Under the pressure of a rising number of strikes that were violently repressed, Congress passed, and President Franklin D. Roosevelt signed, a bill designed explicitly to protect workers' efforts to form unions. The Wagner Act established the National Labor Relations Board (NLRB) to mediate disputes between business and labor, to block unfair labor practices by employers (such as harassment of pro-union employees), to

oversee elections in workplaces to determine if workers wanted a union, and to obligate employers to bargain in good faith with the leaders of unions so chosen.

One of the new industrial unions was the autoworkers' union. After it was organized in 1935, tire and auto industry workers in Flint, Michigan, engaged in a new form of strike, a *sit-down strike*, in 1936–37: Instead of walking off the job (and out of the factory) to go on strike, the workers simply sat down and occupied the factory. With newly won legal rights for labor, organizers of the mass-production industries—steel, auto, textiles, electrical products, rubber tires, and others—succeeded in establishing a number of *industrial unions*, and these coalesced into the new Congress of Industrial Organizations, or CIO.

THE DECLINE OF LABOR UNIONS Despite the fact that public sector unions (those representing federal, state, county, and local government workers) have grown rapidly in recent years, since the 1960s there has been a significant decline in the share of the labor force belonging to unions. As the lower half of Figure 7.1 shows, union membership peaked at a bit more than a third of the workforce in 1950 and has been declining since then. In 2015 only 11.1 percent of all wage and salary workers belonged to a union—and the number was just 6.7 percent for private-sector employees. The figure also shows that the decline in union membership as a percent of all wage and salary employment (double line) after about 1979 roughly paralleled a decline in U.S. manufacturing employment as a percent of all nonfarm employment (the dotted line) in that same period.

The top half of Figure 7.1 shows changes (corrected for inflation) in various categories of wages over the course of the twentieth century. The data for coal miners (for which we show continuous wage data going back to the beginning of the Great Depression) illustrate how workers have benefited or been set back as the power of unions has grown and receded. Similarly, the wage data for all (non-supervisory) production workers have tracked the rise and fall of unions and of manufacturing employment.

Although most U.S. workers do not belong to a union, both the existence and the strength of unions affect how well nonunion workers fare with *their* employers. Nonunion employers who resist wage increases should not be surprised if their workers try to join an existing union or to form a new one. Hence, some employers—Delta Airlines is one—have paid relatively high wages in order to dissuade their employees from joining unions.

Figure 7.1 also shows the real value (the purchasing power, after correcting for inflation) of the federal minimum wage from its establishment in the late 1930s to the present. The real minimum wage increased steadily until the 1970s. It then declined, along with all production workers' wages. This decline flowed from the fall in union membership, as well as the movement of some production operations abroad, throwing fear into the remaining workers that their jobs, too, would flee the country. Another cause was the increased political influence of business owners relative to that of labor.

In many other countries, however, workers' organizations thrived for decades as U.S. unions shrank. In 1960 union membership as a fraction of the labor force ("union density" is the term used) was about the same in Finland as it was in the United States (about 30 percent), but in Finland it rose to 81 percent in 1993 before it receded. In Sweden, union density rose from a low of 66 percent in 1963 to a peak of 87 percent in 1994, and by 2013 was still 67 percent. Among more than thirty member countries of the Organization for Economic Cooperation and Development, the Nordic countries—including Denmark and Norway—have consistently had the

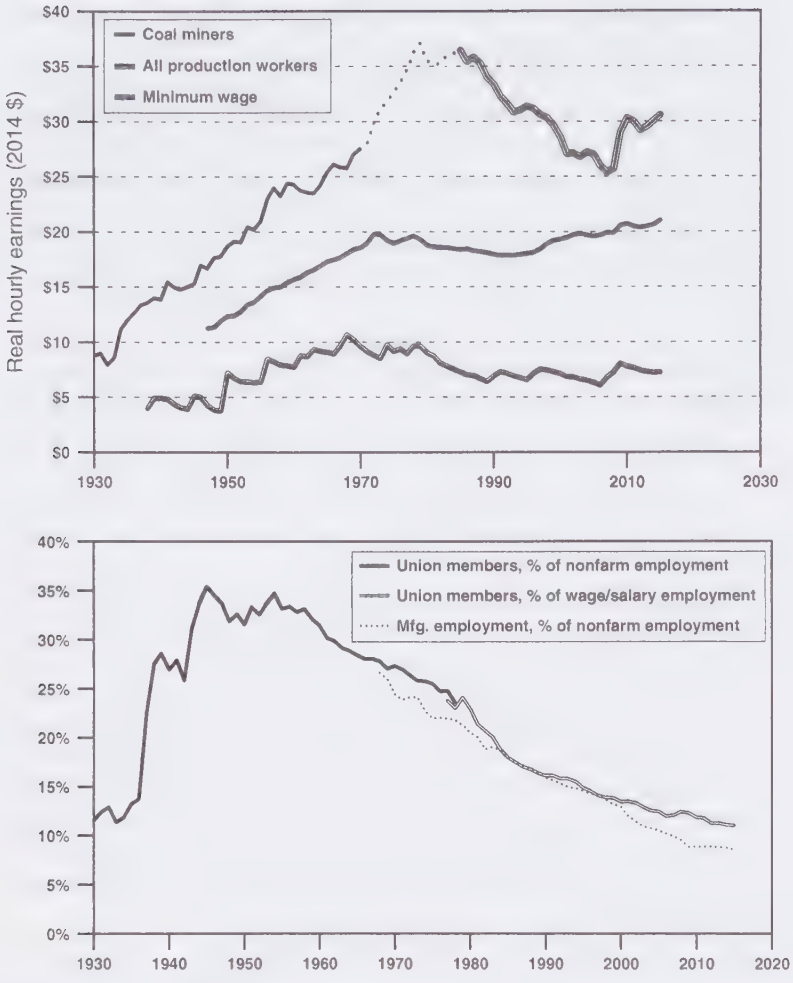


FIGURE 7.1 U.S. wages and union membership, 1930–2015.

The bottom graph shows that union membership in the United States grew in the 1930s (after the Wagner Act) and during World War II, but soon began a long decline that paralleled a fall in manufacturing's share of total employment (dotted line). The top graph shows that real wages rose along with union power through the 1970s, but generally stagnated or declined thereafter. The top line shows coal miners' wages. "Real" means that nominal wages are adjusted for inflation using the Consumer Price Index. Data for coal miners' hourly earnings come from three partially overlapping series which are not entirely consistent because of changes in which groups of miners were included; the dotted line imputes the likely trend for coal miners' wages in years for which data cover a broader group of miners. The middle line shows hourly earnings of production and nonsupervisory workers, about 80 percent of all those employed. In the upper graph, the lowest line shows the real (inflation-adjusted) value of the federal minimum wage.

Sources: Mining wages: U.S. Bureau of the Census, *Historical Statistics of the United States: Colonial Times to 1970* (Washington, DC: US GPO, 1975), p. 170: Series D 811-817, Earnings and Hours for Bituminous Coal-Lignite Mining and Class 1 Steam Railroads: 1890 to 1970; U.S. Bureau of Labor Statistics (BLS) Current Employment Statistics, Series SIC 10-14 (EEU10000006): Average hourly earnings of production workers, mining, 1947–2003; BLS, CES NAICS 2121 (CEU1021210008), Average hourly earnings of production and nonsupervisory workers, coal mining, 1985–2015; deflated using the CPI-U from bls.gov. Production workers' hourly earnings: Lawrence Mishel et al., *The State of Working America*, 12th edition, Economic Policy Institute, Ithaca, NY: ILR Press, Tables for p. 185, Figure 4B: Real hourly earnings and compensation of private production and nonsupervisory workers, 1947–2011. Real federal minimum wage: "History of Federal Minimum Wage Rates Under the Fair Labor Standards Act, 1938–2009." https://www.dol.gov/whd/minwage/chart.htm; deflated with the CPI-U. Union density (membership): Barry T. Hirsch and David A. MacPherson, unionstats.com. Manufacturing employment share of total employment: BLS series CES0000000001 (total nonfarm employment) and CES30000000001 (manufacturing employment), 1968–2015, www.bls.gov.

highest union density, with well over half of workers unionized, while since 1970 only France has had lower union density than the United States.

In the 1930s and 1940s almost all American union members were blue-collar private-sector workers. Although unions have continued to gain some new private-sector members, especially where low wages and stringent employer controls have been imposed on white-collar and service workers, on balance these gains have not offset losses among other workers. Indeed, if it were not for the 35.2 percent union membership among public-sector workers, reflecting the high unionization rates of police officers, firefighters, teachers, sanitation workers, clerical staff, hospital aides, and other state and municipal employees, the 11.1 percent overall union membership rate (in 2015) would be even lower. The gains of public-sector unions, along with the recent rise of a national nurses union, represent two of the only significant advances for the union movement since the 1940s.

Figure 7.2 shows how people in various groups and occupations differ with regard to their rates of union affiliation. African Americans, workers thirty-five or older, and people employed in utilities, construction, transportation, education, protective services, and the public sector (teachers, police, and others) are more likely than the average worker to be union members.

THE RISE AND FALL OF THE LABOR ACCORD After the intense class conflict of the 1930s and 1940s, some large corporations and unions developed an uneasy but nonetheless enduring working relationship. This "accord," or truce, was never an explicit, formal agreement. Rather, it consisted of a set of understandings and expectations that unions and corporations shared. The most important of these shared understandings was that through collective bargaining, unions would obtain for their members relative job security and steadily rising real wages, sharing with employers the gains resulting from increases in productivity. In exchange, the corporations would retain the freedom to introduce new technology, to reorganize production as they saw fit, and to invest their profits wherever they pleased. The accord was intended to give workers a fair deal and employers a free hand.

Neither side would have chosen voluntarily to make the concessions embodied in the post-World War II accord. Corporations were forced to compromise because the unions were too strong to be destroyed. As a huge wave of strikes in 1946 showed, if employers were to have any sort of labor peace they would need to come to some kind of an agreement with the unions. The unions were also forced to compromise because, while they had grown larger and more powerful in the 1930s and 1940s, they still could not match the economic and political clout of the business community. And for those workers (a majority of the labor force) who remained unaffiliated with unions, there was no labor accord. Although a few nonunion corporations—IBM and Polaroid, for example—introduced employee protections similar to the ones achieved by unions through collective bargaining, these were the exceptions, not the rule. Indeed, it is clear that such corporations did these things to discourage their workers from joining unions.

For various reasons the postwar labor-management accord began to break down in the 1970s. One key factor was that employers moved some production operations abroad, to Mexico or to other countries, especially as new legislation provided incentives to do so. Besides reducing employment, this intimidated workers in plants that had not moved, pressing them to back off from demands for higher wages and better working conditions. If in the face of globalization unions were less able to achieve gains, that meant there was less reason for workers to support unions. This and other

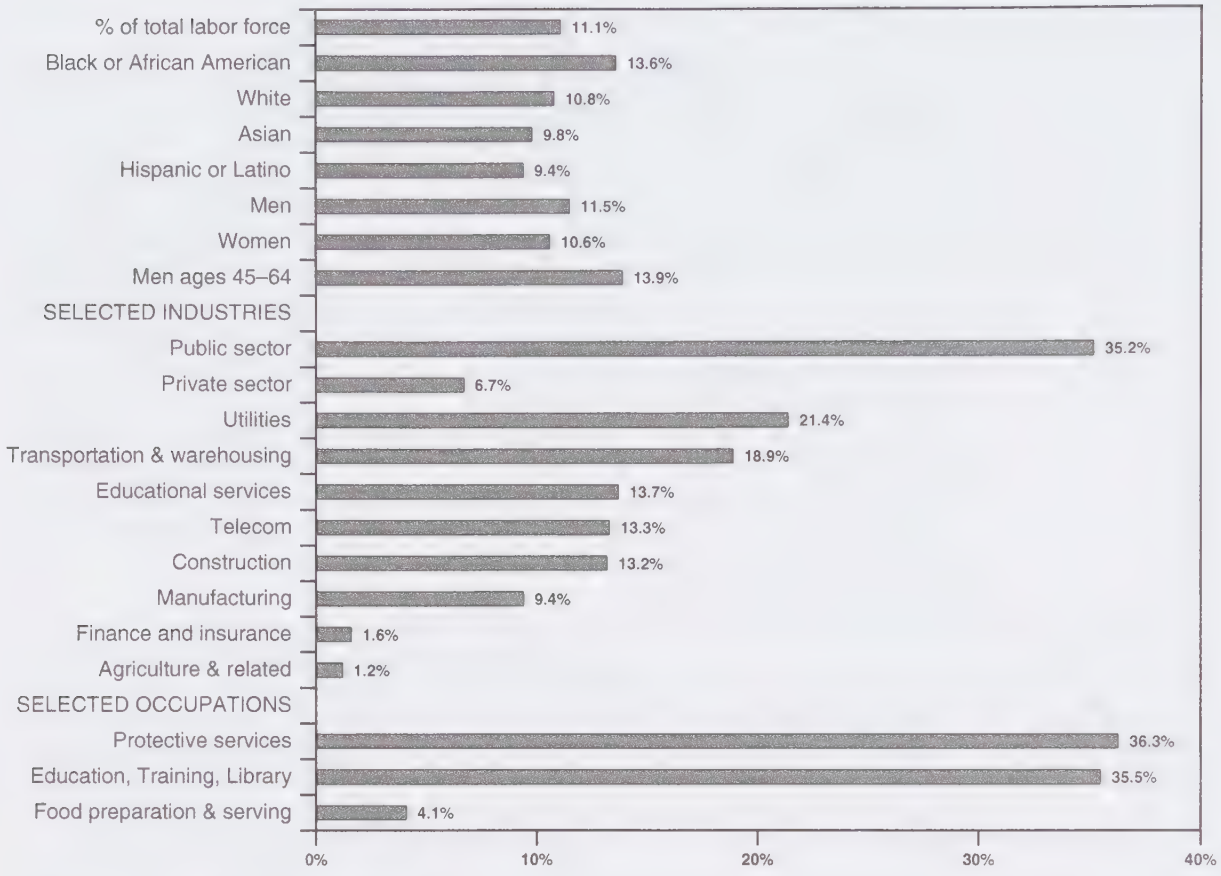


FIGURE 7.2 Who belongs to unions in the U.S.?

This figure shows the unionization rates for various categories of American workers in 2015. Each bar represents the number of union members, or members of an employee association similar to a union, as a percent of the total number of wage and salary workers employed in each group, industry, or occupation.

Sources: U.S. Bureau of Labor Statistics, "Union Members Summary: Union Members–2015," available at http://www.bls.gov/news.release/union2.nr0.htm; "Table 1: Union Affiliation of Employed Wage and Salary Workers by Selected Characteristics, 2014–2015 Annual Averages," available at http://www.bls.gov/news.release/union2.t01.htm; and Table 3: Union Affiliation of Employed Wage and Salary Workers by Occupation and Industry," available at http://www.bls.gov/news.release/union2.t03.htm.

factors altered the balance that had made the accord possible. Increasingly, employers succeeded in doing business without unions instead of with them.

Labor rights were also eroded by change in laws, judicial decisions, and enforcement. Although the NLRB was intended to be a neutral body, favoring neither business nor labor in any disputes, appointments to the board since the presidency of Ronald Reagan have given it a decidedly pro-business bias.

Segmented Labor Markets

The limited coverage of the labor accord reflects another important institutional feature of contemporary U.S. capitalism—*segmented labor markets*. Labor markets are described as *segmented*—that is, divided into separate, or distinct, markets—when the demanders and suppliers of labor in one market do not compete with demanders

and suppliers in other markets. Instead of having one large market for all jobs and all workers, there are, in effect, different market "segments"—each with its own demand for labor (by employers) and supply of labor (by workers). Labor markets may be segmented either because workers in some jobs are *shielded* from competition with other workers or because workers are excluded from competing for certain jobs. The following describes the kinds of labor market segments that existed when the concept of segmented labor markets was first developed. Segmentation persists, but some of its forms have changed.

The development of unions and the postwar labor accord tended to create segmentation by shielding some workers from wider competition. Unionized workers, through strikes and collective bargaining, won increased job security for themselves. Union contracts obligated employers to limit arbitrary firings, to grant workers increased job security as their seniority increased, and to abide by grievance procedures. Although there remained substantial competition for entry-level jobs (the jobs people get when they are first hired), unionized workers in higher-level jobs (usually with more seniority) were largely shielded from outside competition. Union contracts prevented employers from firing more experienced workers in order to replace them with new hires.

Unionized workers had won important protections from labor market competition, but since only a fraction of the labor force was unionized, only a fraction enjoyed these protections. A segmentation, or division, was created between union jobs that provided protections to workers, and nonunion jobs lacking such protections.

Even in sectors of the economy in which workers enjoyed union protections, large firms came to rely more and more on *bureaucratic control*, organizing their labor processes by stratifying their work forces, establishing job ladders, and awarding seniority-based pay and promotions according to explicit rules. Although these changes were often improvements over the arbitrary and sometimes capricious practices of some employers, they contributed to a widening gulf in work experiences and pay levels between workers in large firms and all other workers.

Job discrimination was another cause of segmented labor markets, since it prevented some workers from competing for certain jobs. For instance, female job applicants have sometimes been considered for clerical work but not for production or managerial jobs. Similarly, racial discrimination has sometimes channeled black or Hispanic workers into certain jobs while barring them from others. In this way, job discrimination increased, and still increases, labor market segmentation. To this day barriers still exist for women as well as black and Latino males seeking top-level corporate positions usually held by white males.

Discrimination, the emergence of a labor accord, and the spread of bureaucratic control eventually produced three relatively distinct labor markets. Two of these markets, the *independent and subordinate primary markets*, included the jobs of workers who were covered by the labor accord or were organized bureaucratically in some way. The remaining segment, the *secondary market*, encompassed the jobs of all the workers who were neither covered by the accord nor organized bureaucratically, and who were therefore not shielded at all from labor market competition.

The **independent primary labor market** includes those jobs with highly elaborate bureaucratic or professional career patterns; it contains mainly the jobs of craft, technical, professional, and lower-level supervisory workers.

THE INDEPENDENT PRIMARY LABOR MARKET The **independent primary labor market** included those bureaucratically organized jobs that offered stable employment, often with considerable job security, clearly defined career paths, and relatively high pay. It included bookkeepers, technicians, scientists, engineers, lower-level supervisors and managers, commercial artists, and craft workers such as electricians, telephone linemen, and machinists.

Independent primary jobs typically required, and still require, skills learned in apprenticeships or in some type of high-level or specialized schooling. Workers may need graduate school degrees, special licenses, or other credentials to qualify for these jobs, and formal credentials are highly rewarded. Blue collar independent primary jobs may require an apprenticeship of several years' duration or more. Some independent primary workers are organized in craft unions, although others, such as machinists, are members of industrial unions that include a wide range of occupations. In any case, success in their jobs is often judged according to occupational or professional standards (not just the employer's approval) that establish what constitutes good work. Such jobs are labeled "independent" because they frequently permit, and sometimes require, independent initiative or self-pacing of work.

The **subordinate primary labor market** includes most government employment, as well as private sector jobs in workplaces organized according to collective bargaining agreements; it includes mainly the jobs of the traditional unionized industrial working class.

THE SUBORDINATE PRIMARY LABOR MARKET The **subordinate primary labor market** included, and still includes, most government employment as well as the jobs of the traditional unionized industrial working class: autoworkers, truckers, railroad workers, underground coal miners, steelworkers, dockworkers, and so on. These private sector jobs differ from jobs in the secondary labor market mainly by the fact that many of these workers are unionized. Because in some industries, such as steel-making, there exist both unionized and nonunionized firms, it is mostly the fact that wages and working conditions are negotiated by a union that places it in the subordinate primary market. However, in industries where workers have both skill and bargaining power, and are hard to replace in the short term, unions are able to negotiate higher wages and better working conditions than in industries where workers are more replaceable and there is intense competition between unionized and nonunionized firms.

Subordinate primary labor market jobs are generally better paid than are jobs in the secondary market and have typically included some medical and retirement benefits. However, union pensions in private sector employment have declined sharply since the early 1980s, while most federal and state employees, and some local government employees, continue to have government pension plans.

As long as the firms in which subordinate primary labor market jobs are located have been growing, they have generally involved long-term, stable work with prospects for pay raises and some job security. Seniority has usually been highly rewarded. The work itself, however, is distinguished from independent primary jobs in that it is repetitive, routinized, and often subject to machine pacing. The required skills can be learned fairly quickly and are often acquired on the job, but the jobs themselves provide little autonomy and offer few opportunities for the exercise of independent initiative.

A **layoff** refers to a firm's temporary or permanent dismissal of workers for any of several reasons.

Despite their protection by union-negotiated grievance procedures from being fired arbitrarily, workers in subordinate primary jobs have remained highly vulnerable to layoffs. The term **layoff** refers to a firm's temporary or permanent dismissal of workers to reduce its workforce either due to a drop in sales, automation of some operations, or a shift of some or all production to another location. Millions of unionized workers employed in the old "rustbelt" manufacturing industries (steel, auto, and the like) have been laid off in recent decades, and many will never return to their old well-paying jobs.[2]

---

[2]Clare Ansberry, "Laid-Off Factory Workers Find Jobs Are Drying Up For Good: It's Not Just the Slowdown—Structural Changes Strand Many With Basic Skills," *Wall Street Journal*, July 21, 2003, p. A1.

With the decline of private-sector unions from the 1980s on, the subordinate primary labor market shrank. In such industries as auto, steel, and other heavy manufacturing, some production remains unionized, but newer plants have been set up without unions, and some manufacturing has been moved overseas or displaced by competition with firms abroad. Many U.S. workers have been laid off in the subordinate primary market and some have been reemployed in the secondary market.

The **secondary labor market** includes jobs in workplaces that lack the formal organization (such as collective bargaining agreements, bureaucratic control, or professional or craft patterns) of primary markets; it contains jobs such as those of service and retail workers, clerks in the private sector, seasonal workers, and nonunionized employees of small businesses.

THE SECONDARY LABOR MARKET The **secondary labor market** includes most of the remaining workers. It is highly diverse, unified only in that it is the preserve of workers in dead-end jobs who have few protections from economy-wide labor market competition. These workers have neither the rights enjoyed by unionized workers, nor the benefits of a system of bureaucratic control imposed by the employer. Here the relation between boss and worker is simple and direct, and workers continually face the threat of being replaced.

The secondary market includes the jobs of blue-collar workers in nonunion factories; nonunion janitors, waitresses, hospital orderlies, messengers, guards, retail sales clerks, data entry personnel, file clerks, and record keepers; seasonal or migrant farm workers; and many other employees of small businesses. Union organizing has made inroads among some of these workers, particularly those in state and local governments. Thus it is mainly the nonunion private sector workers in such jobs who are in the secondary labor market. What marks these as secondary labor market jobs is that they lack the features that shield workers in primary labor markets from economy wide competition. Secondary jobs are usually less secure and less well paid. Secondary market workers also typically lack a designated career path: They have no guarantee that they will progress up a job ladder or acquire more skill, pay, job security, and responsibility as they gain seniority. Neither schooling nor seniority is highly rewarded in these jobs. In short, they are dead-end jobs.

U.S. Capitalism Today is Transnational

Even while dividing the labor market into distinct segments, the accumulation process has created an economy that is more integrated with the rest of the world. Increasingly, U.S. firms have moved into the world economy, and the rest of the world's products and some of its peoples have been brought into the United States.

For much of its history the sheer size of the United States and its distance from the large population centers in Asia and Europe isolated the U.S. economy from the rest of the world. No more. Over the past one hundred years, the cost of transporting goods over long distances has dramatically fallen. The nature of goods has changed, too: some products can be "shipped" electronically at virtually zero cost.

Increasingly, firms based in the United States operate on a world scale; the accumulation process is now a global phenomenon. Figure 7.3 shows that the fraction of all U.S. profits reportedly made abroad has been increasing. Half a century ago, for every $20 of American profit, just $1 was reported to have been made outside the United States; now that number is about $8 out of every $20. Some of this shift, however, is not a change in where the goods are produced and sold at a profit, but rather a change in accounting practices: firms deliberately shift reported profits to places such as tax havens where they will be taxed lightly, if at all. Shifting profits abroad also allows delaying paying the taxes on those profits to the IRS for years or

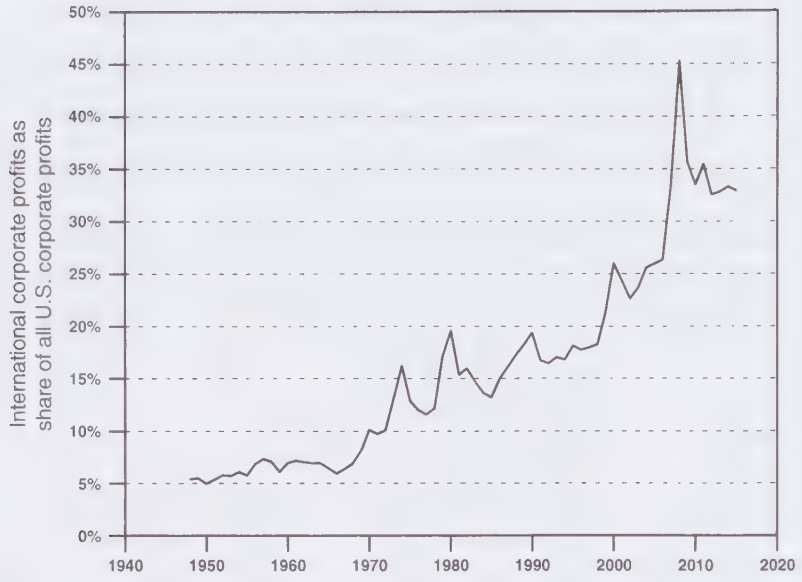


FIGURE 7.3 The growing importance of international profits, 1948–2015.

This figure charts the profits received by U.S. corporations from economic activity in other countries as a percent of all U.S. corporate profits. U.S. corporate profits coming from abroad include: (a) dividends (which include distributed earnings and other private receipts) and (b) undistributed profits (which include reinvested earnings and income of U.S. commonwealths and territories, for example, Puerto Rico and the U.S. Virgin Islands). These are before-tax corporate profits. This means that corporate profits coming from abroad are figured after payment of taxes to foreign governments, but before payment of U.S. taxes.

Source: U.S. Department of Commerce, Bureau of Economic Analysis, National Income and Product Accounts, Tables 6.16B, 6.16C, and 6.16D: Corporate Profits by Industry, Line 6: Receipts From the Rest of the World, Divided by Line 1: Corporate Profits with Inventory Valuation and Capital Consumption Adjustments.

even decades. We return to the globalization of the accumulation process in Chapter 15, where we address the question: When U.S. firms seek greener pastures abroad, where do they go and why?

While U.S. businesses benefit from profit-making opportunities around the globe, they and their U.S. employees face stiffer competition from goods produced abroad (including goods produced by U.S. companies operating abroad). Although both consumers and businesses may benefit from the lower prices of imported goods, U.S. firms and workers are often hard hit when they have to compete with low-cost producers in other nations. More than a third of the cars and trucks sold in America are produced elsewhere, even if often by U.S. transnational corporations, or by joint ventures between such U.S. transnationals and automakers based in other countries. As for other products, the United States is still a major exporter of corn, wheat, and other grains, and still produces much of its own meat, but since the 1990s has been importing a growing share of other foods. From 1990 to 2013, imports of plant-based foods, including beer and wine, as a share of all domestic consumption of such foods, doubled from 12.8 percent to 26.8 percent; meanwhile, the import share of animal-based foods in domestic consumption was lower and grew more slowly, from 8.5 percent to 11.9 percent.

Immigration

While the internationalization of the U.S. accumulation process is new, the fact that the United States is a global nation in its ancestry is as old as the Pilgrims, conquistadores, and slavery. As high school history textbooks explain, many immigrants came to America fleeing religious intolerance and oppression, and these "huddled masses, yearning to breathe free" are celebrated by the Statue of Liberty in New York Harbor. However, as Figure 7.4 shows, the fraction of U.S. residents born in another country has varied considerably over the past 150 years.

The big upsurge in immigration in the latter half of the nineteenth century was not simply the result of people fleeing from intolerance and oppression, however. It was also driven by an early episode of globalization quite different from the globalization we see today. In the latter part of the nineteenth century, what happened was virtually the reverse of the process as we now know it: U.S.-produced goods were putting people in other countries out of business.

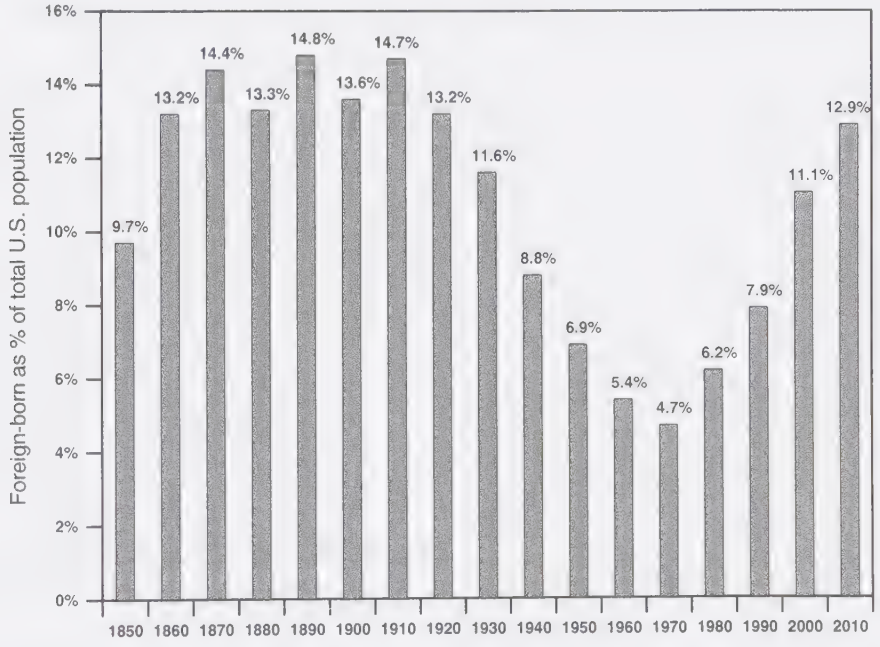


FIGURE 7.4 Percent of U.S population born in other countries, 1850–2010.

This figure shows, in each year, what percent of the U.S. population was born outside the U.S. From 1860 to 1920, the percent of the whole population that was foreign-born hovered between 13 and 15 percent, reflecting the large flow of immigrants from Europe in that period. After 1920, immigration into the U.S. was restricted, and the percent of foreign-born Americans declined. U.S. immigration laws began to be relaxed again in the 1960s. As a result, immigration increased and the percentage of Americans born in other countries rose from 4.7 percent in 1970 to 12.9 percent in 2010, with most immigrants coming from Latin America and Asia.

Sources: Campbell J. Gibson and Emily Lennon, "Historical Census Statistics on the Foreign-Born Population of the United States: 1850–1990," Table 1, "Nativity of the Population and Place of Birth of the Native Population, 1850 to 2000," http://www.census.gov/population/www/documentation/twps0029/twps0029.html; U.S. Bureau of the Census, *Profile of the Foreign-Born Population in the United States: 2000*. Current Population Reports, Special Studies; available at http://www.census.gov/prod/2002pubs/p23-206.pdf; Current Population Survey, March 2012 Detailed Tables: "Table 1.1: Population by Sex, Age, Nativity, and U.S. Citizenship Status: 2012," https://www.census.gov/data/tables/2012/demo/foreign-born/cps-2012.html

With the construction of railroads across the American and Canadian Midwest, and with the sudden decline in ocean freight rates following the replacement of sail by steam power, cheap North American grain flooded into Europe. Just as the growing availability of inexpensive foreign cars and electronic devices (often produced in foreign affiliates of U.S. multinational firms) has eliminated the jobs of those Americans who were unlucky enough to be producing similar products, imports of cheap North American grain devastated European farmers in the second half of the nineteenth century. Some of these farmers were able to switch to other ways of making a living where they were, but many migrated to the United States (or to Argentina or Canada).

The decline in the fraction of the American population born outside the United States following World War I was the result of a deliberate policy to restrict immigration. This tightening of U.S. immigration laws has been relaxed somewhat during the past few decades, but as Figure 7.5 shows, the latest wave of immigrants has come primarily from Latin America and Asia, not from Europe. One result is that the United States is becoming truly a global nation.

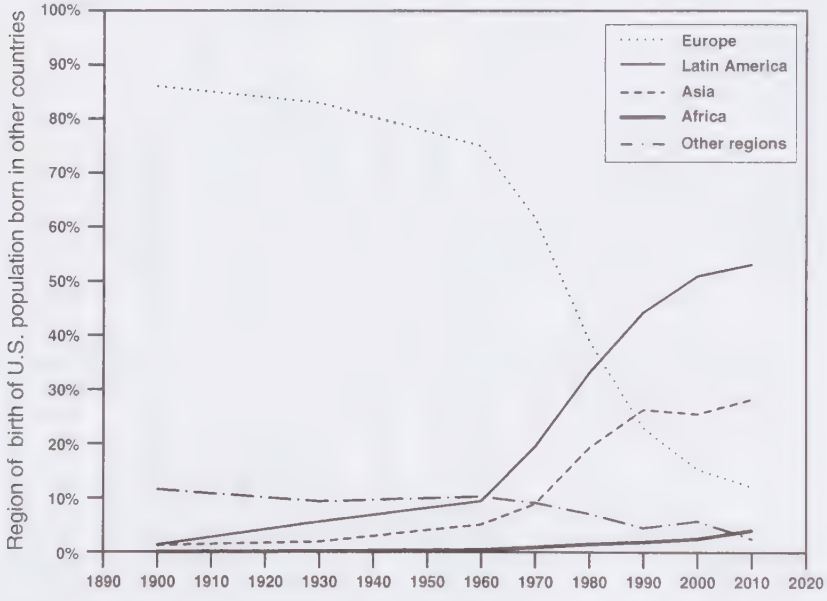


FIGURE 7.5 Region of origin of U.S population born in other countries.

Where do the foreign-born in the U.S. come from? This figure traces trends from 1900 to 2010, a period over which the composition of the foreign-born changed significantly. In 1850 more than 90 percent of the first-generation U.S. immigrant population were European-born, and they remained the majority of the foreign-born population in the U.S. until 1980. In 2010, however, the European-born U.S. population was barely 12 percent, mostly not recent immigrants. Since 1930 there has been a rapid growth in the number of immigrants coming from Latin America and Asia, so Europeans are now in third place, far behind people from these two areas.

Sources: U.S. Bureau of the Census, "Table 2: Region of Birth of the Foreign-Born Population: 1850 to 1930 and 1960 to 1990," Internet release date March 9, 1999, http://www.census.gov/population/www/documentation/twps0029/tab02.html, based on Campbell J. Gibson and Emily Lennon, "Historical Census Statistics on the Foreign-born Population of the United States: 1850–1990"; U.S. Bureau of the Census, *Profile of the Foreign-Born Population in the United States: 2000*, Current Population Reports, Special Studies, P23-206, December 2001, p. 11, Figure 2-2, "Foreign-Born Population by Region of Birth: Selected Years, 1850 to 2000," and p. 19, Figure 6-2; Elizabeth M. Grieco et al., "The Foreign-Born Population in the United States: 2010," American Community Survey reports, ACS-19, May 2012, https://www.census.gov/prod/2012pubs/acs-19.pdf.

Shifts in What U.S. Labor Produces

The never-ending competition for profits has changed the structure of the U.S. economy in other ways, too. When people think about the economy, they often visualize smokestacks, mineshafts, and fields of waving grain. Think again. Visualize point-of-sale terminals, restaurants, keyboards, call centers, classrooms, and hospital beds. Driving through an unfamiliar town, you may have wondered: what do people *do* here to make a living? What do they *produce?* They may in fact *produce* no physical product. Nowadays, there are far more people in the United States *selling* things than making them. Moreover, most Americans neither sell nor make goods: they provide *services* by educating, healing, waiting on tables, transporting goods, and doing many other useful things.

As Figure 7.6 shows, the number of Americans producing goods—in agriculture, mining, construction, or manufacturing—has never exceeded 25 million, a level it almost reached in the year 2000. From 2009 to 2015 it stayed below 20 million.

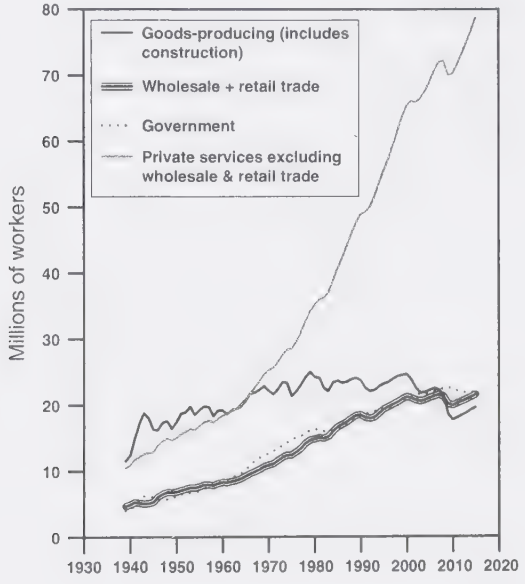


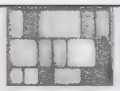
FIGURE 7.6 The changing composition of employment in the U.S., 1939–2015.

The composition of employment in the United States, like that in other rich countries, has shifted toward employment in service-producing firms (including wholesale and retail trade), while government services have also grown with the establishment of Medicare and the growth of the total and the over-65 population. Goods-producing workers here include construction workers. Some such changes result from changes not in the nature of work but in the organization of work. For instance, manufacturing plants used to hire custodial staff directly, but now often contract the work out to custodial service firms, reclassifying their workers into the service-producing instead of the goods-producing sector, although the actual work has not changed. A similar shift has occurred in bookkeeping, accounting, graphic design, food service, and many other kinds of services; computer programming services are typically contracted out as well.

Source: U.S. Bureau of Labor Statistics, *Employment, Hours, and Earnings from the Current Employment Statistics* Survey, "Establishment Data, Table B-1. Employees On Nonfarm Payrolls by Industry Sector and Selected Industry Detail." Available at http://www.bls.gov/webapps/legacy/cesbtab1.htm

Meanwhile, the number of people employed in selling things and providing services has been rapidly growing. Workers in private services excluding wholesale and retail trade already exceeded goods-producing workers in 1965; workers in wholesale and retail trade surpassed goods-producing workers in 2009. Together, the number of workers in services and wholesale/retail trade together exceeded 100 million in 2015, five times the number who were producing goods in that year. One in three U.S. workers produced goods of some sort fifty years ago; today about one in six does.

Why has goods-producing employment remained flat in the United States for the past thirty years, and fallen recently? One reason is that productivity growth in the production of goods has enabled the same number of people to produce increasing quantities of grain, steel, cars, and most other things, allowing more people to move into—or take their first jobs in—sales and service occupations. But there have been other factors as well. U.S. companies have shifted their manufacturing



Globalization: U.S. Winners and Losers

Globalization has winners and losers, with higher-wage workers often losing jobs to lower-wage workers in other countries, or accepting wage cuts to prevent this from happening. Three cases show the harm that U.S. workers can suffer.

One detailed study was done of workers who lost their jobs at a Smith-Corona plant in Cortland, New York, when it was moved to Tijuana, Mexico, in 1992. A sample of 159 laid-off workers was interviewed three years later. Men had lost an average of a quarter of their previous income; women had lost 36 percent. The "after" earnings were not only lower on average, but were also much more unequal, with almost a third of the workers earning somewhat more after the plant closing, but many earning far less.

When a plant shuts down, it is not only the plant's own workers who lose their jobs: there is also a domino effect on other workers. When General Motors closed its auto assembly plant in Moraine, Ohio, in 2001, it laid off 2,170 workers. But researchers from a research institute at the University of Michigan found that, right in the area where the plant was situated, there were another 10,850 layoffs. On top of that, yet another 3,334 lost their jobs among those who worked for firms in GM's supply chain. This included 2,120 layoffs in four Delphi plants that supplied parts to the Moraine GM plant, and other layoffs among other suppliers. In all, the number of indirect jobs lost because the Moraine plant was closed totaled 27,520.

The third group that is affected is those still working in similar plants, who are unnerved by seeing the effects on their co-workers and fellow union members. There is solid evidence that U.S. unions bargaining with firms that have plants outside the United States have had less success in raising wages (see box "Hardball: Owners, Workers, and Taxpayers in the Global Economy" in Chapter 13).

Of course, there are winners as well as losers: The workers in the places to which the jobs moved—like the new Smith Corona plant in Tijuana—needed the jobs, too. But because the Mexican workers' pay was lower, the stockholders also benefited.

Source: Lourdes Beneria and Luis E. Santiago, "The Impact of Industrial Relocation on Displaced Workers: A Case Study of Cortland, New York," *Economic Development Quarterly* 15, no. 1 (2001): 78–89; James Heintz, Low-wage manufacturing and global commodity chains: a model in the unequal exchange tradition," *Cambridge Journal of Economics* 30, no.4 (2006): 507–520; Richard McCormack, "A GM Factory With 2,100 Workers Closes, And 33,000 Other People Lose Their Jobs—Impacting 120,000," *Manufacturing & Technology News* 17, no. 1, January 17, 2010, available at http://manufacturingnews.com/news/10/0112/GM.html; Roland Zullo and Kallol Mukherji, "Regional Employment Effects of Motor Vehicle Industry Job Loss in Michigan Counties: 2001 to 2008," Institute for Research on Labor, Employment, and the Economy (IRLEE), University of Michigan, February 2011.

operations abroad, often to countries where the wages are lower and the regulations less stringent. (See box "Globalization: U.S. Winners and Losers," and accompanying figure.) By one estimate, the United States had 2.7 million fewer manufacturing jobs in 2011 than in 2001 due just to its growing trade deficit with China.[3]

The sales and service sectors of the U.S. economy differ in many ways from the goods-producing ones. Of course, there are large differences between "good" manufacturing jobs—unionized jobs in the automobile and steel industries, for example—and "bad" manufacturing jobs—certain jobs in food-processing plants and clothing sweatshops (yes, sweatshops still exist in the United States). But the differences between "good" and "bad" jobs in the service and sales sectors are far greater. For example, many jobs in the financial services industry are among the best paid in the world, while in the United States people who perform "domestic" and janitorial services are paid very little. Many sales workers are also very poorly paid because few of them are represented by a trade union, so they have little bargaining power and minimal on-the-job rights. As has been widely publicized, some Wal-Mart employees are paid so little that they are eligible for public welfare, and as a group these workers receive $6.2 billion in benefits—so that in effect taxpayers are subsidizing Wal-Mart's profits. Selling real estate, on the other hand, is quite lucrative, and many government and transportation workers (both classified as being in the service sector) belong to unions and have secure jobs and relatively high pay.

Fragmenting Global Production

We date the current SSA from 1991 for several reasons. Other economists have argued that it began as early as 1981, when the Reagan administration began deregulating parts of the financial sector, and favoring the wealthy by slashing income tax rates on the highest incomes. We judge, however, that on a number of counts, the SSA was not fully consolidated until the 1990s, although the shape it took in the 1990s certainly had roots in the 1980s. Before and during this period, an increasing number of U.S. corporations became transnational by establishing plants and other affiliates (various types of corporate entities) in other countries. In the 1990s and later, many firms that were already transnational increased the number of their foreign affiliates and moved more of their operations abroad, particularly as China opened up to foreign investment.

Trade also increased, largely because transnational firms further fragmented their production, having components produced in various countries and then assembled into subassemblies that were then shipped to another country for further assembly. Why did production become so fragmented? One reason is that the electronic age made many parts smaller, lighter, and therefore cheaper to ship. Another is that trade barriers such as import tariffs were lowered. New advances in electronics also made it easier for management to coordinate production among affiliates and suppliers in many nations, a strategy that also alerted suppliers that they needed to remain competitive or they would lose out to rival firms in other countries. Such strategies also reduced the risk that unanticipated and disruptive political or economic changes, earthquakes or typhoons in any one country would interrupt a firm's global production process.

---

[3]Robert E. Scott, "Trading Away the Manufacturing Advantage," Economic Policy Institute, EPI Briefing Paper #367, September 30, 2013, p. 26, available at http://www.epi.org/publication/trading-manufacturing-advantage-china-trade/

Rules for the Global Economy

Transnational corporations also benefited from expansion of their rights and privileges under new international agreements. For decades there had been a move toward freer trade under the General Agreement on Tariffs and Trade (GATT), in a series of negotiations during the decades following World War II. But in the 1980s, in a new round of GATT negotiations among member countries under that pact, a new set of rules was agreed upon, much broader in scope. These new rules were enshrined in the World Trade Organization created in 1995. Some of the rules implied freer trade, but others implied the opposite, especially new rules protecting and expanding intellectual property rights (IPRs: monopoly patents and copyrights, and the like). These IPRs benefited patent holders such as pharmaceutical corporations, and copyright holders such as software producers, as well as film and other entertainment companies.

Seed companies gained a new right to patent life forms, and so they patented seeds they had developed. This meant that farmers who for centuries, even millennia, had reserved part of each grain harvest for seed for the next planting season (as in the "Grain Technology" box in Chapter 4, no longer had the right to do so. Now, every planting season these farmers had to pay patent royalties (fees paid to the patent holder for permission to use the patented seeds) to the seed company—even if the farmers had saved the seeds out of the crop they themselves had grown out of patented seeds. Firms like Monsanto gained massively, not by producing something of great value, but by securing a new property right—to life forms like seeds—which before had been no one's property. The new, expanded IPRs also made patented medicines less available to developing countries. Governments of those countries, particularly in southern Africa, that were suffering from an HIV/AIDS pandemic needed affordable medicines to treat the disease. So, during over a decade of WTO negotiations beginning in 2001, they sought to expand the right of countries, in cases of public health emergencies, to produce or buy needed generic medicines without paying royalties to patent owners. They had only mixed success.

Some of the new WTO rules in 1995 also weakened the power of national governments to regulate foreign investment in the national interest, a rule that especially affected resource-rich developing countries in their dealings with oil and mining companies. In both resource extraction and manufacturing, this has made it harder for developing countries to ensure that their workers and managers learn new technologies from foreign firms. For example, before it joined the WTO in 2001, China had a strong policy requiring foreign investment projects to include specific plans to transfer technology to local managers and workers. When China joined the WTO, it was compelled to formally renounce that policy. In the case of China, however, its size and attractiveness as a location for foreign investment allowed it to work around the technology transfer rules and make it difficult for foreign investors who did not share technology. But for smaller countries the WTO rules are harder to get around, and limit their governments' ability to choose policies that best foster development.

The ultimate effect of new WTO rules was to strengthen the hand of developed countries relative to developing countries and, in developing countries, the hand of transnational corporations relative to national corporations and national governments. Transnational capital has more freedom than it had before, and national governments have less freedom to choose the policies that benefit their populations. Other bilateral and multilateral agreements, including the North American Free Trade Agreement (NAFTA), went further, giving foreign investors rights that even

national firms did not have, such as the right to sue national governments. Chapter 19 will spell these rights out further and discuss their implications.

Tax Motives for TNCs to Go Global

Ever since just after World War I, one of the reasons for firms to establish foreign affiliates has been to reduce the amount of corporate income taxes they owe to the IRS, as well as the amount they actually pay in any given year. If a U.S.-based transnational firm has income from operations abroad, it has to file an income tax return stating that income and figuring the taxes it owes. But it has the option of deferring (postponing) actual payment of those taxes until it "repatriates" (returns to the United States) the income it earned abroad. It is allowed to keep the income abroad for many years, and invest those funds and earn a return on them, all the while deferring payment of the taxes the corporation has already admitted it owes. Not surprisingly, many corporations show no eagerness to return the earnings to the United States and pay the taxes on them. It has been estimated that $2.4 trillion in income earned abroad by U.S. transnational firms is being held abroad, and that the taxes owed on that sum may amount to close to $700 billion.[4]

There are also at least two ways in which transnational corporations can end up legally *never* paying most of the taxes they owe. One is to persuade Congress to declare a tax holiday, allowing firms to repatriate their earnings in a given year and fulfill their tax obligations on those earnings by paying a small fraction of the taxes they said they owed. Such a tax holiday was declared by Congress in 2004, and many TNCs took advantage of it in 2005, bringing back several hundred billion dollars in total, and so (legally) fulfilling their obligations by paying only 15 cents on each dollar of the taxes whose payment they had deferred. The other method has been for a firm to move its headquarters outside the United States (such as by performing a "corporate tax inversion"—selling itself to an overseas firm—or some variant of that). A corporate inversion puts the firm's past and future earnings out of reach of the IRS, since the firm is no longer a U.S. resident—though in 2015–2016 new regulations were issued to discourage such moves. Chapter 19 will explore these issues in more detail.

The Transnational Capitalism SSA: Deregulation and Financialization

The transnational aspects of the current SSA are important, but are just one of its features, and some of its other facets are equally central. Let us explore some of these. One feature has been deregulation of the financial sector, which began in earnest in the early 1980s under the presidency of Ronald Reagan. Some of the deregulation was described in Chapter 6: lifting the cap on interest rates and relaxing the rules governing savings and loan institutions. These measures led to a major crisis in the S&Ls, and the bankruptcy of many of these institutions, and eventually prosecution and conviction of some of their executives and personnel for fraud.

Financial deregulation that was begun by the Reagan administration in 1981 (described in Chapter 6) was extended in the 1990s. For example, legislation in 1999

[4]Citizens for Tax Justice, "Fortune 500 Companies Hold a Record $2.4 Trillion Offshore: They May Be Avoiding up to $695 Billion in U.S. Taxes," March 3, 2016, available at http://ctj.org/pdf/pre0316.pdf.

tore down the remaining wall between commercial banking and investment banking, and prohibited regulation of certain risky financial products called derivatives (to be explained in Chapter 18), some of them new, whose effects on the economy were poorly understood. This opened a Pandora's box of financial speculation that ignited the financial and economic crisis of 2007–2009.

Deregulation, along with innovation, allowed domestic financial sector corporate profits to rise as a share of total domestic corporate profits. As shown in Figure 7.7, the ratio of domestic financial sector corporate profits in the United States to all domestic corporate profits rose, particularly in the 1990s and later. Financial sector salaries also grew faster than those in the rest of the economy, and the net product (value added) in the financial sector began to account for a rising share of the whole economy's value added.

In addition, the financial sector began to lobby Congress heavily and to escalate its campaign donations, particularly targeting members of Congress who sat on banking and financial committees in both houses. By 2014 the financial sector

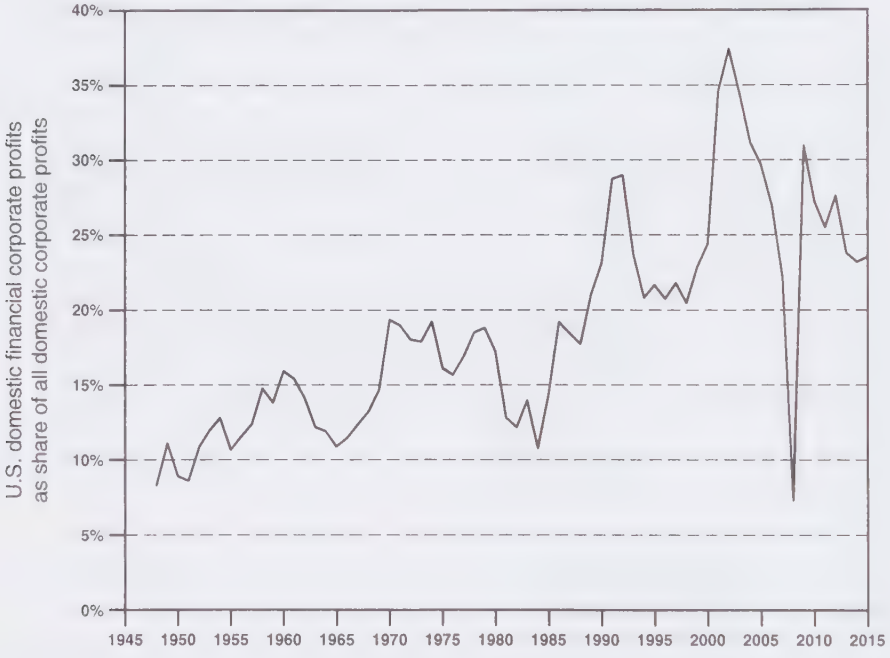


FIGURE 7.7 The financial sector's share of all reported U.S. domestic corporate profits, 1948–2015.

Financial profits have become an increasing share of all corporate profits in the U.S., rising from about 10 percent after World War II to over 30 percent in some years in recent decades, with a dip in 2008 due to the financial crisis that year. Some part of this trend may be explained by nonfinancial transnational firms using accounting devices to shift their reported profits from the U.S. to overseas affiliates to reduce their tax liability. Shifting reported profits abroad also reduces the taxes they actually pay in any given year, because as long as their earnings abroad are kept outside the United States, they can legally defer (postpone) payment of the taxes for which these firms have already declared they are liable. If Congress declares another tax holiday for repatriated profits (as happened in 2004), it is possible that firms can also escape much of even those deferred taxes.

Source: U.S. Bureau of Economic Analysis National Income and Product Accounts Table 6.16B: Corporate Profits by Industry, Line 3: Corporate profits (with IV and CCAdj), Domestic Industries, Financial, divided by Line 2: Domestic Industries (last revised August 6, 2015).

was spending well over a million dollars a day on lobbying at the federal level, an extraordinary sum given that there are only 535 members of Congress altogether. As Chapter 18 will recount, the rewards from such generosity were considerable. Much of the financial sector was handsomely bailed out from the 2007–2009 crisis, and its executives generally escaped prosecution, although the institutions themselves eventually paid monetary settlements in the billions of dollars.

Deregulation of the financial sector has been by no means the only important change in rules. Other rule changes have affected the balance of power between owners of capital and labor, as well as the benefits each class gets from government, and the costs it pays. For example, in earlier decades a social safety net was in place that prevented layoffs or high medical bills from turning family misfortune into catastrophe. Unemployment benefits replaced some of lost wages; families reduced to poverty could go on welfare, and if they were truly drowning in unpayable bills, such as from expensive medical care, they could file for bankruptcy, wipe out their debts, and start over. But these safety nets were weakened. In the mid-1990s, the welfare program Aid to Families with Dependent Children was replaced with the much less generous Temporary Assistance to Needy Families. TANF limited the number of years during a lifetime that a family in need could get assistance, so that needy people who had used up all their benefits could easily become destitute. As a result of this change in the rules, government assistance to poor children dropped sharply in the two decades after the AFDC program was ended. In 1995, 62 percent of children living in poverty were receiving government assistance (AFDC benefits or food stamps, mostly). But in 2014 a mere 17 percent of children in poverty were getting government assistance (TANF or SNAP).[5]

And in 2005, not long before the subprime crisis hit, Congress also made it harder for deeply indebted families to file for bankruptcy and make a new start. When the law took effect in 2006, bankruptcy filings suddenly dropped. Nationwide, the number of bankruptcies filed per 1000 of population during the *three* crisis years of 2007–2009 (when debts became overwhelming to many) was *less* than the number filed in the *two* boom years of 2004–2005 (when millions more people were still employed).

Another factor that allowed capital owners an edge in the conflict with labor was the demise of the Soviet Union in 1991. This event shifted the balance of power in favor of capital. The collapse of the first country that had established a form of socialism after the 1917 Russian Revolution diminished the threat that workers in other countries angry with their own employers or governments might rise up and form a workers' government that would seize capital from its owners. Past situations in which the Soviet Union had sometimes given monetary and technical aid to such uprisings were no longer possible. Removal of this threat made it that much easier for employers to keep wages down, to defeat unions, and to keep workers powerless in precarious employment, with irregular work hours, hazardous working conditions, and no job security.

---

[5]Child Trends Data Bank, "Child Recipients of Welfare (AFDC/TANF)," updated December 2015, available at http://www.childtrends.org/?indicators=child-recipients-of-welfareafdctanf.

Corporate Stock Buybacks and Falling Productive Investment

In 2014–2015, there were a hundred CEOs in the United States who were paid at least $20 million a year; ten of them made $40 million or more.[6] And these numbers do not include hedge fund managers, who typically made far more: the highest-paid twenty-five make an average of about half a billion dollars a year, and the very highest-paid among them make several billion dollars a year.

Such astronomical pay for top executives was not always the case. CEOs were paid about twenty times their average worker in the 1960s, and about sixty times by the beginning of the 1980s. Not until the 1990s did these numbers grow to over three hundred times the pay of the average worker. In fact, the increase in CEO pay accounts for a big share of the overall increase in income inequality in the United States over the last several decades.

Why did CEO pay grow so fast? It was partly because stocks became an important part of CEO compensation packages. In turn, this was in large part because some economists made a persuasive case that if a CEO's pay package depended directly and heavily on the company's stock price, then he or she would manage the company more in the interest of the stockholders. Without stock prices as a direct incentive, it was thought, CEOs would tend to build their own empire within the company, trying to benefit themselves instead of acting in a way that produced long-term growth in the value of the firm. The idea of basing a big chunk of CEO pay on stock was that a good CEO, motivated to make the company's value rise by owning a large block of stock, would use the company's earnings to invest wisely in research and development to improve products or production processes, or to open up new markets. Sales and profits would grow, more investors would buy the stock, and the stock price would rise, benefiting the stockholders as well as the CEO.

In practice it was not quite that simple. The snag was that there is a much simpler way to drive the stock price up, free from the diligence, thought, and sheer hard work required to improve the quality, distinctiveness, or appeal of the product, or the technology for producing it. CEOs could just use the company's earnings to buy back its own stock, directly driving the price up in the short term. Before 1982 such stock buybacks were illegal: the rules of the Securities and Exchange Commission did not allow executives to manipulate their own company's stock price. But in 1982 the SEC changed the rules, making it legal for a corporation's executives to buy back the company's own stock. As long as they did not buy back too much in one day, they need not worry that the SEC would charge them with manipulating their own stock price.

The result was a tidal wave of stock buybacks by large corporations in recent decades. Over the ten years 2004–2013, the 454 firms publicly listed in the S&P 500 Index spent $3.4 *trillion* on stock buybacks, or a total averaging $340 billion a year.[7] Often these were done at the expense of carrying out the kind of productive investment that would put the corporation on a solid track to future competitiveness. Many economists and business analysts, as well as stockholders, have found this trend deeply troubling. Some call for a return to the old rule outlawing stock buybacks; others propose alternative solutions.

---

[6]AFL-CIO, "100 Highest-Paid CEOs," available at http://www.aflcio.org/Corporate-Watch/Paywatch-2014/100-Highest-Paid-CEOs. Where 2015 pay totals were not yet available, 2014 pay totals were used.
[7]William Lazonick, "From Retain-and-Reinvest to Downsize-and-Distribute," Center for Effective Public Management, The Brookings Institution, April 2015, available at http://www.brookings.edu/research/papers/2015/04/17-stock-buybacks-lazonick.

An example is what happened when in 2004 Congress approved the tax holiday mentioned above, so that TNCs could bring back the profits they were holding abroad, and pay only about one-seventh of the taxes they would normally have to pay on the funds. Advocates of the legislation had sold it to Congress on the grounds that the repatriated funds would be used for productive investment that would create jobs and make voters happy. Numerous firms did take advantage of the arrangement in 2005 and brought back hundreds of billions of dollars. But a Congressional Research Service study found that some of the corporations that took the most advantage of the tax holiday laid off thousands of workers rather than creating jobs. Other academic studies found that in the same year that firms brought their profit earnings back, they also spent substantial amounts on stock buybacks. (By law, the funds themselves were not allowed to be used for stock buybacks, but accountants worked their magic and recorded the transactions in a way that was evidently technically legal.)

Of course, in principle any company's CEO is expected to run the firm in a way that increases the firm's value, and so benefits the stockholders, even without any exceptional incentives to do so. Linking CEO pay to stock prices, therefore, should in principle just be icing on the cake. Instead, in many corporations it has led to massive use of earnings for stock buybacks and dividend payouts, leaving little or nothing—and in some cases less than nothing—for investment in research and development, for updating workers' skills or acquiring new machinery, indeed for *anything* that would improve the product or the method of producing it. Of the twenty-five corporations that had the largest value of stock buybacks in the decade 2004–2013, for fifteen of them the earnings spent on such buybacks, plus the earnings paid out in dividends to shareholders, added up to *more than 100 percent* of total earnings in that decade so that in effect none of current earnings were available to spend on productive investment. And of the other ten, five firms had no more than 20 percent of their earnings left for productive investment after buybacks and dividends.[8]

During past SSAs, as explained earlier in this chapter, capital owners were compelled to invest in order for their firms to survive. But in recent decades the link between competition and investment seems to have been broken, at least for many firms in the United States, undermining the effectiveness of the SSA in supporting accumulation of capital.

Conclusion

A capitalist economy normally has a certain intrinsic dynamism. Competition compels firms to innovate, to invent new devices and methods for producing existing products better and more cheaply, and to invent new products. Developments in smartphones and apps alone in recent years testify to the power of capitalism to promote innovation. Yet a viable social structure of accumulation is also needed to sustain a healthy rate of investment and a dynamic economy. In the United States, the current SSA has relied for its dynamism on innovation in computers, programming, electronic devices, telecom, and related sectors, as well as on creation of new types of financial assets and earning strategies. The ability of these strategies to sustain accumulation appears to be decaying. At the same time, inequality of income and wealth means that markets for mass consumer products are growing only slowly, or even shrinking. The current SSA appears to be in a process of decay, and the next SSA has not yet emerged.

[8]Lazonick, "From Retain-and-Invest."

Suggested Readings

Joyce Appleby, *Capitalism and a New Social Order: The Republican Vision of the 1790s* (New York: New York University Press, 1984).

Samuel Bowles and Herbert Gintis, *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life* (New York: Basic Books, 1976).

Alan Dawley, *Class and Community: The Industrial Revolution in Lynn* (Cambridge, MA: Harvard University Press, 1976).

Eric Foner, *Thomas Paine and Revolutionary America* (New York: Oxford University Press, 1976).

Philip S. Foner, *Organized Labor and the Black Worker* (International, 1976).

David M. Gordon, Richard Edwards, and Michael Reich, *Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States* (Cambridge: Cambridge University Press, 1982).

Greg Grandin, *The Empire of Necessity: Slavery, Freedom and Deception in the New World* (Picador, Reprint edition, 2015).

Thom Hartmann, *Unequal Protection: The Rise of Corporate Dominance and the Theft of Human Rights*, 2nd edition (San Francisco: Berrett-Koehler, 2009).

Morton J. Horwitz, *The Transformation of American Law, 1780–1860* (Cambridge, MA: Harvard University Press, 1977).

Nancy Isenberg, *White Trash: The 400-Year Untold History of Class in America* (Viking, 2016).

Peter Lindert and Jeffrey Williamson, *Unequal Gains: American Growth and Inequality since 1700* (Princeton: Princeton University Press, 2016).

Grant McConnell, *The Decline of Agrarian Democracy* (Berkeley: University of California Press, 1953).

Terrence McDonough, Michael Reich, and David M. Kotz (eds), *Contemporary Capitalism and Its Crises: Social Structure of Accumulation Theory for the 21st Century* (Cambridge, UK: Cambridge University Press, 2010).

Samuel Rosenberg, *American Economic Development Since 1945* (New York: Palgrave Macmillan, 2003).

C. Vann Woodward, *The Strange Career of Jim Crow* (New York: Oxford University Press, 1957).

Erik Olin Wright, *Class Counts: Comparative Studies in Class Analysis* (Cambridge: Cambridge University Press, 1997).

Howard Zinn, *A People's History of the United States* (New York: Harper & Row, 1990).

Michael Zweig, *The Working Class Majority: America's Best Kept Secret* (Ithaca, NY: ILR Press, 2011).

Supply and Demand: How Markets Work



Ibn Battuta, the great fourteenth-century Arab geographer, reported that long-distance trade occurred as follows along the Volga River in what is modern-day Russia:

> Each traveler . . . leaves the goods he has brought . . . and [the travelers] retire to their camping ground. Next day they go back to . . . their goods and find opposite them skins of sable, miniver, and ermine. If the merchant is satisfied with the exchange he takes them, but if not he leaves them. The inhabitants then add more skins, but sometimes they take away their goods and leave the merchant's. This is their method of commerce. Those who go there do not know whom they are trading with or whether they be *jinn* [phantoms] or men, for they never see anyone.[1]

Herodotus, the ancient Greek historian, describes similar exchanges—called *silent trade*—between the Carthaginians and the people of Libya in the fifth century BC. After having left their goods, Herodotus reports, the Carthaginians withdrew, and the Libyans "put some gold on the ground for the goods, and then pull back away from the goods. At that point the Carthaginians . . . have a look, and if they think there is enough gold to pay for the cargo they take it and leave." Herodotus describes how the process continued until an acceptable price was hit upon, remarking with surprise that "neither side cheats the other . . . [The Carthaginians] do not touch the gold until it is equal in value to the cargo, and the natives do not touch the goods until the Carthaginians have taken the gold."[2]

Silent trade is but one of the many ways that people have devised to engage in the process of exchange. Transfers of goods among strangers can range from gifts at one extreme, through mutually advantageous exchanges, to what might be called plunder at the other extreme. The potential gains from trade are often greater, the more distant geographically or socially are the parties to the exchange. The fact that the parties to a silent trade did not meet face to face helped to reduce the chances of outbreaks of violence among the often heavily armed traders.

---

[1]Ibn Battuta, *Travels in Asia and Africa: 1325–1354* (London: Routledge and Kegan Paul, 1929), 151.
[2]Herodotus, *The Histories* (New York: Oxford University Press, 1998), 300–301.

Other kinds of trade are anything but silent. The trading floor of one of the world's stock markets is a din of offers and bids, and a fruit market in modern Nigeria resounds with the almost musical call of market women advertising their wares and the hum of haggling over prices. Other modern markets are as silent and anonymous as the silent trade. When you buy a shirt or a book online, the only sound you hear is the mouse click when you hit the shopping cart icon on your screen. Similarly, you can buy an entire basket of groceries at a supermarket without saying a word, and, in contrast to the Nigerian fruit market, the only need for verbal communication occurs when the checker asks if you want plastic or paper bags for your goods! If you buy an item online through an auction market (such as e-Bay), you will experience an entirely different kind of market: there you will compete with others in posting a price for a good, the sale going to the highest bidder.

Goods and services are exchanged in many different ways. Families exchange gifts at holiday times, individuals work at jobs in return for money, and one member of a couple cares for the kids while the other works for the wages that pay the rent. In each case, who gets what in return for what is determined in a particular way, sometimes by custom, sometimes by law, and sometimes by the competitive forces of supply and demand. The main idea of this chapter is that *competitive markets—an important form of exchange—can be analyzed using the concepts of supply and demand.* This idea is expressed in two key points:

1. *Competitive markets* consist of many potential buyers and sellers, each acting independently, with no one participant having enough power to dictate terms to any other.
2. The key concepts in understanding how competitive markets work are *supply curves*, *demand curves*, and *market-clearing.*

The Nature of Markets

A **market exchange** is a transfer to another party of the title to a good, or of the right to receive a service, in return for some form of payment on mutually acceptable terms.

A **market** refers to all the buying and selling activities of those persons willing to trade a good or service as long as the price is within an acceptable range; a market consists of suppliers potentially wanting to sell and demanders potentially wanting to buy.

A **market exchange** occurs when the owner of a good or service sells it to someone else. Selling it usually means that it is exchanged for money: the seller transfers ownership of the item to the buyer and receives money in return. The other side of the transaction is that the buyer pays money to acquire ownership of the good or service. If money is not involved in a transaction—that is, if one product or service is traded directly for another product or service—this is called *barter.*

A market exchange is thus a transfer to another party of title to a good, or the right to receive a service, in return for some form of payment on mutually accepted terms. To say that the terms of a market exchange are mutually accepted means that under the circumstances, both parties would rather make the exchange on the terms that are being offered than make no exchange at all.

People always have the right to refuse a market exchange. Therefore, by definition, all market exchanges that take place can be said to be *voluntary* in nature. But sometimes the right to walk away from an exchange does not mean much. For example, in principle a severe illness in your family may force you to sell much of what you own to pay the medical bills. It is not that someone is holding a gun to your head and telling you to sell your possessions, but sometimes circumstances can force people into exchanges they would otherwise avoid. (Recall Marlon Brando in *The Godfather* saying to one of his henchmen: "Make him an offer he can't refuse.")

The term **market** refers to the buying and selling activities of all those who are interested in trading (buying or selling) a particular good or service if the price is

within an acceptable range. Market activities are sometimes, but not always, concentrated in one location. Examples of markets with specific locations are the New York Stock Exchange and the Tokyo fish market. For other markets, however, there is no single specific place where you can "see" the market. For example, the Chicago labor market includes all the potential buyers and sellers of labor time who are meeting and coming to terms (or not coming to terms) anywhere in the Chicago area. A market, then, is not a place but rather a set of buying and selling activities.

Markets work to determine two basic economic outcomes: the *price* at which a good or service is exchanged and the *quantity* of it that will be bought and sold. These two outcomes affect many other aspects of society. The labor market, for example, determines not only wages (and hence living standards) but also the amount of employment (and hence also the number unemployed).

Each market has two types of participants: *demanders*, or those wishing to buy the good or service over some range of prices, and *suppliers*, or those wishing to sell the good or service over some range of prices. A market may comprise, say, two potential demanders facing three potential suppliers. This might occur in a local real estate market, if three people wish to sell their homes, and two other people are independently interested in buying any one of the three. Or the market might have a small number of suppliers and millions of demanders, like the market for computers. Some markets have thousands of suppliers and just a few demanders. An example is the labor market of a town with just a few large employers and considerable unemployment: many people will be seeking work (supplying their labor), but only the very few employers will be offering jobs (demanding labor).

In this chapter we focus on markets with large numbers of potential demanders and suppliers. Following in the tradition of Adam Smith, such markets are termed **competitive markets** since the rivalry of the different participants—each one competing to make an advantageous purchase or sale—greatly affects the actions of all the others. Many markets in the United States and throughout the world are not competitive in this sense. We explain the workings of markets with smaller numbers of competitors in Chapter 11.

Competitive markets are those with many actual or potential demanders and suppliers.

The most important consequence of having large numbers of participants in competitive markets is that no one of them is powerful enough to influence the price at which goods will sell. If there were just one seller, for example, a large corporation, it could gain a higher price for its product by making less of it available. But this strategy is ruled out in competitive markets.

Demand and Supply

We can understand how markets work by looking at the interaction of demanders and suppliers. We do this with the help of demand curves and supply curves.

Demand

A *demand curve* is a graphical representation of the buyers' side of the market. It shows, for a given time period, such as a day, how much of a particular commodity the demanders of this product will want to purchase at each possible price, given their taste for the product and the amount of money they have at their disposal. Each point on the curve represents a particular combination of a price (measured on the vertical axis) and the corresponding quantity demanded (measured on the horizontal axis).

A **demand curve** indicates, for each possible price, how much of the good or service demanders are willing and able to buy.

Consider, as an example, the market for a certain item, say, beer in Iowa City, Iowa. (In the remainder of this chapter we assume that the beer referred to is all of the same type and quality; in the jargon of economics, we assume that beer is a *homogeneous* commodity.) Imagine that we asked every person in Iowa City (and all those who might travel to Iowa City to buy beer), "How many bottles of beer would you buy today if the price were $2 per bottle?" We would then add up all the answers. If the total came to 1,040 bottles, we would have one point on the demand curve: at a price of $2, buyers will demand 1,040 bottles on this day.

We might then repeat the survey, asking buyers how many bottles they would buy, first, if the price were $1 and, second, if it were $0.50 per bottle. Suppose we obtained answers of 2,000 bottles at a price of $1 and 3,760 at $0.50. We would then have two more points on the demand curve for beer in Iowa City on this day.

In Figure 8.1 the demand curve D shows the various quantities of beer that buyers in Iowa City will demand on a certain day at all the possible prices, including the prices of $2, $1, and $0.50 for which we obtained answers in our survey. It is important to remember that both the demand curve and the supply curve present answers to *hypothetical* questions. In the case of the demand curve, the question is, "If the price were to be _____, what quantity would you buy today?" As we will explain shortly, most of the combinations of price and quantity on the demand curve and the supply curve will not actually be chosen.

Demand curves are almost always thought of as sloping downward to the right, or having a negative slope, as does D in Figure 8.1. The economic reason for this is

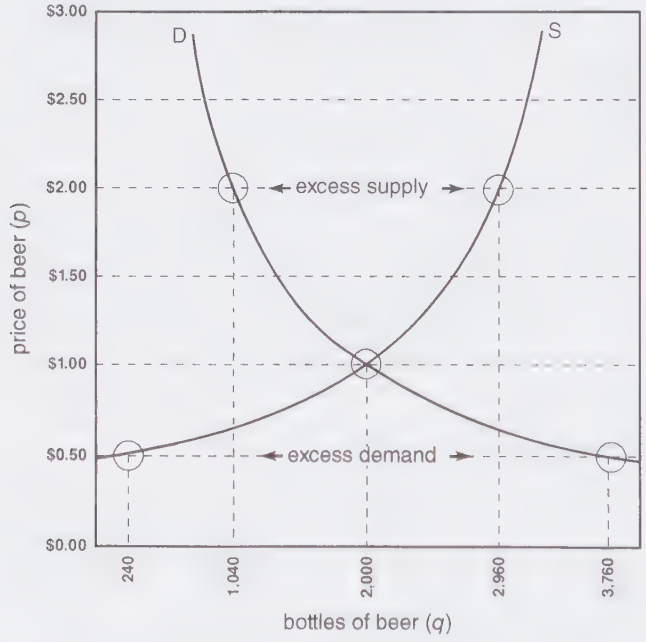


FIGURE 8.1 The demand curve and the supply curve.

The demand curve D provides the following information. If the price of beer per bottle is p dollars, the amount demanded by consumers will be q bottles a day. For example, if the price is $2 a bottle, consumers will buy 1,040 bottles a day. The supply curve S provides similar information about what producers will offer on the market. If the price is p dollars per bottle, suppliers will offer q bottles per day on the market. For example, if the price is $2 a bottle, beer suppliers will want to sell 2,960 bottles a day. Normally, demand curves slope downward (to the right) and supply curves slope upward.

that, in general, the lower the price, the more of the good buyers will want to buy. In our beer example (Figure 8.1), if the price is high, say, $2 per bottle, then consumers will want to buy relatively few bottles per day. If the price is low, say, $0.50 per bottle, they will buy a larger quantity each day.

Note that the demand curve does not represent what buyers *need*. It reflects only what they *want* and are *able to purchase*, given the *price* and their *incomes*. Adam Smith defined what is reflected in a market as "effectual demand" (saying that a poor man's wish for a carriage drawn by six horses would not be an effectual demand), and modern economists have followed in his footsteps, defining "demand" simply as a want backed up by money.[3] We cannot tell from D whether the buyers of beer are desperately thirsty after performing arduous labor or whether they have already had more beers than they should have had. Indeed, if there are some people in Iowa City who desperately want beer but have no money to buy it, their wants will not be expressed in D at all, since the demand curve expresses only what people are willing and able to buy. All the demand curve tells us is how much beer consumers will buy at any given price.

How much beer people want to buy will depend on many considerations besides the price. As we have noted, it will depend on the buyers' incomes. If everybody suddenly got a pay raise, people might want to buy more of many things, including beer. Another factor is the number of potential buyers. If new people came to town, that would tend to increase the demand for beer. More importantly, demand will change with the strength of people's desire—or, as economists put it, their "preference"—for a product. Thus, a consumer in Iowa City may want to buy more beer if she recently watched an effective beer ad on TV, or if the weather has just turned hotter. A final variable is the price of a competing product. For example, people might buy less beer if the price of wine suddenly dropped. Such products are termed *substitutes*.

A demand curve, then, expresses how much the buyers are willing and able to buy at the various possible prices, *assuming* that nothing else affecting their demand changes. If nothing else changes, the demand curve allows us to say how a change in the price will affect the quantity demanded. For instance, in Figure 8.1 we can see that if the price drops from $2 to $1 per bottle, *and if nothing else changes*, the quantity demanded will rise from 1,040 to 2,000 bottles a day.

Supply

A **supply curve** indicates, for each possible price, how much of the good or service suppliers wish to sell.

The **supply curve**, by contrast, represents the sellers' side of the market. It depicts the suppliers' willingness to sell beer at different prices, and this willingness to supply beer will depend to a large extent on the costs incurred by beer-producing firms. In Figure 8.1 the supply curve S shows, for a particular day, what quantity of beer sellers will supply to the market at various prices. For instance, if the price were to be $2 on some day, suppliers would try to sell 2,960 bottles that day; at a price of $0.50, suppliers would try to sell 240 bottles; and so on.

Supply curves are almost always thought of as sloping upward to the right (having a positive slope), as does the supply curve S in Figure 8.1. When prices are high, suppliers will want to sell a lot of beer compared to when prices are low. At higher prices it will pay to put on extra shifts of workers at the brewery. The high price may also attract new suppliers from nearby cities where sellers are not able to get such a high price. When the price is low, on the other hand, some suppliers in Iowa City may try to find other cities in which to sell their product; some of them might even stop producing beer.

---

[3]Adam Smith, *Wealth of Nations*, Book I, Chapter VII, eighth paragraph.

Other factors besides price influence how much of a product suppliers will want to sell at each price. The cost of producing beer compared with the rewards available from producing and selling other items will affect how much beer the suppliers will want to sell at each price. For example, if a labor-saving technical change occurs in the brewery industry, the cost of producing beer will fall, and the profits to be made in brewing and selling beer will rise. As a result more firms will be attracted into the industry and the amount supplied will increase. Similarly, if the grain used in making beer becomes more abundant, its price will drop, the costs of brewing will fall, and the quantity of beer supplied at each price will increase.

A supply curve, then, represents the quantities of a commodity that sellers will supply to the market at various possible prices, assuming that everything else affecting its supply remains unchanged. If nothing but the price changes, the supply curve tells us how the quantity of the good supplied will change when the price changes. In Figure 8.1, for example, if the price of beer rises from $1 to $2 per bottle on a certain day, suppliers will want to increase the amount they offer in the market that day from 2,000 to 2,960 bottles.

Some of the key factors determining the position of the supply curve include the technologies available for producing the good; the costs of the various inputs and how they vary with the amounts of the inputs used; and the number of firms producing the good, which may vary as the profitability of producing and selling beer varies relative to alternative business activities. (In a later section we discuss in detail what causes the supply curve to shift.) You may want to test your understanding of the supply curve by asking yourself how changes in any one of the above factors will shift the supply curve in Figure 8.1 either to the right or to the left. For example, what would happen to the supply curve if a new technology for producing beer were to be invented, one that would allow the same quantity of output to be produced with less labor?

Marginal cost is the increase in the total cost incurred by a firm when it increases its total output of a commodity by one unit.

MARGINAL COST All the influences that affect the position of the supply curve combine also to determine the **marginal cost** of a commodity's production. "Marginal cost" is defined as the *increase in the total cost incurred by a firm when it increases its total output of a commodity by one unit.*

To grasp the significance of the concept of marginal cost consider the following step-by-step reasoning: Small movements up a supply curve show, at each point along the curve, how much the price of a good must be increased to induce some supplier of that good to produce and offer for sale in the market one additional unit of the good. Even when the market price of something goes up, however, we know that no firm will produce an additional unit of the good in question unless the higher price covers the cost of producing an additional unit of it. Since the cost to a firm of producing an additional unit of a good is, in fact, its marginal cost, we can say that a supply curve shows not only the amount of a good that will be supplied at each price but also what the marginal cost of the good is for at least one firm in the market.

Of course, what happens in a market is not based on the decisions of a single firm. Indeed, the amount supplied in a market on any particular day will be the amount that results when the outputs of all the firms in the market are added together. The point here is simply to establish the idea that there is a marginal cost of producing and offering for sale one additional unit of a good. This concept will be important in Chapter 9, where we discuss the circumstances under which the market by itself produces what is best for society.

Average cost is the total cost of producing a certain number of units of a good or service divided by the total number of units produced.

AVERAGE COST The marginal cost of producing a product differs from the **average cost** of its production, which is defined as *the total cost of producing a certain number of units of a good or service divided by the total number of units produced.* While, at a particular level of output, the marginal cost is how much would be added to the total cost by producing just *one* additional unit, average cost is a measure of the cost of producing *all* of the units being produced at any given time.

In many cases both the marginal and the average costs of producing a commodity *rise* as more units of it are produced (beyond some minimum level of output). This is because, as production expands, crowding of existing facilities can raise the marginal cost of labor, making it more expensive to produce additional units of the commodity. For this reason the marginal cost curves for many commodities will be upward sloping to the right on a graph that has the number of units of output (per time period) on the horizontal axis and the corresponding marginal cost for each level of output on the vertical axis.

The reason average and marginal costs may rise with increasing output is easiest to see in agriculture or other natural resource–based industries. In these cases there is a limited amount of good land (or easily exploited natural resources), so production will be more costly at higher levels of output. This is because poorer quality land—or deeper mines or more remote forests—must be used.

Economies of scale exist when an increase in the number of units of output produced—an increase in the *scale* of production—brings about a fall in the average cost, that is, a fall in the cost per unit of output.

However, there are also situations in which average cost *falls* as more is produced. Such a fall in average cost will result whenever there are **economies of scale**. Economies of scale exist when an increase in the number of units of output produced—an increase in the *scale* of production—brings about a fall in the average cost of a unit of output. Thus, as explained in Chapter 3, economies of scale produce *decreasing average costs* (as the quantity of output increases), and the two terms, since they refer to the same phenomenon, can be used interchangeably. To avoid confusion, however, we will generally use the term *economies of scale*, or *scale economies.*

Situations in which firms experience falling average costs (because of economies of scale) are common and important throughout the economy. Surprisingly, however, this familiar situation—in which average costs are falling—is discussed only briefly, if at all, in conventional economics textbooks, and we will see why in the next chapter. But first we must carry the discussion of supply and demand a bit further.

Demand and Supply Interacting

We can now join the two strands of the story by explaining how supply and demand together will determine both the price of a good and the amount of it that will be traded (the *price* and the *quantity*). It will be useful here to refer once again to the hypothetical beer market in Iowa City and in the process to consult Figure 8.1 one more time (since it combines in a single graph both the supply curve and the demand curve for beer in Iowa City).

Of course, neither the buyers nor the sellers see the supply and demand curves. These are just analytical tools that *we* use to understand what *they* do. In most markets the sellers, not the buyers, set prices. Each seller sets a price assuming that a higher price will mean more profits per unit sold and that a lower price will mean more units sold. Depending on the demand curve, a move in either direction (that is, moving the price higher or lower) might increase the total amount of profits.

The price that maximizes profits will depend on what the other sellers are doing and how strong the demand is, two pieces of information about which a seller can guess

but cannot know in advance. The one thing that sellers do know and may act on is what happened in the recent past. If goods have piled up on their shelves, they may wish that they had been charging a lower price, and they will most likely consider reducing the price. Conversely, if they sold out before the end of the day (or the month), or if they have accumulated a waiting list of eager customers, they probably will be considering a price hike. Other sellers are engaged in the same trial and error method to get the price right.

To see how this process works, look at Figure 8.1 again and suppose that the average price of beer in Iowa City is $2 per bottle. What will happen? As pointed out earlier, suppliers will wish to sell 2,960 bottles at this price, but demanders will want to buy only 1,040 bottles. The difference of 1,920 bottles is referred to as **excess supply**. Those suppliers who can find buyers at $2 a bottle will be happy, while those who cannot find buyers at this price will be dissatisfied, and the second group will then offer slightly lower prices, say, $1.75 or $1.50 per bottle, in an effort to attract customers.

Excess supply exists at a particular price when at that price more of some good or service is supplied than is demanded.

As long as excess supply persists, some suppliers will cut their prices in order to try to gain customers, and this will exert downward pressure on the market price. When the prevailing price has fallen to $1 per bottle, the quantity that suppliers wish to sell (2,000 bottles) will be just equal to the quantity demanders wish to buy (also 2,000 bottles), hence there will be no more excess supply. Price cutting by suppliers will therefore stop at this price.

Now consider the opposite situation. If the initial price in the market had been $0.50 per bottle, there would have been **excess demand**. As noted earlier, the demand for beer at this price would be 3,760 bottles, but the supply would be only 240 bottles, leaving an unmet (or "excess") demand of 3,520 bottles. The supply of beer would fall short of the demand by this amount, and those unable to buy would tend to bid up the price. Excess demand would be eliminated only when the market price reached $1, which, of course, is the same figure we arrived at in the analysis of excess supply.

Excess demand exists when at a particular price more of some good or service is demanded than is supplied.

Summing up, we can say that competition in the market for beer pushes the whole market toward a **market-clearing price**—the price at which sellers want to sell exactly the quantity that demanders want to buy. At such a price neither excess supply nor excess demand will exist, and the market is said to "clear." In Figure 8.1, the market-clearing price in the Iowa City beer market is $1, for at this price the quantity of beer supplied is precisely equal to the quantity demanded (2,000 bottles).

The **market clearing price** is the price at which buyers want to purchase exactly the quantity that sellers want to sell.

Figure 8.1 also shows that the market-clearing price and quantity are located at the *intersection* of the supply and demand curves. For this reason—and, as we shall see, only in markets similar to the one for beer in this example—it can be said that supply and demand *determine* both the quantity sold and the price at which it is sold, meaning, more precisely, that the particular positions of the supply and demand curves (and, of course, the factors that themselves determine the positions of these curves) determine the market-clearing price and quantity.

In a market such as the beer market, market clearing is often described as an **equilibrium** situation, and the concept of equilibrium is important in economic reasoning. It is used to describe *a situation in which there are no forces internal to the situation pushing it to change.* This concept is borrowed from physics, and it can be illustrated with a physical example: if one drops a marble into a bowl that is sitting on a table, the marble will roll around for a while, eventually stopping at the bottom of the bowl. The result is an equilibrium, for nothing internal to the situation (the location of the marble in the bowl and the shape of the bowl) will cause it to change. If one were to tilt the bowl or push on the marble, of course the marble would move, but these would be forces external to the situation.

Equilibrium refers to a situation—a price and quantity exchanged—in which there are no forces internal to the situation pushing it to change.

Economists reason the same way. In the beer market the price and the quantity sold will remain at the market-clearing price and quantity until something from the

outside changes them. A change coming from the outside would be something like an increase in people's taste—or, as economists like to say, their "preference"—for beer relative to other things such as wine. Such a change would bring about a rightward shift of the demand curve D, as will be more fully explained in the next section. Similarly, the adoption of a new technology for producing beer that reduces the cost of producing a bottle of beer would shift the position of the supply curve S downward. The resulting movements in these curves would change the market-clearing price and the amount of beer sold. But as long as the demand and supply curves remain in their present positions, the equilibrium price and quantity will not change.

It is a remarkable fact that that there are some markets, including some very important ones, in which the equilibrium price is not the same as the market-clearing price. This means that in equilibrium the demanders have reasons why they do not offer a higher price than the current one, and suppliers likewise do not offer to sell at a lower price than the current one, and yet the quantity offered for sale is not the same as the quantity actually purchased. Nevertheless, the stability of an equilibrium situation (barring the appearance of any external, or *exogenous*, source of change) means that nothing an individual buyer or seller may try to do can change the equilibrium price or quantity.

In our Iowa City beer market example, the equilibrium price and quantity *are* the same as the market-clearing price and quantity. In that situation, none of the buyers or sellers of beer in Iowa City can benefit from any possible change in their behavior, given what all the other market participants are doing. For example, a buyer might like to pay less than the going price for a bottle of beer. But if such a buyer tried offering a lower price, no supplier would sell her or him any beer. Similarly, a supplier might like to sell beer at a price higher than the market-clearing price. But if any company raised its price, its sales would fall drastically, since similar beer would be for sale in the market at a lower price and this particular company's customers would switch to other suppliers, especially if the company persisted in charging a higher price for a long period of time. (Recall that throughout this chapter we are assuming that any bottle of beer in the Iowa City market is exactly the same as any other bottle; although this assumption is somewhat unrealistic in this case, there are other commodities, such as wheat, corn, and milk, that are more like the *homogeneous* product in our example.) Thus, the prevailing market price (since it is available to anyone in the market) limits what any individual buyer or seller can do. This is the way a competitive market works.

In reality, of course, individual sellers will try changing their prices to see if they can do better. Even when there is no excess supply or demand, therefore, not all prices of a good will be the same. This may be confirmed by pricing beers at a few local stores or by checking the price of a book both at Amazon.com and Barnesand-Noble.com. But it is not likely that prices of the same good will differ very much if there is a high degree of competition in the market for that particular good.

An important result of the analysis of the interaction of supply and demand in competitive markets is that *when a competitive market is in equilibrium, the price of the good will be equal to its marginal cost.* Another way of putting this is to say that in equilibrium $P = MC$ (where P stands for the price and MC the marginal cost). We will have more to say about this idea later, but the logic of it is as follows: If P is *not* equal to MC the amount supplied will change, so the market cannot be in equilibrium. To see this, imagine that P is greater than MC for some firm. A firm in this situation can gain by producing one additional unit, increasing its revenue by P at a cost of only MC. Similarly if P is less than MC for some firm, that firm can gain by producing one unit less (reducing its costs by MC but reducing revenues by only P). So the amount supplied by a firm will not change only if $P = MC$. Moreover, for the

market as a whole to be in equilibrium (recall the definition of an equilibrium) P will have to be equal to MC for every firm in the market.

Shifts in Demand or Supply

So far we have been considering how price and quantity are determined when the supply and demand curves are in a particular position. We have looked at each curve and asked how the quantity demanded or the quantity supplied would change in response to a certain change in the price. We will now consider what happens if something else besides the price changes.

For example, suppose that the demand curve D in Figure 8.2 represents the demand for beer in the middle of a certain semester at the University of Iowa. As the semester comes to an end and students leave campus for home, the situation will change, and smaller quantities of beer will be demanded in Iowa City at every possible price. This change is represented by a leftward *shift* of the whole demand curve from D to D* (see Figure 8.2). On the other hand, a heat wave during the semester would have the opposite effect: it would shift the demand curve rightward by bringing about an increase in consumers' preferences for beer.

The demand curve plots how each price level determines the quantity demanded—but in a situation that is also determined by other factors that are not shown on the diagram.

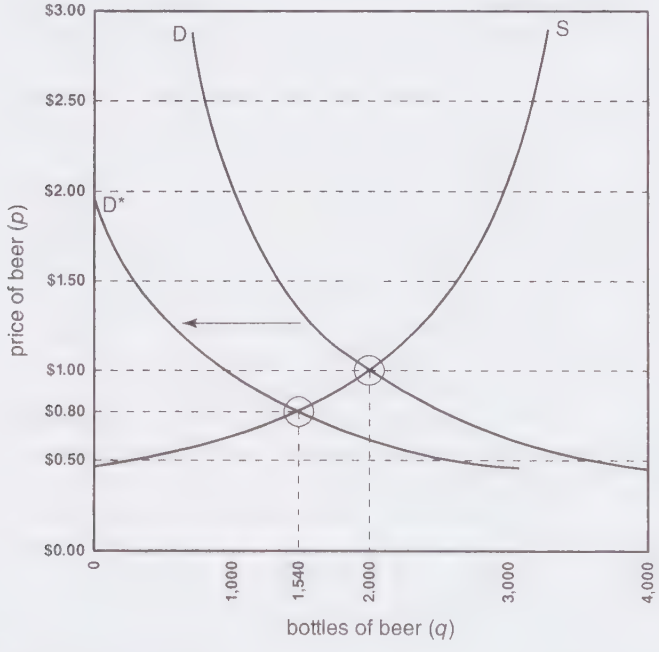


FIGURE 8.2 A shift of the demand curve.

When something other than the price changes, there will be a shift of the demand curve, either to the right or to the left. A *shift* of the demand curve is different from a *movement along* the curve, which is what happens when only the price changes. In this figure, D* shows the position of the demand curve after it has shifted to the left, indicating that there is now less demand for beer at every price. It is also possible that, with a different change in one of the influences on the demand curve, it would shift to the right.

Thus a change in the price alone produces a *movement along* the demand curve, whereas a change in one or more of the conditions underlying the demand for the product (and therefore the location of the demand curve) produces a *shift* of the demand curve. As can be seen in Figure 8.2, the shift of the demand curve from D to D* changes the market-clearing price from $1 to $0.80 per bottle and the quantity sold from 2,000 to 1,540.

Similarly, the supply curve for beer will shift if there is a change in something affecting supply other than the price. Suppose, for instance, that the price of the grain used in making beer rises. What will happen? The additional cost will reduce profits in the beer industry, and this will lead some suppliers to withdraw from the beer market, perhaps to utilize their resources in a more profitable industry. The higher costs will also cause other beer producers to scale back their operations in the hope that this will restore their previous level of profits. These two developments will shift the whole supply curve from S to S* (see Figure 8.3). As can be seen in Figure 8.3, this shift in the supply curve, with D unchanged, will cause the market-clearing price to rise from $1 to $1.20 per bottle and the quantity sold to fall from 2,000 to 1,680.

As with the demand curve, a change in the price of the good creates a movement along the supply curve, whereas a change in something else affecting supply causes a shift of the entire curve. Shifts of a supply or a demand curve occur whenever one of the determinants of these curves' positions changes—whether it is, in the case of our beer market example, a new brewing technology, cheaper grain, a successful advertising campaign, or a change in the size of the relevant population. The determinants of the positions of supply and demand curves are summarized in Table 8.1.

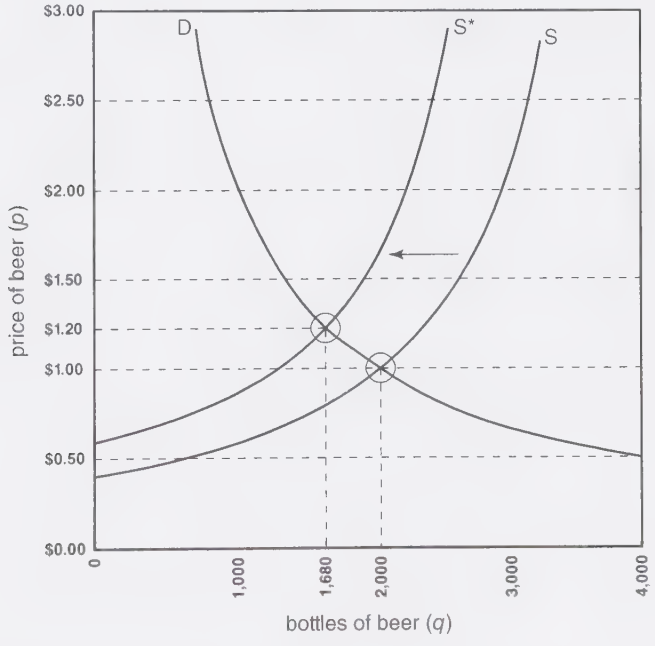


FIGURE 8.3 A shift of the supply curve.

When something other than the price changes, there will be a *shift* of the supply curve, either to the right or to the left. A shift of the supply curve is different from a *movement along* the curve, which is what happens when only the price changes. In this figure, S* shows the position of the supply curve after it has shifted to the left, indicating that there will now be a smaller amount of beer supplied at every price. It is also possible that, with other changes in one of the influences on the supply curve, it would shift to the right.

TABLE 8.1. DETERMINANTS OF THE POSITIONS OF SUPPLY AND DEMAND CURVES

THE SUPPLY CURVE	THE DEMAND CURVE
Technology	Consumers' tastes or "preferences"
Costs of inputs, including the costs of obtaining the necessary permissions to use patented ideas, copyrighted material, etc.	Consumers' incomes and their distribution: more income at the high end will mean more demand for luxury goods; more at the low end will mean more demand for basics
Alternative opportunities for profit available to suppliers	Number of alternative products, and their prices, available to consumers
Number of potential producers	Number of potential buyers

Conclusion

Markets provide a way of coordinating economies by means of buying and selling, or exchange, relationships. Competition, both among buyers and among sellers, tends to produce a market-clearing price at which the quantity demanded equals the quantity supplied. The result is that market interactions determine both the prices of goods and the quantities of them that will be bought and sold.

But what does it mean to say that the equilibrium price of beer and the quantity of it that is sold are "determined by supply and demand"? This is a little like saying that a murder was committed by the gun. What actually determined the price and the quantity was whatever determined the positions of the supply and demand curves, because these determinants are what made a particular combination of price and quantity necessary to clear the market.

Supply and demand curves themselves do not do anything: they are even less involved than the gun. Rather, they are ingenious devices that help us understand and explain the many and complex influences on prices and quantities. The invention of supply and demand curves did not change the way markets work. However, it did lead to a better understanding of how markets function.

To summarize: Price and quantity are determined by all the factors that determine the positions of both the supply curve and the demand curve. The process of competition, in turn, works to make actual prices and quantities move toward equilibrium prices and quantities.

As we will see in later chapters, competition does not always cause markets to clear. There are some kinds of markets in which something like equilibrium prevails, and yet there is a persistent gap between the quantity demanded and the quantity supplied. In the labor market, for example, excess supply of labor (unemployment) can persist for long periods of time or even indefinitely. In Chapter 12 we explain exactly how the labor market differs from the market for beer, so that reductions in wages do not have the effect of clearing the labor market (that is, ensuring that everyone who wants a job has one).

When markets do not clear, supply and demand will influence the price and the quantity but can hardly be said to determine them, for there is obviously some other important influence at work if the market is not clearing. Our next chapter extends the analysis of supply and demand, explaining how markets may, under certain circumstances, coordinate the economy in beneficial ways, while, under other circumstances, they may fail to do so.

Suggested Reading

Alfred Marshall, *Principles of Economics*, 8th ed. (London: Macmillan, 1920).

Competition and Coordination: The Invisible Hand



In the late 1980s the governments of Poland, the Soviet Union, Czechoslovakia, and other Communist countries took a dramatic turn. Not only did they begin the process of granting voting rights and democratic freedoms to their citizenry, they also began to place more emphasis on markets rather than centralized economic planning as a way of coordinating their economies. The transition to market-oriented, mostly capitalist economies continued through the early 1990s, while the Soviet Union broke up into Russia and many other countries, Czechoslovakia divided into the Czech Republic and Slovakia, and Yugoslavia also fragmented into many small countries.

For more than seventy years in the Soviet Union and for forty years in the other countries, the government had made most important economic decisions. A system of economic planning had collected information and then implemented decisions on such issues as how much steel and other investment goods to produce, what technologies to use, how many schools to construct, what consumer goods to produce and how much of each, and how much workers, managers, doctors, and others should be paid.

These centrally planned economies initially had achieved major improvements in education, health, and other aspects of living standards, particularly for the less well-off members of their societies. But during the 1980s their governments began to implement major economic and political reforms in response to popular dissatisfaction not only with the lack of democratic rights but also with the slowdown or even reversal of economic growth. The most important economic reform was to allow private companies and individuals to make more economic decisions, and this meant relying on markets to coordinate millions of decisions.

The changes in economic approach in eastern Europe and the former Soviet Union dramatize the main idea of this chapter: *markets provide a way for individuals and firms to coordinate their many and complex economic activities, with no one in particular directing the process.* This idea is expressed in two main points:

1. By rewarding success and punishing failure, competitive markets provide a decentralized system of *motivation*; through market prices they also transmit *information* about the relative cost to society of various goods and services.

2. Under certain circumstances markets *coordinate* the economy in ways that are generally beneficial. They do this when the individual's decision to buy a good is closely linked to the total benefit for society of having that good produced and used. But under other circumstances markets fail to perform this function well, and we describe such situations as *market failures*.

Adam Smith and Laissez-Faire Economics

In Chapter 2 we introduced Adam Smith as the eighteenth century political economist and philosopher whose book *The Wealth of Nations* (1776) was the most influential book on economics ever written. Smith developed the idea that the self-interested actions of individuals would create a system of competitive markets translating into material progress of all society. Certainly Adam Smith would have been pleased with the reforms in the former Communist countries, since they followed his inclination to let markets, not the government, coordinate production. In *The Wealth of Nations* Smith identified a basic challenge to society: how to coordinate the independent activities of large numbers of economic actors—producers, transporters, sellers, consumers—often unknown to one another and widely scattered across the world. This problem of coordination arises because no person is self-sufficient: everyone's livelihood requires a multiplicity of goods and services produced by others. This is much more true today than it was in Smith's era.

Smith developed the idea—radical in his day—that society could leave the coordination of the division of labor up to the individual self-interest of the economic actors themselves. This idea was radical because it asserted that a rational order might arise without any person or institution consciously attempting to create or maintain that order. The English philosopher Thomas Hobbes, writing a century before Smith, had advocated a powerful government as the only way to make order out of the chaos of the self-seeking activities of large numbers of people. Smith argued that markets could do the job as long as two conditions were in place: First, property rights must be well defined, to make it clear who owned what property, and therefore had the right to exchange it. Second, there must also be enough competition among economic actors so that no market was monopolized.

Smith explained how a system of competitive markets could translate the self-interested actions of individuals into results beneficial for society. This was a truly radical claim at a time when selfish behavior was regarded as immoral. Smith argued that while the self-interested activity of the farmer, the baker, or the barber may be based on greed, the pursuit of that greed will, under the right conditions, benefit all, even if most people care little or nothing for the well-being of their neighbors. The right conditions to produce socially beneficial results are well-defined property rights and competitive markets. Smith coined the phrase *the invisible hand* to refer to the tendency of markets to guide the economy toward the best use of its human and natural resources.

The key to all this is that the price of goods is a signal to producers. If the price of bread rises (for example, due to an increase in demand—a rightward shift of the demand curve), bakers are signaled that baking bread is now more profitable than before, and if they bake more bread, they will earn more money. If the price falls (due, for example, to an increase in supply—a rightward shift of the supply

curve), that signals bakers to bake less bread. No other announcement, instruction, nudging, or command is needed.

In Smith's view, then, markets would coordinate and regulate the economy, harnessing self-interest to achieve material progress for the whole society. Therefore, he thought the government, instead of trying to direct the economy, should leave well enough alone. This is roughly the meaning of the French expression **laissez-faire**.

Laissez-faire is an approach to economic policy that advocates a very limited role for the government, confining its activities to national defense and the enforcement of laws and contracts.

Smith was partly right in one important sense and partly wrong in another. Under certain circumstances, letting prices guide economic activities through markets works fairly well. But in the more than two centuries since he wrote, we have learned that the circumstances under which markets work well are quite a bit narrower than he realized in the eighteenth century. We first look at how markets work when they work well, and then consider the conditions under which markets fail to work well.

Coordination

The "production-reproduction cloverleaf" presented in Figure 4.1 is like an aerial photograph of the economy. It represents the horizontal dimension of the economy as a complex circulation of goods and services—and people. Things produced in one location end up being used in another. People born and raised in the home leave and take up employment in factories, offices, or their own businesses; sometimes they work in their own or other people's homes. The circulation of goods, services, and people is called the horizontal dimension because from this perspective things and people move from one place to another, a motion which is horizontal in nature. Understanding the horizontal dimension of the economy means explaining such movements of things and people, answering questions such as: Why do some people work at certain jobs and some at others? What determines who and how many will raise children, pour concrete, or make shoes? How do the shoes get from the shoe worker to the shoe wearer?

In various societies over the course of human history and around the world today, questions like these have been answered in very different ways. Consider, for example, a self-sufficient family farm on the American frontier in the early nineteenth century like the one described at the beginning of Chapter 7. Here most of the labor processes and movements of products shown in Figure 4.1 took place *within* a single household. Both the production of necessary inputs and the reproduction of people all took place (more or less) under one roof. Tools were made and repaired, draft animals were tended, a new fence was put up, food was prepared, firewood was collected, children were born and raised, and clothing was made—all of which meant that the farm was largely *self-sufficient*. What products were produced and how they were used was coordinated by a combination of custom, necessity, and patriarchal authority. Tasks were assigned according to age and sex. Though not common today, this is one distinct way of determining who will do what labor and how the resulting output will be used.

If each family does not produce everything its members use (if the family unit is not self-sufficient)—and this is true in most of the modern world—the situation is much more complicated. Most families, or their working family members, will specialize in a particular type of work, and families will have to engage in some form of *exchange* to get the goods and services they need. Various social arrangements will determine how labor will be specialized and how outputs will be distributed.

In India, for thousands of years the caste system has determined what kinds of work people do from birth. And although government affirmative action programs over decades have made some headway toward offering opportunities based on merit, not caste, nevertheless caste is still a factor in determining one's occupation. As mentioned earlier, in countries like the Soviet Union, it was central planners that determined the allocation of jobs and the distribution of output until the collapse of the Soviet-type system. Historically, in some societies, another method of organizing production and distribution was through an elaborate process of gift-giving; in still others theft or tribute played an important role. Caste, custom, plan, gift, theft, and tribute are all ways of determining what will be produced, for whom, and by whom. They are all different methods of *economic coordination*. While one or more of these factors play a part in most economies, by far the most important methods of economic coordination in the modern world are markets and planning. We refer to these as "coordination by rules" and "coordination by command."

Coordination by Rules and by Command

Seldom in history has anyone proposed so startling a notion as Adam Smith's concept of the invisible hand—that the economy, even if it is quite complex, does not need to be run by anyone at all. Another way of putting this is to say that the economy will run itself—if it is simply subjected to the right rules.

Imagine someone telling you that an invisible hand will coordinate the landings and takeoffs of planes at O'Hare Airport in Chicago (where there are, on an average day, around one hundred landings and takeoffs per hour). The advocate of the invisible air traffic controller might say, "We don't need the control tower. Pilots should just keep right and yield to traffic on their left or below them." Given this method for coordinating air traffic at O'Hare, we suspect that most people would decide to avoid air travel to Chicago.

Even though the amount of activity at O'Hare is awe-inspiring, the interactions that occur in a whole economy are infinitely more complex. In the United States, for example, the economy involves the interactions of approximately 30 million businesses, 125 million households, and 225 million adults, each of which make decisions every day that affect at least some of the others. Add the complexity of global economic interdependence into the picture, and one can easily see how daunting the task of coordination is—and how radical it was of Adam Smith to suggest that the economy need not be coordinated by anyone at all.

Coordination by rules takes place when interactions are governed by general principles of behavior.

But Smith's notion of the invisible hand is not as preposterous as it sounds. Many of our interactions are, in fact, coordinated without a coordinator. Consider another traffic problem, this time for automobiles. In the United States we follow a simple rule—drive on the right—and it does a pretty good job of coordinating the interactions of millions of drivers passing each other every day.

Coordination by command takes place when interactions are governed by orders specifying precise behavior.

The point is that coordination can be achieved by either of two means. One is **coordination by rules**, with no one dictating anyone else's precise behavior, but everyone observing a set of rules. The other is **coordination by command**, with someone (or perhaps more than one) directing the behavior of others. The basic difference comes down to obeying rules versus obeying orders. A rule specifies a range of behaviors appropriate in a given situation (drive on the right) without

specifying particular behaviors (where and when to drive). An order specifies a particular behavior (United flight 407 is to make a 90-degree left turn and land on runway 14A).

Which works better? Obviously, we need both types of coordination in our economy, by rules and by command, and the most appropriate method will vary with the situation. Adam Smith advocated the invisible hand because of his dissatisfaction with the particular kind of coordination by command that prevailed in eighteenth-century Britain. This involved, among other things, the government setting wages and prices, as well as creating large monopolies like the British East India Company. But even Smith acknowledged that the government should protect the nation from external enemies and assure internal justice with police and a system of courts. He also advocated that the government invest in bridges, roads, canals, and other "public works" (such as free education for all children) and proposed taxing alcohol to discourage drunkenness. And he was acutely concerned with the sometimes negative human consequences of the pursuit of economic gain. He worried, for example, that England might turn into a country of mindless robots if something were not done to alleviate the oppressive and mind-numbing conditions in its factories.

The Limits of Coordination by Command

We can understand how markets work if we first review two problems with coordination by command (planning), one having to do with *information* and the other with *motivation*. The individuals giving the commands (the planners) may not have enough information to do the job well, and those who are supposed to carry out their commands may have little motivation to do so. It is even possible that the planners themselves may have little incentive to do the job well.

Air traffic coordination by command works well because all relevant information is available to the controllers on radar screens and computer monitors, and the pilots have a powerful motive to obey the controllers' orders: both their own and their passengers' lives depend on following the commands precisely. And, certainly, the controller has every reason to do the job well: a mistake might result in the loss of many lives as well as the loss of his job.

But in other situations neither the information nor the motivation is adequate. The central planners in the Soviet Union could not know the consumer tastes of its nearly 300 million inhabitants, so they could not make adequate decisions about what consumer goods to produce. Nor could they accurately determine the output capacity of each factory, mine, and office, so they could not assign production targets efficiently.

The problem in a large centrally planned economy is not exactly a *lack* of information: the consumers know more or less what they want, and the plant managers know more or less how much they can produce. So the information exists. The problem is that it is not in the right place: the relevant information is not readily available to the planners (the decision makers). Those who have the information may have an interest in keeping it from the planners or in lying to them. Consumers may wish to exaggerate their needs in hopes of getting more, and plant managers may want to understate their production capacities so that they will not have to produce so much.



Hunger Vs. Compulsion: Coordination by Rules And Coordination by Command

In England in Adam Smith's time, local governments were responsible for feeding the poor. In return for their food, poor people were required to work at particular jobs. This system, regulated by what were known as the Poor Laws, is an example of coordination by command.

Joseph Townsend (1739–1816), a geologist, collector of fossils, sometime physician, longtime Anglican parish priest, and prolific eighteenth-century writer, was one of the most severe critics of the Poor Laws. He thought it would be better to let hunger do the job of getting people to work. In 1786, a decade after the publication of Smith's *The Wealth of Nations*, Townsend wrote the following words in his essay *A Dissertation on the Poor Laws:*

> *Hope and fear are the springs of industry. . . .* [But] *what encouragement have the poor to be industrious and frugal . . . when they are assured that if, by their indolence and extravagance* [and] *by their drunkenness and vices, they should be reduced to want, they shall* [then] *be abundantly supplied not only with food and raiment but with their accustomed luxuries, at the expense of others. . . . In general it is only hunger which can spur and goad them on to labour. Yet our laws have said* [that] *they shall never hunger. The laws . . . have likewise said* [that] *they shall be compelled to work. But then legal constraint is attended with too much trouble, violence, and noise* [and it also] *creates ill will and never can be productive of good and acceptable service. Whereas hunger is not only a peaceable, silent* [and] *unremitt*[ing] *pressure but* [is also] *the most natural motive to industry and labor, it* [therefore] *calls forth the most powerful exertions. . . . The slave must be compelled to work, but the freeman should be left to his own judgment and discretion, should be protected in the full enjoyment of his own* [wealth], *be it much or little, and* [should be] *punished when he invades his neighbor's property. By recurring to those base motives which influence the slave, and trusting only to compulsion, all the benefits of free service, both to the servant and to the master, must be lost.*

Anticipating by three centuries the arguments of some of today's advocates of "welfare reform," what Townsend was actually proposing in his *Dissertation* was a system of coordination by rules. The rules he was putting forward for consideration were: (1) you can eat only what you grow or buy and (2) you have no right to take anyone else's property, no matter how little you have or how hungry you are. The subtitle of Townsend's book was *By a Well-Wisher to Mankind.*

Source: Joseph Townsend, *A Dissertation on the Poor Laws* (Berkeley: University of California Press, 1971 [1786]), 23–24. Some punctuation in the quoted passage has been slightly altered—and the words in brackets have been added—to assist the reader in following Townsend's train of thought.

The difficulties with coordination by command do not end with problems of information. The *motivations* of the planners and the other economic actors may also be a problem. In practice, so it was said, in the Soviet Union furniture factories would produce extremely heavy sofas, because their production quotas set by planners were measured by weight, not by the number of sofas. There was no lack of technological expertise—after all, the Soviets launched a man into space before the United States did—but there was not a strong motive for enterprises to serve the public conscientiously. As long as production quotas were met, managers seem to have been assured of their jobs. Even a good plan will not be implemented in the



The Key Role of Information

The late philosopher-economist Friedrich A. Hayek posed the challenge of how best to organize an economy as a problem of how to make the best use of information:

> *Which of these systems* [central planning or competition] *is likely to be more efficient depends on the question under which of them can we expect that fuller use will be made of the existing knowledge. And this, in turn, depends on whether we are more likely to succeed in putting at the disposal of a single central authority all the knowledge which ought to be used but which is initially dispersed among many different individuals, or in conveying to the individuals such additional information as they need in order to enable them to fit their plans in with those of others.*
>
> —F. A. Hayek, "The Use of Knowledge in Society," *American Economic Review*, September, 1945.

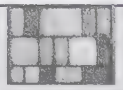
way that best serves society unless both the plant managers and the workers have sufficient incentives to carry out the plan.

Given the problems associated with coordination by command, we need to examine the other main method of coordination in the modern world, coordination by rules. Adam Smith's idea was that markets can take the place of planning as long as two rules govern the economy: competition and private property. In an economy in which private property prevails, and in which most people are not self-sufficient, then, market activities—buying and selling—will have to play a major role in the economy. But under what circumstances will individual decisions about what to buy automatically benefit society?

The Invisible Hand

Adam Smith was not interested in the price of beer in Iowa City. He was interested in how the British economy should be organized: should it be run by royal decree, or should most economic outcomes be determined by the interactions of millions of buyers and sellers in competitive markets, with nobody in particular making the key decisions? He advocated the latter, a system of coordination by the rules of the competitive market. Since Smith's time his argument has been considerably refined, and some of its shortcomings have been clarified. The gist of it is quite simple, but to understand it at a deeper level we have to see what markets really do.

We are, however, not interested in what *particular* markets do: fish markets make fish available to consumers; housing markets make apartments and houses available. Rather, we want to know what markets do in general. As noted earlier, markets perform two important functions: they transmit economically important information and they provide the motivation to act on the information. Under ideal circumstances, then, markets address the two main shortcomings of coordination by command: they overcome the difficulties involving information and motivation.



In Adam Smith's Words . . .

[E]*very individual . . . employs his capital* [and] *endeavours . . . to direct* [it in such a way] *that its produce may be of the greatest possible value. . . .* [Thus] *every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it . . .* [H]*e intends only his own gain, and he is in this . . . led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.*

—Adam Smith, *The Wealth of Nations* (New York: Random House, 1937), Book IV, Ch. II, 421.

The motivation the market provides is of two kinds. First, markets encourage consumers to try to meet their needs with goods that cost less than other goods that could also meet the same needs. People shop around for the best buy, seeking the good that will satisfy a particular need at the lowest available price. If, for example, a person would be just as happy with a hamburger as with a tenderloin steak, the relatively lower price of the hamburger will induce them to satisfy their hunger in the way that takes the smaller toll on society's resources.

Second, the market encourages producers, either companies or individuals, to economize on the use of inputs, by constantly seeking the lowest-cost inputs that will do the job. Since that is what society needs—to produce what it wants using no more of resources than necessary—society benefits when individuals or firms make such decisions.

Thus, both consumers and producers will seek to do something very sensible, namely, to conserve what is costly and use what is less costly, where we mean the cost to society. Nobody will require them to do this. They will do it simply because it is in their personal interest to do it. As Adam Smith implied, the market achieves its results behind the backs of the participants. This is the basic argument that supports his notion of the "invisible hand."

The Invisible Hand in Action

The case for the invisible hand rests ultimately on the claim that even if all of the economic actors behave only with regard to their own self-interest, markets can allocate scarce economic resources in a desirable way. Advocates of laissez-faire (limited government) argue that competitive markets not only address the problems of information and motivation, but in so doing offer a method of coordination that is superior to central planning (coordination by command).

The beer market in Iowa City offers an example of how a competitive market can induce producers and consumers both to respond to a change in tastes and to economize on society's scarce resources. First, as in Chapter 8, imagine that the beer market is in equilibrium and that the market-clearing price is $1 a bottle. Suddenly the U.S. surgeon general comes out with a report saying that drinking wine causes baldness. What will happen in the beer market?

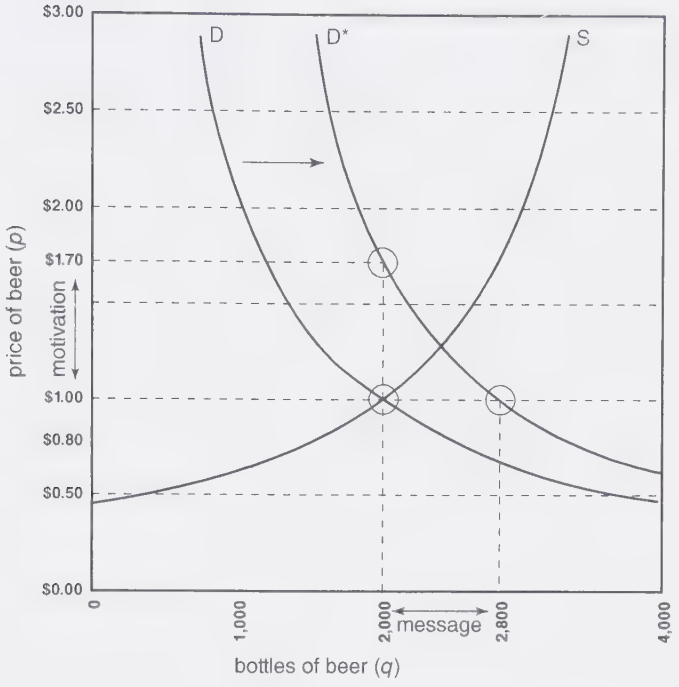


FIGURE 9.1 Market messages and market motivation.

The shift of the demand curve to the right (from D to D*) initially creates excess demand in the hypothetical Iowa City beer market (see Chapter 8). As long as the price per bottle remains at $1.00, excess demand of 800 bottles will now exist because the supply is only 2,000 bottles while the demand is for 2,800. The 800 disappointed customers will be willing to pay more for a bottle of beer, and the suppliers will be willing to produce more if the price rises. As the price rises, the market is sending the *message*: "More beer!" Suppliers can increase their total profits by selling more beer than before, so they will be motivated to supply more beer. Thus, the rising price is the *motivation* provided by the market.

As people switch from wine to beer, more beer will be demanded at each price, shown by a rightward shift in the demand curve for beer, from DD to D*D* (see Figure 9.1). The result is an excess demand for beer: At the price of $1, producers are still willing to supply 2,000 bottles, but now the public demands 2,800 bottles, so there is excess demand of 800 bottles. If the price remains at $1 at first, then producers only supply 2,000 bottles per day, and there will be a line of 800 disappointed customers, some of whom would have been happy to pay up to $1.70 per bottle. But suppliers will be willing to produce more beer if the price rises. That is what it means that the supply curve is upward sloping: A higher price motivates suppliers to produce more bottles (for instance, perhaps because to produce more, they will have to pay some workers overtime for working extra hours, and the higher price will allow the beer producer to do this). Demanders who worry about not getting beer at all will also be willing to pay a higher price than $1. The price therefore rises, and suppliers are then able and willing to supply more beer. Thus, in the form of the rising price the market is sending a message: "More beer!" Suppliers who respond to this signal can sell more beer and make more money.

The motivation to expand the production of beer comes from the opportunity now available to beer producers to make additional profit by selling beer at a higher price. The profit seeking of the beer suppliers will lead them to do not only what is

in their own interest, but also what is in the interest of consumers. Moreover, insofar as this process leads to the best possible use of the economy's resources, it may also be said to promote the interests of society as a whole.

To summarize: If excess demand exists, the market offers producers a *carrot:* higher total profits await those who grasp the meaning of the market's information and expand production. When excess supply exists, the market provides producers with motivation in the form of a *stick*: Some firms will need to go out of business and invest their resources in some other kind of business or other location. If they do not, all firms will suffer falling profits. In this way the market directs self-interested producers to do what is in both their interest and the interests of consumers, even if they are consciously only interested in their own success. This is the invisible hand in action.

How did the invisible hand overcome the two main shortcomings of coordination by command, namely, the planners' lack of relevant information and the absence of incentives either to plan well or to carry out good plans should they happen to occur? The answer is *decentralization*. Prices are the signals. They provide information about consumers' wants and producers' capacities and costs, and this information is communicated directly among all the market participants without first having to go through a central planning office. Also, the motivation to "do the right thing"—with regard to the efficient allocation of resources—is based simply on the self-interest of all participants. When all the players in the market seek the best possible deal for themselves, everyone benefits.

TABLE 9.1. PRICES ARE A DECENTRALIZED SYSTEM OF INFORMATION AND MOTIVATION

PRICES AS *INFORMATION*
• to *consumers:* the price measures how much it costs to produce an additional unit of a commodity.
• to *producers:* prices measure how much demand there is for an additional unit of the commodity and how much it costs to acquire the necessary inputs.
PRICES AS *MOTIVATION*
• to *consumers:* prices, in conjunction with the need to stay within one's budget, motivate consumers to satisfy their wants as cheaply as possible.
• to *producers:* prices, in conjunction with the need to make money in order to stay in business, motivate the lowest-cost production of goods and services that consumers want.

TABLE 9.2. THE INVISIBLE HAND: ASSUMPTIONS AND CONCLUSIONS

If the prices of goods, as they are sold to consumers, measure the ability of the goods to satisfy human needs, and
If the costs of producing goods, as measured by firms, take into account the social costs of acquiring and using the goods,
Then the profit made on each unit of a good (the price minus the cost per unit) will measure the social contribution made by producing each good, and
Hence the pursuit of self-interest (firms seeking greater profits and consumers trying to maximize their satisfactions) will result in a socially desirable allocation of our human and natural resources.

If all this sounds a little too good to believe, it is because we have not yet asked how well or poorly the invisible hand works if the ideal circumstances we have been assuming are not, in fact, present.

Problems with the Invisible Hand

A **market failure** is said to take place when the spontaneous interactions of buyers and sellers on markets each pursuing their own objectives results in outcomes that are different from the decision that society as a whole would make if it had all the information it needed to make the choice.

For the invisible hand to work well, the individual's choice of which product to buy must also be a decision that benefits society. What links these two? The answer is that if the price of each of the several products among which the consumer is choosing accurately reflects the cost to society of producing and using that product, then the individual will make the same decision that a well-informed planner (with good motives) would have made, and society will benefit. In that case, the link between individual decisions and benefit to society is strong, and the invisible hand works well; coordination through markets alone delivers benefit to society.

Unfortunately, as economists have learned, there are many kinds of situations in which markets do not perform so well. Individuals make decisions, but for a variety of reasons these decisions do not automatically benefit society. When this happens, we say a **market failure** exists. We will clarify these ideas shortly with a series of illustrations. But first, it is worth noting that if the invisible hand counts dollars spent on goods and services as votes for what society should produce, then people with more money have more votes. The box, "Voting with Dollars" points out a problem with this arrangement.



Voting with Dollars

It is sometimes said that markets are like elections, in which consumers "vote" with their dollars for the commodities they want. If a large number of dollar "votes" are cast for yellow shirts, a large number of yellow shirts will be produced. In a capitalist economy competition for profits will see to it that resources are allocated in such a way as to produce commodities in the proportions determined by dollar votes.

Voting for commodities in markets is an unusual kind of election, however, because some people vote more times than others. If every dollar of household income had been cast as a vote in 2014, the average household in the richest fifth of the U.S. population would have had more than sixteen times as many votes (194,053) as the average household in the poorest fifth (11,676). Rather than the one-person, one-vote principle of democracy, this is more like an economic version of ballot-box stuffing.

Source: U.S. Census Bureau, Carmen deNavas-Walt and Bernadette D. Proctor, *Income and Poverty in the United States: 2014*, Current Population Reports, p60–252, September 2015, Table A-2: Selected Measures of Household Income Dispersion: 1967 to 2014, available at: https://www.census.gov/content/dam/Census/library/publications/2015/demo/p60-252.pdf

Market Failure

To understand when the invisible hand works well and when it does not, we need to look at how individual economic decisions are linked—or not—to the benefit of society. What circumstances make individual decisions closely linked to social benefit, so that the market works well in coordinating economic activity? A key to the answer is our old friend, *information*. If the buyer chooses what and how much to buy based on all the information available to the whole society, then the invisible hand will work well. In many situations, however, the information on which the buyer bases her decision does not reflect the effect the decision will have on society.

First, the buyer herself needs to have accurate information. When I buy something, I need to know exactly what I am buying. How else can I know how much it is really worth to me (how much I should be willing to pay for it) and therefore how much resources society should devote to producing it? Everyone sooner or later buys something that turns out to be disappointing: a restaurant meal that causes indigestion, an electronic device that works poorly if at all, a medicine that has bad side effects, clothing that falls apart in no time, a car that is a lemon. In such cases, buyers have made bad choices because they lacked the information they needed to make good choices. Through no fault of their own, their spending tells society to use up resources producing something that is not worth the price paid for it. This is one kind of *market failure*: misinformation means that individual economic decisions do not automatically benefit society.

But in a capitalist economy, sellers often have little incentive to tell potential buyers all that is wrong with things they hope to sell, unless the law requires that they do. Information is therefore often a problem. Much advertising focuses on stirring emotions, not conveying information. On TV ads, serious and possibly fatal side effects of drugs are announced—as required by law—while actors smile broadly, fondle their pets, and glow with health, so that viewers pay little attention to any dire warnings.

There is another important way in which a mismatch can exist between what informs my individual buying decision, and what information should be the basis for society's decisions about how to use resources. Even when I have good information about how *I* will be affected, my buying decision may affect others (that is, society as a whole), in ways that my buying decisions will normally not take into account. In late 2015, Associated Press journalists exposed slavery in catching and processing of seafood (fish, shrimp, and so on) in southeast Asia. Some workers were literally kept in cages, and were physically prevented from leaving or even communicating with their families. They were beaten when they tried to complain, rebel, or escape.[1] Before these revelations, clearly buyers of seafood were unaware of these facts, and so did not consider slave labor in deciding whether to buy seafood. But society was most definitely harmed by these brutal and inhuman methods of producing seafood. (This story is told here in the past tense, but although about two thousand slaves were released and the reporters won a Pulitzer prize, there is no certainty at all that slavery has been eradicated, even in the southeast Asian seafood industry.)

The buyer pays for the seafood, but the cost to society is both the cost of physically producing the seafood (with low labor cost because the labor is enslaved), and the cost in the form of the misery suffered by the slaves and their families. Since the cost to

[1]Public Broadcasting System Newshour, "How the AP Uncovered Secret Slavery Behind the Seafood in Your Supermarket," April 20, 2016, at 6:20 PM EDT, available at http://www.pbs.org/newshour/bb/how-the-ap-uncovered-secret-slavery-behind-the-seafood-in-your-supermarket/.

society is much higher than the cost to the seafood buyer, from the standpoint of society the price of seafood is too low, and society is putting too much of its resources into the slave seafood industry (it should not put *any*). Society would be better off putting its resources into some other economic activity. This is a market failure of a common kind. It also has a special name: an *externality*. It is so named because the purchase of the seafood has an "external" effect on someone who is neither directly the buyer nor directly the seller—in this case the enslaved workers. In fact, U.S. law prohibits imports of goods made with slave labor, but clearly the law was not being adequately enforced.

The invisible hand does not work well when the price facing the individual buyer does not reflect the cost to society. If the cost to society is higher than the cost to the buyer, then from the standpoint of society too much of the good is being bought, and so too many resources are being put into producing it. These resources could produce more benefit if used to produce something else.

Why, then, does the invisible hand of the market fail, in many situations, to coordinate economic activities in ways that produce desirable results? The general answer is that people affect one another's well-being in countless ways, and market prices often fail to take into account all the effects of our actions on others. In small groups, such as families and friendships, we typically consider the costs and benefits of our actions not only with regard to ourselves but also as they might affect the others we care about. Such consciousness, also known as altruism, is an essential part of social life, but it obviously does not prevail when we are interacting with large groups of strangers, about whom we may know little and care less. Still, a concerted campaign to get people to systematically refuse to buy products made with slave labor, or products with strongly negative effects on society, can sometimes be successful. If people believe others are going to act out of conscience, they will sometimes do so, too.

The underlying assumption of the invisible hand is that *if prices are right*, all the effects of people's actions on one another will be taken into account in the prices of goods and services. Thus, Adam Smith's theory (including its modern version) holds that—if we assume (as Smith did) that the existing distribution of wealth is acceptable—coordinating an economy with markets will bring about an optimal allocation of society's resources. Proponents of the invisible hand theory offer examples such as the following to support the theory: When a certain consumption choice, say, ordering a tenderloin steak, uses up a lot of society's resources, that consumption choice will be appropriately paid for in the price of the steak. This is because the consumer will be paying a price that is exactly the amount it cost to produce that tenderloin steak.

The problem is that *prices are often not right*. Consider the tenderloin steak. Most industrially fattened cattle are routinely treated with antibiotics as a way to prevent disease among hundreds or thousands of cattle crowded together in a feedlot. But using antibiotics constantly on a massive scale causes bacteria to evolve that are resistant to antibiotics. Already they are becoming more common. A likely future scenario (not necessarily that far away) is that bacteria resistant to *all* known antibiotics will exist, spread, and fatally infect humans, because there is no antidote. The one who buys the tenderloin steak may not know this, and it may not affect his choice to buy it. But society will pay the price for future illness and mortality.

Negative Externalities as Market Failures

In the case of the steak, we say the market for steak has a *negative externality*. The buyer and the seller of the steak are the two people directly involved in the sale of the steak, and neither one is directly hurt (much) by the fact that the transaction has made

a little more likely the chance of a future outbreak of drug-resistant disease. It does this by signaling feedlot owners that using antibiotics routinely on cattle will result in saleable meat. But society will definitely be affected by an outbreak that will cause pain and suffering to many. These people, and the effects on them, are *external* to the transaction and have no say in it. Hence we say the market has an *externality*, and since the higher likelihood of disease is a negative outcome, we call it a *negative externality*. Of course, individual consumers could make the decision to stop buying meat altogether, or only to buy antibiotic-free meat. But fish farms also liberally use antibiotics; and many cattle are also given hormones, which European Union authorities, at least, clearly believe carry negative externalities, and therefore ban such meat.

As is now widely known, the decision to buy and burn gasoline also has a negative externality: it produces air pollution, which harms asthma sufferers and others, imposing both medical expenses and lost work time on them and their caregivers. It also produces greenhouse gases that contribute to global warming (see Chapter 1). Global warming, in turn, increases extreme and destructive weather patterns, like floods, drought, wildfires, and windstorms—and these cause damage, death, and injury. The increased lung disease, and the injuries and deaths and destruction caused by worsening weather, are all *external* effects of people's decisions to fill up their gas tanks and drive. These costs are imposed on society, but at most only a tiny bit of that cost falls on the buyer or seller of gasoline. Here again, the individual decision is delinked from the benefit to society by the fact that a negative externality exists. Buying and burning the gasoline is far more costly to society than it is to the buyer. In this situation, society (through the individual decisions of many buyers of gasoline who do not take account of the costs to society) is putting far more resources into the production of gasoline than it should.

In all these cases there is a discrepancy between (a) the costs borne or benefits received by the decision maker and (b) the costs and benefits experienced by all the members of society. The costs and benefits accruing to the decision maker are referred to as the **private costs** and benefits of the activity in question, whereas when added together all the costs and benefits experienced by everyone (again including the decision maker) are termed the **social costs** and benefits.

Private costs are the costs borne by the user of a good or service (a person or a company), while the total costs borne by all members of a society are termed **social costs**. At best, prices only measure private costs.

How can this be fixed? If the government places a large enough tax on a gallon of gasoline, one equal to the damage caused by producing and burning that gallon, then the individual buyer will have to take account of the full cost to society, because the individual buyer will have to pay the tax equivalent of the social cost. Those who advocate this actually call for a tax not just on gasoline but on anything that produces carbon compounds such as carbon dioxide and other greenhouse gases. Such a *carbon tax* would "internalize the externality," that is, force buyers to pay a cost equivalent to the costs to society, so that they take into account these costs in deciding how much gasoline to buy. The tax can make the private cost equal to the social cost. For instance, currently jurisdictions at various levels in the United States are being urged to pass carbon taxes, as a way of slowing global warming. This more or less would restore the effectiveness of the invisible hand, provided it were also done with *all* negative externalities, not just the one from buying and burning gasoline.

Positive Externalities as Market Failures

There is also such a thing as a *positive externality*. The clearest way to understand this is to consider the *benefits* of a transaction. If you buy a college education, you expect to get a great deal of benefit out of it in the form of future earning power,

along with the joy of learning and of making lifelong friends. But does society also benefit? Certainly—because educated people contribute to the development of technology, that usually benefits everyone. They also contribute to public debate and to the general availability of sound information on which to base public policy decisions. So those not involved in the decision to pay for an education, namely, the general public, enjoy the *positive external effects* of your decision to get an education. Thus we say that education has *positive externalities*. The social benefits exceed the private benefits.

Similarly, if you get a vaccination against a communicable disease (like measles or whooping cough), those who have not gotten the disease are then a little less likely to get it. This is a positive externality. To internalize positive externalities like college educations and vaccinations, they should be subsidized, that is, sold at a price lower than the cost of producing them—because without the subsidy, their private cost is higher than their social cost.

Thus an externality can be either positive or negative. It is positive if the external effects of a transaction are beneficial to others. It is negative if the external effects impose costs on others.

Important sources of positive externalities are education, the production of knowledge, the introduction of new technologies, and vaccinations. Important examples of negative externalities are pollution, other kinds of environmental degradation, and automobile traffic congestion in metropolitan areas. Table 9.3 lists some examples of positive and negative externalities.

TABLE 9.3. POSITIVE AND NEGATIVE EXTERNALITIES

ACTION	EXTERNALITY	PRICE NOT EQUAL TO SOCIAL COSTS/BENEFITS
Drinking too many beers	The costs of alcohol abuse on families, friends, and medical providers	The price of the beer does not reflect such negative externalities
Driving a car	This can contribute to traffic congestion, and it also causes environmental degradation from both the production and the use of fossil fuels	The price of the gas used does not reflect these negative externalities
Research	The benefits of any good ideas produced will most likely be enjoyed by other people or companies	Any increase in the profits of the firm that supported the research—or rise in the salary of the individual who conducted it—may not take into account all the benefits of the resulting knowledge enjoyed by others
Education	The benefits of one person's education are enjoyed by neighbors and workmates	The individual benefits (higher earnings) do not capture the positive externalities enjoyed by others
Training employees	Some employees will move to other firms, the owners of which will benefit from the training paid for by the firm that did the training	The training firm's profits do not reflect the benefits flowing to other firms with the workers who move
Wearing a $500 watch (luxury consumption)	This may have the effect of lowering the (relative) status of others, creating envy	The $500 price of the watch does not include the status and envy costs imposed on others

Monopoly as a Market Failure

We have already learned that for the invisible hand to work well, the information that informs the individual buyer's decision must coincide with the information that would inform the whole society's decision about what to produce and therefore how to allocate productive resources to do so. We know that if the individual decision maker has misinformation about the quality of the product, that alone can break the equivalence between individual decision and social benefit.

Another way in which that link, or equivalence, can be broken is if the price that faces the individual buyer exceeds the cost of producing one more unit of output. This happens if the market has just one monopoly seller, or a few large firms. The invisible hand argument assumes that markets are competitive, with many actual or potential buyers and sellers in every market. Given this assumption, no producer can charge more than the marginal cost of producing a particular commodity. In the real world, however, many markets are not competitive.

In the case of a *monopoly*, a single producer with no actual or potential competitors can raise the price buyers must pay for its product without fear that its customers will switch to buying from another supplier. The only ones who will buy it are those who are both willing and able to pay a high price for it. An obvious example is patented drugs for which the patent owner—typically a giant drug company—charges a price many times the cost of producing it. A monopoly has *market power* since it can raise the price of its product merely by producing (and selling) less of it. This will create an artificial scarcity in the market for the product and drive up its price. In this situation the price charged will be higher than the marginal cost, and therefore less of it will be bought—and produced—than would be true if society decided how much to produce, based on the cost of producing an additional unit of it.

Of course, if there were other firms that could produce the same product, they would enter the market, would compete for customers by charging a lower price, and would continue doing this until the price of the product fell to its marginal cost. It is this process that leads to the result that in a competitive market, the market-clearing price will be equal to the marginal cost ($P = MC$) in every firm. But with barriers to the entry of new firms, which is what makes a monopoly a monopoly, a monopolist does not have to fear competition from new firms. The workings of markets that are not perfectly competitive are discussed in Chapter 11.

Economies of Scale as a Market Failure

Prices often exceed marginal cost for yet another reason, one of great importance in the electronic age. To see this, ask yourself what is the marginal cost of an eBook. Chapter 8 explained that marginal cost is just the increase in the total cost that a producer incurs by producing one additional unit of output. Since it costs almost nothing to give one more person access to an existing electronic document, the marginal cost of an eBook is very close to zero, possibly a matter of just pennies. Even if the publisher makes a royalty payment to the author for each copy sold, the marginal cost will still be low—considerably less than the *average* cost of production of the book.

Recall that the average cost is the total cost divided by the number of units produced. The total cost has two distinct parts. The larger cost is for all the labor and materials that are used in just producing the very first copy of the book. This includes the cost of writing, designing, editing, typesetting, and producing the book, including the cost of professional-looking artwork as well as the required payments

for permission to use any copyrighted material, and any interest paid to the bank on loans, rent for the use of production facilities, and the costs of advertising and distribution (which need to be set into motion before the book is actually published). Such costs do not vary with the number of electronic units produced.

The invention of the printing press was an enormous advance in making information widely available. The fact that the *marginal cost* of reproducing and distributing the electronic document is very close to zero is another great boon to society. It means that it is far cheaper to make the book available to thousands or millions of people than before. However, it also creates a problem for the publisher.

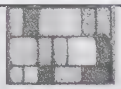
For society, the cost of producing/distributing one more book is nearly zero. But the publisher who sets the price at marginal cost will never recoup the costs incurred in producing the first copy: writing, typesetting, and putting the book into the electronic form that works with eBook software. Setting the price at marginal cost will make the publisher lose money. The publisher needs to recover at least the *average cost* per copy of the eBook. If the up-front costs are $100,000 and the cost of making the eBook available to one more customer is $1, then if the publisher sells 10,000 copies of the eBook, the break-even price (the average cost) would be $11 ($100,000/10,000 + $1); in order to make a profit the publisher would have to charge perhaps a dollar more.

An industry like this with large up-front costs and a marginal cost that is constant, or at least not rising, has *economies of scale*. The average cost falls as those large up-front costs are distributed among a larger and larger number of units sold. If there are economies of scale in an industry, the average cost always exceeds the marginal cost (because it includes the marginal cost plus a piece of the up-front costs).

The eBook industry, with its economies of scale, is an example of a market failure. Society could get the benefit of the low marginal cost if the publisher could sell at marginal cost without going bankrupt, or if the publisher were a state-owned enterprise and could pay for its costs partly out of tax revenue, so that bankruptcy was not a danger. But the need to make a profit makes it impossible for society to get the full advantage of a production process with economies of scale. In effect, *the invisible hand is not working* in this case.

Notice that the production and sale of the eBook is also happening within a monopoly, namely, the copyright that the publisher holds, which legally protects the publisher's sole right to reproduce it unless the copyright holder grants that right (usually in exchange for money). If it were legal for others to copy the eBook, then they could do so for the marginal cost, and the publisher could no longer succeed in charging a higher price—and with the prospect of losing money, might not undertake to publish the eBook at all. And if electronic copies circulated at such a low price, and most people acquired works in that form, then the publishing industry might find itself struggling to survive, which itself could affect society in a way that many would regret.

The phenomenon of economies of scale provides but one more way of explaining why markets fail. If we put all of the explanations of market failure together—inadequate pricing, externalities (spillovers), incomplete contracts, market power, and economies of scale—we arrive at a more complete understanding of the concept of market failure. We can regard as a market failure any situation in which the market interactions of buyers and sellers result in outcomes that are undesirable either to individuals or to society as a whole. The accompanying box ("The Invisible Foot . . .") provides a list of some of the more common types of market failure.



The Invisible Foot: When Markets Fail

Market failures occur when the market interactions of buyers and sellers result in outcomes that do not make society as a whole better off, even though individual buyers believe they are making good decisions for themselves. Market failures result

- when markets are controlled by a small number of buyers or sellers. (For example, when there is a single monopoly seller, the price charged for a good will exceed the cost to the firm of producing another unit of the good.)

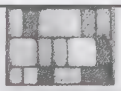
- when environmental degradation or other negative externalities resulting from either production or consumption occur. If it is in production (like air pollutants emitted from factories), then the cost to the firm of producing another unit of the good will be lower than the cost to society; the private cost will be lower than the social cost.

- when positive externalities are present. Here, the benefit or cost to the individual consumer will not accurately measure the benefit or cost to society as a whole. An example is when the education one person gets has beneficial effects on society by making public discussion better informed.

- when people's needs are not reflected in market demands. This may happen when individuals, such as homeless people, do not have enough money to purchase necessities (like housing) for themselves in the market.

The type of market failure that stems from economies of scale leads at least some countries to avoid relying on privately owned companies for the supply of such things as electric power generation, transportation networks, and phone systems, all of which are usually characterized by economies of scale. Such countries choose, instead, to have either their governments or regulated enterprises carry on these types of economic activities. Whether these solutions work better in practice than private production without regulation depends on the nature of the government in question.

The fourth category of market failure listed in the "Invisible Foot" box raises, again, the issue of efficiency and income distribution discussed in Chapter 3. The question is whether one can say that an economy is efficiently allocating a society's resources if some people have huge incomes while others do not have enough income to provide for their most basic needs. The box on "Sleeping Sickness" provides a concrete illustration of the issue that was brought out earlier in the box "Voting with Dollars."



Sleeping Sickness: "It Really is a Failure of Capitalism."

Sleeping sickness is a horrible disease common in Africa; it attacks the brain, driving sufferers insane before killing them. Spread by the tsetse fly, in the late 1990s it struck over a quarter of a million people each year. A key drug in curing one variety of the disease is eflornithine, combined with another drug in the second stage of the disease. It is so effective that even comatose sleeping sickness patients have been revived, earning it the name "resurrection drug."

Wonderful news. A triumph of modern medicine? Not exactly. Eflornithine began as an anti-cancer

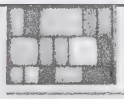
drug in the 1970s that failed. By 1990 eflornithine was known to be good for treating sleeping sickness, but production ceased in 1995 because a drug that saves mainly poor people is unprofitable; the people who need it cannot afford it. It was saved from oblivion only when the Bristol Myers Squibb Company found a profitable use for it as the main ingredient in Vaniqa, a cream that removes women's facial hair.

Maladies like the 17 Neglected Tropical Diseases (NTD) identified by the World Health Organization, including sleeping sickness, river blindness, and guinea worm disease, ruin the lives of many in poor countries, yet the market does not deliver remedies. Pharmaceutical companies worldwide sell only 1 percent of what they produce in Africa, but three-quarters of their output to the United States, Europe, and Japan, with less than a fifth of the world's population. And only a tiny portion of private firms' medical research is focused on diseases endemic to poor countries. Just 1 percent of drugs licensed worldwide between 1975 and 1997 were for tropical diseases.

Where markets failed to solve this problem, the Neglected Tropical Diseases initiative, a public-private partnership announced in 2012, has stepped in. The World Health Organization, a leader in it, has been implementing a plan to combat such diseases, aided by drugs donated by major drug companies. Guinea worm disease (dracunculiasis), for instance, is now down to just twenty-two reported cases in 2015, with hope of eradication by 2020.

The Bill and Melinda Gates foundation also funds efforts to reduce the incidence of disease in Africa. Over a decade ago, on his PBS show *Now*, Bill Moyers asked Bill Gates, "What does it say to you that 11 million children, roughly, die every year from preventable diseases? What does it say to you that of the 4 million babies who die within their first month, 98 percent are from poor countries? What do those statistics tell you about the world?" Gates replied, "It really is a failure of capitalism. You know, capitalism is this wonderful thing . . . But in this area of diseases of the world at large, it's really let us down." Moyers then countered: "But markets are supposed to deliver goods and services to people," and Gates responded, "and when people have money it does . . . Here what we have is . . . not only don't the people with money have the disease, but they don't *see* the people who have the disease. If we took the world and we just reassorted each neighborhood to be randomly mixed up, then this whole thing could get solved. Because you'd look out your window and you'd say, you know, there's [a] mother over there whose child is dying. You know, let's go help that person."

Sources: Michael Kremer, "Pharmaceuticals and the Developing World," *Journal of Economic Perspectives*, Fall 2002, pp. 67–90; Donald McNeil, "Cosmetic Saves a Cure for Sleeping Sickness," *New York Times*, February 9, 2001, p. A1; *NOW with Bill Moyers*, PBS, May 9, 2003; "Dracunculiases Surveillance Programmes Respond Strongly to Rumors, While Cases Plummet," available at http://www.who.int/neglected_diseases/news/surveillance_programmes_respond_to_rumours/en/.



Private Incentives, Public Benefits

Until early 2003 the congested traffic in central London crept along at a snail's pace. Facing near gridlock, Mayor Ken Livingstone tried a radical solution: charge those driving private cars for the congestion they impose on others. In February 2003 the city imposed a fee of $8 a day for driving a car in the central part of London except central London residents. A computer system kept track daily of who had paid (one could pay by texting, or in other ways). License plate number-recognition devices installed throughout the central part of the city nabbed scofflaws.

Traffic congestion costs that are imposed on others are examples of negative externalities (like the

continues

costs of pollution). Livingstone's plan forced drivers to take these costs into account. Economists call this "internalizing the externality," and this is the remedy that most economics textbooks recommend: a tax or fee equal to the cost imposed on society.

Livingstone's critics thought his plan would worsen the traffic problem and hurt businesses in the affected area. To the surprise of many, however, it worked. Traffic flowing into central London fell by 20 percent, and delays were cut by almost 30 percent. Average traffic speeds jumped from 9.5 mph to 20 mph. On an average day in March 2003, about 100,000 motorists paid the congestion fees. The revenue from fees—and from the hefty fines paid by those who foolishly tried to beat the system—brought over $1 million a day to the city government. A third of the companies in the affected area said that Livingstone's new policy was helping them; only 5 percent said it was hurting their business. After six weeks of operation, half of Londoners liked the policy, while only a third disliked it. Livingstone's popularity ratings hit an all time high.

Source: "Ken's Coup," *The Economist*, March 22, 2003, p. 39.

Coordination Failure

Coordination failure occurs when markets or other types of coordination by rules fail to coordinate an economy in such a way as to produce outcomes that are desirable.

When markets fail to coordinate an economy in such a way as to produce outcomes that are desirable, economists call this a *coordination failure*. The term *coordination failure* refers to *any* situation in which the self-interested behavior of individuals results in an outcome that is less beneficial to them than one that might have been achieved by better-coordinated, or cooperative, behavior. Two parables can be used to introduce the concept of a coordination failure. The first is known as the "Prisoner's Dilemma," and the second, the "Tragedy of the Commons."

The Prisoner's Dilemma and the Benefits of Cooperation

This widely circulated story makes the following point: individuals who *cooperate* may come closer to achieving their personal objectives than do those who pursue their self-interest without regard to the well-being of others. The theory of how markets coordinate an economy—conveyed in Adam Smith's metaphor of the "invisible hand"—shows that *under certain conditions* competition based on self-interest but coordinated by markets will bring about a desirable allocation of economic resources. The story of the Prisoner's Dilemma, however, dramatizes the fact that *under other conditions* lack of cooperation results in a coordination failure.

The story begins when two individuals who have committed a crime are picked up by the police and placed in separate cells. The police have enough evidence to convict either or both of the prisoners (who we will call "A" and "B") of a lesser crime for which the sentence is two years in jail, but they wish to get at least one of the prisoners to confess to the crime and implicate the other one, which will result in longer sentences. Each prisoner has to choose either to confess to or deny involvement in the crime. The consequences of all the possible choices are explained below and summarized in Table 9.4. A and B both know all the four points below, but because they are not allowed to communicate, each must decide independently whether to confess or deny guilt.

The possible outcomes are these, and Table 9.4 shows each outcome in a shaded box numbered to correspond to the numbers below:

1. If both A and B deny involvement in the crime, each will be convicted of a lesser offense (regarding which the prosecutor has ample evidence), and each will be sentenced to two years in jail.
2. If A denies guilt, but B confesses and implicates A, then A will be sentenced to four years in jail, and B to only one year.
3. If A confesses, implicating B, but B denies guilt, B will be sentenced to four years in jail, and A to only one year.
4. If both prisoners confess, they will get three-year sentences.

First, assume that both prisoners make their choices based only on their own interests. What will A and B each decide to do? It turns out that under this assumption each prisoner has the incentive—*regardless of whether the other prisoner confesses or denies*—to confess to the crime and implicate the other partner. Let us see why.

Suppose you are prisoner A. If your partner confesses and implicates you, then denying guilt will get you a four-year jail term (shaded cell 2), while in that situation confessing will get you a three-year jail term (shaded cell 4). So if your partner confesses, you are better off confessing. On the other hand, if your partner denies involvement in the crime, you are still better off confessing: If in this case you deny guilt you will get a two-year jail term (shaded cell 1), but if you confess you will only get a one-year jail term (shaded cell 3). So if your partner does not confess, you are still better off confessing. In short, whatever your partner does, if you confess to the crime you will end up better off than if you deny your role in the crime.

Notice that we have been assuming that you must act to protect yourself, because you cannot coordinate with your partner by agreeing that neither of you will confess. This is because you are isolated, each of you in separate interrogation rooms. So you are led to assume that your partner is going through the same thought process, and will also decide that confessing is the best strategy. This means both of you will be put away for three years (shaded cell 4). Notice, too, that there is a better outcome, namely, for neither of you to confess, so that each of you gets only two years in jail—but without an ironclad agreement between the two of you that you each are confident that your partner will stick to, you will both act independently and end up with three-year terms.

Now imagine instead that you and your partner are very close, and that each one of you truly believes that the other will not confess. Under these assumptions *cooperative* behavior can prevail, with both denying guilt, and both therefore serving two

TABLE 9.4. PRISONER'S DILEMMA

		PRISONER B	
		Confesses	Denies
Prisoner A	Confesses	A and B both get 3 years (4)	A gets 1 year (3) B gets 4 years
	Denies	A gets 4 years (2) B gets 1 year	A and B both get 2 years (1)

years in jail (shaded cell 1). This outcome is clearly better for each of you than the one resulting when you both act based only on your own interest (shaded cell 4). This could happen, for example, if both you and your partner are people who keep promises (whether or not keeping them seems to be in your own interest) and if, before committing the crime, you each promise to claim innocence if arrested.

The moral is that *in some situations the pursuit of self-interest by all parties leads to outcomes in which none of the participants benefit.* These situations are just the opposite of the situation described by the invisible hand.

The Prisoner's Dilemma can help answer questions such as: why is it so difficult to prevent destruction of the environment or to establish minimum standards of workplace health and safety among nations? In these cases the choices are not to confess or deny, but to "adhere to environmental (or workplace) standards" or "violate the standards."

THE PRISONER'S DILEMMA AND GLOBAL WARMING. As Chapter 1 mentioned, a major obstacle to agreeing on a plan to stop global warming is that the government of each country prefers to see all *other* nations limit *their* emissions of greenhouse gases like carbon dioxide while avoiding the politically unpopular task of limiting its own citizens' freedom to pollute. Many people in rich countries resist limits on activities, like driving a car, that pollute the environment; and many people in poor countries believe that limiting greenhouse gas emissions in their countries would hamper industrialization and make it harder to raise their living standards to those of rich nations.

The countries of the world thus face a coordination problem: if each nation's government acts independently and seeks to serve only the perceived interests of its own citizens (or its most powerful ones), nations are not likely to adopt the measures needed to avert global warming. To see why this is so, imagine that there are just two countries, called North and South, and just two options, called Nothing (meaning do nothing to prevent global warming) and Protect (meaning adopt protective measures to slow or halt global warming). Each country would like the other one to reduce its greenhouse gas emissions while avoiding the costs of doing so itself. Thus, for each the best alternative is to do Nothing while the other does Protect. The worst alternative for each is to Protect while the other does Nothing. The second-best alternative for both is that both Protect, and the third-best alternative for each is that both do Nothing.

The options available to the players are shown in Table 9.5. This is a Prisoner's Dilemma situation, and as we have seen earlier, what is best for each is worst for all. If each country makes its decision independently and they both do what their own people see as being in their own self-interest, both countries will do Nothing. They would both be better off, however, if they both chose to Protect.

TABLE 9.5. NORTH'S ACTION/SOUTH'S ACTION

		NORTH'S ACTION	
		Nothing	Protect
South's Action	Nothing	Third best for both	Best for S, worst for N
	Protect	Best for N, worst for S	Second best for both

The best outcome can be achieved, however, only if the two countries agree to Protect, perhaps accomplishing such an agreement by signing an international treaty committing themselves to implementing protective measures. But two questions need to be answered. First, who will enforce the agreement? There is no world government or any other body that could force each nation to abide by its agreements. The second problem is that different ways of protecting the global environment result in differing costs and benefits for each country. Why, for example, would the poor countries of the world agree to pay an equal share of the costs of reducing greenhouse gas emissions, most of which are now caused by the high-income countries?

The poor countries might argue that each nation should have the right to pollute in proportion to its population. In this case any solution to the problem would require drastic reductions in carbon dioxide emissions in the rich countries (where pollution per capita is high), while leaving the poor countries relatively free to industrialize in a way that increases pollution. While this solution may seem fair to many people, it is likely to be politically unpopular in the rich countries.

In practice, as it turns out, there is one factor that has encouraged nations to take action against global warming on their own: the truly noxious clouds of air pollution from burning coal smothering a number of cities in China and elsewhere recently, driving China's own population to rebel, and causing massive disruption to economic activities. Such events have helped to spur the government to adopt environmentally conscious policies, including seeking alternatives to burning fossil fuels, such as passive solar (using the sun to heat water for household uses).

Extending the lesson of the Prisoner's Dilemma more generally to economic systems, we can say that reliance on competitive markets may fail to coordinate an economy in a desirable way: the pursuit of self-interest by individuals may result in outcomes that are not optimal from the point of view of any member of the society. This is what is meant by the term *coordination failure*.

The Tragedy of the Commons

Another illustration of a coordination failure is referred to as "the tragedy of the commons," a parable set forth in 1968 by Garrett Hardin in a now-classic article in *Science* magazine.[2] This parable is about the possibility that environmental destruction will result from the uncoordinated pursuit of individual self-interest.

Imagine a large lake, its shores dotted by the houses of people who fish in the lake to earn their livings. No one owns the lake: it is the common property of all the members of the surrounding community. The situation described here, one that involves communal ownership of an important shared resource, is similar to arrangements that were widespread in many of the early settlements in New England (it has, of course, also existed in other parts of the world at various times in history), and it takes its name from the shared grazing land for cattle and other livestock that was referred to as the town "common."

In the parable of the lake, each person decides independently the number of hours to fish each day (or, alternatively, how many fish to catch) before heading back to shore. As self-interested people, they fish as long as the additional benefit of another hour (or another fish) is greater than the inconvenience or discomfort of the additional time and effort spent fishing.

---

[2]Garrett Hardin, "The Tragedy of the Commons," *Science* 162 (Dec. 13, 1968).

However, as in the case of the Prisoner's Dilemma, what is rational for one is not beneficial for all. The more each person fishes, the more difficult it is for the others to catch fish. The reason for this is simple: there are a limited number of fish to be caught, and as each person catches more of them, fewer are left to be caught by the other lakeside residents. Each person would like to see limits put on the fishing of the others while remaining free, herself, to fish without limit. As long as there are no limits placed on anyone's fishing, there will be overfishing. The end result is predictable: there will soon be hardly any fish left in the lake, and, eventually, the lake will be entirely "fished out." That is the tragedy.

This story is about fishing in a lake, but it might just as well be about overgrazing open pastureland, dumping sewage into a river, or polluting the air. The moral of the story is that the pursuit of individual self-interest can be highly irrational because it can have very negative, possibly irreversible, consequences.

A real-life example of the tragedy of the commons is the case of what happened with a particular effort to harness geothermal energy for the generation of electricity. A potentially cheap and environmentally clean way of generating electricity involves tapping the steam that shoots in geysers out of the earth. In the years since this method of producing electric power became practical, a huge field of geysers 70 miles north of San Francisco has been regarded as a promising place to try it out. When energy prices suddenly rose in the mid-1970s, the geysers north of San Francisco became the focus of intense but largely uncoordinated energy development. However, as more and more electricity-generating plants were built there, the amount of steam available for earlier-installed plants fell. (In such a field of geysers there is a common but limited amount of steam deep beneath the earth's surface.) Eventually, so much steam was being tapped that none of the electricity-generating plants could operate at full capacity. By the early 1990s it became clear that because the number of plants had not been limited, the geysers were no longer an efficient source of energy.[3]

How can tragedies like this be avoided? There are many answers to this question, but all involve finding some way to take others' interests into account when making decisions. The most obvious solution to the problem would be simply to regulate access to the common resource. In the fictitious case described above, the fishing people around the lake could have met, ascertained how many fish might reasonably be taken from the lake each week, and decided collectively on a fishing limit for each member of the community that would have allowed the weekly target to be met. In fact, there are fishing villages in Japan, Turkey, and elsewhere that engage in this practice.[4]

If, on the other hand, the problem was overgrazing, those with the animals could figure out how much grazing the common land could support and then decide on the number of animals each herder would be allowed to graze. Solutions to the problem such as the ones mentioned so far all involve what is termed *social regulation of the commons*.

An alternative, which might be termed the *private property solution*, would be for the commons to be owned by a private individual or company. Such an owner would, say (in the cases mentioned above), not only employ the lake residents to fish or the animal herders to graze their livestock, but also would, in self-interest, limit

[3]Richard Kerr, "Geothermal Tragedy of the Commons," *Science*, 253 (July 12, 1991): 134–35.
[4]F. Berkes, D. Feeny, B. J. McCay, and J. M. Acheson, "The Benefits of the Commons," *Nature*, 340 (July 13, 1987), 91–94.

the use of the common resource in order to prevent its deterioration. Obviously, the owner would not want to see the lake overfished or the pasture overgrazed. A problem with this "private property solution" to the tragedy is that the common resource is often too large to be owned by a single individual or company. A tragedy of the commons was averted when depletion of the high-altitude ozone layer was halted by a 1989 ban on ozone-depleting chemicals. World oceans are a similarly threatened common resource. For such large commons, private ownership is obviously not possible, so the social regulation approach is all there is. However, as noted earlier (see "The Prisoner's Dilemma and Global Warming" above), social regulation is not easy to work out in practice.

One approach is the system under which the government sells or otherwise distributes pollution rights, or "emissions permits," that allow companies to emit a certain quantity of pollutants into the atmosphere, soil, or water during a particular time period. Once such rights have been acquired, they can be exchanged among companies in what is, in effect, a market for pollution rights. In such a market pollution rights can be traded for anything from cash to pollution-absorbing forests. What generally happens, however, is that a company builds a state-of-the-art environment-friendly plant and therefore has a surplus of pollution rights—rights it does not need to use because its new plant emits less pollution than most older plants. This company can then sell its unused pollution rights to another company whose plant exceeds the pollution standard set by the government. Under this system the more a company pollutes the more it will have to pay. Moreover, if the price of a pollution right is set correctly, a company will have to take into account in a precise way the harm it does to others when it decides what technology to use when it is planning, say, the construction of a new plant. With this system, its proponents argue, profit and loss calculations will bring about a desired amount of pollution reduction in the most efficient possible way. However, pollution reduction from such schemes has been nowhere near sufficient to avert global warming.

Because millions or billions of dollars as well as life and death matters of public health are affected by policies to correct market failures, the special-interest groups that stand to benefit or lose from their adoption often dominate debates about these policies. An as yet unanswered question is: how can ordinary people build up enough influence in policy-making centers such as Washington, DC, to equal or outweigh the influence that lobbyists, representing a small number of individuals or corporations, exert on the legislators and others who determine which policies will be adopted—or not adopted—to correct for the various types of market failures discussed in this chapter?

Suggested Readings

James K. Boyce, *The Political Economy of the Environment* (Cheltenham, UK: Edward Elgar, 2002).

Allen E. Buchanan, *Ethics, Efficiency, and the Market* (Totowa, NJ: Rowman & Allanheld, 1985).

John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: The Invisible Hand* (New York: Norton, 1989).

Milton Friedman, *Capitalism and Freedom* (Chicago: University of Chicago Press, 1962).

F. A. Hayek, *The Fatal Conceit: The Errors of Socialism* (Chicago: University of Chicago Press, 1988).

David Jenkins, *Market Whys and Human Wherefores: Thinking About Markets, Politics and People* (London and New York: Cassell, 2000).

Naomi Klein, *This Changes Everything: Capitalism vs. the Climate* (New York: Simon & Schuster, 2014).

Robert Kuttner, *Everything for Sale: The Virtues and Limits of Markets* (New York: Knopf, 1997).

Marie Lavigne, *The Economics of Transition: From Socialist Economy to Market Economy*, 2nd ed. (New York: Palgrave, 1999).

Charles E. Lindblom, *The Market System: What It Is, How It Works, and What To Make of It* (New Haven, CT: Yale University Press, 2001).

Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge, UK: Cambridge University Press).

Robert Pollin, *Greening the Global Economy* (Cambridge, MA: Boston Review Books, MIT Press, 2015).

Michael J. Sandel, *What Money Can't Buy: The Moral Limits of Markets* (New York: Farrar, Straus and Giroux 2012).

Amartya Sen, *Poverty and Famines: An Essay on Entitlement and Deprivation* (Oxford, UK: Oxford University Press, 1981).

Cass R. Sunstein, *After the Rights Revolution: Reconceiving the Regulatory State* (Cambridge, MA: Harvard University Press, 1990).

Capitalist Production and Profits

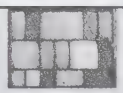
10

Capitalism is often said to be a system of profit and loss. And so it is—when viewed from the standpoint of a firm. Yet capitalists as a group generally see profits every year, not losses. This was true in the United States for all of the twentieth century and the twenty-first century so far—except 1932 and 1933, the two worst years of the Great Depression. Even in those two years, owners of capital goods who received their income in the form of interest or rent did not suffer losses. Despite the misfortunes of particular companies, the capitalist class as a whole makes a profit every year.

Moreover, total profit in the United States is typically large. In the United States in 2014, for instance, profit was $2,073 billion, and total property income—profit, proprietor's income, interest, and rent—amounted to $3,889 billion, more than one-fourth of the nation's total income in that year. (See the box "Calculating Total Property Income" in this chapter to find out how this calculation is done.) Such income is not wage or salary income, but it is real money received by investors, banks, stockholders, bondholders, landlords, pension funds, and retirees, among others.

Why, in a system of profits and losses, does the capitalist class as a whole usually make profits, not losses? And what determines the amount of its profits? These are the questions we begin to address in this chapter and continue to explore in the next three chapters. Understanding the profit rate illuminates why and how owners of capital control the surplus product and what perpetuates the class system. Later in the book we show how the profit rate is the concept that best brings together the horizontal, vertical, and time dimensions of economic life—competition, command, and change.

The profit rate is affected by competition among rival firms as well as by conflicts between employers and workers. In Chapter 11 the profit rate is central to explaining how firms compete and collaborate with other firms in their attempts to increase profits. In Chapters 12 and 13, we focus on how employers' efforts to increase their profit bring them into conflict with workers. In Chapter 19 we look at how the actions and policies of governments may affect the profit rate, a topic we that was also touched on in Chapter 6.



Business and Profit

Business functions around one predominant organizing principle, profitability.

—Robert E. Rubin, former U.S. treasury secretary, in a speech to the Economic Club of Chicago, February 22, 2001

The main idea of this chapter is that *the profit rate is the basic tool for analyzing how capitalism works, and understanding the profit rate requires understanding how the capitalist class prevails in its conflicts with other claimants on the product of economic activity.* This main idea is expressed in four main points:

1. *Profit* is the remainder, or surplus, left for employers after their output has been sold and the inputs—materials used, wear and tear on machines and equipment, and labor employed—have been paid for. Profit would not exist if the payments for materials, machine use, and wages and salaries added up to the value of the entire output of the firm. A similar idea applies to the whole economy: if the proceeds from selling output are entirely used up in paying for inputs, then there is no surplus—in capitalism, no profit.
2. The *profit rate* is defined as the amount of profit divided by the value of the capital invested in the firm, which we assume here is approximately equal to the value of the capital goods owned by the firm.
3. How high the profit rate is for the capitalist class as a whole depends on how successful that class is in waging a three-front war: to extract high levels of effort from *workers* at a low wage cost, to get services from *the government* at a low tax cost, and to acquire *raw materials and other inputs* at a low cost.
4. The *determinants of the profit rate* constitute a list of the various ways in which outcomes of the conflicts between the capitalist class and other groups, as well as within the capitalist class, influence the profit rate.

What Are Profits?

The search for profit, and for a high and rising profit rate, are the keys to understanding how capitalism works. But what exactly are profits?

One way to make a profit, as Adam Smith observed, is to "buy cheap and sell dear"—in other words, to sell something for a higher price than you paid for it.[1] Buying cheap and selling dear results in a profit even if you have not improved or in any other way changed the item in question. Ticket scalping for World Series games or rock concerts is one example. Buying gold in the hope that the price will go up—which is called speculation—is another. When a merchant makes a living from

[1]Adam Smith, *The Wealth of Nations* (New York: Random House, The Modern Library, 1937 [1776]), Book 4, Chapter 2.

buying cheap and selling dear on a regular basis, the merchant's gains from buying cheap and selling dear are called **commercial profits**.

For a merchant, **commercial profits** result from selling something for more than it cost to purchase ("buying cheap and selling dear").

Centuries ago, spice merchants bought spices in Asia and transported them, by camel caravan and other means, to market in Europe, where, by weight, at times they sold for a higher price than gold. But it was only because wealthy Europeans were willing to pay a price high enough to cover the cost of transport that the spices were transported and sold at all.

The distinguishing characteristic of commercial profits is that these profits do not result from a labor process, except sometimes the labor of transporting them from one place to another: profits are made possible by the fact that the price of a good differs at two different times or places. Profits or gains made from buying and selling assets, such as real estate, stocks, or national currencies, account for some of the world's largest fortunes. But aside from the sums involved, they are not very different from the profits made from scalping, since trading in real estate, stocks, and currencies does not *produce* anything.

In a certain sense "buy cheap, sell dear" is also the way profits are made in firms that produce a physical product. But there is an important difference. What is "bought cheap" is the inputs, and what is "sold dear"—at a price higher than the total cost of inputs—is the output, which is a different good altogether. In between the buying and the selling, a production process takes place. Still, we need to know why it is possible to sell something for more than the cost of its inputs. The short answer is that sometimes it is possible, and sometimes it is not; not all enterprises succeed in making a profit. This and the next few chapters explain why it is sometimes possible for capitalist enterprises to make a profit by selling output for more than the cost of inputs.

Capitalist profits are profits that result from a production process.

In this book, what we mean by **capitalist profits** are profits that result (directly or indirectly) from a production process. Unlike commercial profits, they do not depend on the existence of different prices for the same item at different times or places. Something is produced, and profit results because it is possible to sell the product at a price greater than the cost of the labor time and other inputs used to produce it. Our focus in this chapter is on capitalist profits.

Profits from a Production Process

There are many kinds of capitalist firms that can make profits, but it is easiest to understand one that produces a physical product. Imagine Sew Fine, a garment factory that makes shirts. The factory has capital assets, which in this case are mainly its sewing machines and tools. Its production process is a kind of cycle from money to shirts and back to money, like the process shown in Figure 10.1.

The figure depicts the process of making profit as a kind of sandwich: two exchanges with a production process sandwiched in the middle. Sew Fine starts with money that it uses to buy fabric, thread, and buttons, and it also hires labor; that is the first exchange. Then comes the production process: Workers cut and sew the inputs into shirts and package them for sale, under the watchful eyes of the owner or manager and floor supervisors—a system of command. Finally, Sew Fine sells the shirts, usually to a wholesaler, who normally will sell them to various retail stores. Selling the shirts to the wholesaler is the second and final process of exchange from the standpoint of the factory. Of course, the idea is that the money collected from selling the shirts will normally be more than the money that was used to buy the inputs; otherwise the business will steadily decline in value and eventually go bankrupt.

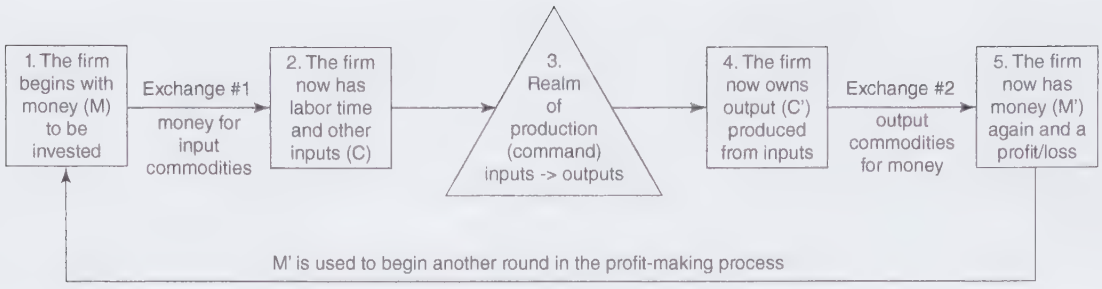


FIGURE 10.1 Exchange and command in the profit-making process.

This figure shows that profit making is a process involving five stages. In stage 1, a firm has a certain amount of money (M) to be invested. To get to stage 2, it carries out the exchange (#1) shown in the figure: It purchases the labor time, materials, and machines (capital goods)—the commodities (C)—required for production. The next move is from stage 2 (where the firm has the necessary inputs) to stage 3, the process of production itself (P). Production results in commodities (C′) that the firm now owns (stage 4). To get from stage 4 to stage 5, the output must be sold (exchange #2). Assuming that it is sold, the firm gets a sum of money (M′) that is either larger or smaller than the original amount of money (M). If it is larger the firm will have made a profit—if smaller, a loss. Whichever it is, at least a part of the proceeds from the sale of C′ is likely to be reinvested, and the entire process will begin again.

In the two buying and selling parts of the process (between stages 1 and 2 and between stages 4 and 5), the economic actors come together in markets, and their activities are regulated by supply, demand, competition, and other market conditions. We think of these as horizontal relations: people and firms buying and selling on a more or less equal footing. In the production part (stage 3), though, relations are vertical: The economic actors (employers and workers) come together in workplaces, and interactions take place between supervisors who issue commands and subordinates who carry them out.

Just as with commercial profit, money goes into the process (in step 1), and a larger sum of money comes out of it (by step 5). The difference is that with commercial profit there is no production process between the first and second exchange. Notice, too, that for the merchant making commercial profit, the entire process is horizontal in nature, with the merchant making exchanges on the open market on a more or less equal footing with the sellers from whom the merchant buys, and the buyers to whom he sells. The process of commercial profit making occurs entirely within the (horizontal) relations of the market: it does not involve (vertical) class relationships. In contrast, the capitalist production process, unlike the exchanges that come before and after it, is not an exchange between equals, but a process of *command* with the employer giving instructions and the workers carrying them out in stage 3. (Of course firms engaged in the business of making commercial profits might well have vertical relationships within them.) Moreover, commercial profit making is not a process that is found only in capitalism. It occurred in various societies long before—as well as since—the advent of capitalism.

The point of Figure 10.1 is that turning money into more money tends to become the main preoccupation of the firm's owners—not making shirts. Capitalism tends to create businesspeople who—even when overseeing the production of shirts—care less about shirts and more about turning money into more money. In principle, any production process could suffice to make that happen. It does not have to be shirts. In fact, it does not even have to be production of tangible goods as such: other processes like lending at interest, buying financial assets like stocks or bonds, or

speculating in real estate are ways to turn money into more money—ways to get a good return on invested capital.

Figure 10.1 merely depicts the stages in a capitalist profit-making process, however; it does not *explain* the existence of profits in a capitalist firm or economy. Explaining the existence of positive profits requires explaining why various prices are at certain levels, including the "price" of labor time (the wage rate), the prices of materials and capital goods, and the price of the output commodities. There will not be positive profit—profits will not "exist"—unless the total revenues generated by the sale of the output commodities are greater than the total costs of production, including wages, materials costs, depreciation, and so on. And total revenues cannot exceed total costs unless the average cost of producing a unit of the output commodity is less than the price of it.

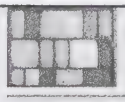
Rent refers to payments that firms are required to make to the owners of land, office space, buildings, or other facilities in return for the right to use or occupy them.

Calculating Profit and Other Property Incomes in the Whole Economy

Let us look more closely at how profits are calculated, both at the level of the firm and at the level of the whole economy. There are several kinds of payments that resemble profits in the sense that they are payments to owners of property that is used in production processes. The total amount of property income made in the economy as a whole includes a number of different types of payments that are made to the owners of capital assets such as the capital goods used in production. These types of payments include *rent*, *interest*, and *profits* of firms. In turn, profits can be divided into *dividends* and *retained earnings*.

The first type of payment is *rent*. Among the costs that a firm deducts from its sales revenue in order to calculate its profit is **rent** paid to owners of land, office space, buildings, or other facilities used by the firm to carry out its operations. A second type of payment that is also a cost deducted from revenue is the *interest* that the firm pays to bankers who have made direct loans to it, or the *interest* (coupon payments) to owners of corporate **bonds** that the firm has sold

A **bond** is an "IOU" that contractually commits the issuer of the bond (a government or a corporation) to making regular "coupon payments" to the owner of the bond, of a specified amount, at specified time intervals, as well as to make a final payment of the face value of the bond on the maturity date of the bond.



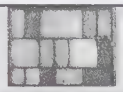
Two Cheers for Profits

[L]et us recognize an unassailable fact of economic life. All Americans will benefit from more profits. More profits fuel the expansion that generates more jobs. More profits mean more investment, which will make our goods more competitive in America and in the world. And more profits means that there will be more tax revenues to pay for the programs that help people in need. That's why higher profits in the American economy would be good for every person in America.

—Richard Nixon, October 7, 1971

Texas Instruments exists to create, make, and market useful products and services to satisfy the needs of our customers throughout the world. Our ability to meet those demands is determined by our innovative skills and measured by our profits. But that profit is not an inherent right. We are permitted to operate by the societies we serve and any profit we do make is our incentive as well as our reward for doing our job well. Society will pass judgment on our value. If we do not meet genuine demands we will not make a profit and will cease to exist.

—Mark Shepherd, Jr., CEO, Texas Instruments, testimony to Congress, 1980



Calculating Total Property Income

How much of the U.S. national income takes the form of profit or other property income? That is, how much is income other than wages and salaries? The answer to this question can be arrived at in various ways. Here is how we calculated it for 2014:

Type of Income	$ in Billions
Corporate profits	2,072.9
Proprietors' income*	673.4
Rental income	610.8
Net interest	532.3
Total property income	**3,889.4**
National income	15,076.5
Property income/National income	25.8%

* Proprietors' income is actually reported as $1,346.7 billion. We divided this number in half on the assumption that one-half of total proprietors' income should be counted as wages paid to proprietors in return for the labor they perform in the businesses (proprietorships) that they own.

Source: U.S. Bureau of Economic Analysis, *National Income and Product Accounts* (NIPA), Table 1.12 (lines 1, 9, 12, 13, 18) (https://www.bea.gov/iTable/index_nipa.cfm), July 30, 2015 update. In this table, "corporate profits" includes not only profits from corporations in the United States, but also those received by U.S. corporations and U.S. residents flowing from corporate profits abroad; and it includes both financial and nonfinancial corporate businesses. The first three categories include a capital consumption adjustment, and the first two categories (corporate profits and proprietors' income) also include an inventory valuation adjustment.

A share of **stock** represents a share of ownership of a corporation and entitles its holder to receive a **dividend**—a payment per share decided by the board of directors—whenever a corporation decides to pay out a portion of its total profits as dividends to its stockholders. The amount of the dividend varies

to investors to borrow money from them (see Chapter 5). After the firm deducts all costs from revenue, including the costs of rent and interest, so that what remains is before-tax *profit*, taxes on such profit must be deducted. Finally, out of the remainder, the *after-tax profit*, the corporation's board of directors decides how large a share of the profits it will distribute as **dividends** to shareholders. The dividend is defined per share of **stock**, and each shareholder receives the dividend per share times the number of shares she owns. The part of after-tax profit that is not paid out as dividends is set aside by the corporation as **retained earnings**, to be used for future investment or other purposes. Thus, the total amount of profit in an economy (R) is the sum of all the payments made to people or institutions that own, directly or indirectly, any private property used in production processes.

When looked at from the standpoint of the economy as a whole, total property income (shown for 2014 in the box, "Calculating Total Property Income") can be seen as the monetary value of the economy's surplus product. This is because profit is what remains of total sales revenues after all costs of production have been paid,

Retained earnings are the part of total profit that a corporation does not pay out as dividends, but keeps for future investment or other purposes.

including the depreciation of capital goods as well as the costs of the materials and labor used in production.

Profit does not, of course, include wages, salaries, or other income received as a result of labor. It is sometimes difficult to draw the line between different forms of income, especially when the same person is receiving both a salary and a return from ownership and control of productive assets, but income from labor is certainly different from property income and is therefore not a part of profit. There are other forms of personal income, such as gifts, winnings from gambling, and the proceeds of theft, but they are not regarded as part of capitalist profit, either. The box, "Calculating Total Property Income" shows that property incomes of various types listed made up about one-fourth of national income in 2014. Over half of that ($2.1 trillion out of $3.9 trillion) was corporate profit.

Profit in a Grain-Growing Capitalist Economy

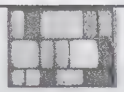
When we discussed the concept of the surplus product in Chapter 5, we introduced a grain model with grain-growers producing a grain surplus. But we presented this model without connecting it to any particular economic system or class structure. We now bring back the grain model, this time applying it to a capitalist society in which capitalists own all the land and draft animals, as well as any grain storage facilities or animal enclosures. Grain is the only commodity. Thus, in order to survive, each family must sell its grain-growing labor time for wages to a capitalist. The total output of grain will also belong to the capitalist. (See box "Capitalism Comes to Grain Growing.")

Because profits are simply the surplus product measured in monetary terms (rather than bushels of grain), the amount of profit is determined by the same factors that determine the size of the surplus product. However, profit in a capitalist economy differs from the surplus product in our simplified grain example in the following ways: The capitalist economy, like most economies, produces a variety of different outputs, not simply grain, and workers are, obviously, not paid in grain. This means that we will not be able to measure all the outputs and inputs in bushels or any other physical measure. We will have to measure them in terms of their monetary, or market, values. Thus, profits are necessarily calculated in *monetary* terms. In the grain-growing economy, profits are $2,000 for every worker employed by the enterprise.

Calculating the Rate of Profit

The **rate of profit** is the total amount of profit divided by the value of the capital goods owned by a firm.

The most important thing we are interested in is the *ratio* of the amount of profit to the value of the capital invested in a firm, which we take to be approximately the value of the capital goods owned by the firm. This ratio, which is the amount of profit divided by the value of the capital assets (which we take here to be the value of the capital goods), is the **rate of profit**. It is important to distinguish this ratio from the *amount* of profit R. A big firm will typically make a large amount of profit, while a small firm will make a smaller amount. The rate of profit indicates



Capitalism Comes to Grain Growing

Suppose that a farm produces only grain and that the average producer on this farm works 1,000 hours a year to produce 100 bushels of grain. Of this, 30 bushels are used to replace the seed, feed the draft animals, and take care of other maintenance activities, leaving a net output of 70 bushels a year per producer. (This resembles the grain model in Chapter 5).

Now suppose that this farm is part of a capitalist society and that its producers are workers hired by a capitalist. Assume that their wage is $5 per hour, that the market price of grain is $100 per bushel, and that a typical worker spends an annual income of $5,000 by purchasing 50 bushels of grain from the capitalist. These 50 bushels represent the customary level of consumption for workers.

The net output of 70 bushels per worker is owned by the capitalist. Apart from the 50 bushels of this which each worker buys, the other 20 bushels are exported for $100 per bushel, and the $2,000 received is spent by the capitalist on imported consumption goods. Total annual sales revenue is $7,000, of which $5,000 is paid to the workers, leaving the capitalist with a profit of $2,000 per worker.

These relationships can be summarized as follows, where all calculations are for just one average worker's production (and consumption) in this enterprise:

In Bushels		In Dollars	
Total output	100 bu.	Total revenue (assuming that all the output is sold)	$10,000
Grain to be set aside for materials and capital goods for next year	30 bu.	Materials used up (feed and seed)	$3,000
Net output	70 bu.	Value of net output	$7,000
Consumption of producers	50 bu.	Wages paid	$5,000
Surplus product	20 bu.	Profit	$2,000

how well each firm is doing *per dollar of capital goods owned*. In mathematical terms, the rate of profit is:

(10.1)
$$r = \frac{R}{K}$$

where r = rate of profit
R = amount of total profit (dollars per year)
K = value of capital goods owned (dollars)

In short, the profit rate is an important indicator of the degree of success in the profit-making process. It indicates how much profit was made *in relation to* how much money was invested in capital goods to make the profit.

Equation 10.1 can be used to calculate the rate of profit of a particular company. For example, Philip Morris International stated in its annual report for 2014 that it owned assets of various types worth $35.2 billion, and in the same report it noted

The **after-tax profit rate** is the profit rate made by firms after they have paid all required taxes.

that it had made a profit, before taxes, of just under $10.7 billion. Thus, its rate of profit before taxes was 30.4 percent. Subtracting the $3.1 billion in taxes it paid that year, Philip Morris ended up with an "after-tax" profit of $7.7 billion in 2000, providing it with an after-tax rate of profit of 21.9 percent.

Similarly, Equation 10.1 can be used to calculate the profit rate (r) for an entire economy, such as that of the United States. In this case R is the total amount of profits, and K the value of the economy's total capital stock. Most economic activity in the United States is conducted by *corporations*. They account for about three-fourths of the nation's total payrolls if we count among corporations those businesses organized as "S-corporations," a legal form of organization designed for small businesses. Here we calculate the profit rate for the U.S. corporate business sector.

In 2012 total corporate profits before taxes (R) in the United States amounted to $1.6 trillion. In that same year American nonfinancial corporations had private nonresidential fixed assets (K)—including equipment and software as well as structures—valued at $15.2 trillion. Thus the overall profit rate of U.S. corporations in 2012 (calculated using Equation 10.1) was 10.7 percent.[2]

The Corporate Profit Rate in the United States

Figure 10.2 shows how the U.S. corporate profit rate, both before and after taxes, varied in the second half of the twentieth century. The trend was generally downward, with the corporate profit rate before taxes hovering at around 13 percent in the two decades following World War II and then declining to an average of 8 percent after 1970. After-tax profits were of course lower, averaging 7 percent to 1970 and 5 percent thereafter. From about 1970 on, however, the largest firms were increasingly becoming multinational by moving some of their operations outside the United States. In part because of their accounting practices (discussed in Chapters 7, 15, and 19), this tended to reduce both their reported before-tax and after-tax profit in the United States. So the decline shown in the average profit rate does not necessarily represent a decline in the fortunes of U.S.-based multinational corporations, because there is some evidence that profits were greater in overseas operations, which are not included in Figure 10.2.

What is shown in Figure 10.2 is the rate of profit (r) of the *average* company in the U.S. corporate business sector. This rate of profit is calculated by dividing the profit (R) made by all American companies each year by the total value of their capital stock (K) in the corresponding year. Thus, the profit rates shown in Figure 10.2 are averages, from year to year, for the whole corporate business sector. In any given year, of course, some companies will do better than the average and others will do worse.

Notice in Figure 10.2 that the distance between the before-tax profit rate and the after-tax profit rate shrank between 1948 and 2012, reflecting a decline in the effective tax rate on corporate profits during this period. However, even though the profits do not include profits on foreign operations, the taxes *do* include taxes paid to foreign governments, and both federal, state, and local taxes in the United States. Chapter 19 will discuss taxes in more detail.

---

[2]Total corporate profits here is a different category from corporate profits in the box, "Calculating Total Property Income." These data are from the U.S. Bureau of Economic Analysis NIPA accounts, Table 1.14, line 27, and Table 4.1, line 17.

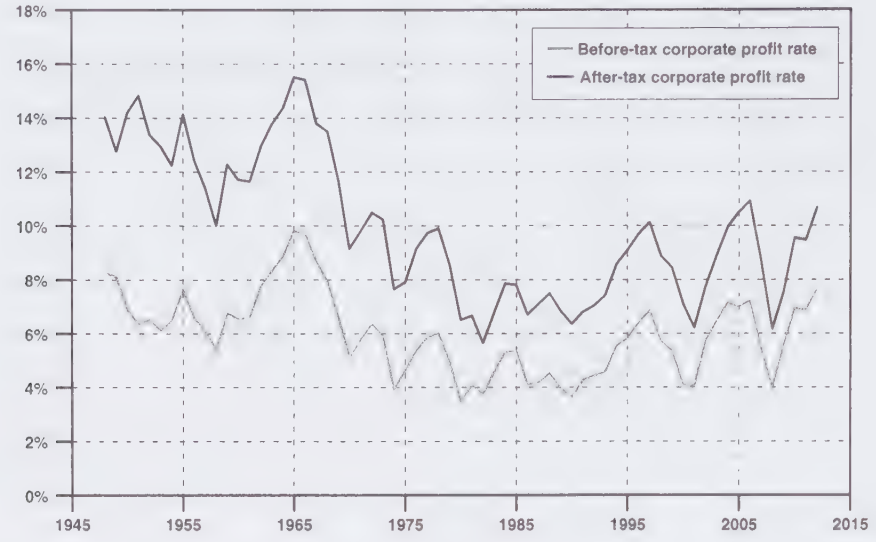


FIGURE 10.2 The U.S. corporate profit rate, 1948–2014.

This figure traces the rate of profit, before and after taxes, from 1948 to 2014. In each year, it is the total profit made by the corporation in the U.S., divided by dollar value of the capital stock in the same year. This includes both nonfinancial and financial corporate businesses in the United States. Excluded from the data are profits received by U.S. corporations on their foreign operations, as well as the value of capital stock owned outside the U.S. There is some evidence that corporate profits recorded by U.S. firms on foreign operations have often been higher than the profits they have reported on domestic operations. This is evidently in part because corporate accountants are able, within limits, to shift reported profits to countries or jurisdictions where they will be less heavily taxed. The drop in reported profit rates after the mid-1960s may therefore be at least partly due to the increase in the foreign operations of U.S. corporations, and the use of such accounting practices.

Sources: U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts (NIPA)*, https://www.bea.gov/iTable/index_nipa.cfm; profit data from Table 1.14: "Gross Value Added of Domestic Corporate Business in Current Dollars and Gross Value Added of Nonfinancial Domestic Corporate Business in Current and Chained Dollars," lines 11 and 13, May 29, 2015 release; capital stock data from Table 4.1: "Foreign Transactions in the National Income and Product Accounts," line 9, May 29, 2015 release.

Corporations and Other Businesses: Who Gets the Profit or Bears the Loss?

In describing what businesses do, we often use the generic term *firm* to cover both corporations and unincorporated businesses. In fact, as explained in Chapter 6, there are several ways a firm can be legally organized, and to understand profit it is important to learn something about these different legal forms and how they operate.

In one form or another, for thousands of years, groups of people, small or large, have carried out commerce, produced handicrafts and other tradable products, and offered services for sale. But it was in the nineteenth century that corporations emerged as a legal form accompanied by legal rights and status that fostered their rapid growth. As Chapter 6 explained, by the mid-nineteenth century in the United States, corporations gained the great advantage of limited liability. By late in that century, corporations also secured status as "legal persons" in the United States, meaning that they were granted legal rights similar to the legal

rights of human beings ("natural persons"). A similar evolution happened in other developed countries.

Corporations

Today the majority of business is done by corporations. A corporation, by definition, issues stock, and an owner of a share of stock owns one piece of the corporation, with the right to share in any profits it distributes to stockholders, as well as to vote on matters that are put before the stockholders. Yet in the United States, and indeed in most countries of the world, the majority of businesses still are small enterprises that are not incorporated. A very few giant firms are also not corporations, that is, they are "closely held," owned by an individual or a few individuals; these include a few multinational grain traders, for example, and at least one large U.S. company that produces oil and products derived from oil.

A corporation is one of three main ways in which a business can be organized. Firms owned by single individuals have long been called *proprietorships*, and more than 70 percent of the 33 million firms in the United States are proprietorships. Although relatively easy to set up, proprietorships usually have difficulty raising money to finance their operations. Once established, however, this type of business benefits from a direct connection between ownership and control: every dollar made above costs is a dollar earned by the proprietor, so there is a strong incentive for him or her to work hard and diligently. Moreover, a proprietor legally has total control over every aspect of his or her enterprise, although at times there may be obstacles to exercising real control.

Two or more people who go into business together often form a *partnership*. Partnerships are less numerous and more complicated than proprietorships. About 10 percent of U.S. firms are partnerships, and they must operate on the basis of a legal document (a "partnership agreement") that specifies what each partner's contribution to and reward from the business will be. Thus, each partner's share of the profits is agreed upon in advance. Partnerships are commonly used by professionals—from lawyers to singing groups—to organize their business activities.

One thing that proprietorships and partnerships have had in common in the past is that in both cases the firm's owners or partners bear *unlimited personal liability* for paying back all of its debts. If worse comes to worst, a creditor—for example, a bank—has the right to seize and sell the assets of the firm *and the personal property of the proprietor or partners* to recoup the value of an unpaid debt. However, see the discussion below of the "limited liability company" (LLC), a form introduced in the United States in the twentieth century, which also offers the advantage of limited liability, but makes it available to smaller businesses.

LIMITED LIABILITY. A *corporation* differs from an ordinary partnership or sole proprietorship, in that its investors have only *limited liability* for its debts or other obligations. At most, the owners of a corporation—people who own shares of its stock—can lose only the value of their stock; their other assets, business and personal, cannot be appropriated by the firm's creditors.

A newer form of business, the "limited liability company" (LLC), introduced in most states in the United States by the end of the twentieth century, is a kind of cross between an ordinary partnership or sole proprietorship and a corporation, which offers a form of limited liability (and often some tax advantages). Here we are more interested in corporations.

The "limited liability" aspect of a corporation gives it a distinct advantage over proprietorships and partnerships. Whereas the capital assets of the latter types of businesses are generally limited by what the proprietors and partners (and their friends) can bring to the enterprise, a corporation is in a position to raise money for its operations by selling shares of stock to outside investors. As a result, corporations tend to be larger than proprietorships and partnerships; though they comprise less than 20 percent of U.S. businesses, they account for over 80 percent of U.S. business revenues. On a global scale the largest one thousand corporations produce about four-fifths of the world's marketed output.

Corporations that sell stock to outside investors are said to be selling shares to "the public." Hence, the terms *public* and *publicly-owned* are used to describe such enterprises. In fact, however, "public" corporations are not directly accountable to the public. Corporations have long been required to obtain a *charter* from some political authority, and the political authority granting the charter has had the right to supervise and regulate the practices of the chartered corporation. Recall that the first European settlers in America were sponsored by the Massachusetts Bay Company, a corporation chartered by the king of England. In the United States today a corporation must obtain its charter from a state government. In the early nineteenth century it was somewhat difficult to get such charters—and state governments generally kept a fairly close eye on the practices of the corporations they chartered. These days, however, state governments rarely look into the affairs of corporations. Corporations shop around among states for the best bargain, and many end up incorporating in states such as Delaware that offer exceptionally favorable rules for corporations. To be sure, there are federal laws governing the activities of corporations, such as those subject to enforcement by the Securities and Exchange Commission. For example, these laws establish requirements that corporations report on their activities in certain ways that privately held (unincorporated) businesses are not required to do.

FINANCING BY ISSUING BONDS OR STOCK. Corporations have two main ways of raising money for financing their activities. One is to borrow from the public by selling *bonds*. Someone who buys a corporate bond whose face value is $1,000 is lending money to the corporation in return for a fixed schedule of fixed repayment amounts. For example, the corporation might promise to pay the bondholder $100 per year as a "coupon payment" for each of ten years, and then pay $1,000 after ten years are up ("after the bond matures"). During this time, the bond can be resold on the *bond market*, for a price that usually will differ from the $1,000 face value listed on the bond.

The second way a corporation can raise money is by selling shares of stock to the public. Anyone can buy stock, including those who work for the corporation—and that can, and often does, include those who owned the firm before it became incorporated. In this way they can try to maintain control of the firm, because only the major shareholders are likely to have much influence on what the firm does. For each share of stock a shareholder owns, she is entitled to receive whatever dividend payment the corporation decides to pay out of its profits in a given year. If a fifty-cent dividend is paid, a holder of twenty shares gets a dividend payment of $\$0.50(20) = \$10$.

However, usually investors do not have a personal interest in a corporation in which they invest. They hold its stock in order to make money, either from the

dividends the corporation may pay or from a rise (*appreciation*) in the value of the stock. Though obligated to pay interest to their bondholders, corporations do not have to pay dividends to their stockholders. When they decide to do so, they pay the dividends out of their after-tax profits.

MANAGEMENT SEPARATED FROM OWNERSHIP. The most important difference between a corporation and the other two forms of business organization (proprietorship or partnership) is the separation of a corporation's management from its ownership. In this context the term *management* refers to the directors of a corporation as well as the managers they hire; ownership lies with the stockholders. Although ownership of a share of (voting) stock confers a legal right to participate in the governance of a corporation, the diffusion of share ownership among a large number of people generally makes this right meaningless except for large stockholders. And although a stockholder has the right under certain conditions to submit a proposal for consideration at the annual meeting of a corporation, the management's collection of "proxies" normally gives it the power to defeat all proposals other than its own. Thus, unless one holds a substantial fraction of a company's stock, the only effective way to express displeasure with the management of a corporation is to sell one's shares in a stock market or to attend the stockholders' meeting and stand up and say something.

The separation of management and ownership in early corporations led Adam Smith to warn: "The directors of such companies . . . being the managers rather of other people's money than their own, it cannot well be expected that they should watch over it with the same anxious vigilance with which the partners in a private co-partnery frequently watch over their own. . . . Negligence . . . must always prevail, more or less, in the management of the affairs of such a company."[3] Smith's words still have the ring of truth today.

The directors of a corporation are supposed to represent the interests of its stockholders—they are, in fact, elected, at least formally, by the stockholders—and the directors and the managers they hire are together responsible for maximizing the profits of their companies. The term *corporate governance* refers to the system used to run a corporation. It involves, in particular, the relationships among directors, top managers, and stockholders. In recent decades corporate governance became a hot topic as the accounting and governing practices of companies such as Enron and WorldCom came under scrutiny in the late 1990s. Questions have been raised about the independence of directors from their managers and of accounting firms from the corporations they audit.

In many corporations the chief executive officer (CEO) is also the chair of the board of directors. In addition, ties of friendship or business interests often connect a majority of the directors with the CEO. In such situations the directors might not exercise independent judgment with regard to the practices of the top managers of their companies, and they might have a hard time firing the CEO even if the company is being poorly run. Similarly, if the accounting firm hired by a corporation to audit its financial statements has lucrative consulting contracts with it in other areas such as legal services, the accounting firm will not want to risk its other business with the

---

[3]Adam Smith, *The Wealth of Nations* (New York: Random House, The Modern Library, 1937 [1776]), Book 4, Chapter 2.

corporation by pressing it to issue accurate financial statements. Thus the large size of a corporation can give it some power (command) over other businesses whose services it contracts for. Chapter 18 describes some examples of how this contributed to the devastating subprime mortgage crisis of 2007–2009.

When the Enron Corporation went bankrupt in 2001 after numerous irregularities in its accounting practices had been revealed, there was a widespread loss of trust in the governance and financial reporting of U.S. corporations. The ensuing decline in stock prices was at least partly due to a rising skepticism among investors about their ability to estimate the profitability of corporations. In 2002 the U.S. Congress passed the Sarbanes-Oxley Act to institute reforms such as one that requires corporate boards to have a majority of "outside" directors. The jury is still out on the effect of the Sarbanes-Oxley reforms, but the corporation, in one form or another, will continue to be the dominant form of business organization.

Determinants of Total Profit and the Profit Rate

The **determinants of the profit rate** are the things upon which the profit rate depends; they determine how high the profit rate will be.

The profit rate depends on (is determined by) the **determinants of the profit rate**. Knowing the value of each of these determinants will allow us to understand why the profit rate is high or low. Even when we do not know their values, but do know which determinants are rising and which are falling, this helps us to know whether the profit rate is rising or falling.

We already know some of the determinants of the profit rate from our analysis of the surplus product in Chapter 5. To see exactly how these determinants affect the profit rate and to discover what other determinants play roles, we need to derive the precise relationships between the profit rate and its determinants. For nearly all of this chapter, we regard the value of the capital stock K as constant. This allows us to examine what determines the total profit R, because when K is constant, the profit rate $r = R/K$ will always move in the same direction as the total profit R. We therefore focus initially on the determinants of the total dollar amount of profit R. We will eventually return to the profit rate.

In Chapter 5 we saw that the total product of a whole economy, whether or not that economy is capitalist, can be divided into three parts. One part is used to replace the materials used up in production, including the wear and tear on any capital goods (tools, machinery, equipment, structures, and so forth). A second part goes to workers to sustain themselves and their families. Whatever remains is the *surplus product*. While those categories were for a whole economy, we can apply almost exactly the same categories to describe a single capitalist firm's output. But in the capitalist firm we measure everything in money terms, so we give the categories slightly different names, as shown in Figure 10.3. Instead of *total product* we refer to the *value of total output* (in a year, for example), which equals the price of output P times the number of units of output Z. For example, the value of total output for a shirt factory producing 100,000 shirts per year of one single variety, for \$20 per shirt, would be $PZ = \$20(100{,}000) = \$2{,}000{,}000$. We will refer to the *value of total output* interchangeably with *sales revenue*, on the assumption that all output is actually sold.

The dollar value of a firm's total output PZ can be divided into the three parts described in the previous paragraph, and shown in Figure 10.3: *material inputs*, *labor cost*, and *profit*. The *cost of material inputs* consists of the costs of both (a) materials that are used up in production and (b) capital goods replacement (depreciation), since

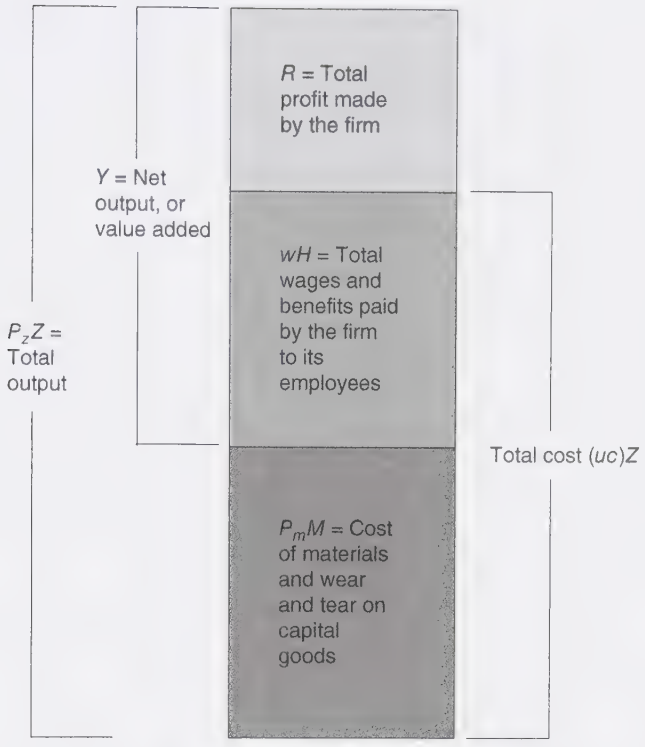


FIGURE 10.3 Division of a firm's total output into material input costs P_mM, wage cost wH, and total profit R.

This shows the case in which the only kind of cost is costs per unit of output produced. Chapter 11 analyzes the case in which another category of cost, called *fixed cost*, is added, and this will change the picture somewhat. The diagram also shows that *net output* Y is total output minus the materials and portion of capital goods used up in production, and that *total cost* $(uc)Z$ is the sum of labor and materials costs, which equals the unit cost uc times the number of units produced Z.

in economic terms these are both *intermediate inputs*, that is, physical inputs into a production process. In shirt production, the thread and fabric used are materials used up in production (they show up in the final product, although things truly used up, like energy, also count in the same category)—they are not "final" goods, like the shirts themselves that are produced for people's consumption. The *cost of material inputs* is written as the price of the material inputs P_m times the quantity M that is used up, that is, P_mM.

The (hourly) **wage rate** is used here to refer to total wages and benefits paid to a worker, divided by the total number of hours worked.

Similarly, the total *labor cost* (W) is the **wage rate**, or the hourly wage w, times the total number of hours H worked by all workers in the firm in a year. The "hourly wage" here actually includes not only the wage, but all of *employee compensation*, including the monetary value of benefits provided. These can include such things as paid holidays, paid vacation days, paid sick leave days, and paid parental leave. In the United States, benefits sometimes also include the value of the employer's contribution, if any, to the worker's healthcare costs. (Most other high-income countries, and some middle-income countries, have some form of universal healthcare system paid for by the government out of taxes, a system that reduces administrative costs considerably; but in the United States, employer contributions to health care can still be an important element of total compensation.) The monetary value of all these

benefits is added to total wages in the firm, and then that total is divided by total hours worked to find w. The total labor cost is written as wH.

What is left over from sales revenue after both non-labor and labor input costs are paid is the *profit* R, in dollars. Putting all this together, we have (as in Figure 10.3):

Total value of output = Cost of material inputs + Labor cost + Profit

(10.2) $$P_z Z = P_m M + wH + R$$

where

P_z = price of output, in dollars
Z = number of units of output produced per year
$P_z Z$ = value of total sales (dollars per year), or *total revenue*
P_m = average price of intermediate inputs used in a year, in dollars
M = amount of intermediate inputs (materials + capital goods) used up in a year
$P_m M$ = cost of all intermediate inputs (materials used up + wear and tear on capital goods) in a year
w = hourly total compensation (including both wages and benefits per hour)
H = total hours worked in a year by all those employed in the firm
wH = total labor cost in a year
R = profit, in dollars

We can group the three main elements of the total value of output in two different ways, shown in Figure 10.3. The bracket on the left side labeled *Net output, or value added*, identifies the value newly created (denoted Y), which is

Net output = Total revenue – Materials costs = Wages + Profit

or

(10.3) $$Y = P_z Z - P_m M = wH + R$$

The value newly created is the total value of output minus the cost of previously existing material inputs; and because the newly generated income is paid out partly as wages and partly as profits, it is also the sum of wages and profits (and other property incomes, which we leave out to keep things simple). This newly created value is the **value of net output** (Y) or **value added** (or simply net output, for short).

The **value of net output** is the dollar value of the total output minus materials and machine costs.

Another way to group the same three elements is indicated by the bracket on the right hand side of Figure 10.3. Its purpose is to emphasize that a capitalist firm requires positive profit, at least on average, to be viable, something which only happens if sales revenue is larger than the sum of costs. To make this point we group the elements of the total value of output as follows:

Profit = Total revenue – Total costs
= Total revenue – (Materials costs + Labor costs)

or

(10.4) $$R = P_z Z - (P_m M + wH)$$

In this equation, and in the rest of the book, we use M inclusively to refer to the costs of both (a) materials used up and (b) capital goods replacement (depreciation), since, in economic terms, these are both *intermediate goods*. These are considered to be "intermediate" because they are inputs into a labor process (like cotton used in the production of shirts and the wear and tear on the sewing machines used to produce the shirts) as opposed to "final" goods that are produced for people's consumption (like the shirts themselves). We also assume for now that all output of the economy is sold, which lets us use $P_z Z$ to represent both total output and sales revenue.

Example: The Good Cod Fishing Company

To illustrate the concepts introduced here, imagine a fishing enterprise, the Good Cod Fishing Company, that owns ten nets and employs ten workers, with each worker using one net. The workers catch fish by throwing the nets out from the shore, and each worker brings in an average of 25 pounds of fish per hour. To simplify things we assume that there are no other capital goods or materials used apart from the nets. The workers are paid $18.50 an hour, and they each work 100 hours a year for Good Cod. (We assume here that the fish come close to the shore only 100 hours each year and that the workers have some other job during the rest of the time.)

Each net costs Good Cod $1,000, and it wears out after 200 hours of use, which in the case of Good Cod means after two years of normal use. Half of Good Cod's nets (that is, five) wear out each year and are replaced at the beginning of each year when Good Cod buys five new nets at a total cost of $5,000. These are added to the five one-year-old partly worn nets it already has, which are now worth $500 each, half their original price. So Good Cod pays $5,000 each year for material inputs $P_m M$. Finally, Good Cod is able to sell its fish for $1 per pound, so that with 10 workers each working 100 hours per year and catching 25 pounds of fish per hour, total sales revenue is ($1/lb. fish)(10 workers)(100 hrs./worker)(25 lb. fish/hr.) = $25,000. Notice that most of the units in this expression (lb. fish, workers, hrs.) cancel out.

Notice too that the $5,000 of materials used up annually is a *flow* of inputs into the production process, and, like all flows, is measured per unit of time, in this case per year. In fact, the one quantity that is a *stock*, not a flow, is the $7,500 of capital invested in the firm (K), that is, the new nets ($5,000) plus the one-year-old nets ($2,500). That capital just sits there, tied up and unavailable to serve any other purpose. By the way, we have assumed—to keep matters simple—that the capital stock of Good Cod is not changing over time, although of course in principle a firm can always add to its capital stock by investing in new equipment. So we assume that each year when Good Cod buys new nets, it buys just enough to replace those that have worn out.

How much total profit does Good Cod make each year? The calculations are shown in Table 10.1. We just calculated that total sales revenue ($P_z Z$) is the price of fish P_z ($1 per pound) multiplied by the 25,000-pound annual catch, or a total of $25,000. We also found that $P_m M$ = $5,000. The value of net (sorry!) output $Y = P_z Z - P_m M$ is therefore $25,000 − $5,000 = $20,000. Good Cod pays $18.50 an hour in wages to each of 10 workers for 100 hours each (1000 hours annually), so wH = $18.50(1000) = $18,500. Total profit ($R = Y - wH$) is therefore

(10.5) $R = Y - W = (P_z Z - P_m M) - wH = \$25{,}000 - \$5{,}000 - \$18{,}500 = \$1{,}500$

TABLE 10.1 TOTAL PROFIT *R* AND THE PROFIT RATE *r* FOR THE GOOD COD FISHING COMPANY

VARIABLE	DEFINITION	PRICE	QUANTITY	VALUE = PRICE X QUANTITY
Value of output	P_zZ	$1 per pound of fish	25,000 pounds	**$25,000**
Value of materials used up + wear and tear on capital goods	P_mM	$500/net per year	10 nets	**$ 5,000**
Wages and benefits paid	wH	$18.50/hr.	1000 hrs. (10 workers × 100 hrs. each)	**$18,500**
Net output (Value added)	$P_zZ - P_mM = Y$			$25,000 − $5,000 = **$20,000**
Profit	$P_zZ - P_mM - wH = Y - wH = R$			$20,000 − $18,500 = **$ 1,500**
Value of capital stock	K	new net: $1000 1-year-old net: $500	5 new nets + 5 one-year-old nets	**$ 7,500**
Profit rate	$R/K = r$			$1,500/$7,500 = **20%**

Finally, since the value of capital goods owned (K) is $7,500, Good Cod's rate of profit is:

(10.6)
$$r = \frac{R}{K} = \frac{\$1{,}500}{\$7{,}500} = 20\%$$

Profit per Unit of Output

To make it easier to examine the individual determinants of profit, it turns out to be enormously useful to transform the profit Equation 10.4 into an equation about *profit per unit of output*. We do this by dividing through both sides by the number of units of output Z to get an expression for R/Z, which indeed is the profit made from each unit of output, on average. When we get to Chapter 11, this step will pave the way for an easy and natural transition to issues we discuss there, including the all-important issue of what advantages large firms have over small firms, and why some sectors are dominated by large firms.

Dividing both sides of Equation 10.4 by Z, we get:

(10.7)
$$\frac{R}{Z} = P_z - \left(\frac{P_mM}{Z} + \frac{wH}{Z}\right)$$

This says that the *profit per unit of output* equals the output price P_z, minus the **unit cost** or costs per unit of output—costs that include both the **unit materials cost** or materials cost per unit of output P_mM/Z and the **unit labor cost** or the labor cost per unit of output wH/Z. Using standard terms for these costs, we can write:

Profit per unit of output = *Price* − (*Unit materials cost* + *Unit labor cost*)

or

(10.8)
$$\frac{R}{Z} = P_z - (umc + ulc)$$

Unit materials cost is the portion of the average cost of producing each unit of output that is spent on materials used in producing the output, plus wear and tear on the machines and tools used. It is therefore the total cost of materials and wear and tear on machines, divided by the number of units of output produced.

TABLE 10.2. PROFIT AND ITS DETERMINANTS PER UNIT OF OUTPUT, FOR THE GOOD COD FISHING COMPANY

VARIABLE	DEFINITION	PRICE	QUANTITY	VALUE = PRICE × QUANTITY PER UNIT OF OUTPUT
Price of output	P_z	$1.00		$1.00/lb.
*Unit materials cost (**umc**)*	$P_m M/Z$	$500/net	10 nets/25,000 lbs. fish	$0.20/lb.
*Unit labor cost (**ulc**)*	wH/Z	$18.50/hour	1000 hrs./25,000 lbs.	$0.74/lb.
Profit per unit of output	$P_z - P_m(M/Z) - w(H/Z) = R/Z$			$1.00 – $0.20 – $0.74 = **$0.06/lb.**

Unit labor cost is the labor portion of the average cost of producing each unit of output; it is the total wage bill (hourly wage times total hours worked) plus benefits, divided by the number of units of output produced.

where

umc = unit materials cost, or materials cost per unit of output = $P_m M/Z$
ulc = unit labor cost, or labor cost per unit of output = wH/Z

Let us interpret Equation 10.8 for the Good Cod company, by finding its profit per unit of output, to get a feel for how it works. We can convert the numbers in Table 10.1 to a "per unit of output" basis, shown in Table 10.2, as follows. The unit of output is one pound of fish, and as we know, the price per unit of output is $1.00 per pound of fish. How much of this dollar do costs eat up, and how much profit is left?

First, $P_m M/Z$ is the unit materials cost (the cost of materials, including wear and tear on capital goods, that are used up in making a unit of output). We know that the materials cost $P_m M$ is the $5,000 spent on replacing worn-out nets at the beginning of each year, and this cost is spread out over 25,000 pounds of fish (which is Z), so the materials cost per pound of fish is $P_m M/Z$ = $5,000/(25,000 lbs.) = $0.20 per pound.

As for unit labor cost, the total wage bill wH is $18,500 (see Table 10.1), and when this is divided by the annual output of 25,000 pounds of fish, we get $0.74/lb. as the labor cost per unit of output. (Alternatively, we could just look at an hour's wages and output, since every hour is the same, and find that $18.50/25 lb. = $0.74 in labor cost per pound of fish.)

So of the $1.00 value of a pound of fish, what remains as profit per unit of output, after costs are deducted, is $1.00 − $0.20 − $0.74 = $0.06 profit per pound of fish. The same number could also be found by taking the $1,500 per year total profit and dividing it by the 25,000 pounds per year total output of fish: $1,500/25,000 lb. also equals $0.06/lb. of fish.

Notice, by the way, that the profit *rate* is the total profit divided by the value of the capital stock, which is $1,500/$7,500 (see Table 10.1). Be careful: although only six cents of total revenue are profit, do not be fooled into thinking that the profit rate is $0.06/$1.00. Remember that the profit rate is the total profit *divided by the value of the capital goods owned by the firm* (*not* divided by total revenue).

Conflict over the Profit Rate

A firm owner or manager looking at Equation 10.7 or its simplified form, Equation 10.8, might quickly, and almost correctly, conclude that there are three main ways to try to increase the firm's profit rate: (1) increase the price of output, (2) reduce the

materials cost per unit of output, or (3) reduce the labor cost per unit of output. We discuss these in turn.

A firm can always try increasing its selling price, but typically this will affect other determinants of the profit rate. For example, raising the selling price will reduce the quantity of output sold, which in some industries can increase the labor cost per unit of output. Suppose, for instance, that it takes two hours to set up the machines and tools to run a certain job (producing a certain item)— during which time nothing is being produced. Suppose the hourly wage is \$10. Suppose also that 50 items can be produced per hour once the job is running. This means that to produce 200 items will take 6 hours (2 hours set-up, 4 hours run time). The unit labor cost (*ulc*) for this production run will be 6(\$10)/200 = \$0.30 per item. But now, if a price increase reduces sales to 150, producing this smaller quantity will take 5 hours (2 hours set-up, 3 hours run time), and the *ulc* will be *higher*: 5(\$10)/150 = \$0.33. This is an example of one type of *economies of scale*: producing more units of output will reduce the average cost (producing fewer will raise it). Chapter 11 will elaborate on the nature and implications of economies of scale.

Intermediate Inputs per Unit of Output as Profit Rate Determinants

Profit is what remains after costs are paid. So any increase in a firm's costs cuts into the amount of revenue left as profit, and any decline in costs will increase profit, as long as no other determinants of the profit rate change at the same time.

For example, if a corporation has to use more materials or machines (M rises), or has to pay more for them (P_m rises), its profit rate will fall. If a new law requires the corporation to install "scrubbers" in its smokestacks to protect the environment from harmful emissions, this requires the firm to use up more capital goods (increasing M), thereby reducing its rate of profit. (It also likely increases K, the total value of capital goods, which increases the denominator of the profit rate $r = R/K$, and so also helps reduce the profit rate.)

Similarly, if there is a rise in the price of the oil used as an input (increasing P_m), the firm's profit rate will also fall. Conversely, if it can find a cheaper source of the materials or machines that it uses—or find a new technology that uses smaller amounts of them to produce the same amount of output per hour of labor—its profit rate will rise. As usual, we mean that this happens as long as other determinants of the profit rate do not change.

This is true on a "per unit of output" basis as well. If a firm can find a way to reduce the total value of material inputs used in producing a unit of output (P_mM/Z)—that is, the *unit materials cost* (*umc*)—the firm can raise its profit per unit of output, provided there are no changes in other determinants of output that offset the profit increase.

Firms typically search for lower-cost inputs, either by finding the same inputs at a lower price, or by switching to substitute materials that may or may not be of similar quality. Another strategy is to reduce waste of inputs in the production process, for example through better engineering such as computer optimization of patterns cut from fabric for garment production, or finding a use for fragments of fabric.

But some strategies for cutting input costs run a high risk of provoking conflict with workers, customers, suppliers, or others. Sometimes firms seek lower-priced versions overseas of the same input materials they have always used. In the case of crude oil and valuable minerals, this search has sometimes been aided by armed U.S.

intervention abroad, which sometimes creates greater access to such resources for U.S.-based firms. But such actions can be expensive; the report *War at Any Price?* by the Joint Economic Committee of the U.S. Congress in 2007 estimated that the wars in Iran and Afghanistan would impose a total cost of around $50,000 on the average U.S. household by 2017, and later research confirms the validity of this estimate.

Using cheaper tools and machines, or less of them, sometimes means not providing necessary safety equipment, and so endangering worker health or safety. In such cases, it can create conflict with workers, who naturally object to being exposed to such dangers. In other cases, consumers can be the ones affected, as in reported cases of adulterated baby food in China, or dangerous toys produced by firms cutting corners to keep down costs. Lowering intermediate input costs in an attempt to increase profit may not always generate conflict with workers, consumers, or others, but it does have the potential to do so.

Unit Labor Cost as a Profit Rate Determinant

Lowering the cost of labor per unit of output, or *unit labor cost* (*ulc*), is another common strategy that firms use to try to make their operations more profitable. As Table 10.3 shows, the *unit labor cost* is defined as the hourly wage, times the total number of hours worked, divided by the number of units of output produced, or wH/Z. This expression can be rewritten by dividing the numerator and denominator by H, to get $w/[Z/H]$, which is simply the wage per hour, divided by the output produced per hour. We denote Z/H by just z to emphasize that, like w, it is measured per hour:

(10.9)
$$\text{unit labor cost} = \frac{wH}{Z} = \frac{w}{Z/H} = \frac{wage/hour}{output/hour} = \frac{w}{z}$$

The "hour" units cancel out, and we are left with the wage (or labor cost) per unit of output, that is, the *unit labor cost*. This suggests either of two strategies by which the capitalist might try to increase profit: one is to try to lower the hourly wage without any loss of output per hour; the other is to try to increase output per hour while the wage per hour stays the same. Either of these, if it can be achieved, will cause unit labor cost to fall and profit per unit of output to rise. However, both strategies risk rebellion from workers, which can limit their success. Chapter 12 explains how the employer can find the wage at which profit is maximized.

Although firms sometimes do try to lower the wages (or benefits) paid to their current workers, this kind of wage cut may change other determinants of profit in ways that counteract the effects of the wage cut and prevent the profit rate from rising. Think about how you would feel if one day your boss told you that he wanted you to keep working with the same care and speed, but was going to cut your pay by 5 percent. Would you keep on working hard and well? History tells us that workers often respond angrily to such moves, either by working more slowly or less carefully than before, or by going on strike; or by individually seeking other employment. It is therefore rare that firms directly cut the wages of current workers; more often they cut the wages for new entry-level workers yet to be hired, or try to replace permanent with temporary lower-paid workers, or replace full-time with part-time workers who are not paid full benefits.

For some tasks firms may contract out services, such as nonproduction custodial and security functions, to nonunionized lower-wage contractors. They may also

outsource some operations to workers in lower-wage countries, where possible, or they may move production operations abroad to such countries (or even to lower-wage areas of the United States). They sometimes use the threat of relocation to extract wage concessions from existing workers. Many universities and colleges have been moving toward increasing use of part-time or temporary faculty, at far lower pay rates than tenure-track faculty, to reduce average pay levels. Chapter 13 will explore these issues in more detail.

Firms could also try to cut the quantity of labor used per unit of output by replacing operations now done by humans with operations done by machines, if the machines indeed cost less than human labor per unit of output. Chapter 13 discusses technology and the labor process.

Work effort refers to a worker's level of exertion (intensity of work) on the job during a certain period of time such as an hour.

WORK EFFORT AND THE EFFICIENCY OF LABOR. In the expression for the unit labor cost (w/z), the output per hour term (z) merits closer attention, because the two factors that determine it play a central role in Chapters 12 and 13, where conflicts between workers and employers are discussed. These two factors are (1) how much **work effort** is exerted by workers per hour (denoted by e), and (2) the quantity of output that results from a unit of work effort (denoted by f):

$$ulc = \frac{w}{z} = \frac{w}{ef} \tag{10.10}$$

The output produced per hour of labor (z) is the product of the level of work effort e and the efficiency of labor f. It is not difficult to understand the concept of *work effort*. Anyone who has held a job knows that the intensity of work, or work effort—how hard and carefully a person works—can usually be varied, at least to some extent. Although hard to measure, work effort, defined as the amount of effort put out by a worker during an hour, is an important variable in determining the profit rate. Greater work effort per hour e will lower unit labor cost $w/(ef)$ and so increase profit and the profit rate, if nothing else changes.

The **efficiency of labor** refers to how much output can be produced as a result of a certain level of work effort.

The **efficiency of labor** (f) refers to how effective a person's work is in producing a product, *given his or her level of work effort*. It is expressed in units of output per unit of work effort. It will depend on the technology—the type, quantity, and quality of capital goods and materials—used in the production process. These factors will determine how much output (z) is produced with any given amount of work effort (e) put forth during an hour of labor.

The first of these transformations is the basic challenge facing an employer in the workplace: how to get workers to put out as much work effort as possible during all the hours they have been hired for. The second transformation also depends on decisions made by the employer, namely, those that determine—by means of *investment*—the quantity and quality of the capital goods provided for workers to work with.

Many economists do not distinguish between the two components e and f that determine output per hour, assuming instead that there is a fixed relationship between the number of hours worked and the amount of output that results (for any given quantity and quality of capital goods). In making this assumption they ignore the fact that how hard an employee works (e) depends on such things as the way in which work is organized. Again, see Chapter 12, which focuses on the relationship between wages and effort, and Chapter 13, on technology and the efficiency of labor.

To illustrate these concepts we return to the Good Cod story, adding just one more piece of information. The new information is that the 25 pounds of fish caught

per hour result from the worker throwing his or her net into the water five times each hour and, on average, bringing in five pounds of fish with each throw of the net. The advantage of this example is that, unlike in many real world situations, we can clearly separate the work effort e (number of throws of the net per hour) from the efficiency of labor f (number of fish caught per throw).

To show the importance of examining the components of z, namely e and f, we ask what happens to Good Cod's profit rate if the workers decide to throw their nets into the water four times instead of five times per hour ($e = 4$ rather than 5), with each throw of the net still catching 5 pounds of fish. Each worker now catches 20 pounds of fish per hour, which is \$2,000 of fish per worker per year (in 100 hours of work). Since there are still ten workers, Good Cod's total revenue is now \$20,000, and total costs are still \$23,500 (wages are still \$18,500 and the non-labor inputs M still cost \$5,000); assume that the nets wear out exactly as fast as before. With only \$20,000 of revenue the new rate of profit is negative:

$$r = \frac{R}{K} = \frac{P_z Z - P_m M - wH}{K} = \frac{\$20{,}000 - \$5{,}000 - \$18{,}500}{\$7{,}500} \tag{10.11}$$
$$= \frac{-\$3{,}500}{\$7{,}500} = -47\%$$

So, because of workers' more relaxed pace of net throwing, Good Cod's rate of profit goes from *plus* 20 percent to *minus* 47 percent (i.e., to a 47 percent *loss*), though nothing has changed other than the intensity of labor (the work effort per hour). No wonder firms place a high priority on getting their workers to work hard!

Understanding the Profit Rate

Despite the impression the preceding discussion may have left, the profit rate is not determined by mathematics. The math just allows us to see *what the profit rate depends on*. Understanding the profit rate means understanding the things that it depends on. The math allows us to break the profit rate into parts, enabling us in the chapters ahead to look at each of the profit rate determinants one by one. But the basic idea in all of what follows is simple. Whether profits are higher or lower will be determined to a large degree by the outcomes of conflicts over how the product of production processes is to be divided. Behind each of the profit rate determinants are people—employers, workers, consumers, retired people, people in the rest of the world, government officials, and so on—trying, often in conflictual ways, to get the things they want in life.

We have already discussed both how profit is determined (using three different measures: profit per unit of output R/Z, total profit R, and the profit rate R/K). We end this chapter by putting the key concepts into Table 10.3 and reviewing the conflictual *relationships* that correspond to each determinant of profit.

Our strategy for understanding the profit rate has been straightforward: after having developed Equation 10.4 for total profits and Equation 10.8 for profit per unit of output, we examined each of their key elements in greater depth. Table 10.3 summarizes the determinants of profit per unit of output, along with the strategies a capitalist might adopt to try to increase such profit by affecting one of the determinants of the profit rate. In it, each of the broad elements that determine the profit rate per unit of output is listed in the first column: price, unit materials cost, and

TABLE 10.3 CONFLICTS OVER THE PROFIT RATE

		DETERMINANTS OF PROFIT PER UNIT OF OUTPUT R/Z; (8) DETERMINANTS OF PROFIT R AND PROFIT RATE R	SOME WAYS FIRMS MAY RAISE PROFIT PER UNIT OF OUTPUT R/Z, AND HENCE PROFIT RATE R	SOME WAYS WORKERS, CONSUMERS, SUPPLIERS, OR OTHERS MAY COUNTER EFFORTS TO RAISE THE PROFIT RATE
	P_z	1. Price of output (+)	Gain monopoly or near-monopoly power to raise price but not lose many customers; avoid raising price above the point that profit starts to fall.	Competing firms weaken monopoly power, drive prices down, limit profits. Consumers buy cheaper substitute goods.
$P_m M/Z$	***Unit materials cost =*** **Cost of intermediate inputs used up per unit of output**			
	P_m	2. Price of intermediate inputs (materials and capital goods) (–)	Find lower-cost input sources; a large firm can bargain input prices down. Support armed intervention abroad to control cheap oil and mineral deposits.	Domestic suppliers push for tariffs on imported inputs; citizens oppose military intervention abroad to secure cheap inputs.
	M/Z	3. Quantity of intermediate inputs used per unit of output (–)	Develop new production methods; reduce waste of materials and breakage of tools.	Rival firms learn and use similar methods; wide adoption of new method lowers P_z and R/Z again.
$w(H/Z)$	***Unit labor cost*** **= Cost of wages and benefits per unit of output = w/z**			
	w	4. Wages and benefits per hour (–)	Use cheaper labor, like recent immigrants or workers abroad; decertify unions, reduce rights to form unions.	Workers organize to form unions, bargain for higher wages, extend organizing efforts to lower-wage countries.
	$z = Z/H$	5. Output per hour in physical units (+)	See below	See below
$z = ef$	e	6. Work effort (+)	Speed up work; hire more supervisors to control work pace.	Workers unionize, resist working faster.
	f	7. Efficiency of labor (+)	Introduce more efficient labor-saving technology.	Workers bargain to save jobs.
Z	***8. Total output***			
	R/Z R r	R/Z is determined above. Multiply by Z to get total profits R. Divide by value of capital goods K to get profit rate r. $R = (R/Z)Z$ $r = R/K$	Rising output Z can realize *economies of scale*, with gains noted in Ch. 11. Attain large size for power over suppliers. Acquire or bankrupt competing firms.	Competing firms limit gains from scale economies, until merger, buyout, or bankruptcy. Antitrust action could break up firms.

unit labor cost. The next column lists the elements within each type of cost, and the second column from the right describes some ways in which capitalists might try to reduce such costs. Finally, the right hand column explains what kind of resistance may be offered to various capitalist strategies to increase profit. The table is far from complete; it is only a brief summary of issues that will be discussed in more detail in the rest of the book, and about which much more could and will be said.

The + or − sign following a determinant indicates how profit will change if that determinant changes, provided there is no change in any other determinant. A + sign means that the profit rate will tend to move in the *same* direction as the direction of change in the determinant. For example, if the price of output rises, and nothing else changes, profit will of course increase. Conversely, a – sign means that

the profit rate will tend to move in the *opposite* direction of the direction of change in the determinant. As an example, if the cost of materials increases, and nothing else changes, profit will fall. Note that for the situation assumed in the table, which is that the value of the firm's capital goods K is unchanged during the period being considered, the total profit and the profit rate will either both rise, or both fall.

However, the + or − sign can be quite misleading, because quite often a change in one determinant will reliably cause changes in other determinants. We can see this even for the case mentioned in the previous paragraph: When the price of a good rises, the amount of it that is sold (that is, Z, the number of units sold) will normally fall—and if it falls far enough (as in some cases it will), profit will actually *fall*, not rise. If a change in a determinant causes changes in other determinants, the + or − sign in the table does not really tell us what will actually happen to profit. We must follow through all the ripple effects on other profit determinants to discover the full impact on profits. Chapter 11 will shed further light on the question of when a firm will be better off raising the price of output, and when it will be better off lowering it. The point is that the reader should not take the + or − sign as a statement about what *will* happen, but rather as a first step in understanding what *could* happen under limited circumstances. Usually more thought is needed to see how the determinants will interact with one another.

The middle column of Table 10.3 lists some of the ways in which capitalists, either as owners of a single company or as members of a class, may attempt to change the determinants in such a way as to bring about increases in their profits. Similarly, the column on the right lists ways in which consumers, workers, suppliers, or other groups may resist efforts by capitalists to raise their profit rates. Can you think of strategies for raising the profit rate—or ways to resist such strategies—other than the ones listed?

Notice that some of the strategies listed in Table 10.3 may be inconsistent with other strategies, even for an individual employer. For example, lowering the price of a product might increase the firm's sales and therefore its output, so it might actually increase profit even though the price of output is falling. Chapter 11 will consider this possibility.

Moreover, the efforts by some employers to increase *their* rates of profit may have the effect of reducing the profits of *other* employers. For instance, suppose one employer is in the business of selling a material—say, oil—that is needed in the production processes of other employers. If oil sellers raise their output price (strategy 1), their profit rate will increase, other things remaining constant. But the profit rates of other firms will fall, because for them the rise in the price of oil is an increase in P_m.

Similarly, suppose all employers are able to cut their wage rates (strategy 4). Would this increase their rates of profit? Possibly it would, but much of the market demand for employers' output comes from consumers who earn their incomes as workers. Thus, if wages were cut across the board, many employers would find that they could no longer find customers for as much of their output as before. Their total sales revenue could fall, and if it did, their total profit and profit rates could then decline as well.

Finally, it is important to understand the bottom row of Table 10.3. Every row of the table except the bottom row is based on profit per unit of output. The bottom row of the table explains how to convert profit per unit of output to total profit, and then to the profit rate. To convert the profit per unit of output (which is R/Z) into total profit, we multiply it by Z, the number of units of output, and get R. So for the Good Cod

company, profit per unit of output was \$0.06 per pound of fish. To get total profit, we multiply that by 25,000 pounds of fish sold per year, and get \$1,500 in total profit. If we then want to find the *profit rate*, we divide R by K, which is \$7,500 for Good Cod, and get a profit rate $r = 20\%$ (see Table 10.1).

This means that as long as a firm's K—the value of its capital goods—is unchanged, then a rise in total profit will mean a rise in the profit rate, and a fall in total profit will also mean a fall in the profit rate. What about the profit per unit of output? Any change in any determinant listed in rows 1 through 7 in Table 10.3 that increases the profit per unit of output and does not reduce total output will increase the total profit and the profit rate. *However*, if a change in some determinant increases the profit per unit of output *but reduces the total number of units produced and sold*, it might or might not increase the profit rate. One such change will be discussed near the beginning of Chapter 11: raising the price of output will generally reduce the number of units sold (and produced), and might either raise or lower profit and the profit rate. Fortunately, it is possible to gain some clarity on when this is likely to happen.

Conclusion

We now have a foundation for understanding how the rate of profit is determined, since we know something about all the determinants listed in Table 10.3. In subsequent chapters we ask how each of these determinants is itself determined. In Chapters 12 and 13 we look at the vertical dimension of the capitalist economy (command) and examine how e, the amount of work effort per hour, is influenced in various ways. In Chapter 11 we focus on both competition and concentration. Competition is the horizontal dimension of the economy and is about the ways in which capitalist firms compete with one another in their attempts to maximize their rates of profit. Chapter 11 also discusses concentration: the process by which large firms come to dominate some industries in which large size confers advantages over smaller firms with regard to determinants of the profit rate.

Suggested Readings

Gérard Duménil and Dominique Lévy, *The Economics of the Profit Rate: Competition, Crises and Historical Tendencies in Capitalism* (Brookfield, VT: Edward Elgar, 1993).

Duncan Foley, *Understanding Capital: Marx's Economic Theory* (Cambridge, MA: Harvard University Press, 1986).

David M. Gordon, *Fat and Mean: The Corporate Squeeze of Working Americans and the Myth of Managerial "Downsizing"* (New York: Free Press, 1996).

Francis Green and Robert Sutcliffe, *The Profit System* (Middlesex, UK: Penguin, 1987).

John Micklethwait and Adrian Wooldridge, *The Company: A Short History of a Revolutionary Idea* (New York: Random House, The Modern Library, 2003)

Ted Nace, *The Gangs of America: The Rise of Corporate Power and the Disabling of Democracy* (San Francisco, CA: Berrett-Koehler Publishers, 2003).

Competition and Concentration



Apple Computer introduced its first Macintosh product in 1984. Combining in one very small plastic box a central processing unit, a display (screen), and a strikingly original, user-friendly interface that relied on "icons" rather than complicated commands, the Macintosh was a bold innovation. Because its new product was so different from anything else on the market, Apple was able to achieve, in setting the price of the Macintosh, a much higher "mark-up" over costs (a higher profit margin) than could other computer companies. As a result, this small, young, new company was able to reap enough profit to finance a large research and development operation, and this enabled it to continue to bring out innovative products.

Apple's mistake was that it failed to target its products toward business users. This failure allowed IBM and other PC manufacturers to maintain their dominance in the corporate sector. In the end Apple secured only a small share of the total computer market, and, predictably, the larger computer manufacturing firms, supported by their software-producing partner, Microsoft, were soon imitating Apple's user-friendly interface, icons and all, thus ensuring that their customers would not be tempted to switch to a Macintosh.

Microsoft itself, having achieved dominance in the computer software market, was found guilty in 1999 of violating the antitrust laws of the United States. In one of the most publicized court trials of its kind in the twentieth century, the federal district court judge in charge of the case found that the company "enjoys monopoly power in the relevant market." The court noted that it had so few competitors that it could raise the price of its Windows operating system without having to fear a reduction of the demand for it. The verdict was that to achieve its monopoly power Microsoft had stifled innovation, used heavy-handed and illegal tactics against its competitors, erected significant barriers to entry into its market, and, in the process, had acted against the interests of consumers. The judge concluded: "Microsoft has demonstrated that it will use its prodigious market power and immense profits to harm any firm that insists on pursuing initiatives that could intensify competition against one of Microsoft's core products."

Competition is often depicted in economics textbooks as a game played by countless numbers of small, powerless, anonymous firms—all ultimately serving

the expressed wishes of consumers. This characterization is only approximately accurate for a very few markets, if any.

As noted in Chapter 1, the largest one thousand companies in the world produce about four-fifths of its industrial output. The concentration of economic power is also evident in the United States, where twelve hundred corporations, the leaders of which could fit comfortably into a large concert hall, produce about half the nation's output. These twelve hundred large companies produce roughly as much as the other 28 million U.S. businesses combined.

That so much of the U.S. economy is run by so few people should not be taken to mean, however, that the most potent American corporate leaders can do whatever they want, that they all agree, or even that they have the same interests. The most powerful corporations based in the United States have to compete in the global marketplace with large enterprises based in other countries. With the lowering of barriers to trade and the increasingly rapid movements of capital and people from one country to another (discussed in Chapters 1 and 7), corporations based in Japan, South Korea, China, and western Europe are challenging the market dominance and profits of large American firms. Thus, competition, even for the patronage of U.S. consumers, is becoming more and more global, and the increasingly global character of competition makes it necessary for us to develop further the idea of *competition for profits* introduced in Chapter 7.

Business owners invest money in order to make more money. The prospect of profits spurs them to action, but competition channels their activity. Competition limits what any one of them can do but shapes what all of them (considered as a group) are driven to do. And competition for profits generates constant change in the economy.

Neoclassical economists depict competition as producing equilibrium rather than change. They ask the student to imagine a marketplace—say, a hypothetical open-air fruit market—where there are many sellers and many buyers. As explained in Chapter 8, economists believe that there will be a tendency in this type of competitive market situation for the market to move toward equilibrium. Equilibrium will occur if competition can drive the price of fruit to a level at which suppliers want to sell exactly the quantity that demanders wish to buy. In this case supply and demand will have determined the price and quantities of fruit exchanged, and the market will clear. The next day will dawn, the sellers will return with a new supply of fruit to be sold, the buyers will bring a new demand for fruit, and the competitive market process will be repeated. In the conventional view, then, equilibrium is achieved every day, and competition itself provides little impetus for change. In fact, even an open-air fruit market may or may not really operate in this way. But the real point is that economists imagine an ideal type of competitive market, and tend to assume that a fairly big part of the world resembles it enough so that it is the main type of market that needs to be analyzed.

The conventional view of market relations is not wrong, but it is limited and therefore misleading. Once equilibrium is achieved in the conventional model, nothing changes until something from the outside—an "exogenous" force—intervenes.

In Chapter 8 we explained the idea of equilibrium using the example of a marble coming to rest at the bottom of a bowl. Here we add the time dimension and offer a different example: a bicycle can stay upright only if it is moving forward. The process of competition for profits is more like the bicycle than like the marble in the bowl. With competition for profits, the economy is in perpetual motion (like the bike), not at rest (like the marble). The metaphor of the bicycle captures what economists and others refer to as *dynamic equilibrium*, while the marble, when it has settled at the

bottom of the bowl, portrays *static equilibrium*—a given way that things are. Both are useful ideas, but one cannot understand capitalism without the bicycle.

The conventional approach is one-dimensional in that it focuses on the horizontal (competition) dimension of the economy but considers neither the vertical (command) nor the time (change) dimensions. In this chapter we present a three-dimensional view of competition. Such an approach supplements the conventional analysis by showing how power (command) relations affect competition and how something internal to the economy—the process of competition itself—brings about change and development.

The main idea of this chapter is that *as business owners search for profits, competition drives them to change their operations. The competitive scramble limits what any one firm can do, shapes what they all are driven to do, and produces enormous pressure for continual change in the economy as a whole.*

This main idea is expressed in six key points:

1. Business owners compete for profits by seeking to *manipulate the determinants* of their own profit rates. They are limited in what they can do, however, by the forces of competition.
2. Business competition takes three principal forms: (a) efforts to achieve a price advantage (*price competition*), (b) efforts to create new situations in which potential competitors are at least temporarily left behind (*breakthroughs*), and (c) efforts to eliminate competition (*monopoly power*). For all three types of competition, firms must *invest to compete*. Investment is the primary way that firms achieve price advantages, breakthroughs, and monopoly power.
3. In industries with substantial economies of scale, large firms dominate largely because they can achieve lower unit cost by spreading out fixed costs over many units of output.
4. Because firms must invest to compete, and investment disrupts the prior situation, competition imparts a powerful dynamic tendency to the economy as a whole.
5. The competitive scramble produces different profit rates among firms, but, conversely, it also tends to equalize profit rates. Whether there is an overall tendency toward *equalization of profit rates* will depend on the balance between the opposing forces.
6. The dynamics of competition produce both economic *concentration* (which results when large firms displace or acquire smaller firms) and *less concentration* (resulting from new firms entering the market or smaller firms increasing their market share). Whether there is or is not a *trend toward economic concentration* will depend on the balance between these opposing tendencies. Recent years have seen waves of mergers and acquisitions creating greater concentration.

Competition for Profits

Entrepreneurs seek profits, and for this purpose they invest. But how can they tell which projects will be most profitable? And, having invested, what can they do to ensure high profits?

Most investments entail *risk* because they involve long-term commitments and there is no way for the investors to be sure what future profit rates resulting from them will be. Still, they can know what to look for. In Chapter 10 we learned that a firm's rate of profit depends on eight determinants (see Table 10.3).

As businesses search for ways to influence and improve (from their point of view) the determinants of their profit rates, they will run up against competition from other businesses, and this will limit their range of options. For instance, a business may not be able to increase its profits simply by raising its output price. Why? Because if one seller tries to set a price higher than the prices being charged by other suppliers of the same product, some or all of the customers of the price-raising business will buy the product from the other firms. The result will be a decrease in sales for the price-raising firm, and at least some of its output will remain unsold.

Similarly, in a rather competitive market, a business that is not exceptionally large usually cannot reduce the price it pays for materials, capital goods, or labor. Again, the reason is competition: suppliers of these inputs will sell them to firms offering higher prices.

And there are other factors besides competition that limit profit making. For some of the profit determinants, a business is limited by the present state of knowledge or technology. For example, it may be impossible to raise the efficiency of labor (f), or to cut down on either the quantity of materials used, or the amount of wear and tear on machines, per unit of output (umc) until new technologies are developed.

The social interactions that occur within firms must also be taken into consideration. For example, the intensity of labor—the amount of work effort per hour (e)—can be increased only if employees can be induced or compelled to work harder. But if a firm attempts to speed up production by coercion, such actions are likely to be met with opposition from its employees, and the firm may have difficulty attracting the best-qualified people to work for it. Moreover, if the employees of a particular firm are pressed to work harder than workers normally work, those employees can quit their jobs at that firm and seek employment in other, more conventional firms.

In all these ways, then, businesses are limited by competition. Moreover, one business's profits are often another's loss. For instance, the quadrupling of oil prices after OPEC (the Organization of Petroleum Exporting Countries) in 1973 refused to sell oil to the United States increased the profits of multinational oil companies by billions of dollars, but it also reduced the profits of firms in energy-using sectors of the economy, such as the automobile and steel industries, by a comparable amount.

Just as competition limits what any one business can do to change its profit determinants, changing those determinants may erode the conditions that sustain its current rate of profit. Every time an investment raises the rate of profit for a particular firm, other firms may be enticed to enter that firm's market. Thus, current success can lead to greater competition—from more competitors—in the future.

Profit making is a never-ending struggle by each firm to break out of the limits set by competition. The ensuing warfare among the combatants will sort out the winners from the losers.

Because competition produces incessant change, a capitalist economy is ever changing. If a static equilibrium happens to occur—like the marble in the bowl—competitive investments by profit-seeking firms will soon throw out of balance the underlying conditions of that equilibrium. A lasting equilibrium is as unimaginable as the bicycle remaining upright even when brought to a stop. Accordingly, the idea

that an equilibrium might continue until some external force disrupts it is misleading: competition itself will immediately put it out of existence.

Moreover, the dynamic of competition continues *without end or limit*. After all, the end of one period of competition is but the beginning of the next period. Firms that do not take advantage of every opportunity for growth will be left behind by firms that do. There is no rest for the weary, and there is none even for the victors in a capitalist economy.

The Forms of Competition

There are three principal forms of, or strategies for, competition. Each offers a way to stay even with, or get ahead of, one's competitors.

Price competition is a form of, or strategy for, competition in which firms attempt to attract customers primarily by offering lower prices.

Price competition occurs when firms attempt to attract customers by offering them products equivalent to those offered by other firms but at a lower price. A firm following a price-competitive strategy will continue to pursue its current activities while making small improvements here and there as opportunities arise. Price competition may lead an employer to try to cut wages, speed up the pace of work, eliminate waste in production, or otherwise cut costs in order to be able to reduce the price of the output and still make a profit. Many other firms may also be engaging in price competition; success will go to those firms that are able to offer the lowest prices to their customers for goods or services of similar quality.

A **breakthrough** occurs when a firm discovers or develops a new method of doing business, such as a new way of organizing work, a new product, or a new market.

Breakthroughs occur when firms discover or develop something totally new—a new source of inputs, a new method of organizing production, a new product, or a new market. In this case a firm can earn greater profits, often big profits, not because it is the best at the old game but rather because it is the first player in an entirely new game.

If a firm with a breakthrough has developed a new product, it may have little competition at all. If, on the other hand, its breakthrough creates a large cost advantage in producing a product that is already being made by other firms, the breakthrough firm may be able to cut the price of its product and still make more profit than it was making before.

A firm that consistently achieves breakthroughs is more likely to attract skilled personnel, develop superior production know-how, reduce costs, and obtain credit more easily and thus have a better chance to grow. Being first gives a firm relief from competition, and this lasts until other firms catch up or new firms enter the market with similar products or production processes. Big profits are sooner or later likely to attract competitors who want to capture some of these profits for themselves.

Monopoly power is the ability of one or a few firms in an industry to exercise substantial control over the market price and other aspects of competition, usually by excluding other firms.

Monopoly power exists when firms are able to exercise significant power in a market by excluding some or most potential competitors. A firm can rarely exclude all other firms (this would be "pure monopoly"), but it can often, together with two, three, or four other firms, control a major portion of a particular market and thus achieve what is called "monopoly power." When there are only a few firms in a market, the situation is referred to as **oligopoly**, or **shared monopoly**.

Oligopoly, or **shared monopoly**, is a market situation in which several firms together, but no one firm by itself, can exercise substantial monopoly power.

For monopoly power to last for a significant period of time, it must be supported either by some enduring economic advantage or by barriers that prevent other firms from entering and competing in the market. In either case firms are able to charge higher prices for their output, pay lower prices for their inputs, or have sole access to superior technologies or methods of organization.

Price Competition

Price competition confronts each firm with a dilemma about how to set its prices. This is best explained by an example.

Think about a bakery near campus (call it Sweets) that makes brownies. The owner of Sweets needs to decide what price for the brownies will result in the most profit. It might seem that a quite high price would be most profitable, but a moment's thought will make clear that this may not be so. At too high a price, the bakery will lose so many customers that its total profit will be low, even though its profit per brownie is high. But it is also true that at too *low* a price, the bakery will be making so little profit per brownie that even with an enormous number of customers, its total profit will be small. Somewhere in between the devil and the deep blue sea, there is a price level that will result in a larger profit than any other price level.

To find the price that yields the highest profit, it would help to know how many brownies will be bought at each price. Suppose we know this much about how prices affect average daily sales:

Brownies sold by Sweets Bakery

Price	Quantity sold
$1.00	400
$1.50	300
$2.00	200
$2.50	100
$3.00	0

If we graph this we see that all five points lie on a straight line, and we can draw a line through them and call it our estimate of the demand curve for brownies that Sweets faces. (A demand curve need not be a straight line—but it does simplify the algebra.).

It is crucial to realize that we are *not* talking about the demand curve facing *all* the bakeries in town taken together. Together, they sell far more brownies than does Sweets. We are only talking about the demand curve facing Sweets: how many brownies will this bakery alone sell at various possible prices? Of course, this depends on what prices other bakeries are charging for brownies, so we must (and do) assume that other bakeries are not changing their prices. We need not assume all other bakeries are charging the *same* price. A bakery far from campus might charge less, both because it needs to persuade customers (student customers, who may be on foot) that it is worthwhile to travel that distance, and because far from campus the bakery may pay less to lease space in a commercial building. And a bakery that makes a deluxe brownie with extra expensive ingredients might charge more.

Notice: The fact that there is a downward-sloping demand curve facing Sweets means that the market for brownies is not a perfectly competitive one with identical very small firms selling identical products. This means that if a bakery raises its price above the prevailing or average local price for brownies, it loses only some of its customers, not all of them. This is usually the situation in the restaurant and take-out food sector (and in many other sectors). Not all brownies are the same, and everyone has their special favorites, which they will likely continue buying as long as the price does not rise too much.

Location also matters. People might buy close to where they live rather than go way across town just for a brownie, even if the one across town might taste better or cost a little less. So all bakeries can count on keeping some of their customers even if they do raise their price (or rival bakeries lower their prices).

Now let us review the key equations from Chapter 10 that will help us to find the price that brings the most profit in these circumstances. Recall equation 10.4 from Chapter 10:

(11.1) $$R = P_z Z - (P_m M + wH)$$

where R = total profit (in dollars)

P_z = price of a unit of output

Z = quantity of output

P_m = price of materials used up and wear-and-tear on capital goods

M = quantity of materials used up and quantity of depreciation of capital goods

w = hourly compensation, including hourly wage and benefits

H = total numbers of hours worked by all employees

As Chapter 10 explained, we can simplify Equation 11.1 considerably by thinking in terms of the *unit materials cost* and the *unit labor cost*. Let us review these concepts. In Equation 11.1 the cost is the part that is enclosed in parentheses. Dividing the total materials cost $P_m M$ by the quantity of output Z gives the materials cost ($P_m M/Z$) of a unit of output, or the *unit materials cost* (*umc*). Likewise, dividing the total labor cost wH by Z gives the labor cost of a unit of output wH/Z, which we refer to as the *unit labor cost* (*ulc*). The sum of the *unit materials cost* and the *unit labor cost* is simply called the *unit cost* (*uc*):

(11.2) $$\textit{Cost per unit of output} = \frac{P_m M}{Z} + \frac{wH}{Z} = umc + ulc = uc$$

where

umc = unit materials cost: the cost of materials and wear and tear on capital goods used in producing one unit of output

ulc = unit labor cost: the cost of labor used in producing one unit of output

uc = unit cost: the combined materials and labor cost of producing a unit of output

We can therefore rewrite Equation 11.1 in this very simple way:

(11.3) $$R = P_z Z - (uc)Z = (P_z - uc)Z$$

This says that the total profit R equals the profit per unit of output (the price P_z minus the unit cost *uc*), times the number of units of output Z.

We do not *in general* assume that the *unit cost* (*uc*) (which is the same thing as the *average cost*) is constant; it may vary as total output Z varies. However, in the case of Sweets we assume that *uc* is constant, and that it costs Sweets \$1 to produce each brownie: *uc* = \$1. Now, using Equation 11.3 and the table of price and quantity

demanded on page 238 we can find the profit that results from a price of $1.50: $R = (P_z - uc)Z = (\$1.50 - \$1.00)(300) = \$150$. By similar calculations (try them yourself to check your understanding), we find that the total profit for each price level listed in the table above is:

Price	Profit
$1.00	$ 0
$1.50	$150
$2.00	$200
$2.50	$150
$3.00	$ 0

It certainly looks like the best price is $2.00, or somewhere near. If we do the math to check which price yields the maximum profit (see Appendix to this chapter), it turns out that indeed, a price of $2.00 does result in a higher level of profit—namely, $200 (per day)—than any other price for brownies.

Figure 11.1, part (a), shows the demand curve for brownies facing Sweets that is implied by the table on page 238. The equation of the demand curve is $P_z = 3 - 0.005Z$. In part (b) on the right, the curve traces, for each price on the vertical axis, the dollar amount of profit, measured in the horizontal direction as a distance from the vertical axis. It shows clearly that a price of $2 yields a profit of $200 per day, and that this is higher than the profit made at any other price.[1]

In fact, it is also possible to track the change in total profit by looking at Figure 11.1(a). Notice that in this example the total cost is represented by the shaded rectangle whose left side ranges from $0 to $1 and whose bottom side ranges from 0 units of

[1] See Appendix to this chapter for the mathematical details.

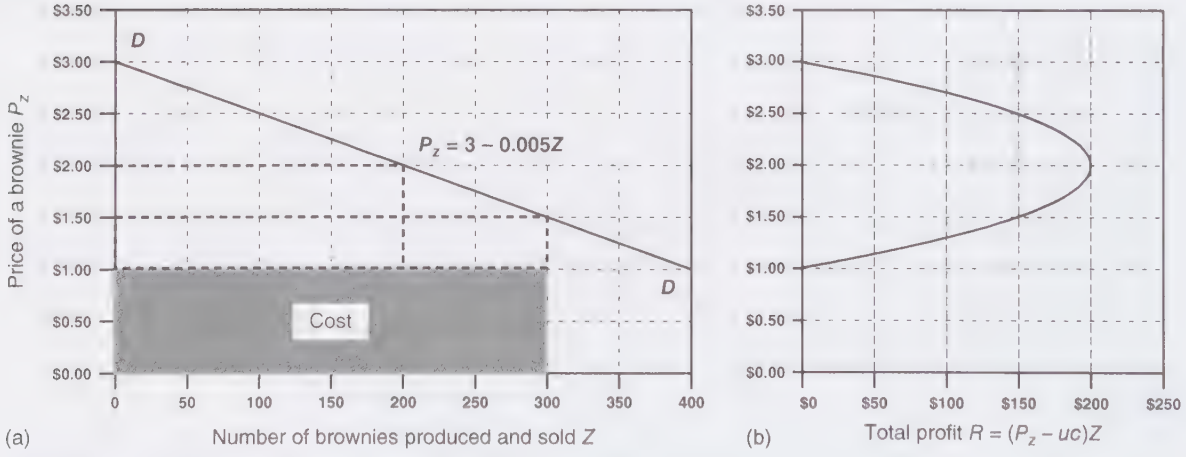


FIGURE 11.1 Finding the selling price at which profit is highest.

A business faces a downward-sloping demand curve for its output if when it raises its price it loses some, but not all, of its customers. This figure shows (a) the quantity of output (brownies) a bakery can sell per day at each price, as well as (b) the total profit, measured along the horizontal axis, that the bakery expects to make at each price level it could set for its brownies. The demand curve facing the firm may change constantly, though, as rivals offer new products, set new prices, or open new stores. In practice, setting prices often involves trial and error.

output to 300 units. To calculate total profit, we subtract cost from total sales revenue, so to find the area that represents profit, we must omit cost and only look at the area above the darkly shaded cost rectangle. When the price is \$1.50, sales are 300 units, and the rectangle that represents profit has its vertical dimension from \$1.00 to \$1.50 and its horizontal dimension from 0 to 300 units. Since (\$1.50 – \$1.00)(300) = \$150, the area of the rectangle, which represents total profit at a price of \$1.50, is \$150.

Now look at what happens when the price is raised to \$2.00. The height of the profit rectangle is now (\$2.00 – \$1.00) and the width is (200 – 0), so its area (which is total profit) is \$1(200) = \$200. To check your understanding, try to work out in Figure 11(a) what the profit is when the price is further raised to \$2.50.

The main point here is simply that *there does exist a price that will bring in more profit than any other*, under current conditions. A business owner or manager may not know what that price is, but by trial and error it is possible to home in on it to some extent. In any case the economic environment of a business will keep changing, so that the price that maximizes profit will itself be constantly changing, although perhaps not by very much.

In the interest of simplifying the situation enough to make it understandable, we have made some assumptions here that may not be valid in every case. It turns out that, despite these simplifications, the main point that there is a profit-maximizing price in any situation still holds (though it might only be findable through trial and error). We have assumed that the demand curve is a straight line; it may not be. We have also assumed that the cost per unit of output is constant; it may not be. None of this really matters; it is still true that f*or each firm and each product, there is some price at which profit is a maximum*; a firm that has set a higher price than that should lower its price to reap more profit, and a firm that has set a lower price than that should raise its price to reap more profit. In other words, in competition, much change occurs, but it is not total chaos; there is some order to it, and there are principles for navigating through it. But there will still be winners and losers from competition among firms.

For a firm that has already found the profit-maximizing price, there is no point in lowering price unless costs are lowered first.

Will Sweets benefit from finding a way to lower unit cost (*uc*) without lowering quality? It seems so from equation 11.3. Lowering unit cost without changing the selling price certainly means that more of the total sales revenue will end up as profit, because less of it needs to be paid out in costs. As the Appendix to this chapter explains, with a lower unit cost, the profit-maximizing price will also be lower, and yet Sweets' total profit will be higher than before.

Wait, though: This conclusion is based on the existing demand curve. But what if other bakeries lower the price of *their* brownies (perhaps after implementing the same cost savings method)? A change in rival firms' prices will shift the demand curve that Sweets faces, altering its own profit-maximizing price and possibly leaving *none* of the bakeries with higher profit than before Sweets' cost reductions. As we will see shortly, however, in some industries—those with substantial economies of scale—large firms can gain a lasting advantage over smaller firms.

Price as a Markup over Cost

The intensity of price competition shows up in how much a firm can raise its price above the cost of producing the product. This idea is expressed in the *markup* of

price over cost. For example, if a certain laptop producer can build laptops for $1,000 apiece, and then sets its selling price at $1,500 each, the business has "marked up" the price over cost by 50 percent: ($1,500 – $1,000)/$1,000. The profit per unit is $500, and the cost per unit is $1,000; the *markup* equals (**unit profit/unit cost**). Thus the markup tells us what percent profit is of cost. How does this fit into the algebra of Chapter 10? We already know that:

(11.4) **price of a unit of output** = **materials cost plus wear and tear on machines per unit of output** + **labor costs per unit of output** + **profit per unit of output**

We also know that we can add the two types of costs (unit materials cost [*umc*] and unit labor cost [*ulc*]), to get unit cost (*uc*), which gives us the simple equation:

(11.5) **price = unit cost + unit profit**

which we can write simply as

(11.6) $$P_z = uc + \frac{R}{Z}$$

Here, *R/Z* is profit per unit of output, which we also refer to as *unit profit*. The equation defining the **markup** is simply

(11.7) **price = (unit cost)(1 + markup)**

where the **markup = (unit profit)/(unit cost)**. In our example, $1,500 = $1,000 (1 + 0.50). This means that the markup is the unit profit divided by the unit cost, which is:

(11.8) $$\textbf{markup} = \textit{unit profit/unit cost} = \frac{\left(\frac{R}{Z}\right)}{uc}$$

The firm, in setting the price, first calculates its unit cost and then marks it up to yield the desired amount of profit per unit. As we have seen, however, marking up the price and increasing the profit per unit might either raise or lower the total profit, depending on the demand curve facing the firm. When price competition is fierce, markups are kept down because the firm that raises its price above the prices charged by other firms will lose most of its customers. On the other hand, where one or more large firms win the competitive battle and dominate an industry, they are typically able to increase their markups and make higher profits than they could in an intensely competitive market with no dominant firms.

What does this analysis add to our understanding? It is mainly another way of looking at the same conclusions we reached for Sweets Bakery. Looking at a firm's ability to have a high markup is equivalent to looking at a firm's ability to charge a high price relative to its cost—and we already analyzed the decision about P_z. However, in some contexts the concept of a markup can be useful.

If the laptop maker chooses the markup (here, 50 percent) without knowing what will happen to sales, it may find that it ends up with either too few or too many customers. It may need to adjust the markup (which means adjusting the price) up

or down to make the most profit. One alternative is to hold the markup steady, but to take other actions (such as stepping up the advertising and sales effort) to try to increase the number of units sold.

Market competition tends to establish prices that reflect current levels of supply and demand. If such prices allow a lower profit rate to be realized than could be made producing other products, at least some firms will leave the industry they are in, and some may be induced to switch to another industry in which they can make at least the average rate of profit. If the price of the product in a particular industry is so high that its profit rate is above the economy wide average rate, new producers will be attracted to that industry. Thus, the movement of firms—or "capital"—among industries will tend to push the rate of profit in each industry toward the average rate for the economy as a whole.

Economies of Scale and Price Competition

Although small firms like bakeries are very common, it is large corporations that produce the majority of goods and services. One question whose answer is central to this chapter is why, in quite a few industries, a very few large firms control production and sale of most of the output. In price competition, large firms can have a considerable advantage in certain kinds of industries, namely, those in which there are fairly large *economies of scale.*

Chapter 9 mentioned that in industries where there are significant economies of scale, the average cost per unit of output falls as the number of units produced rises. It is now time to be more precise about what exactly this means. Price competition among small businesses might not lead to any permanent or substantial advantage. But if a firm in an industry that has large economies of scale is able to grow to a size that gives it a lasting advantage over rivals, it can often raise its profit rate. In fact, this helps to explain the emergence of monopolies, to be discussed later in the chapter. Later in the chapter we will discuss other ways in which large firms have advantages over small ones.

In Bahia, in northeast Brazil, sits the large, modern Ford Camaçari auto assembly plant which opened in 2001. It is not the largest auto plant in the world; that distinction belongs to the Hyundai plant in South Korea, producing 1.5 million cars per year. But it is quite large. It was initially built with the capacity to produce 300,000 cars per year, and it cost $1.9 billion to construct; more than a billion dollars was invested recently to develop it further. Because the plant was designed to take advantage of economies of scale, it can produce cars at remarkably low prices. One low-priced subcompact it produces (more or less in the $10,000 to $15,000 range) is the Ka (pronounced as in "I left my Ka near Hahvahd Yahd").

Why is it cheaper to produce cars in large plants than in small ones? To answer this question we have to acknowledge that we have left something out of the profit equation, which we must now add in. We have been assuming that the unit cost *uc* is constant regardless of the number of units produced, and that it includes everything that goes into the average cost of a unit of output.

Fixed costs are costs that a business pays regularly and that do not vary when the total number of units of output produced varies.

What we have left out till now is one important type of cost: *overhead.* Overhead refers to **fixed costs** that a business has to pay regularly, regardless of how many units of output it produces. They include such costs as the lighting bill, the heating/cooling bill, and the cost of the security system—in short, all those things

that change very little or not at all when the number of units of output produced rises or falls. Another key fixed cost, and often the largest, is repayment of the money borrowed to build a plant and acquire and install the machinery. This consists of scheduled loan payments of interest and principal on these funds. A firm's total fixed cost is a regular payment that does not vary (or varies very little) when the total number of units of output produced varies.

Suppose that the Ford plant has a fixed cost consisting of payments of interest and principal on $500 million of funds it borrowed to build the plant. (Assume the rest of the construction cost was paid for by government subsidies or out of retained earnings.) Let us assume that this is the plant's only fixed cost, and further, that these loan payments amount to $60 million a year for each of ten years, after which the loan is fully paid off. This means that every year the plant's *fixed cost* F is $60 million. It is fixed in the sense that the number of vehicles produced does not affect it at all. Even if the plant shuts down for maintenance for several weeks and produces zero cars, the fixed cost must still be paid. Notice that it is the $60 million *flow* of payments on the debt that we are regarding as the fixed cost (not the $500 million stock of the debt itself, which of course will decline over time as it is paid off).

If the plant produces 200,000 cars per year, how much must it add to the price of each car to cover the $60 million in total annual payments on the debt? Since Z is the number of units of output produced, we need $F/Z = \$60{,}000{,}000/200{,}000 = \$300$, so the portion of fixed cost that must be added to the cost of each car will be $300. But if the plant happened to produce only 100,000 cars in any given year—because it was only able to sell that many—then the payments would still have to be made—so each car would need to have $600 added to its price to cover the payments, instead of $300.

At this point we need to clear up any possible confusion about the different categories of cost that we have mentioned. When there is a fixed cost F, the average cost of producing a unit of output has two pieces. One is the marginal cost (MC) the combined materials and labor cost of making one more unit of output. The other piece is each unit of output's share of the fixed cost, which is F/Z. Thus

$$\text{Average cost of a unit of output} = uc = MC + \frac{F}{Z} \tag{11.9}$$

Figure 11.2 shows how, in this situation, the average cost of a car falls as the number of cars produced increases. This is entirely due to the decline in the fixed cost per car F/Z as the level of output Z increases. The advantage of building a giant auto plant is that even an enormous fixed cost in the tens or even hundreds of millions of dollars can be spread out over hundreds of thousands of units of output, so that the fixed cost will not add much to the average cost of a car.

How does this change our previous expression for profit? Simple: we subtract one new cost term, namely the fixed cost F, from the expression for profit R in Equation 11.3:

$$R = P_z Z - (uc)Z = P_z Z - [(MC)Z + F] \tag{11.10}$$

where MC = *marginal cost.*

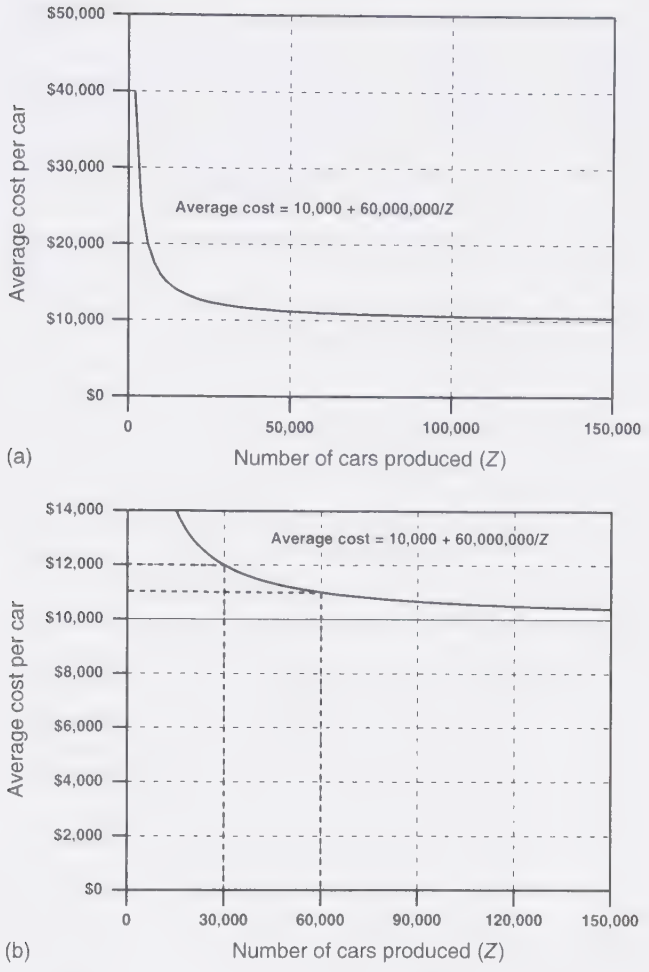


FIGURE 11.2 Average cost curve for an auto plant with marginal cost $10,000 per car and fixed cost $60 million per year.

There are large economies of scale in the auto industry, which result in decreasing average cost as the number of cars produced (Z) increases, as panel (a) shows. Panel (b) shows a close-up of the exact same average cost curve, highlighting the fact that if the plant produces 30,000 cars a year its average cost will be $12,000 per car, but if it produces 60,000 cars, its average cost will be only $11,000. With 100,000 cars produced per year or more, the average cost falls to $10,600 or less. This allows a large-scale producer either to charge the same price as its competitors producing in smaller plants, but make more profit per car sold, or else to charge a lower price, sell more cars, and gain larger market share.

The version of the profit equation that is most useful in explaining *economies of scale* is the "per-unit-of-output" version, which looks like this, after dividing Equation 11.4 through by the quantity of output Z:

$$\frac{R}{Z} = P_z - uc = P_z - uc$$
$$= P_z - \left(MC + \frac{F}{Z} \right) \tag{11.11}$$

Now look back at Equation 11.1. While the *marginal cost* (*MC*) of the materials and labor that go into each unit of output is constant because the production process is the same for each unit of output, each car's share of the fixed cost shrinks as *Z* grows, so the average cost (the *unit* cost) also shrinks. That is why Figure 11.2 (both parts) shows a downward-sloping average cost curve. Notice, however, that the average cost cannot fall below the marginal cost of $10,000, no matter how many cars are produced.

Figure 11.2(b) shows that if a plant produces 60,000 cars a year, then the average cost will be $11,000. But if a plant the same size only produces 30,000 cars a year, then the average cost of a car will be $12,000. Thus the plant that produces 60,000 cars a year can charge the same price as the plant producing 30,000 cars a year (with the exact same equipment), but can get $1,000 more in profit per car. Alternatively, it can charge a lower price and accept a smaller profit advantage with the reasonable hope of getting a larger market share. Either one gives it a competitive advantage. While a smaller plant may also have a somewhat smaller fixed cost, nevertheless the conclusion is that *large plants and large firms tend to gain economies of scale that equip them to dominate the market in industries where fixed costs are high.*

Notice that by dividing up average cost (or unit cost) into the two categories of *marginal cost* and the *share of fixed cost*, we have left behind the other division of *unit cost* that we have previously used, namely, into *materials cost* and *labor cost*). These two methods of dividing up the average cost of a unit of output are distinct. The *marginal cost* actually includes *both* some labor *and* some materials cost, since both are needed to produce one more unit of output. Likewise, each unit's share of the *fixed cost* usually includes some materials cost (like repayment of loans for equipment) and some labor cost (like salaries of security guards). Each way of dividing up average cost has its uses, depending on the issue of interest.

In some sectors fixed costs are enormous. In petroleum refining, the average cost per barrel for refining crude oil is much lower if a large refinery is built to process enormous quantities of crude oil and turn it into fuel and other products, than if the crude oil is refined in small refineries. (This has partly to do with the fact that refineries are mostly pipes, and doubling the diameter of a pipe roughly doubles its cost, but increases its capacity to carry fluid by four.) The oil industry is dominated by giant firms in part as a result of this and other types of scale economies—such as scale economies in shipping the oil to market (large tankers can operate more cheaply per barrel of oil than small ones), or the need to have deep enough pockets to tolerate the risks of drilling for crude oil, when some drilling holes yield no usable crude oil.

The situation is similar in production of engines a rule of thumb has been that an engine plant needs to produce something like 100,000 or more engines per year to spread out the overhead cost enough so that the share of overhead cost allotted to each engine is small, and the cost of each engine is therefore only a little higher than its unit materials and labor cost. Part of the Camaçari complex in Bahia is a very new engine plant, built in the last several years, that can produce 200,000 vehicle engines per year. Similarly, there are large economies of scale in production of auto bodies, so that it is best to produce at least 100,000 units per year to keep unit cost low.

In a world of scale economies, those that establish a large plant first can dominate the market by selling at a low but profitable price and making it hard for new, smaller plants with initially higher costs to compete. Small firms will at best be left with a razor-thin profit margin, at worst lose money. That is why, when developing countries began trying to establish their own auto plants, only very large countries like Brazil were able to do so (because Brazil's large population offered a

large enough domestic market to sell most of the cars produced), and even then only with considerable help from government subsidies and policies that protected local plants from import competition. Even now, the world's biggest automakers dominate Brazil's auto industry. But at least Brazil did succeed in establishing an auto industry, even if a foreign-dominated one, and the auto industry has had an enormous impact on the economy, not least as a driver of Brazil's "miracle" growth period from 1967 to 1973 when average annual growth rates topped 10 percent. South Korea, a much smaller country, developed its own quite successful auto industry only after a government-supported expensive all-out effort by fledgling auto assemblers to learn the technology.[2]

On the other hand, in some industries, like the garment industry, the economies of scale are not large at all. Each worker uses a relatively inexpensive sewing machine, and there are usually no large, costly machines for other operations like fabric cutting, even though cutting can be automated to some extent. A relatively small plant can easily get started, and competition among different plants is fierce, relying in large part on low pay and paying workers by the number of pieces they produce instead of by the hour. Thus small developing countries often are able to have a garment industry, but not an auto industry.

Let us review what we have learned so far. One form of competition is price competition. Firms that face a downward-sloping demand curve because they have some customer loyalty will tend to choose the price that yields the highest profit. Cutting cost lowers that price and increases profit, but only if rivals do not cut costs and prices. Improving quality may work better.

In industries with scale economies, a firm can cut unit cost by investing in a large-scale plant and running it at near full capacity. But this only works if the additional output will be sold.

Other Advantages of Large Size

The Ford Camaçari plant in Bahia also illustrates some other advantages of bigness. An unusual feature of it is that numerous suppliers of auto parts are actually located within the same complex and integrated into the production process, although they are not subsidiaries of (not owned by) Ford. They employ their own workers and run their businesses as independent enterprises, yet they are clearly deeply interdependent with the whole Ford assembly process. The engineering efficiency of the whole process is said to be enormous; parts roll off the lines of suppliers and onto the Ford assembly lines seamlessly, with a minimum of both time and labor.

How is this integration of suppliers related to Ford's large size? Just that the large size of the plant made it worthwhile for suppliers to establish production units there in the first place, and then to stay there. In addition, there is a port right next to the plant where large ships can dock, load, and transport the finished autos to market. Notice that although there might have been competition among suppliers for the chance to locate in the plant, once each supplier is there, the competition from outside suppliers not located in the plant is presumably minimal.

Yet another advantage of bigness is the ability to bargain with local, state (or provincial), or national authorities for subsidies to the cost of building and operating plants. Even Brazil's 1950s efforts to build production of engines and auto bodies

---

[2] Lin-Su Kim, "Crisis Construction and Organizational Learning: Capability Building in Catching Up at Hyundai Motor," *Organization Science* 9, no. 4 (1998): 506–21.

included offering government subsidies or tax breaks to foreign automakers. And in the United States in the last several decades, local and state jurisdictions have been offering subsidies for such investment. In Bahia, Ford received multi-year subsidies for between 68 and 75 percent of the $1.9 billion initial construction cost for the plant that opened in 2001, or roughly $1.3–1.4 billion.[3]

Unfortunately, different states in Brazil compete to attract such projects, and so the benefits to the public—the jobs and economic development—are likely to be largely offset by taxes paid to fund the subsidies. To what extent the public nevertheless benefits, after deducting the cost of subsidies from the economic development benefits, is unclear. The researchers who calculated the total subsidy to the Ford plant did not arrive at a definitive assessment of the overall relative size of costs and benefits. However, the Bahia region was poor, and has been developing a significant industrial base.

To recap, some of the benefits to a firm's profitability from establishing large size plants can include:

- lower average cost based on spreading out the fixed cost over many units of output; this allows the firm to undersell smaller firms with higher average costs, and still make a profit;
- ability to persuade supplier firms to locate nearby, making their products more immediately available on short notice, and eliminating the cost of transporting them to the assembly plant;
- ability to bargain with governments for tax breaks and subsidies.

Large size of a *firm* (not just a plant) also enables hiring a substantial lobbying team to press governments to adopt laws and regulations that help the firm make higher after-tax profits; such policies might include tax reductions, weakened labor laws, protection from consumer lawsuits, and so forth. Large firms might also be invited to play a central role in negotiating trade and investment agreements with provisions favorable to business, especially multinational business.

Where in all this is the public interest? Does the *public* benefit from concentration of economic activity in large firms? While economies of scale can make some products, like cars, cheaper, there are aspects of large firm size that can also work against the public interest. Lobbying by large firms can be a problem, especially when elected representatives and appointed officials become more responsive to corporate interests than to the public interest. Chapters 18 and 19 will have more to say on this point.

Capacity Utilization

Another way to think about competition is in terms of *capacity utilization*. A plant is considered to be utilizing 80 percent of its capacity if the output it produces in a time period (say, a month) is 80 percent of what it could produce with the resources (such as capital goods) it currently has available. The capacity utilization index for all of industry (manufacturing, mining, and the like) measured by the Federal Reserve system has fallen below 70 percent only once in the last forty-eight years, and that was in the depths of the Great Recession in early 2009. But it also has never risen above 90 percent, and only came close once in the early 1970s.

---

[3] Luis Ricardo Cavalcante and Simone Uderman, "The Cost of a Structural Change: A Large Automobile Plant in a Brazilian Less Developed Region," Regional Economics Applications Laboratory, Federal Reserve Bank of Chicago and University of Illinois, Discussion Paper REAL 04-T-5, April 2004, available at http://www.real.illinois.edu/d-paper/04/04-t-5.pdf.

The *disadvantage* of having capacity utilization below 100 percent is that the firm may be paying for capacity which it is not fully using. But the *advantage* of having unused capacity is that the firm is able to meet a sudden surge in demand, at least one of moderate size. If a plant were working at full capacity, then a surge in demand would force it to tell customers, "Sorry! Please go to our competitors if you need this product right away." For most plants, increasing productive capacity by buying new equipment (and possibly leasing or building more space) cannot be done from one day to the next. Much machinery is built to order, and this takes time. For this reason, firms have a reason to "build ahead of demand" to ensure that they can meet any surge in demand and not let their customers get used to buying from their competitors.

The organization of work also can limit the ability to fully use capacity. Imagine a professor being told that, since her office desktop computer is idle when she is teaching a class, she should arrange to let a student into her office to use it while she is busy elsewhere, so that the computer's capacity can be fully utilized. As should be obvious, for reasons of confidentiality and security, this is not going to happen. Unused capacity is in some cases far preferable to the alternative.

When an enterprise is using all of its resources, including all of its capital goods (all its machines, all the space in its buildings, all its durable tools), let us call its output Z^*, the maximum possible output with the resources it has. If at some other point in time it has the same resources but its actual output is Z, and Z is less than Z^* because some of the firm's resources are not being fully used, then we say that the firm's rate of **capacity utilization** at that point in time is Z/Z^*.

Capacity utilization is the ratio of the quantity of output a firm is actually producing to the quantity of output it could produce if its resources were fully in use.

For example, if the firm is producing 85 percent of what it could produce by using all its resources, then capacity utilization is 85 percent. Capacity utilization for all U.S. industry taken together has recently been lower than in the past. For most of the last fifteen years, it has been measured at between 70 and 81 percent.

Measuring Z^*, however, is actually tricky, and of course measuring Z/Z^* depends on accurately and consistently measuring Z^*. If a plant has *never* produced at full capacity, how do we really know what *is* its level of output at full capacity? And, for example, if a plant has never run more than one daytime shift, should we consider its full-capacity *daytime* output as 100 percent, or should we consider it to be 50 percent, with the idea that adding a full 8-hour second shift could double its output? This is not an easy question, and the answer is somewhat arbitrary. Because of difficulties in measuring Z^*, and therefore in measuring capacity utilization, for many purposes it may be more useful simply to look at Z itself, particularly when Z^* is not changing. The analysis is simpler, and we stay away from sticky measurement problems.

However, if a firm is thinking of expanding Z^*, that is a different story. Then it is quite important to think about capacity utilization. Expanding a plant's capacity in a situation where demand for the product is not expanding quickly will saddle the firm with extra capacity for which it is paying, but which it is not using, or not using efficiently. One example of this problem is the Ford Camaçari plant itself. Just as it was expanding the plant in 2014–2015, the Brazilian economy was slowing down and sliding into recession. In early 2016 the plant ended its night shift, evidently because the sales of its output were insufficient to make it profitable to run the plant twenty-four hours a day. During better economic times the night shift will return; but Ford is also has plans to invest in building new plants elsewhere or upgrading existing ones, including in China and the United States.

Another case in which capacity utilization is important is the case of too many hospitals in a metropolitan area buying the same expensive diagnostic equipment. If one of several private hospitals competing in a city of moderate size buys an

expensive computerized tomography (CT) scanner, then sometimes a second hospital gets the same equipment in an effort to share in the revenues from selling this high-priced service. But this often results in neither piece of equipment being fully utilized. If hospitals are able to raise the price of these services, they may be able to continue recouping the cost of the scanner. If not, they may both lose money, because both hospitals now have excess capacity in CT scanning.

In such cases, planning—or a government-supported hospital that does not have to make a profit—might allow acquiring just one machine and using it efficiently by keeping it busy much of the time. (To some extent the same problem exists with nonprofit hospitals as with for-profit hospitals, because both have to at least cover their costs.) Much concern has been expressed in the last five years that doctors are overprescribing computerized tomography (CT) scans—a problem both because of the expense and because the diagnostic radiation increases the chance of getting cancer in the future. A *Consumer Reports* study in 2014 noted that a study in the *Archives of Internal Medicine* found that 29,000 future cancer cases in the United States could be linked to the CT scans done in just 2007. Other research suggests that for-profit medicine is one of the reasons for overprescribing of such diagnostic procedures, particularly when the doctors themselves have a financial stake in the outfits that provide the services.[4]

Breakthroughs

Firms want to break out of the limits set by price competition. For that purpose they search for breakthroughs and monopoly power, both of which introduce elements of command (or power) in the relations among owners of firms and between them and other groups such as consumers and labor.

A *breakthrough* made by a firm is something new that gives the firm a competitive advantage and enables it to be first in a market. Most often it is a new product or technology, but anything innovative that gives one firm an advantage over its rivals can be a breakthrough. In some situations a new way of recruiting labor can be a breakthrough, though not if it is equally available to both a firm and its rivals. Breakthroughs change the terrain on which the competitive struggle is fought. If price competition is like trench warfare, breakthroughs are like tank warfare.

The auto industry is on the verge of marketing extraordinary new vehicles: the self-driving car, and the car that avoids accidents by braking even if the driver does not step on the brake. Coming within a few years of the commercially practical electric vehicle and hybrid vehicles, these inventions promise to inspire increased car purchases and give a temporary advantage to automakers that introduce them (if they are shown to meet safety standards).

A breakthrough creates a competitive advantage because it enables a firm to be first in a market. Other firms may attempt to follow it, but until they can catch up, the firm with the breakthrough will achieve a higher profit rate. Of course, if the breakthrough is patented, the firm that developed it will have a monopoly over its production for twenty years from the date the patent was applied for. Copyrights give a monopoly for longer. Thus achieving a breakthrough that is patentable or copyrightable assures the firm of some advantage for at least two decades. However, the protection in some

[4] "Dangers of Having Too Many CT Scans," *Consumer Reports*, April 2014, available at http://www.consumerreports.org/cro/2014/04/dangers-of-ct-scans/index.htm; and Atul Gawande, "Overkill," *The New Yorker*, May 11, 2015, available at http://www.newyorker.com/magazine/2015/05/11/overkill-atul-gawande.

cases is narrowly defined, so that rival firms can sometimes still invent and produce (and patent) a device or process that is fairly similar. Invention of truly new pharmaceutical products is often followed by invention and patenting by another pharmaceutical company of a drug that does something similar. Many of the most widely sold drugs—blood thinners, drugs for erectile dysfunction, and others—are available in at least two different brand names from at least two different companies.

Some breakthroughs may affect other profit rate determinants besides the price of output. Take the wage rate (w), for example. Many corporations have moved plants to, or opened plants in, lower-wage countries and achieved breakthroughs involving w. A 2002 *Wall Street Journal* article originally titled "Auto Makers Look for Another Edge Farther from Home," described how large automakers in rich countries were learning that their cars could be produced in lower-wage countries as well as—or better than—at home. The article reported, "Wages in Africa, Thailand, and Latin America often are less than a tenth of what workers earn in developed [country] markets. That more than offsets the added costs of shipping key parts to the remote factories and bringing them back. Adding plants in developing countries to its global production network 'will dramatically increase our competitiveness,' says Honda Chief Executive Hiroyuki Yoshino."[5] The U.S. share of global vehicle production fell from 42 percent in 1961 to 22 percent in 1981 to just 14 percent in 2014. The 1961 to 1981 decline was largely driven by the meteoric rise in Japanese auto production. But since then U.S. auto firms themselves have been increasingly producing abroad, so the 1981 to 2014 decline at least partly reflects the growing production outside the United States in plants owned by U.S. automakers. (Figure 15.5 in Chapter 15 will show how shares of global auto production in various countries have evolved.)

The ease with which companies can shift production to lower-wage developing countries often allows them to discourage union demands for higher wages in the United States just by threatening to move their plants abroad. In this way employers can achieve "breakthroughs" in paying low wages to their employees, although the same tactic is usually available in principle to their competitors. (See box "Hardball: Owners, Workers, and Taxpayers in the Global Economy" in Chapter 13.) Another strategy for a firm is to seek to "decertify" a labor union (hold an election that removes it as the representative of the workers in a plant). That, too, will usually allow it to pay lower wages and become more competitive, but without moving its production facilities abroad.

Breakthroughs, then, raise a firm's sales revenues or reduce its costs sufficiently to produce a higher-than-average profit rate. This competitive advantage lasts only until other firms catch up. Apple brings out new versions of iPhones in quite rapid succession, driven by the need not to let Samsung catch up. In the meantime, however, the temporary relief from competition may produce substantial profits. And while the other firms are catching up, the firm with the breakthrough not only makes more profit than its competitors but, as a result, can spend more money searching for a new breakthrough. This is the attraction of breakthroughs.

Capitalist competition holds not only the promise of breakthroughs but the threat that a rival firm will achieve a breakthrough and undercut one's own profits. Even an employer content with his or her firm's current rate of profit must fear other firms' potential breakthroughs, for any new competitive disadvantage could well wipe out current profits. Here, as always, competition provides both the carrot and the stick. Firms must keep running, as if on a treadmill, just to stay in the same place.

---

[5] Todd Zaun, Gregory L. White, Norihiko Shirouzu, and Scott Miller, "Auto Makers Look for Another Edge Farther from Home," *Wall Street Journal*, July 31, 2002, p. A1.

Monopoly Power

Capitalists can also make higher profits by achieving monopoly power. Much like a breakthrough, monopoly power enables a firm to escape competition or at least to insulate itself from the effects of competition. If a firm can use monopoly power to erode or eliminate competition, it will be able unilaterally to influence its profit rate determinants in ways that raise its rate of profit.

The term *monopoly power* can be used to refer either to the situation of a single firm (perfect monopoly) or to the power of a small group of firms that are able to dominate a particular market (oligopoly, or shared monopoly); in either case, if some firm or firms can exclude others from a market, monopoly power exists.

Breakthroughs and monopoly power are quite similar: both generate a higher profit rate for the breakthrough or monopoly firm by giving it an advantage over other firms. Indeed, the twentieth century economist Joseph Schumpeter understood breakthroughs as creating "temporary monopolies," and his insight points to the main difference between the two concepts: breakthroughs are temporary, lasting only until other firms catch up; monopoly power is longer lasting and is usually based on some economic, legal, or institutional obstacle that excludes other firms. Such obstacles are called **barriers to entry**. In an industry with large economies of scale, these scale economies themselves act as a barrier to entry, for the reasons explained earlier: large-scale producers can afford to charge prices that are low enough to rob their higher-cost competitors of most of their profits, and so drive them out of business (or keep them quite small).

Barriers to entry are obstacles that make it more difficult or costly for new firms to enter a market; examples include technical secrets, initial investments that are very large, and exclusive marketing arrangements.

Monopoly power is most often used to set the price of a product (P_j) above the price that would prevail in a competitive market. Monopoly power may also be used to change the other determinants of the profit rate, but setting prices is the most common strategy.

Firms can achieve monopoly power—whether pure or shared—in three ways: first, they can be or become the only seller of a particular product; second, they can collude (conspire) with other producers of the same product to eliminate competition; or, third—and this is the most common way in which monopoly power is exercised—they can establish informal or tacit relations with other companies in their market to limit competition. In principle the first and second options are illegal in the United States and many other rich countries, though these strategies have sometimes been used successfully anyway.

One example of the only-seller case is the local electric power company. If you want electricity service in your home, in many areas you have little choice but to buy it from your regional power company. Since electric power is frequently provided by a monopoly, the government often regulates the price of electricity. Local cable providers in some communities are also monopolies, because they are the only seller in the local area. This not only may keep prices high, but may reduce the incentive to provide excellent service.

The second way of exercising monopoly power is referred to as "collusion." Although outright collusion is against the law in the United States and in some other places, the impulse to collude is common among large corporations. When the U.S. Department of Justice was prosecuting the large agricultural products company Archer Daniels Midland (ADM) for price fixing, one of the top executives of ADM admitted that the company's view was that "our competitors are our friends; our customers are our enemies."

A classic instance of collusion involved General Electric (GE) and Westinghouse in the decades before 1960. GE, Westinghouse, and some other companies in the industry were convicted of fixing the prices of electrical equipment. And in December 2001 the world's two largest auction houses for valuable art, jewelry, and furniture, Christie's and Sotheby's, were found guilty of colluding to fix the commissions they charged their sellers, a practice they engaged in for six years and that resulted in overcharges of hundreds of millions of dollars. The two companies agreed to pay $537 million to settle price-fixing lawsuits brought against them in the United States by their customers, and the former chairman of Sotheby's was sentenced to pay a $7.5 million fine and spend a year in prison even though he was seventy-eight years old. Since the defendant showed no remorse for his actions, he was ordered to pay the cost of his incarceration, a judgment that cost him $21,601.

Monopoly power also exists when a few large firms dominate a particular market, a situation referred to as shared monopoly. This is not only the most common method for asserting monopoly power, it is also the most complicated. In the breakfast cereal, steel, airlines, and tobacco industries, for example, a few firms have dominated their markets for many years. The firms in these industries "share" monopoly power in the following sense: without engaging in direct conspiracies or acts of collusion, they have developed ways of behaving that permit them to avoid price competition and thus to share in the higher profits associated with monopoly power. In other words, each shared-monopoly firm has learned to live by certain rules, or ways of doing business, that all similar firms accept, and their competition is governed by these rules.

In some parts of the world and in some periods of history in the United State, shared monopoly power has been achieved by direct cooperation. When this happens, the cooperating firms or countries are referred to as a **cartel**.

A **cartel** is a combination of business firms or states operating in concert to regulate the production, pricing, and marketing of goods by its members.

Chapter 6 recounted briefly the early story of monopoly in the U.S. oil industry, in the form of John D. Rockefeller's Standard Oil Trust, formed in 1882. Formation of trusts led to passage of the Sherman Anti-Trust Act and eventually to breaking up the Standard Oil empire in 1911. But by the late 1920s an international oil cartel was formed, with Standard Oil of New Jersey, Shell Oil, and the Anglo-Persian Oil Company (in which the British government was a majority owner, and which later became British Petroleum). The cartel signed the As-Is Agreement, which pooled world oil markets and then divided them up, assuring each firm of maintaining its existing market share. Although this evidently violated antitrust laws, The American Petroleum Institute argued for exempting oil from antitrust prosecution, on the grounds that it was in the national interest to do so. And indeed, oil companies were not prosecuted, though neither were they apparently granted an official exemption. Much later, in the 1960s, the Organization of Petroleum Exporting Countries (OPEC) was formed to assert the interests of oil-rich developing countries. By jointly regulating its members' supply of oil, OPEC has at times been able to determine, or strongly influence, the price of oil on the world market.

Shared monopoly is like an uneasy coalition among potential enemies. It holds together only as long as each partner perceives that the agreement not only works to advance the interests of all members but also serves the individual interest of each. The higher profits to be gained from cooperation motivate individual firms or states to continue cooperating in the shared monopoly arrangement. In principle a failure to honor the cartel's decisions can be very costly for the rule-breaking firm or state because the most powerful members of the cartel might retaliate, say, by price

cutting in such a way that the misbehaving member is severely damaged. However, OPEC members (especially the most powerful among them) have sometimes gone against the wishes of the majority without suffering significant retaliation.

If a shared monopoly operates on the basis of a set of tacit or informal rules to which all the large firms in an industry subscribe, what are these rules? The two most important are (1) shared monopoly firms develop ways to avoid price competition, and (2) such firms engage in **nonprice competition**, for example, sales promotion.

Nonprice competition involves sales efforts, style changes, or other marketing strategies to increase the appeal of a product without lowering its price.

Price competition is avoided when shared monopolists can develop informal, tacit, or indirect ways of establishing mutually agreeable prices. But if price competition is avoided, other forms of competition attain a new intensity. The demand curve for a product still limits the total sales by all producers at each price. This means that if the price does not change, an individual seller's gain can come only at the expense of the sales of another firm. Thus the main action in the market will be battles for **market share**, with each shared-monopoly firm trying to maximize its market share. As we saw in discussing economies of scale, if the price remains steady (by mutual agreement of the shared monopolists), the firm that achieves greater scale economies and increases output will make higher profits, the more units it can produce and sell. This will drive each company in a shared-monopoly industry to go all out to sell more of its product. However, as we have seen, in the As-Is Agreement, the world oil cartel of the late 1920s explicitly divided up market shares precisely in order to avoid costly and destructive competition.

A **market share** is one firm's sales as a percentage of the total sales in an industry.

All the activities of a firm designed to increase sales of its product are referred to as its **sales effort**. The sales effort includes anything that firms do to persuade customers to buy their products. Corporations recruit armies of sales agents. They hire public relations firms and spend billions on advertising. They continually repackage their products. They spend enormous sums on getting celebrity endorsements and sponsoring sports events. They sometimes bribe foreign governments (and sometimes are convicted of bribing them, under the Foreign Corrupt Practices Act). They contribute to charities and sponsor programs on public television to burnish their image, and so on. Such efforts are aimed at increasing their market share.

The **sales effort** consists of all those activities by a firm that relate to the selling of the firm's product.

Monopoly power, like a breakthrough, makes possible a higher-than-average rate of profit. It relaxes (but does not abolish) the limitations set by competition by eliminating some of the competition. Firms compete to see who can first or most effectively escape competition.

Investing to Compete

One aspect common to all three forms of competition is *investment*. Price competition, breakthroughs, and monopoly power all require that firms invest to achieve the benefits of any of these strategies.

To compete at all in the marketplace, firms must invest. A firm must reinvest its profits from one year if it wants to remain competitive in the next year. Moreover, profits previously earned may provide a firm with the resources it needs if it is to try to get ahead and/or stay ahead of its rivals.

All of the profit determinants listed in Table 10.3 become potential targets for reinvestment. If the firm can discover a way to reinvest its profits so as to change one or more of these determinants in its favor, then it will achieve a higher profit rate.

Conventional economics textbooks portray competition as leading to a static equilibrium in which all firms achieve an average rate of profit and all resources are allocated in an optimal fashion to satisfy existing consumer preferences. Political economy, in contrast, sees firms as engaging in dynamic competition, always seeking to move past the present situation of competition and put themselves in a position to capture higher profits. In dynamic competition no situation is permanent because at least some firms will always be altering the requisite conditions for it. Thus, the real world of competition is a never-ending process in which firms attempt, not always with success, to break out of the limits set by competition itself. Some firms will succeed and achieve high enough profits to compete in the future; others will fail and be driven out of business.

A firm will reinvest profits only if it expects to make more profit as a result of the investment than it would have done without the investment. So every firm, when considering what to do with its profits, must compare the *cost* of investments with the *expected return* from those investments.

For example, consider a firm that is thinking about buying a new machine for which the machine maker charges I dollars for purchase and installation. For the firm to estimate whether this investment will be profitable requires going through three steps.

First, the firm will have to calculate what the return might be on its investment. Suppose, to keep matters simple, that the machine costing I dollars will last just one year. (After producing the goods it was designed to manufacture, the machine is worn out and has no value at the end of the year.) At the end of the year, then, the owner has the goods produced but no machine, and she sells the goods. The revenues from this sale minus the total cost (wage and materials costs plus any fixed cost) are called the *total return on the investment*:

(11.12) $$\textit{total return on the investment} = I + rI$$

with I = the value of the investment
r = the expected rate of profit on the investment

Of course, the investor hopes that r will be positive so that the investment more than pays for itself. But it might not be, in which case the investor would not even get back the amount of money that the machine cost.

The second step a firm must take in order to evaluate the potential profitability of an investment is to determine the *cost* of an investment. There is, of course, the cost of buying whatever it is that constitutes the investment—in this case, the machine—which is I dollars. In addition, however, there is the cost of the money used to make the investment. What is this cost? It can be looked at in two ways.

The **INTEREST RATE** is the cost of borrowing money; for a firm that borrows money, such as from a bank, it is the interest payment divided by the PRINCIPAL, which is the part of the loan amount that has not yet been repaid.

On the one hand, a firm may borrow the necessary funds at the prevailing interest rate (i) from a bank (or other lending agency); the **interest rate** is then the cost of borrowing money, and the amount borrowed is called the *principal*. (See Chapter 5, where the interest rate is defined.) Alternatively, the firm can borrow money from people who are willing to buy bonds from the firm. As explained in Chapter 10, a bond is like an IOU that commits the issuer of it not only to paying back to the purchaser of the bond the face value of the bond but also to paying a fixed series of specified payments on a specified schedule to the owner of the bond over a certain number of years—the life span of the bond, up to its maturity date. (The specified payments are based on a specified rate of interest applied to the face value of the bond.)

On the other hand, a firm may finance an investment using funds available (internally) from retained earnings out of previous profits. In this case it must forgo the opportunity of lending these funds to other firms at the prevailing rate of interest (i). This forgone opportunity is termed the *opportunity cost* to the firm of using its own previously generated funds to make an investment. It is equal to the cost the firm would have had to pay if it had borrowed the funds (externally) from a bank. What **opportunity cost** refers to, therefore, is the cost of the opportunity lost when the firm, instead of lending out its money, uses it to make an investment (in the current example, to buy a machine).

An **opportunity cost** is the value of the best opportunity given up (forgone) in order that whatever activity was chosen could be undertaken.

Whichever way the firm chooses to finance its investment, the **total cost of investment** is the amount of money (I) used to make the investment (in our example, to buy the machine) *plus* the explicit cost—or opportunity cost—of obtaining the money so used. This cost is equal to iI—the cost per dollar (i) times the number of dollars required for the investment (I). Thus,

The **total cost of investment** consists of two parts, the cost of the capital goods purchased and the opportunity cost of the money used to purchase them.

(11.13) $$\textit{total cost of an investment} = I + iI$$

The last step in the analysis of investment profitability is to estimate the **expected profit from an investment** (net of the cost). The profit can be determined from the first two steps: it is the expected total return (defined in the first step) *minus* the cost of making the investment (as determined in the second step). This can be expressed as follows:

profit from an investment = return on the investment − cost of the investment

Put in terms of Equations 11.3 and 11.4, the expected profit from an investment is

(11.14) $$R = (I + rI) - (I + iI)$$

where R = the expected amount of profit
I = the amount invested
r = the expected rate of profit
i = the rate of interest

In other words,

(11.15) $$R = rI - iI = (r - i)I$$

Equation 11.15 allows us to focus on r and i—the rate of profit and the rate of interest—as the key determinants of the total amount of profit (R) resulting from the investment. In Equation 11.15, the amount invested (I) is always a positive number. But what about $(r - i)$? It is positive if r is larger than i, but it is negative if r is smaller than i. If $r > i$, the firm will receive sufficient profit on an investment to be able to more than cover its interest payments to a bank (or to bondholders), if indeed it had to borrow the money to make the investment. After repaying the loans, the investing firm would have money left over, making the investment a profitable one. The same logic would also have to be applied to the case in which a firm finances an investment from its own funds. In this case the investment can be considered "profitable" only if the firm ends up with more money than it would have ended up with if it had loaned out its money instead of investing it.

If r is less than i, the term $(r - i)$ will be negative, and the investment will result in a loss to the firm, because the firm will have had to pay more in interest to the bank (or to its bondholders) than it received in profit. If the firm financed the investment from its own retained earnings, it would still, in effect, suffer a loss. This is because it could have made more money by lending the equivalent of its investment to some borrower at the prevailing rate of interest (i). The point is that no matter what the rate of profit (r), a firm must compare r and i to see if a particular investment is worth making.

Unfortunately for the firm, it cannot compare r and i directly. When a firm undertakes an investment, it cannot know how the investment will work out. The payoff to any investment is uncertain, and therefore making it involves *taking a risk*. A firm usually will know how much an investment will cost (I), and it will also know what the prevailing interest rate (i) is. As noted, however, it cannot know with certainty what the (future) profit rate (r) will be.

The **expected profit rate on investment** is a firm's estimate of the future profit rate that it thinks will be earned on its investment.

In the *absence of certainty* about the future, whenever a firm is deciding whether to make an investment—it must compare the *expected profit rate on the investment*, which may be called *expected r*, with the prevailing actual interest rate i. The expected profit rate is the rate that the firm thinks or expects it will achieve; of course, this expected rate of profit may or may not turn out, in the end, to be the one that is actually achieved.

Based on its expectations, then, a firm will choose to invest only when the expected profit rate on an investment is greater than the prevailing rate of interest:

$$expected\ r > i$$

The decision whether to invest depends on investors' (informed) optimism or pessimism about the future. If they are optimistic and are confident of good or better conditions in the future, they will tend to evaluate investment projects more favorably, seeing the positive possibilities and downplaying the dangers; in this case, *expected r* will tend to be high throughout the economy. If, on the other hand, investors are pessimistic, expecting difficult times ahead, they will be more wary, worrying that investments will not pay off; *expected r* will in this case tend to be low. Investors' optimism or pessimism will lead, therefore, to larger or smaller amounts of investment. Of course, each investor will view the future somewhat differently, and some will have expectations that turn out to be more correct than others.

There is another conclusion that can be drawn from our analysis of the factors that influence a firm's investment decision. Considering how the relationship between r and i (the profit rate and the interest rate) affects the profitability and therefore how desirable making an investment is, one can see that the level of investment in the economy as a whole will normally be influenced by the rate of interest. A low interest rate will tend to stimulate investment, whereas a high interest rate will tend to choke off investment. This places a significant amount of responsibility with a nation's central bank (for instance, the Federal Reserve System, or "Fed," in the United States) because it normally has the power to move interest rates up or down and thus to discourage or stimulate investment.

Finally, the amount of investment a firm undertakes will often depend on its rate of capacity utilization. If it is less than 100 percent (meaning that the firm is not producing as much as it could with the resources it has), any new machines available for purchase will have to be significantly better than the currently owned machines before the firm will decide to invest in them. It obviously does not make sense to buy new machines when a firm is not using all the ones it already owns, unless the

new machines are more efficient. The exception to this might be if there is reason to believe that demand for the firm's products is about to rise sharply, either because of a stimulus to the economy, or for any other reason—and the firm may want to expand its capacity to be able to fill all the orders it expects.

The Dynamics of Competition

We can now see the close connection between investment and competition. *Investment is the firm's way of carrying competition into the future.* And competition, because it forces firms to invest and thereby change the prior situation in some way, is inherently dynamic.

Consider our earlier example (see Chapter 8) of a beer market consisting of a large number of relatively small firms all producing the same kind of beer. Each firm will attempt to make as much profit as possible, but price competition will limit most firms to making just the average rate of profit.

What will be done with the profits of a firm in our hypothetical beer industry? The owner (or owners) of the firm could decide to spend their profits on unproductive activities, such as consumption of luxuries. Alternatively, he, she, or they could reinvest some or all of the firm's profits to improve its competitive position in the future. Owners who decide to reinvest profits will spend their money on things that may reduce the firm's costs of production—things such as research, new machinery, and more effective supervision of workers.

Figure 11.3 shows how future costs of production will differ between (a) firms that do not reinvest their profits and (b) those that do reinvest their profits. We can see that both kinds of firms start out in year 1 with costs of $70 per barrel of beer. But whereas the firms that do not reinvest (shown in the diagram on the left) continue to have costs of $70, those firms that do reinvest (shown in the diagram on the right) see their costs fall as time goes on—to $50 in the second period, $35 in the third period, and so on. The firms with the ever lower costs will have higher profit rates than the higher-cost firms. They will be able to charge lower prices for their product, sell more beer, and increase their rate of capacity utilization. Alternatively, they may use their cost advantage to enjoy a higher markup over costs. In any case they will have a competitive advantage over their rivals, and this will allow them to expand their share of the beer market.

Firms that reinvest may also choose to expand their scale of output. It is likely that profit-reinvesting firms will use their resulting cost advantage to undercut their higher-cost rivals, capture a larger share of the market, and increase their output and sales. The lower-cost firms will have investment opportunities that the higher-cost firms will not have: they can profitably invest in expansion of output, and it will be profitable for them to grow larger.

Dynamic competition will tend to produce uneven results: there will be winners and losers. Some firms will be highly effective or lucky, investing in things that pay off, while others will be unsuccessful or unlucky and never see their higher expected profits materialize. Moreover, not all firms will be equally well situated. Some firms may have better access to needed raw materials (for example, mountain water), be closer to big markets, or simply have better management. The well-situated firms will have higher profit rates than the poorly situated ones. Firms that achieve high profits in one period will have more to reinvest in the next period. The higher the

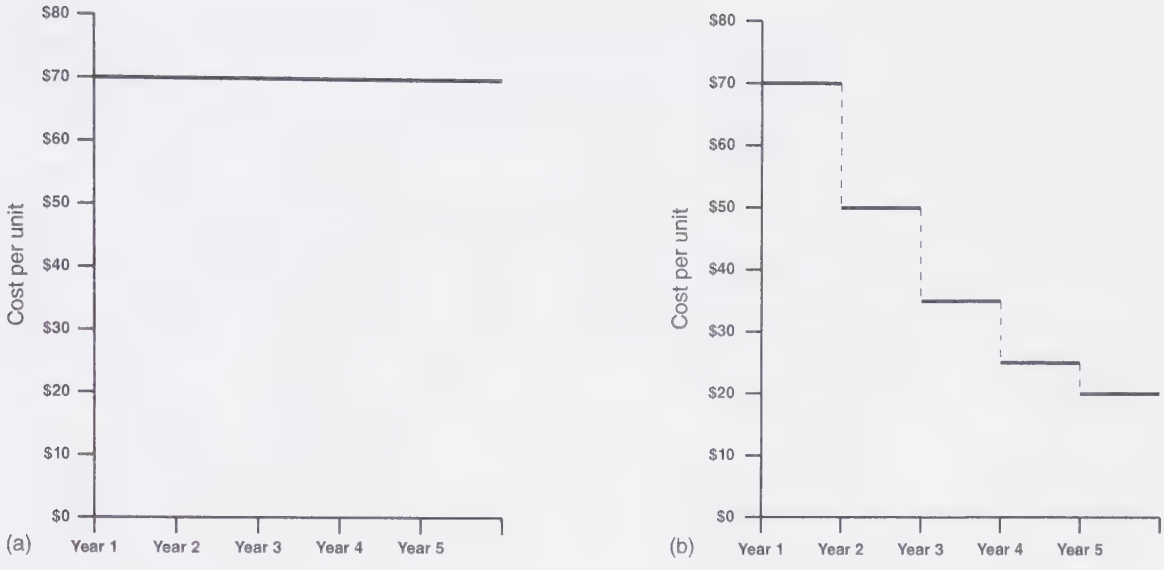


FIGURE 11.3 Dynamic cost advantage.

Panel (a) of this figure shows what happens when firms do not reinvest their profits. Assuming that the prices of inputs (including the wage) remain unchanged, such firms will see their unit costs remain constant from one period to the next. Panel (b) shows that when firms successfully reinvest their profits (in better equipment, for example) their unit costs fall from one period to the next—again assuming no changes in input prices. Lower unit costs will allow firms to achieve higher unit profits or reduce their prices and gain market share. Higher unit profits or lower prices (or both) will give the reinvesting firms competitive advantage over firms that do not reinvest.

profit rate, the more there will be to reinvest; unsuccessful firms have low profit rates and, hence, little to reinvest.

What are the consequences of the constantly changing, dynamic, uneven, disruptive pursuit of profits? On the grandest level, the process drives the constant transformation that is characteristic of capitalist society (see Chapters 1 and 7). More immediately, the process has consequences for firms. It may tend to bring about more equal profit rates among firms, and then again it may not. It may lead to greater economic concentration, and then again it may not.

Toward Equal Profit Rates?

Equalization of profit rates refers to the process by which competitive pressures on firms in different industries, different geographical regions, or different markets push their profit rates toward a common, or average, level.

At any particular time many different profit rates exist in the economy. Some firms will have low profit rates, others high profit rates. Competition produces the divergence of profit rates, and it also tends, over time, to bring them closer together.

The profit rate (r) achieved by a firm during a given year depends on the profit rate determinants affecting that particular firm (see Chapter 10). How effectively a firm manages its inputs of materials and capital goods (M), how much work effort per hour (e) it gets from its workers, and many other factors will determine the firm's profit rate. Some firms will be able to manage at least some of their profit rate determinants well (or they may just be lucky). Other firms will be unable to do so (or they may just have bad luck). As a result their profit rates will be diverse, some low, some slightly higher, some much higher, and so on.

The dynamic of competition, by forcing firms to invest in new and risky projects, continually generates new sets of diverse profit rates. Firms that achieve breakthroughs or monopoly power can achieve exceptionally high profit rates; the firms they compete with have lower profit rates. An advantage for one firm is a disadvantage for its rivals.

In contrast to these ways in which it produces divergent profit rates, competition also involves processes that tend to equalize profit rates. Price competition tends to equalize profit rates. In addition, when firms invest, they will often pull out of markets, industries, or sectors where the rates of return are low and move their operations to markets, industries, or sectors where the rates of return are high. This process will also tend to equalize profit rates. This is often true across countries, regions, or continents as well as across industries or sectors.

The movement of firms and investments from low-profit areas to high-profit areas will have two effects, both of which will tend to make profit rates more equal between the two types of areas. When firms exit from low-profit areas the industry supply curves for such areas will shift to the left. In Figure 11.4 this is shown (in the left diagram) as a shift in a particular industry supply curve from S_1 to S_2, which will raise the market-clearing price (P_2) for that industry or area from P_1 to P_2. However, this may have little effect on the firms that remain. Since the exit of some firms will

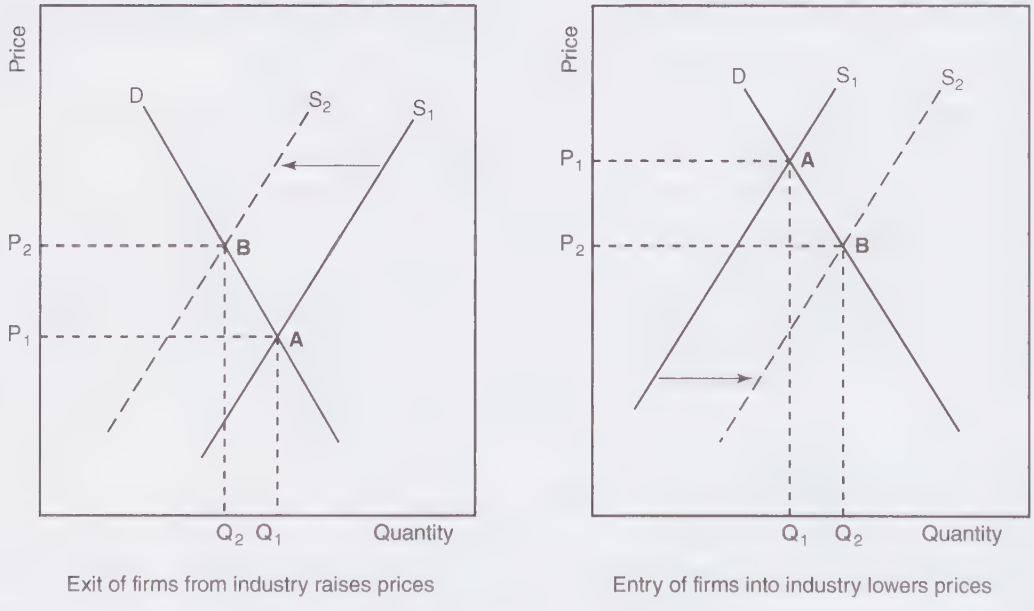


FIGURE 11.4 Exit and entry of firms from and into markets tend to equalize profits.

Low profits will drive some firms out of an initially low-profit industry. This will cause the supply of output in that industry to contract, shown in the left-hand diagram by the leftward shift of the supply curve from S_1 to S_2. This shift will allow the remaining firms to raise their price from P_1 to P_2, thus increasing their profits. Exactly the opposite happens in an initially high-profit industry (the right-hand diagram). High profits will induce new firms to enter this industry, increasing total output, so the industry supply curve will shift to the right, showing that there is more supply at each price. The higher output from entry of new firms forces all the firms in the industry to sell their output at a lower price (P_2), driving their profit rates down. Thus, exit causes the profit rate to rise in low-profit industries and entry causes the profit rate to *fall* in high-profit industries, so the profit rates in the two industries will tend to converge.

reduce the industry's (or area's) overall production capacity, the decline in quantity demanded due to the now higher price for the products of the remaining firms will be divided up among a smaller number of firms than before, and they will suffer little if at all. As a result the higher price (P_2) will have the effect of raising the profit rates of the firms remaining in the low-profit areas.

Second, as new firms enter the high-profit areas, the industry supply curves for these areas will shift outward to the right, as illustrated by the shift of S_1 to S_2 in the right diagram in Figure 11.4. The expanded supply (at every price) will cause the market in the high-profit areas to establish a lower market-clearing price, a movement shown in the right diagram in Figure 11.4 by the price change from P_1 to P_2. And though the lower price (P_2) will attract new customers to the particular industry (or area), these new customers must be shared with the additional firms that have recently entered this market, thus bringing about reduced profit rates in high-profit areas.

As firms' investments shift between areas or industries with widely divergent rates of return, profit rates tend to equalize, as exits raise profit rates in low-profit industries and entries lower profit rates in high-profit industries. If this were all that happened, profit rates would likely converge. But in practice, firms' competitive strategies seem to cause profit rates to diverge as much as to converge in an industry or region. The system is rarely at rest.

Toward Economic Concentration?

Economic concentration is the extent to which the economic activity of an industry or the whole economy is conducted by the largest firms.

Economic concentration refers to how much of an industry or of the whole economy is controlled by a very small number of the largest firms. Consider, for instance, two industries of equal size in terms of the total value of sales. In one industry there are one hundred firms, all about the same size, whereas in the other industry there are only four firms. The second industry is said to be more concentrated than the first. A common way to measure concentration in a sector is by the four-firm concentration ratio. This is measured by the total value of goods produced in a year by the four largest firms in the sector, divided by the total value of goods produced in the sector. For example, if the four firms with the most output of breakfast cereals were to produce 80 percent of the value of all the breakfast cereals produced in the United States, then the four-firm concentration ratio in breakfast cereals would be 80 percent (and in 2007, that is what it was).

Sectors with large economies of scale, like auto, tend to have high concentration ratios. In fact, in any sector in which you can name the top three or four firms in the United States, it is a good bet that the four-firm concentration ratio is high. On the other hand, sectors within the apparel industry, with very limited economies of scale (the main capital goods are sewing machines, which are not that expensive), tend to have low concentration ratios. We will look at some data shortly.

In principle, dynamic competition can produce either more or less economic concentration because, as with equalization of profit rates, competition both generates and erodes economic concentration. Competition generates concentration because greater size usually helps a firm to increase its profit rate, and therefore bigger firms usually can out-compete smaller firms. It does this in three main ways.

First, as we have seen, greater size brings lower costs and greater profits due to economies of scale. Second, greater size increases a firm's ability to achieve monopoly power and breakthroughs. Third, greater size expands the firm's bargaining power with respect to labor, finance, and suppliers, as well as its political and

bargaining power with respect to government. Thus big firms often grow at the expense of smaller firms. We have already established how economies of scale favor large firms, and have given the auto industry as a prime example. What about the other advantages of large size?

Large Size and Monopoly Power

Greater size puts firms in a better position to achieve monopoly power. If a firm is large enough to control the whole market for a product, it is a *monopoly*, and can charge whatever price brings it the most profit. High rates of profit can also be achieved through monopoly power shared by a few firms who collude in making decisions about pricing their products and other matters.

COLLUSION Price-fixing and other forms of collusion are illegal under U.S. and European antitrust law, and price-fixing cartels are a main target of investigations by antitrust authorities. Yet even when executives are given prison terms for violating the law against price-fixing and other forms of collusion, and even when fines that seem heavy are imposed on corporations, cartels continue to arise.

The EU, for example, imposed fines of about €1.5 billion (about $2 billion at the time) on seven corporate groups for fixing the prices of TV and computer monitor tubes from 1996 to 2006. There was nothing tacit about the way these firms operated: for years they met in various locations and explicitly discussed and decided on prices. Finally one cartel member told antitrust authorities about the cartel. The EU has a policy that if one cartel member informs antitrust authorities about a cartel of which the authorities were previously unaware, that member receives 100 percent leniency and is not obligated to pay a fine. That is what happened in this case. This policy has led to many new revelations about cartels, and many new prosecutions, and after U.S. antitrust authorities adopted a similar policy in 1993, it similarly led to revelations. The U.S. decision in the early 1990s to investigate international cartels and cooperate with foreign antitrust authorities also helped step up prosecutions.

Collusion is easier to establish and maintain in an industry in which there are only a few large firms. If there are many firms, it is much harder to establish cooperation. Hence, the presence of only a few large firms in an industry facilitates establishing and maintaining monopoly power and therefore high rates of profit. Yet even in industries with many small firms, collusion has sometimes also been achieved through trade associations: the association itself announces policies that the member firms are expected to follow.

EXPLOITING BREAKTHROUGHS Size also enhances a firm's ability to achieve breakthroughs and, more importantly, to profit from them. Consider technological innovations, for example. We might expect larger firms to spend a higher percent of their budget on research and development and perhaps to be more efficient in their research activities, because such activities themselves might have economies of scale. Neither of these expectations is borne out by the facts; actually, small and mid-sized companies seem to produce at least their share of innovations.

Where big firms dominate is in exploiting the profit potential of a breakthrough: To profit from a breakthrough (say, in personal computers), the new product must be produced in quantity, advertised extensively, distributed widely, and defended against would-be competitors. For instance, Apple and other small (and now forgotten)

companies pioneered the small desktop computer, but giant firms actually captured most of the market for such computers and reaped most of the profit.

A small firm, even one with a great innovation and big marketing ideas, generally has limited ability to do successful marketing, because it has limited profits and limited access to the amounts of credit needed to bring the innovation to fruition. Indeed, a lender may sometimes grant sufficient credit only if the lender is given control of the project. The small firm typically has no existing distribution network, its name is not widely recognized, and it has no funds for large-scale advertising. In this situation a small firm often chooses to sell its innovations to big firms, or to sell itself in its entirety to a large firm. Although the selling price may provide the small firm with a large profit relative to its modest size, the remaining profit to be made from exploiting a breakthrough will be captured by the large firm.

GOVERNMENT SUBSIDIES AND CONTRACTS Yet another way in which greater size may raise a firm's profit rate is by increasing its bargaining and political power. Recall the over $1 billion in subsidies given to the Ford Camaçari plant in Brazil when it was first constructed. Since 1980 in the United States, a total of at least $3.6 billion (more if converted to 2015 dollars) has been paid as subsidies by state and local governments to foreign automakers for locating plants in various states, particularly in the South. Some exceptionally large subsidies include $400 million for the Kia plant in West Point, Georgia, in 2006; $300 million for the Toyota plant in Blue Springs, Mississippi, in 2007; and initially $400 million (reportedly later $577 million) for the Volkswagen plant in Chattanooga, Tennessee, in 2008. In each of these three cases, this amounted to 20 to 60 percent of the announced planned investment in the plant.

Big firms with more bargaining power may gain easier entry into foreign markets, or lower interest rates and easier access to credit from banks. Political power may translate into more access to government contracts, better treatment in tariff or export policy (such as subsidies through the Export-Import Bank), and favorable government loans or loan guarantees. Thus large firms may very well be able to translate political power into higher profits. Perhaps the best examples of this are in the military-industrial complex, the relationship between giant producers of arms and military services and the Pentagon. This relationship keeps lucrative contracts flowing to such firms as Lockheed Martin, General Dynamics, Boeing, and Halliburton.

Notice that only when economies of scale are present will a larger scale of operations *directly* translate into higher profits. In all other cases, greater size will increase a firm's ability to do something else (such as gain monopoly power) that will, in turn, raise its profit rate. Firms may grow because all these factors interact: growth based on scale economies may enhance monopoly power, breakthrough profits may finance a firm's growth and enable it to capture the benefits of scale economies, and so on.

Competition can also work against economic concentration. The high profit rates attained by big firms can attract competitors. Firms seeking new profit opportunities will attempt to enter industries in which profit rates are high, even if the industry is dominated by a few large firms. Then again, entry of new firms can also frighten dominant firms into merging to defend their dominant position.

Aggregate economic concentration is the ratio of large companies' sales, profit, or number of employees to the economy's total sales, profit, or employees. If large firms are able to increase their share of particular markets while other large firms are not losing their share of other markets, concentration in the economy will increase. The extent of concentration in each industry and in the economy as a whole depends,

then, on which set of opposing forces is stronger—those forces (like the advantages of size) that tend to increase concentration or those forces (like the formation of new firms, or the enforcement of antitrust laws) that tend to erode concentration.

Trends in Concentration

Is there a general trend toward increasing concentration in the U.S. economy in recent years, propelled in part by large, highly publicized mergers? One 2016 study found that since the mid-1990s, concentration has increased in over three-fourths of sectors, mainly because large firms have grown even larger through mergers and acquisitions. In sectors where there is more concentration, they also found, on average profits are higher and stock returns are above normal.[6]

Figure 11.5 illustrates the diverse trends in selected sectors between 1997 and 2012, and in particular in some sectors with very high concentrations. What it shows is the four-firm concentration ratio for each sector, measured as the combined share of the top four firms in the sector's total value of output, measured by the value of goods shipped.

In some sectors (near the top of the figure), concentration rose from 1997 to 2012. In others (nearer the bottom), concentration declined. The latter might be because of increased competition through entry of new firms. However, another possible reason is that some large existing firms shifted part of their production abroad, where it was no longer counted in the data. Ironically, such firms thereby probably *gained* market share of *sales* in the United States, but not of U.S. *production*, which is what Figure 11.5 measures. Similarly, the fact that Ford opened a major auto plant in Brazil in 2001 during this period (and other U.S. automakers made similar moves), may show up in Figure 11.5 in the decline in concentration in U.S. auto manufacturing from 79.5 percent in 1997 to 60.2 percent in 2012. The decline in the concentration ratio reflects the relative increase in the share of "smaller" firms in total U.S. auto output—probably mostly a variety of additional foreign-owned auto plants established in the United States (often with the help of government subsidies). Foreign-owned auto plants in the United States now include Toyota, Honda, Nissan, Subaru, Hyundai, Kia, BMW, and Volkswagen.

Increases in economic concentration have tended to happen in waves of mergers and acquisitions. One large wave occurred in the late 1990s, in which the total value of mergers exceeded $1 trillion per year for several years. Deregulation in the financial sector, for example, led to a merger of two giants, Citigroup (banking) with Travelers (insurance and financial services) in 1998–1999. Oil companies Exxon and Mobil merged in 1998 and Chevron merged with Texaco in 2000.

Another major wave of proposed mergers started in 2014–2015. As Assistant Attorney General Bill Baer put it in March 2016 in testimony to the Senate subcommittee that oversees antitrust enforcement (see box "Antitrust"), "The merger wave is back. Big time."[7] He testified that in 2015, sixty-seven mergers were proposed of value $10 billion or more, double the number that size in 2014. Among them were

---

[6] Gustavo Grullon, Yelena Larkin, and Roni Michaely, "Are U.S. Industries Becoming More Concentrated?," paper presented at the China International Conference in Finance, July 7–10, 2016.

[7] Bill Baer, Assistant Attorney General, Antitrust Division [Department of Justice], before the Subcommittee on Antitrust, Competition Policy and Consumer Rights, Committee on the Judiciary, United States Senate, Hearing on "Oversight of the Enforcement of the Antitrust Laws," March 9, 2016.

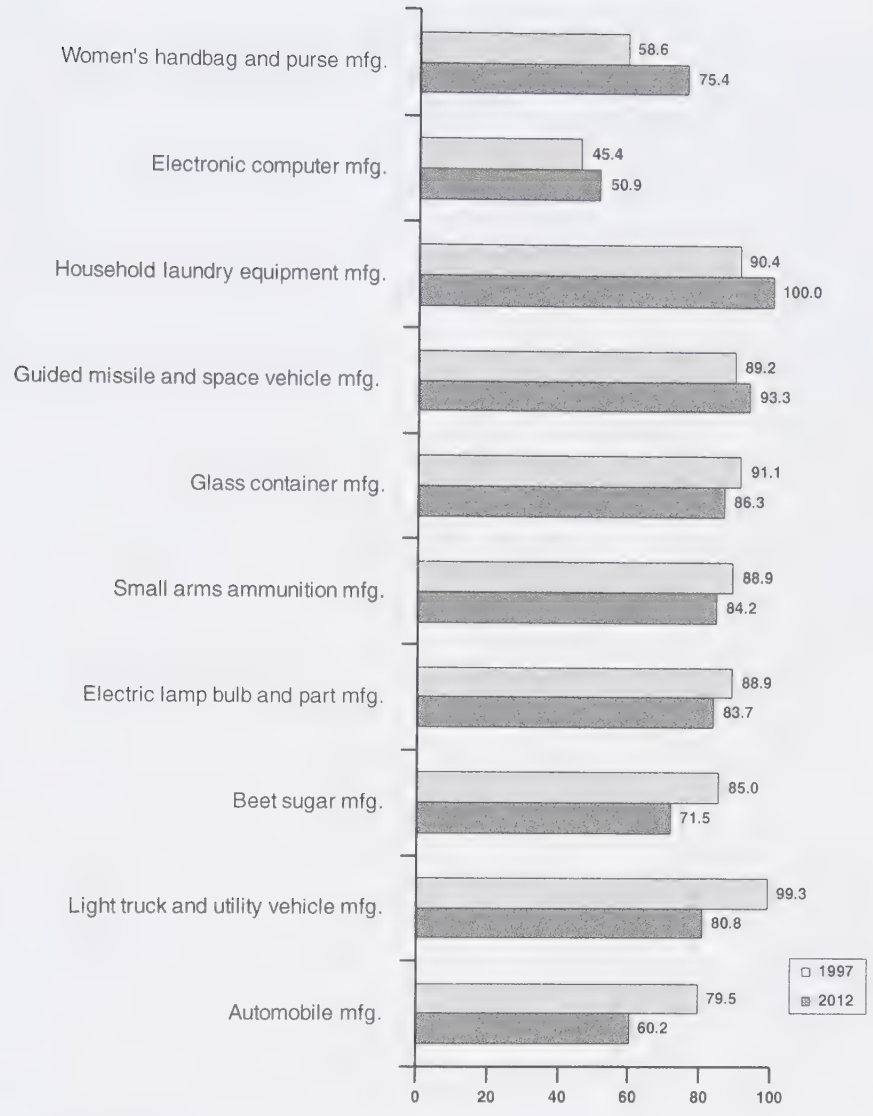
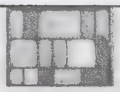


FIGURE 11.5 Diverse trends in industry concentration, 1997–2012.

The degree of concentration, as measured by the four-firm concentration ratio, grew in some sectors and fell in others over the period 1997–2012. The four-firm concentration ratio shown here is the share of the top four firms in the total value of goods shipped over the year by U.S. producers in the sector. Trends vary: sectors whose concentration rose are near the top; sectors whose concentration fell are near the bottom. Sectors were selected, among those for which data were reported in the same narrow (6-digit NAICS) category of manufacturing for both 1997 and 2012, to indicate a variety of products and trends. Many sectors have lower four-firm concentration ratios (e.g., apparel industry sectors), but are not shown. Narrowly defined sectors ("Beet sugar manufacturing" but not "Food manufacturing") tend to have higher concentration ratios.

Source: U.S. Census Bureau, *American Factfinder*: Download the file "EC1231SR2 - Manufacturing: Subject Series: Concentration Ratios: Share of Value of Shipments Accounted for by the 4, 8, 20, and 50 Largest Companies for Industries: 2012" from https://factfinder.census.gov/faces/nav/jsf/pages/download_center.xhtml; and U.S. Census Bureau, *Concentration Ratios in Manufacturing*, 1997 Census, Manufacturing, Subject Series, EC97M31S-CR, Issued June 2001, available at https://www.census.gov/prod/ec97/m31s-cr.pdf.



ANTITRUST

Entry of new, smaller firms into a market may whittle away dominant firms' market share, but it also provokes these large firms to try merge in order to restore their advantage. What sometimes stops such mergers is antitrust enforcement, or other government policy initiatives.

Likewise, firms that do not merge may still act as a shared monopoly—a cartel—by fixing prices, rigging bidding processes, dividing up markets, or colluding in other ways. Antitrust enforcement, if effective, can stop these forms of monopoly as well.

In a very short space of time in the last twenty years of the nineteenth century, in the United States, a series of very powerful "trusts" or monopolies were formed in several major industries—steel, oil, and others. New state and federal antitrust laws created an arsenal of legal weapons used to break up monopolies and restrain cartels.

- But is there really significant antitrust enforcement? Every two years or so, the congressional oversight committee for antitrust regulation takes testimony from antitrust enforcement officials. In March 2016, Bill Baer, assistant attorney general in the Antitrust Division of the Department of Justice gave testimony to a hearing of a senate subcommittee on antitrust and related issues. He reported that the Antitrust Division had
- sued in court to stop Electrolux (a maker of Kenmore appliances) from buying General Electric's appliance business and getting control of 70 percent of the market for affordable kitchen ranges;
- told Comcast and Time Warner Cable what anticompetitive problem it saw with their proposed merger, after which they abandoned merger plans;
- exposed price-fixing by big banks Citicorp, JPMorgan Chase, Barclays, and the Royal Bank of Scotland, which ended up pleading guilty to conspiring to rig the currency spot market;
- filed charges against over one hundred persons for conspiring to keep bids down at local real estate foreclosure auctions;
- prosecuted high-level corporate executives for price-fixing of auto parts.

Baer testified that total criminal antitrust fines collected were over $1 billion in each year from 2012 through 2014, and were $3.6 billion in 2015, a new high. The average prison sentence for antitrust convictions rose from seven months in 1986–1995 to twenty-four months in 2006–2015, and this included company CEOs, presidents, vice presidents, and other executives and managers.

The Department of Justice had long prosecuted price-fixing and other cartel behavior in the United States, but decided also to pursue international cartels beginning in the early 1990s, often cooperating with European antitrust authorities. A key new strategy was to announce that the first member of a previously hidden cartel to inform the authorities of the existence of the cartel would escape being fined—a strategy that the EU antitrust authorities also use. In both regions, the new policy led many more cartel members to reveal the existence of cartels.

Source: U.S. Department of Justice: Bill Baer, Assistant Attorney General, Antitrust Division, before the Subcommittee on Antitrust, Competition Policy and Consumer Rights, Committee on the Judiciary, United States Senate, Hearing on "Oversight of the Enforcement of the Antitrust Laws," presented on March 9, 2016; Margaret C. Levenstein and Valerie Y. Suslow, "Price-Fixing Hits Home: An Empirical Study of US Price-Fixing Conspiracies," *Review of Industrial Organization* 48 (April 2016): 361–79 (published online April 27 2016. DOI 10.1007/s11151-016-9520-5).

proposed big mergers in health insurance, agrochemicals ("biotechnology"), and pharmaceuticals.

In health insurance, among the five largest firms two mega-deals were proposed: Aetna wanted to buy Humana, and Anthem wanted to buy Cigna, making two enormous firms that would overwhelmingly dominate certain specific health insurance markets. In agrochemicals, two pairs among the six largest firms wanted to merge: Monsanto wanted to buy Bayer, and Dow wanted to merge with DuPont, combinations that the American Antitrust Institute said would give the two merged firms control of three-fourths of the market for corn and two-thirds of the market for soybeans.[8]

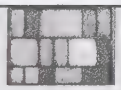
In the pharmaceutical industry there was one proposed mega-deal, the merger of giant drug makers Pfizer and Allergan. However, the enormous tax advantages of the deal were neutralized by changes in U.S. Treasury policy in fall 2015 and spring 2016, and the deal was then called off.

The effects of some mergers and acquisitions may explain some trends in Figure 11.5. For example, Whirlpool bought Maytag in 2006, and from 1997 to 2012 the four-firm concentration ratio in "Household laundry equipment manufacturing" rose from 90.4 to 100.0 percent. Likewise, a series of mergers in the arms industry culminating in Lockheed's 1995 merger with Martin Marietta (the second and third largest arms makers at the time) to form Lockheed Martin—now the largest arms producer—may explain the high concentration ratio (in both 1997 and 2007) in guided missiles.

Antitrust Enforcement

Antitrust enforcers mainly do two things. One is responding to merger or acquisition announcements by investigating whether there are "anticompetitive" aspects to them that could be challenged in court, or told to both firms with the goal of reaching a settlement with them to sell off certain operations to prevent monopoly and remove the anticompetitive problems. The other activity is investigating illegal collusion among firms. The box "Cartels" gives examples of cartels that were found to have committed price-fixing or other antitrust violations.

[8] American Antitrust Institute, "American Antitrust Institute, Food & Water Watch, and National Farmers Union Urge the U.S. Department of Justice to Challenge the Dow-DuPont Merger," May 31, 2016, available at http://www.antitrustinstitute.org/content/aai-fww-and-nfu-urge-us-department-justice-challenge-dow-dupont-merger



Cartels

Cartels operate in many industries to fix prices, or to collude in other ways that are violations of antitrust statutes. More than fifty firms were indicted by the U.S. Department of Justice in the last decade of the twentieth century for price-fixing, and the pace has stepped up since then: In 2013 alone, twenty-one corporations and thirty-four individuals were indicted for various antitrust violations. The table below lists key data on a few of the cases in which firms were charged. The most recent two cases listed are from EU indictments.

continues

EU antitrust authorities have been increasingly active, imposing on average over $2 billion per year in fines during 2005–2014. For the TV and computer monitor tubes cartel alone, €1.5 billion in fines were imposed (around $2 billion). Fines imposed on cartels by the Antitrust Division of the U.S. Department of Justice, often for price-fixing, reached over $1 billion annually in 2012–2014, and $3.6 billion in 2015. Apart from the cases shown, in both the United States and Europe price-fixing cartels are especially frequent in the chemical industry (for instance, such cases were one-sixth of all cases brought by U.S. antitrust enforcers in 1993–2013).

Industry	Period	Country of origin of firms	Number of firms	Market share	Market
Cement	1983–1994	Europe	33	80%	Europe
Sugar	1986–1990	Ireland, UK, Denmark	2	90%	UK
Plastic tubes	1986–1995	Switzerland, U.S.	3	95%	US
Plastic dinnerware	1991–1992	Canada, US	2	90%	US
Vitamins	1990–1999	Canada, Japan, Germany, U.S., Switzerland	3	75%+	world
Nucleotides	1992–1996	Japan, S. Korea	1	45%	world
Beer	1993–1998	Belgium	2	70%	Belgium
Fine arts	1993–1999	U.S., UK	2	95%	world
Tubes for TVs and computer monitors	1996–2006	N/A	7	less than 80%	Europe
Alternators & starters	2004–2010	Japan	3	N/A	Europe

Source: Margaret Levenstein and Valerie Y. Suslow, "Contemporary International Cartels and Developing Countries: Economic Effects and Implications for Competition Policy," 71 *Antitrust Law Journal* 801 (2003–2004). European Commission, EU, 1. Cartel statistics, 1.2. Fines imposed (not adjusted for Court judgments)—period 1990–2016, available at http://ec.europa.eu/competition/cartels/statistics/statistics.pdf. "Antitrust: Commission fines producers of TV and computer monitor tubes € 1.47 billion for two decade-long cartels," available at http://europa.eu/rapid/press-release_IP-12-1317_en.htm?locale=en.

Conclusion

As we have seen, there are market forces tending to create economic concentration, and other market forces that tend to increase competition. Yet the outcome also depends on what rules are in place: what antitrust laws exist, and what regulations and judicial decisions interpret their meaning. It further depends on how well antitrust authorities cooperate with one another, and whether they develop effective incentive structures (like ways to encourage cartel members to turn themselves in first to gain an amnesty from fines). With all these forces and counter-forces, it appears that

waves of proposed mergers are continuing to occur, and yet the degree of economic concentration will remain a contested issue.

An important issue is the degree of concentration in the media industry. Very large firms have come to dominate the industry. However, some attempts to increase market power have ended up being unsuccessful—notably the AOL–Time Warner merger of 2003, which is generally seen as a major failure. Still, the more concentrated the sector becomes, the more the continued existence of many smaller firms providing Internet services, publishing periodicals, distributing films for screening, and so forth, is jeopardized. There are less than half a dozen major book publishers. Although many small presses still manage to survive, increasing concentration means that readers, moviegoers, Internet users, and citizens in general may now have access to a much more limited range of artistic expressions, perspectives, and opinions. Such concentration can limit the variety of information and opinion people need in order to participate in a well-informed way in the political process. And as companies get ever larger, their influence in the political process becomes greater relative to that of individuals.

Suggested Readings

Benjamin Bagdikian, *The New Media Monopoly*, 20th edition (Boston: Beacon Press, 2004).

Michael Best, *The New Competition: Institutions of Industrial Restructuring* (Oxford: Polity, 1990).

John Kenneth Galbraith, *American Capitalism: The Concept of Countervailing Power* (Piscataway, NJ: Transaction Publishers, 1997).

David Kotz, *Bank Control of Large Corporations in the United States* (Berkeley: University of California Press, 1978).

LeClair, Mark S., *Cartelization, Antitrust and Globalization in the U.S. and Europe* (Routledge, 2013).

Margaret C. Levenstein and Stephen W. Salant (eds.), *Cartels* (Cheltenham, UK: Edward Elgar, 2007).

Morgan Lewis (law firm), *2016 Global Cartel Enforcement Report* (published 2017, available at https://www.morganlewis.com/~/media/files/publication/report/2016-year-end-cartel-report-january-2017.ashx).

Barry C. Lynn, *Cornered: The New Monopoly Capitalism and the Economics of Destruction* (New York: Wiley, 2011).

Joseph Schumpeter, *The Theory of Capitalist Development* (Cambridge, MA: Harvard University Press, 1934).

Carl Shapiro and Hal R. Varian, *Information Rules: A Strategic Guide to the Network Economy* (Boston: Harvard Business School Press, 1999).

Helen Shapiro, *Engines of Growth: The State and Transnational Auto Companies in Brazil* (Cambridge, UK: Cambridge University Press, revised edition, 1994).

Stork, Joe, *Middle East Oil and the Energy Crisis* (New York: Monthly Review Press, 1975).

Appendix to Chapter 11

Figure 11.1(a) shows the demand curve for brownies facing the Sweets Bakery. Figure 11.1(b) shows the profit that results from each price level that the bakery owner could choose. This Appendix explains these calculations, assuming that the *unit cost* of making a brownie is uc = \$1.

The equation of the demand curve (see Figure 11.1a) is $P_z = 3 - 0.005Z$, which can be solved for Z and then simplified, as follows:

(11A.1) $$Z = (3 - P_Z)/0.005 = \mathbf{600 - 200P_Z}$$

From equation 11.3 we know that total profit R can be written as $R = (P_z - uc)Z$. Now we replace Z by the expression in equation (11A.1), and use the fact that $uc = 1$:

(11A.2) $$R = (P_z - 1)(600 - 200P_z) = -200P_z^2 + 800\,P_z - 600$$

We can now calculate profit R for values of P_z from \$1 on up, and see (as in the profit table on page 240) that the profit-maximizing price is somewhere near \$2.00.

Or, using calculus, we find that profit is a maximum at the price P_z that makes $dR/dP_z = 2(-200)P_z + 8000 = 0$, namely, the price

$$\boldsymbol{P_z = 2}$$

Plugging this price into equation (11.A.2) gives $R = -200(2)^2 + 800(2) - 600$, or

$$\boldsymbol{R = 200}$$

In general, if the demand curve is $P_z = a - bZ$, we can solve it for Z to get

$$Z = a/b - (1/b)P_z$$

and show by the above method (but using algebra) that, under the assumptions of this example, the price that maximizes profit is

(11A.3) $$\boldsymbol{P_z = \tfrac{1}{2}\,(a + uc)}$$

The profit at that price is $R = (P_z - uc)Z = [\tfrac{1}{2}\,(a + uc) - uc]\,[a/b - (1/b)\,[\tfrac{1}{2}\,(a + uc)]]$ or

(11A.4) $$\boldsymbol{R_z = \tfrac{1}{2}\,(a - uc)^2/b}$$

(Plug in values of a, b, and uc to get $P_z = 2$ as the profit-maximizing price and $R = 200$ as the profit at that price, which we already found above.)

From equation 11A.3 we can see that lowering unit cost uc will lower the profit-maximizing price; and from equation 11A.4 we can see that lowering uc will raise total profit. However, this is true only if no competitors lower *their* prices after Sweets does. If they do (and well they might), *the demand curve facing Sweets Bakery will shift*, and the situation will change—so we cannot be sure what will happen. In many situations, however, the main way to engage in price competition is to find ways to reduce costs.

Wages and Work

12

Our society is often described as a consumer society, one in which people express their creativity and enjoy a sense of freedom by consuming. Thus, we go window-shopping or web-surfing, make our choice of just the right blouse or shirt, and expect that wearing it will say something about who we are. Some of us even go on shopping sprees to fight off boredom or depression.

If consumption is the bright side of the capitalist economy, however, production is often thought of as its dark side. People do not say, "I'm feeling a little down, I think I'll go down to the office and put in a little work." Indeed, in our society work has a bad reputation. The great American novelist William Faulkner expressed our generally dim view of labor in this way: "You can't eat for eight hours a day nor drink for eight hours a day nor make love for eight hours a day. All you can do for eight hours a day is work. Which is the reason why man makes himself and everybody else so miserable and unhappy."[1]

Perhaps Faulkner had a point. Although it is generally a good thing to be employed, a job may be little more than a way of making the money one needs to be able to consume later. Work may be necessary or preferable to doing nothing, but in many cases it does not make an individual feel creative or fulfilled. Instead, it frequently engages people in repetitive tasks and subjects them to the unwelcome authority of supervisors.

Despite its deficiencies, however, work is a very important aspect of people's lives. Of course, we depend on our jobs for income. But work is also, for many of us, a central aspect of our identities: it allows us to say who we are. We spend a major portion of our waking hours working, so whether work is a happy or an anxious experience, whether it fosters boredom, resentment, fear, or pride, it will affect our overall mental state. Where we work may be where we find our "place" in the world, it may provide us with an experience of community, it will certainly be among the places where we get to know other people, and quite possibly it will be a place where we have friends. Moreover, our job may offer us the chance, however limited, to exercise responsibility, show creativity, or rise to the challenge of hard work.

---

[1]Quoted in Studs Terkel, *Working* (New York: Pantheon Books, 1972), ix.

For all these reasons access to work and the conditions of work are enormously important for most of us.

Nevertheless, many people object to the fact that they have very little choice about what they do at work or how they do it. Why is production organized from the top down, with nearly every work process overseen by executives, managers, supervisors, "team leaders," or other authorities? Is the top-down control of work and its fragmentation into repetitious tasks an inevitable result of modern technology? Or is it just "natural" that work is unpleasant?

If one's job is often a problem, not having one can be even worse. Joan Robinson, a twentieth-century British economist, once quipped that the only thing worse than being exploited by a capitalist is not being exploited by one. Moreover, the loss of one's job can be not only a financial disaster but a personal calamity as well. Suicide rates, marital breakups, and mental illnesses are all quite closely correlated with the number of people who are unemployed.

This chapter investigates the organization and remuneration of work. As noted in Chapter 7, the majority of people in the United States work for an employer. We also explained in Chapter 5 that the relationship between employers and workers is a class relationship, with employers and workers constituting different classes. This employer-worker relationship differs from the relationships examined in the last chapter—capitalist-capitalist relationships—and it also differs from the relationships workers have with other workers. Whereas capitalist-capitalist and worker-worker relationships are horizontal in nature, the employer-worker relationship is a vertical one, and involves the command dimension of an economy. Between-class relationships are vertical in the sense that, within them, people relate to each other as supervisors and subordinates, with some being higher up and others lower down in a hierarchy. Of course, most people in a hierarchy are *both* supervisors *and* subordinates, holding power over those below them while at the same time having to obey (or please) those above them.

Many economists do not take account of the ways that employer-worker relationships differ from buyer-seller relationships. These economists assume implicitly that capitalists buy labor in just the same way as they buy materials and machines. They assume, in other words, that workers and employers agree ahead of time on the amount of work effort (e) to be done each hour and that workers always deliver the expected work effort. Thus they treat labor as just another purchased input to the production process. Indeed, Paul Samuelson, one of the most highly regarded economists of the twentieth century, once said: "In the competitive model, it makes no difference whether capital hires labor or the other way around."[2] In fact, it does make a significant difference.

This chapter focuses on the relationship between workers and their employers. The main ideas of the chapter are that (a) *employers and their employees have fundamentally conflicting interests in the workplace,* and (b) *employers organize work hierarchically, in a top-down way, in order to extract work from their employees so as to make profits.* These central ideas are expressed in six main points:

1. The activity of producing things—*work*—is not naturally boring, oppressive, or limiting. Neither is it naturally exciting and liberating. It is not "naturally" anything. What work is like—our experience of it and our reaction to it—depends

[2]Paul A. Samuelson, "Wages and Interest," *American Economic Review* 62, no. 6 (Dec. 1957): 894.

in important ways on how the labor process is organized, how the rest of society is organized, and where we fit into the labor process.

2. A capitalist enterprise—a firm—is a *system of command*, in which power is exercised from the top down. In the marketplace people and firms interact primarily through voluntary offers and agreements; within firms the main form of interaction involves the exercise of authority by supervisors over workers.

3. There is a conflict between what *employers* and *employees* want to get out of the labor process. This conflict is over wages and the pace of work, two of the most important determinants of the profit rate. The conflict between employers and employees arises from their opposing positions in the labor process and is inherent in a system geared toward profit-making.

4. Capitalist firms are organized hierarchically, with control exercised from the top, in order to enhance the ability of employers to *extract work* from their employees. For some kinds of work, people who are independent contractors can contract with the firms to perform tasks specified in a contract. But for regular employees whose tasks are not (and cannot be) fully specified in advance in a contract, work must be extracted because it cannot be secured by contract.

5. Employers are able to exercise power over their employees because good jobs are scarce and because there is almost always an excess supply of labor—*unemployment*—in the labor market. The labor market does not operate like other markets, which tend to "clear," thereby eliminating excess supply. Rather, excess supply of labor (unemployed people who want to work) is a chronic feature of labor markets.

6. Excess supply occurs in labor markets because employers wish to exercise power over workers. To achieve this power they pay wages that are above the minimum that their employees would accept, thus putting the workers in situations in which they have "something to lose" if fired. Similar workers without jobs would be happy to take the employed workers' jobs at lower wages, but the employers are not interested in offering the jobs at lower wages. Hence the workers without the jobs are *involuntarily unemployed.*

In this chapter we investigate the organization of labor processes and examine conflict in the workplace. Why is work organized hierarchically? What exactly is the conflict about, and why does it exist? In the next chapter we explore how employers and employees respond to the conflict.

Work, Sloth, and Social Organization

Though some people enjoy their work, others start living when they get off work in the afternoon and spend their time looking forward to the weekend. Many think of their days as divided between "work" and "leisure"—the first being a "bad" that must be endured in order to be able to acquire "goods" for consumption.

Conventional economists say that people derive "utility" from consumption and get "disutility" from work. This view undoubtedly expresses what many people feel about work. But it is incomplete in two ways. First, it implies that work is

naturally a "bad," something that by its very nature is unpleasant. Second, it ignores many aspects of work that make working an important and positive experience for people.

It is true, of course, that work takes time, and there are only twenty-four hours in the day. Because of this there is always an *opportunity cost* to work. If you are working, you cannot also be at the beach (unless you are a lifeguard). In this respect, of course, work is not different from any other activity (including going to the beach): whenever we spend time doing something, we are spending time doing it that we could have spent doing something else.

Why, then, should work be singled out for disparagement? One reason often given is that humans have a natural tendency toward sloth and therefore dislike doing things that require effort. Another view is that people do not like work because they do not like accepting the authority of others. A third opinion is that people hate work because it requires the repetition of limited tasks. But, in fact, none of these problems is actually rooted in human nature or intrinsic to work itself.

The view of people as naturally slothful is contradicted by the fact that many people get pleasure from exerting themselves in sports or other recreational activities. Most people can think of occasions when they have worked very hard, even at extremely demanding tasks, and greatly enjoyed it. People building a new room on their house, for example, may put in long hours on weekends, working at a pace that would provoke resentment if it were required in the office or at the plant. To cite another example, parents engage in one of the most physically, intellectually, and emotionally challenging types of work, raising children, but they often find the process rewarding.

With regard to the dislike of having to accept authority, it happens that people who resent supervision on the job may enjoy performing under the direction of a talented orchestra conductor or being told what to do by an experienced football coach. Thus, the aversion to accepting authority is hardly universal.

Finally, it is entirely understandable that people do not enjoy work when it involves performing the same task over and over again. But, again, such conditions are not intrinsic to work itself.

The problem, it seems, is not that work is naturally unpleasant or boring but, rather, that some types of work are objectionable. The examples mentioned above suggest that work may be repugnant because of the way it is organized or for whom it is done. Physical exertion, discipline, and responsibility—whether in sports, homebuilding, music, or child-rearing—are often gladly accepted, even welcomed, if the activity is one that a person has chosen, believes in, benefits from, or performs for someone he or she cares about.

The idea of work as a "bad"—or as generating "disutility"—is wrong for another reason. People often derive satisfaction from producing useful things, from developing their skills, or from associating with others at their workplaces. Respect, friendship, pride of accomplishment, and a sense of doing something useful are all, at least potentially, byproducts of human labor.

If many people see work as an undesirable activity, it is because of how production is organized, who controls it, and how the benefits from it are distributed. In the United States today—indeed, in all capitalist economies—most work other than raising children at home or performing other household tasks is done in firms. Understanding work thus requires an understanding of the organization of firms.



On Human Work: Excerpts from a Papal Encyclical

[T]hrough work man not only transforms nature, adapting it to his own needs, but he also achieves fulfillment as a human being and indeed in a sense becomes "more of a human being." . . . [But] *it is well known that it is possible to use work in various ways against man, . . . that work can be made into a means for oppressing man, and that in various ways it is possible to exploit human labor, that is to say, workers.*

It is obvious that when we speak of opposition between labor and capital we are not dealing only with abstract concepts or "impersonal forces." . . . Behind both concepts are people, living, actual people. On the one side are those who do the work without being the owners of the means of production, and on the other side are those who . . . own these means or represent the owners.

Since the concept of capital includes not only the natural resources placed at man's disposal, but also the whole collection of means by which man appropriates natural resources and transforms them . . . it must immediately be noted that all these means are the result of the historical heritage of human labor.

Furthermore, in the church's teaching, . . . [the means of production] *cannot be possessed against labor, they cannot even be possessed for possession's sake, because the only legitimate title to their possession—whether in the form of private ownership or in the form of public or collective ownership—is that they should serve labor and thus by serving labor that they should make possible the achievement of . . . the universal destination of goods and the right to common use of them.*

—From the papal encyclical *On Human Work*
Pope John Paul II
September 14, 1981

The Capitalist Firm as a Command Economy

A capitalist economy is composed of both firms and markets. Markets, as we have seen, involve *exchanges* of commodities for money and of money for commodities, and these exchanges are mostly voluntary (see Chapter 8).

The firm, however, is different. Inside the firm—behind the office door or inside the factory gates—people are generally not engaged in exchanging one activity for another; rather, their interactions are governed by **command relations**: they give and take orders. Accordingly, the capitalist firm can be described as a "command economy" (a term usually used in reference to centralized economic systems such as that of the former Soviet Union).

Command relations are relationships between superiors and subordinates in which the superior exercises substantial power over the subordinate.

Picture a capitalist economy as a sea dotted with many islands, some large, some small. The sea is the market and each island is a firm producing a particular product. The islands exchange their products with other islands, and they also sell products to individuals in the sea. The trades *among* islands and individuals involve (horizontal) market exchanges; but on each island there is a ruler, or chief, or boss, who orders other islanders what to do, so that production is organized by means of (vertical) command relations.

A shorthand way of describing one important aspect of a capitalist firm is to say that it is a social organization of production in which some people try to get other people to do the work. What distinguishes the political economy approach from the view of the economy presented in conventional textbooks is the inclusion of the

social relations of production—treating the firm as a command economy—in the overall picture of the profit-making process.

Just as in political economy (see Chapter 11), conventional economists view profits as being made because some entrepreneurs are able to anticipate better than others where demand will be growing faster than supply, where a new product or production process (a breakthrough or a technical innovation) might provide an advantage over competitors, or where a degree of monopoly power might be achieved. In other words, success is due to ingenuity or skillful management, not to exercise of power. And conventional economists also see profit as a reward for risk-taking on the part of investors and business owners. But the conventional approach tends systematically to ignore command relationships.

It is important to reiterate that, as shown in Figure 10.1, a capitalist economy involves both exchange relations—competition—and command relations. The previous chapter focused on product market competition among firms (exchange 2 in Figure 10.1) seeking to maximize their sales and profits. Here we examine exchange transactions in the labor market (included within exchange 1 in the figure) and command relations in the workplace (the "realm of production" in the figure).

Of course, saying that the firm is a command economy does not mean that all the commands given in a particular firm will be willingly obeyed. Employers generally have the upper hand, but they do not always get what they want.

The Conflict Between Workers and Employers

Workers and employers occupy different positions in the production process. Their interests differ, and this tends to put them in conflict. This is particularly true when the head of a firm really is concerned primarily with making money rather than with producing a particular product well. And, as we saw in Chapter 10, capitalism encourages firm owners to think of how to make money more than how to make T-shirts, or brownies, or steel. Workers perform the work required to produce something and are paid a wage for their labor time. Employers hire workers, direct their activities in the workplace, and have the right to sell the firm's output.

What do employees want from their jobs? They want their jobs to be not excessively tiring; they want their wages to be as high as possible; they want to have interesting work; they want to work in safe and healthy workplaces; they want to have flexible hours, a say in how the workplace is run, long vacations, the right to stay home when a child is sick or another one is born, and so on. In order to keep things simple and to explain the main points, our discussion here will focus on two issues: (a) how tiring a job is, and (b) how much it pays. Employees want e (work effort per hour) not to be excessively high and w (the hourly wage, including benefits) to be as high as possible.

And what do employers want? Some want to build a business that produces a particular sort of product. But most want to maximize the firm's profit rate r, and those who have other goals will often lose the competitive battle to employers who *do* want to maximize their firms' profit rates. There are many ways to do this (see Table 10.3), but this chapter is about ways that just involve the hourly wage w and the effort per hour e. We want to know how both of these enter into both the workers' behavior and the employer's strategy. To focus on just these two variables, we first need to do two things. One is to remind ourselves where w and e fit into the profit

equation. The second is to make clear what we must assume about the situation in order to be able to focus only on w and e.

First, as we know, the employer wants to maximize the profit rate r, and we know too that $r = R/K$. That is, the profit rate is total profit R, divided by the value of the capital goods K owned by the firm. But in this chapter we are not particularly concerned with changes in K. So to simplify the following presentation, we take note of the fact that K is often constant over a period of months or even a few years, because in many firms, changes in K mainly happen when the management decides to do a somewhat major investment project. Since a major investment can disrupt production, it tends to be done all at once. At least several new machines are bought, and the plant layout may also be rearranged, with the aim of increasing efficiency. Thus investment projects usually are done only from time to time, with gaps of months or years in between. Since we are interested here in wages and effort (not changes in K), we confine our analysis to periods when investment projects are not taking place, and so K is not changing.

Since $r = R/K$, holding K fixed ensures that the profit rate r rises whenever total profit R rises; likewise, r falls whenever R falls. That is, the two measures of profit rise or fall together. This in turn means that, to maximize r, the firm owner or manager can simply try to maximize total profit R.

We simplify matters further by leaving out of the picture any fixed costs like those described in Chapter 11. Putting in fixed costs would not make any difference in the conclusions we would reach, but it could make the argument a little harder to understand.

We recall (from Equation 10.4) how total profit R can be expressed as the value of output $P_z Z$ minus all costs, namely, materials costs $P_m M$ plus labor costs wH:

(12.1) $$R = P_z Z - (P_m M + wH)$$

where R = total profit (in dollars)

P_z = price of a unit of output

Z = quantity of output

P_m = price of materials used up and wear and tear on capital goods

M = quantity of materials used up and quantity of depreciation of capital goods

w = hourly compensation, including hourly wage and benefits

H = total number of hours worked by all employees

We said that the employer will seek to maximize total profit R, but it turns out that we can usefully further simplify matters by thinking about production of a single unit of output. (The following discussion is similar to that surrounding Equations 10.1–10.9, as well as Equations 11.1–11.3; it may be useful to turn back to review the pages of these chapters, although most of the concepts are summarized below.) We can convert the equation for total profits R to an equation for profit per unit of output by dividing both sides of Equation 12.1 by the number of units of output Z, as in Equation 10.7:

(12.2) $$\textit{Profit per unit of output} = \textit{Price} - (\textit{unit materials cost} + \textit{unit labor cost})$$

or

$$\frac{R}{Z} = P_z - \left(\frac{P_m M}{Z} + \frac{wH}{Z} \right) \tag{12.3}$$

As in previous chapters, we note that the materials cost per unit of output is typically the same for all units of output, because the same components go into each unit of output, and the same equipment is used in the same way to produce each one. As in Chapter 10, we refer to the *unit materials cost* as *umc*. Similarly, since the same labor process is used to produce all units, we regard the labor cost per unit of output as uniform for all units produced, and we call it the *unit labor cost*, denoted *ulc*. This allows us to write a much simpler equation for profit per unit of output (see also Equation 10.8):

$$\frac{R}{Z} = P_z - (umc + ulc) \tag{12.4}$$

where R/Z = profit per unit of output

umc = unit materials cost (materials cost per unit of output, or $P_m M/Z$)

ulc = unit labor cost (labor cost per unit of output, or wH/Z)

Again, to simplify matters, we are going to view both P_z and *umc* as fixed in this chapter. We regard P_z as fixed simply because strategies that involve changing P_z were already discussed in Chapter 11 (and will not be discussed in this chapter). We also regard the materials cost per unit of output as largely set by the technology and by the prices of materials, so *umc* is fixed as well, at least between major investments in new machinery and equipment, while the capital goods are not changing.

With the price and materials cost fixed, the firm owner/manager can raise unit profit R/Z by lowering *ulc*. Of course, we will need to think about the *ulc* from the workers' viewpoint as well. The unit labor cost can be expressed as the ratio of the hourly wage to the average worker's output per hour:

$$\text{unit labor cost} = \frac{wH}{Z} = \frac{w}{Z/H} = \frac{wage/hour}{output/hour} = \frac{w}{z} \tag{12.5}$$

where ulc = unit labor cost

z = output per hour, expressed in physical units; $z = Z/H$

This says that the unit labor cost is the hourly wage (w) divided by the output per hour (z). Notice that although we have expressed the *ulc* as the ratio of two *hourly* measures, the "hour" units cancel out, and as a result *ulc* is *not* an hourly measure. The *unit labor cost* is simply the amount of money paid to labor for each unit of output produced. Next we rewrite this in a still more useful form.

In Chapter 10's example of the Good Cod Fishing Company, we learned that the number of units of output per hour z can be thought of as the product of (1) *work effort* and (2) the rate at which work effort is turned into output, or the *efficiency of labor*. These variables were defined and measured in terms of *work effort per hour* (e) and *output per unit of effort* (f). We also learned in Chapter 10 that when these

two quantities are multiplied together, we get *output per hour* $(Z/H) = z = ef$. We use this to rewrite the equation for the unit labor cost in a way that involves both the hourly wage w and the effort per hour e:

(12.6)
$$ulc = \frac{w}{z} = \frac{w}{ef}$$

where e = work effort per hour

f = output per unit of work effort

In the Chapter 10 example of the Good Cod Fishing Company, work effort per hour (e) was measured by the number of times per hour a worker threw a net into the water to catch fish, and efficiency of labor (f) was measured by the number of fish caught by one throw of the net. When a worker threw the net five times an hour ($e = 5$ throws/hr.), and the number of pounds of fish caught per throw averaged 5 ($f = 5$ lbs./throw), then the output of fish per hour (z) was:

(12.7) Output per hour $= z = ef = 5(5) = 25$ lbs. of fish per hour

The hourly wage w was \$18.50, so the unit labor cost ulc was w/z = \$18.50/25 = \$0.74/lb. (see Table 10.2). If Good Cod could get its workers to cast six nets per hour, with no change in any other profit rate determinant—and if, in particular, the number of fish caught per cast were unchanged—then they would produce 30 pounds of fish per hour, and the profit rate would rise.

However, bear in mind that changing a variable in an equation and calculating the results might not take full account of what might happen in reality. If you were told to work faster for the same hourly wage, how would you react? Would you think this was fair? Workers might well respond by slowing down instead of speeding up their work, or there might be a sudden deterioration in the quality of their work. Sometimes it is wiser for an employer to use a carrot instead of a stick.

A company can raise its rate of profit (assuming no changes in its other profit rate determinants) by finding ways to reduce its unit labor cost (ulc). As Equation 12.6 shows, this can be done by lowering w, raising e, raising f, or achieving some combination of all three. Thus, employers seeking to increase their rates of profit will try to reduce w and increase e and/or f.

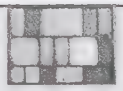
For two of the elements in unit labor costs, employers and their employees want exactly opposite things: employers want lower w and higher e, whereas employees want higher w and lower e. With regard to f, the third variable determining ulc, the interests of employees may coincide with those of their employers; we discuss this point in Chapter 13.

With w and e, then, the conflict between an employee and his or her employer is direct and obvious. A company can raise e only by getting its employees to work harder (more rapidly), and it can reduce w only by paying lower wages.

What determines the levels of e and w? Both are determined in important measure by the relative bargaining strengths of employees—individually or collectively—and their employers. In men's professional basketball, for instance, the National Basketball Players Association collectively bargains with the team owners, represented by the NBA Commissioner, over how much each team must spend on players' salaries.

Collective bargaining occurs when, in negotiating wages and other employment conditions, workers are represented collectively by a union; employers may also be collectively represented by an employers' association.

Consider wages first. For fifteen out of sixteen private sector employees (nongovernmental employees, that is), wage bargaining is mostly a matter between individual employees and their employers, since only one of sixteen private sector workers belongs to a union. When employees are represented by a union, the bargaining takes place between union negotiators and representatives of the employer, and this is called **collective bargaining**. In the coalfields of the eastern United States, for example, the United Mine Workers (UMW) union represents the unionized coal miners in this area, while a mine owners' council bargains for the coal mining companies.



Unit Labor Costs

In 2013, according to industry consultants O'Rourke Group, Partners LLC, the average payment to labor for making one ordinary knit polo shirt for a Bangladesh apparel factory was 12 cents. That was the *unit labor cost* of the shirt. Its retail price was $14.00.

In Bangladesh, the minimum monthly wage is $68 (and was in 2013), while in the United States it is and was about $1,260 for an average month of eight-hour working days. Thus a Bangladeshi garment worker makes something like (68/1,260) ($7.25/hr.) = $0.40 per hour. It evidently takes about 0.3 hours (18 minutes) to make a polo shirt, and 3.33 such shirts can be sewn in an hour.

If we had started out with just the fact that 3.33 shirts were produced per hour and that the wage was $0.40 per hour, we could have calculated the *unit labor cost* of assembling a polo shirt: w/z = ($0.40/hr.)/(3.33 shirts/hr.) = $0.12/shirt

But if the shirt sells for $14.00, who gets the other $13.88, if not the apparel workers? The table to the right gives the answer.

Notice that neither the labor cost of producing the materials for the shirt, nor the cost of "finishing" (applying a coating to make ironing unnecessary, for example) is explicitly included in the table. The garment factory itself pays for the fabric, buttons, and other materials, as well as the finishing process—the goods and services it buys—without knowing or caring how much of that money goes to labor in the textile mill or other suppliers to the garment factory. Such indirect labor costs are not regarded as labor costs by the company making the shirts: the company sees only the cost of a yard of cloth and the cost of a machine that will wear out after being used to produce a certain number of shirts. For this reason the reported unit labor cost understates the actual unit labor cost of producing polo shirts in Bangladesh—though probably only by a few cents.

Costs of a polo shirt made in Bangladesh in 2013:

$ 3.69	materials & finishing
$ 0.12	labor cost
$ 0.07	factory overhead
$ 0.18	agent
$ 0.58	factory margin
$ 1.03	freight/insurance/duties
$ 5.67	**Total cost to retailer**
$14.00	**Retail price to consumer (with ~150 percent markup)**

Source: O'Rourke Group, Partners LLC, maclean.ca, April 2011, available at http://www.ecouterre.com/infographic-how-much-does-that-14-t-shirt-really-cost/14-t-shirt-2/?extend=1

What can be "won" in collective bargaining depends on a number of things that are largely beyond the control of both employees and employers. Employers cannot choose to set just any wage. If the employer's wage offer is too low, the firm will have difficulty attracting and keeping employees. Thus, labor market conditions—the level of wages in similar jobs and, particularly, the rate of unemployment—will set boundaries on the employers' wage offer.

Similarly, a union, no matter how strong it may be, cannot demand or expect to receive just any wage. To take an extreme case, if the wage demanded would make the unit labor cost higher than the price of the product, the employer could not make a profit by hiring labor, even if there were no other costs. In this case the profit-maximizing employer would hire no one at all.

A wage bargain will be struck within a range of possible wages limited by conditions in the labor market, the productivity of labor (z), and output prices (P_z). The wage will be toward the lower end of the range if the employer has the upper hand, and it will be toward the higher end if the employees (usually if they have a union) are stronger. Wage bargains are then formalized in contracts that set the hourly rate of pay (w) and some of the conditions of work (such as the time set aside for lunch).



The Bargaining Power of Owners and Workers

It sometimes happens . . . that a single independent workman has stock [capital] *sufficient both to purchase the materials of his work, and to maintain himself till it be completed. He is both master and workman, and enjoys the whole produce of his own labour. . . . Such cases, however, are not very frequent, and in every part of Europe, twenty workmen serve under a master for one that is independent. . . .*

What are the common wages of labour depends everywhere upon the contract made between these two parties, whose interests are by no means the same. The workmen desire to get as much, the masters to give as little as possible. . . .

It is not, however, difficult to foresee which of the two parties must, upon all ordinary occasions, have the advantage in the dispute, and force the other into a compliance with their terms. . . . In all such disputes the masters can hold out much longer. A landlord, a farmer, a master manufacturer, or [a] *merchant, though they did not employ a single workman, could generally live a year or two upon the stocks which they have already acquired. Many workmen could not subsist a week, few could subsist a month, and scarce any a year without employment. In the long run the worker is as necessary to his master as the master is to him; but the necessity is not so immediate.*

—Adam Smith, *The Wealth of Nations* (1776)

The proprietors of . . . establishments and their operatives do not stand upon an equality; their interests are, to a certain extent, conflicting. The former naturally desire to obtain as much labor as possible from their employees, while the latter are often induced by the fear of discharge to conform to regulations which their judgment, fairly exercised, would pronounce to be detrimental. . . . In other words, the proprietors lay down the rules and the laborers are practically constrained to obey them.

—U.S. Supreme Court (1898)

Sources: Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (New York: The Modern Library, Random House, 1937), Book I, Chapter VIII; U.S. Supreme Court, *Holden vs. Hardy*, 169 U.S. 366 (1898).

Labor Discipline: Carrots and Sticks

Just as with the wage, the amount of effort to be put out by employees each hour (e) is subject to a bargaining process. This bargaining process, however, is one that goes on every day and is never settled by a contractual agreement. Why is this?

An employment contract usually specifies three conditions: (1) the wage rate, (2) the hours of work, and (3) the right of the employer to direct the employee's efforts during work hours. The first two conditions can be stated in advance with precision. Wage rates, for instance, are often specified down to the penny for each hour worked. Also, the hours of work can be stated to the minute, stipulating the amount of time allowed for coffee breaks, lunch, and even, in some cases, trips to the bathroom.

The third aspect of the employment contract, however, the one that gives the employer the right to boss the employee around, cannot usually define in advance precisely how much work effort will be forthcoming from the employee. How hard must an employee work? With what precision must the employee obey the boss's commands? Only the broadest limits can be settled when a worker is hired. At one extreme, employees can put all their energy into doing everything their boss asks them to do. At the other extreme, they may do only as much as they need to do to keep from being fired. The actual outcome within this range cannot be specified in advance of the work itself.

What the employer purchases in the labor market is the right to *potential* labor services. Yet a moment's reflection will reveal that this is not at all that an employer needs. What is required is actual human productive activity—real *work*—not just the right to direct employees during the time they are in the workplace. Only human effort itself will produce the commodities that the firm must be able to sell in order to make a profit.

Labor time measures the number of hours worked; it does not measure how much work gets done, since there are many different levels of work effort (intensities of work) possible.

To put it another way, what the employer purchases in the labor market is **labor time**. This is, as we have noted, not the same as the actual *labor* that has to be performed in order for production to occur. To make a profit the employer must somehow ensure that a certain amount of labor is forthcoming during the labor time that has been purchased. Thus, the employer must be able to *extract work* from his or her employees.

The worker may see the situation from a different perspective. The wage specified in the employment contract will be paid. But the amount of work actually performed by the employee is variable. The amount actually performed depends on such factors as how much the employee likes the work, how resentful he or she may be about having to do the work, and how easy it is for the employer to observe how much work the employee is doing. Depending on such factors as these, the employee may choose to do more than, as much as, or less than the amount of work the employer wants. There are, of course, methods available to the employer for ensuring that employees work hard, and we will examine some of these in the next chapter. The point here is that unless employees want to do as much work as employers expect from them, there will be a conflict between employees and employers over the level of work effort (e).

One reason for the conflict is that employees do the work and employers end up owning what is produced. What would happen if workers owned the materials, machines, and other capital goods used in production? Instead of having to work for someone else, the producers would then have the option of working for themselves. They might do this as independent commodity producers—for example,

just as people in the "old middle class" in the United States did, before the capitalist way of organizing production became dominant (and to some extent still do; see Chapter 7).

Alternatively, people wishing to work for themselves could organize a worker owned and controlled enterprise (see box "Own Your Job. Elect Your Boss" in Chapter 5). Whether as independent producers or as worker-owners, self-employed workers would own not only the capital goods used in production but also the product produced. For this reason they would have an interest in working hard, or at least making sure that their workmates did not shirk. Many observers have found that people who work for themselves—and enjoy the fruits of their own labor—work harder than do people who are paid an hourly wage by an employer.

However, as we saw in Table 7.3, the great majority of people who work outside the home in countries such as the United States are not independent producers or owners of their own enterprises. Rather, they work for employers other than themselves in order to earn a wage or salary. As a result, conflicts between employees and employers are widespread in most highly developed economies today.

The **extraction of work from workers** is the process of transforming the labor time that an employer has purchased into work done.

The process by which an employer transforms labor time into productive work is referred to as **extraction of work from workers** rather than *exchange of work from workers* because it occurs under different circumstances from those that prevail in a market. In order to control labor processes and extract work from their employees, employers organize work hierarchically, with CEOs at the top, vice presidents under them, and a chain of command extending all the way down to the actual production workers. The various layers of bosses, foremen, and supervisors constitute the *management* of a firm, and the management structure is what enables employers to extract work from the workers. Because of the central role played by management in the labor extraction process, the relationships between employers and their employees are sometimes referred to as "labor-management relations."

Of course, management does more than just extract work from employees. Managers and executives must organize the hiring of workers, coordinate the firm's operations, make decisions about investment, and arrange for the output to be sold. Even with respect to labor processes, managers do more than organize the extraction of work. In labor processes involving highly specialized tasks and a complex division of labor there is a need for coordination of the employees. In a shoe factory, for example, the leather cutters might need to be told how much and what shapes of leather the stitchers need. On the other hand, many labor processes do not require bosses for coordination. It is possible, to take the same example, that the stitchers in a shoe factory might be able to communicate their leather requirements directly to the leather cutters.

It is interesting to note that the need for a hierarchical power relationship to ensure the transformation of an input into output exists for only one input commodity, labor. When a ton of steel or a new machine is purchased, one can predict exactly how these inputs may be transformed into output. A ton of coal of a given type and quality will yield a certain number of BTUs of heat in a particular furnace. Steel and machines have no will of their own; when the boss commands, they do not resist. There certainly does not have to be a line of unemployed machines standing outside the factory gate to get the machines inside to work hard!

In the old slogan "a fair day's work for a fair day's pay," both parts of the equation are subject to bargaining and conflict. The wage (the "fair day's pay"), of course, is bargained over and contracted for when the worker is hired (or when

the union-negotiated contract comes up for renewal). What constitutes a "fair day's work," however, cannot be written into a contract and is determined only in the subsequent work process. We call this an **incomplete labor contract** to stress that while it covers the wage (which as a result can be enforced by the courts), it does not cover the amount of effort to be provided by the worker (which must therefore be determined by a daily clash of wills between the two parties).

An **incomplete labor contract** is a contract (explicit or implicit) between an employer and a worker that is incomplete in the sense that it specifies the wage rate but does not specify the exact tasks to be performed or the amount of effort to be provided by the worker.

To avoid the management problem, why do employers not offer to pay only for actual work rather than for potential work? Why would they not pay for labor rather than for labor time? One way to do this would be for the employer to specify carefully in the employment contract all the tasks to be performed. The employee's pay could then be withheld until the tasks were completed. This approach, often referred to as a **contract specifying work to be done**, can be used for well-defined tasks such as having one's lawn mowed and the grass raked up; in such a situation, the pay can be withheld until the tasks are completed. For some such tasks, such as computer programming to accomplish a definite task, a firm may well contract with an independent contractor. Another method is the **piece rate** system in which an employee is paid a certain amount of money (say, $0.50) for each piece of work completed (sewing a collar on a shirt, for example). In both methods (and there are others), the employer would be paying for actual work, not potential work, and there would be no need for command-type relations.

A **contract specifying work to be done** is an agreement between an employer and a worker that specifies payment for actual work activities instead of for work time.

A **piece rate** is a form of wage payment in which a worker is paid for each unit of output produced instead of for work time.

Although the detailed contract and the piece rate methods of labor compensation are sometimes used, there are real problems with each of them. For one thing, the contract method is inflexible and costly. Imagine trying to write a contract specifying every activity required to do a job. It might take a lengthy document just to list all the tasks. And even a lengthy contract cannot list all the things an employee must do to contribute to profits. Indeed, an effective form of employee protest is to "work to the book," that is, to follow work rules and regulations exactly, with the result that production is significantly slowed down.

An additional problem with the contract method is deciding whether the stipulated amount of work has been done. It is easy enough to see whether or not a lawn has been mowed. In many situations, however, determining exactly how much work (and of what type and quality) each employee has done can be a costly process. And who is to decide? Both employees and employers would have their own reasons to argue about it. Moreover, every time a new task needed doing or the employee needed to be reassigned, a new contract would have to be written.

The piece rate system is flawed for some of the same reasons. One problem is that it cannot be applied in many labor processes. If an employee is to be paid a piece rate, it must be possible to separate his or her productive contribution from that of each of the other employees; where many employees cooperate in a complex production process (say, in an automobile factory or a Wal-Mart store), it is not possible to separate one employee's contribution from another's, so the piece rate system cannot be applied. Another difficulty is that a given piece rate can be paid only when an employee performs a certain task over and over again. In more complicated jobs there would have to be a different piece rate for every task. Finally, there remains the issue of the pace of work. It may seem that an employer should not care whether a piece rate employee works fast or slow (since the employer is paying for output, not time). In fact, however, this is not the case. To the extent that employees use capital goods such as buildings that their employer must pay fixed amounts for, the speed of the employees' work will affect the profit rate of the

enterprise. (This is because higher rates of output will allow fixed costs per unit of output to be reduced.) In all such situations, employers will have an interest in making sure that their employees work rapidly—and, accordingly, the labor extraction problem will remain.

For these reasons and others, neither the contract method nor the piece rate pay system is widely used today. The contract system (in the form of subcontracting) is sometimes used to hire electricians, plumbers, computer programming, editing (often with payment per page edited) and other highly skilled workers, but otherwise its application is very limited. Piece rates have been more extensively employed, especially in apparel manufacturing and also in harvesting crops. Use of the piece rate system is, however, still quite limited. In general employers have found that other ways of organizing the labor process are more profitable since they make it easier to transform labor time into work.

Workers, however, typically resist efforts to make them work at a faster pace than they consider normal. In response to unreasonable demands, they may develop individual strategies of resistance, join together in informal work groups, or openly resist through their unions. We will explore workplace conflicts in more detail in the next chapter.

The Labor Market, the Wage, and the Intensity of Labor

A **labor market** is a market in which workers sell their labor time (not work itself) in return for a wage; employers are the demanders, and workers are the suppliers, of labor time.

Employers hire workers, and workers find jobs through the **labor market**. In this market employers are the demanders of labor time, workers the suppliers of it. As in all markets, *voluntary exchange* exists; no one is obligated or forced to sign a contract. But in the workplace, employers must exercise some power over employees in order to be able to extract work from them. Their ability to do this is rooted in the distinct character of the labor market. The labor market shares some features with other markets but differs from them in fundamental ways.

The labor market shares with other markets the features of competition and conflict. As we saw in the previous chapter, commodity markets involve a greater or lesser degree of competition. Labor markets are generally very competitive: there are usually many demanders of labor time and, in most cases, many suppliers of it.

Whereas commodity markets are battlegrounds in which corporate giants fight for competitive advantage vis-à-vis consumers or other firms, the labor market is a different kind of battleground. Here the conflict is between employers and workers, and the issues are wages, the intensity of work, workplace safety, and often other things such as health and retirement benefits.

A key weapon in the employer's arsenal is the capacity to provide—or take away—the job that the worker needs. Although employers and workers both need each other, employers can often last longer without workers than workers can get by without employers. As Adam Smith said, "In the long run, the workman may be as necessary to his master as his master is to him, but the necessity is not so immediate" (see box "The Bargaining Power of Owners and Workers"). If a worker happens to be in that small fraction of the labor force that is organized in unions, he or she may be supported for a short time by a union strike fund, but such funds run out fairly quickly. Thus, workers have to depend primarily on their jobs for income (out of which they may pay union dues), but employers often have other means of support (such as property income from other investments).

It is the ownership of capital goods by employers—and the lack of such ownership by workers—that gives the former a degree of control over the livelihoods of the latter. There are two reasons for this. First, one of the prerogatives flowing from their ownership of capital goods is that employers have the right to grant access to or exclude others from entering the property they own. This prerogative is what gives employers the right to hire and fire workers. Second, because workers generally do not own the tools and materials they would need to go into business for themselves, they do not have this option. Hence, they must work for an employer in order to make a living.

Since the owners of capital goods have the right to hire and fire workers, employers have control of the income of employees. The loss of income that may result from the loss of a job is termed the **cost of job loss** (or *cjl*). How much control an employer will have over an employee depends in important measure on the magnitude of the cost of job loss facing an employee. The higher the cost of job loss, the more power an employer will have over an employee. If an employee has access to another source of livelihood, such as an alternative job possibility or a generous unemployment insurance benefit, the cost of job loss will be lower, and the employer's power will be reduced. Employers and employees thus have a direct conflict of interest over the things that determine how high the cost of job loss will be—for example, the availability of jobs and the availability and level of unemployment insurance.

The **cost of job loss** is the loss of income a worker experiences as a result of quitting or being laid off from a job.

Other factors may also influence the conflict between employees and their employers. If employees belong to a labor union, for example, employers may have to deal with strikes or work slowdowns. Of course, employers cannot convincingly threaten to fire employees who have skills that are essential to the firm's operation and that can be taught to other employees only at considerable cost. It is also true that variations in the wage rate and the pace of work will often be limited by widely shared values concerning what is fair to expect employees to do and employers to pay. Such values are termed *work and pay norms*.

Another important influence on the pace of work and the rate of pay is the government. Government policies set minimum wages, establish standards for health and safety conditions in workplaces, and put in place other rules such as those affecting the rights of employees to organize and join unions. There is hardly a market that is not affected in some way by government policies, and the labor market is no exception. (The relationship between government and the economy is considered at length in Chapter 19.)

Cost of Job Loss

In most situations it is the cost of job loss (*cjl*) that is the most important factor in determining the intensity of work (*e*) and the wage rate (*w*)—and the magnitude of this cost can be estimated with at least some degree of certitude. If an employee is laid off, she will face a period of unemployment and, if the necessary requirements can be met, will receive unemployment insurance payments from the government amounting to some fraction of her previous wage. While unemployed she will naturally search for a new job and, presumably, will find one at some point. Suppose for the moment that she eventually gets a job at the same wage as before. Given this sequence of events, we can express the cost of job loss as follows:

(12.8) $$cjl = (ww - ui)ud$$

where cjl = the cost of job loss
ww = the previous weekly wage (after taxes)
ui = the available weekly unemployment insurance benefit
ud = the expected number of weeks (duration) of unemployment

The meaning of Equation 12.8 is that the cost of job loss (*cjl*) will be equal to the amount by which the previous weekly wage (*ww*) exceeds the current weekly unemployment insurance benefit (*ui*) multiplied by the expected number of weeks (duration) of unemployment (*ud*). When the worker finds another job, assuming that he or she is able to find another job, the cost of losing the previous job will actually be known.

Consider, for example, a laid-off worker who had been earning a wage of $400 a week, after taxes. If this worker is able to qualify for unemployment insurance (eligibility requirements vary from state to state but generally one must have worked for the same employer and earned a certain amount in the previous year in order to qualify, and have been laid off rather than fired for cause), he or she will be entitled to receive, say, a $200 unemployment insurance check every week. (The amount of the benefit is always considerably less than the previous wage had been, and in the United States, for example, the government normally pays such benefits for only twenty-six weeks.) If the worker is able to find another job in twenty weeks, the cost of job loss in this case may be calculated as follows:

$$cjl = (\$400 - \$200)(20) = \$4{,}000$$

If the worker had not been laid off, he or she would have had an annual income (after taxes) of $20,000. With the twenty-week spell of unemployment, however, the same worker's annual after-tax income would be $16,000.

Recall that in the preceding example we assumed that the laid-off worker was able to find another job at the same wage level. One problem with this assumption is that in the United States, for instance, manufacturing sector employment has been shrinking as a share of all employment (see lower portion of Figure 7.1). Hence, workers who have lost manufacturing jobs have often had to take lower-paying jobs in the expanding service sector. For these workers—and, indeed, for all workers who cannot find new jobs at comparable or better wages—the cost of job loss is much greater than Equation 12.8 would suggest. The total cost of job loss would have to include not only the income lost during the period of unemployment but also the ongoing difference between the workers' old and new wage levels.

Our simple example underestimates the cost of job loss in other ways, too. The first is that because of stringent eligibility standards and a twenty-six-week maximum duration of unemployment benefits, a majority of the people who are unemployed at any given time in the United States do not actually receive unemployment insurance payments. Thus, for many unemployed people in the United States, the net weekly loss of income ($ww - ui$) will be exactly equal to whatever their previous weekly wage (*ww*) had been. Second, the unemployment benefit is determined based only on the wage, not on the benefits the worker had been receiving. But many workers will have lost more than the wage itself: any health insurance benefits that may

have been provided by their previous employer will also have been lost (or for a temporary period, they may have the option to pay an absurdly high monthly COBRA fee to have their previous coverage extended for a year or more). Finally, there are the noneconomic costs of being unemployed, including inconvenience, anxiety, humiliation, and personal trauma. A person's self-esteem is closely tied to the holding of a job. Thus, even after the trauma of being fired has worn off, an unemployed person may feel that he or she is not retaining the respect of others. (See box "Unemployment Hurts" in Chapter 16.)

Avoiding the Cost of Job Loss

How much is it worth to an employee to avoid the cost of job loss—even if we consider only the economic cost? One way to determine this is to ask how low the current job's wage would have to go before the employee would not care if he or she lost the job. At such a wage level the employee might well decide to quit. Before deciding to quit, however, the employee would have to estimate the cost of job loss, taking into consideration such things as the length of time it might take to find another job, the potential cost of being unemployed for that period of time (bearing in mind that a person who quit their last job cannot collect unemployment benefits), and the possibility that the new job might pay less, have less generous health benefits, or be less satisfying than the current job. Hypothetically, at least, the employee would be willing to take a pay cut of a certain amount at the present job in order to avoid the estimated cost of job loss. For example, an employee currently earning $16 an hour might be willing to continue working until his or her wage had been cut to $10 an hour, at which point keeping the job would no longer be better than being fired or quitting. The lower wage, $10 an hour, is called the employee's **fallback wage**. The fallback wage reflects, from the standpoint of the employee, his or her prospects (including the cost of job loss) in the absence of the current job. Thus, the fallback wage is a measure of the opportunity cost to the employee of keeping his or her present job.

The **fallback wage** is the wage at which an employee has no preference for keeping his or her current job as opposed to being fired or quitting; it varies with the employee's income prospects in the absence of the current job.

If the employer offers only the fallback wage, the employee will not care whether he or she is kept on or fired. At this point the cost of job loss (*cjl*), as seen by the employee, is zero. And since employees being paid the fallback wage will not mind being fired, they can be as lax or as diligent on the job as they please. (When offered any wage lower than the fallback wage the employee will not bother coming to work and nothing will be produced.) If an employer is paying just the fallback wage and attempts to elicit more work effort (*e*) by threatening to fire the employee, it will have no effect. At $10 an hour the employee would be likely to respond to any threat of dismissal by saying to the employer "Make my day!"

If an employer wants to get more than the minimum amount of work out of an employee, he or she will have to pay more than the fallback wage. Only at a higher wage will the employee have a stake in keeping his or her job. The higher the wage, the more the employee will have to lose, and the more likely it will be that he or she will work as hard as asked to by the employer. Responding to both the stick (of possible unemployment) and the carrot (of the higher wage), the employee will work hard. Indeed, it is possible that the employee will work harder and produce more output with each successive increase in the wage.

This presents the employer with the following dilemma: In the situation described, the wage paid must be higher than the fallback wage to motivate the employee to work hard—and getting the employee to work hard is, of course, essential to

making profits. But the payment of a higher wage will increase one of the main costs that must be subtracted from sales revenues in calculating profits. What wage should the employer offer? He or she must balance the (positive) effect of the higher wage on the intensity of work (e) against the (negative) cost of paying a higher wage rate (w).

An employer can resolve the dilemma of what wage rate to pay by focusing on the objective of minimizing unit labor cost (ulc). Minimizing unit labor cost will have the effect of maximizing the profit rate (see the explanation of this relationship in the paragraphs surrounding Equations 12.1 to 12.4 earlier in this chapter). By focusing on this variable, then, the employer can determine exactly the wage that will maximize the firm's rate of profit.

Recall from Equation 12.5 that unit labor cost (ulc) is equal to the wage rate (w) divided by z, the output per hour of work. Recall also from equation 12.6 that output per hour (z) is equal to the intensity of labor (e) multiplied by the efficiency of labor (f). Therefore, as Equation 12.6 tells us, $ulc = w/z = w/ef$.

How, then, can the employer achieve the objective of minimizing unit labor cost? In seeking the answer to this question we can assume that the efficiency of labor (f) will not change when the wage rate changes. This is a reasonable assumption to make because f depends on the quantity and quality of the equipment with which the employee is expected to work, and this will not change in the period of time under consideration here. Hence, the direction of change in the unit labor cost (ulc) will depend entirely on the relationship between the wage rate (w) and the level of work effort (e).

Imagine a situation in which an employer is paying just the fallback wage, represented by $\underline{w}$. In this situation the employer may be able to gain by raising the wage rate. Why? As the wage is increased there may be a more than proportional increase in work effort (e) and hence a fall in ulc and a rise in the profit rate. Moreover, in this type of situation the employer might continue to gain by further increasing the wage rate. This would continue as long as each increase in the wage caused the intensity of labor (e) to go up more than proportionally with the wage rate (w). As long as this is the case, the increases in w will cause the denominator of w/ef to rise faster than the numerator (because e is increasing faster than w), and this will drive down the unit labor cost (ulc). The result, of course, will be a rise in the firm's rate of profit.

There is a catch, however. Because there is a limit to how hard a person can work—and because the harder one works, the closer one comes to this limit—the additional amount of effort put out by a worker in response to increases in his or her wage will eventually tend to become smaller and smaller as this "effort limit" is approached. At some point it will no longer be in the interest of the employer to continue raising the wage. When raising the wage does not bring about an increase in work effort large enough to reduce unit labor cost, the profit-seeking employer will stop raising the wage: at this point raising w will not increase the profit rate. Exactly at this point the employer will be paying the wage that maximizes the firm's rate of profit. (Some economists call this the "efficiency wage.")

To illustrate the analysis presented here we offer the following (hypothetical) case study. Imagine that a firm has, through years of observation, found that its employees produce more output per hour as their wage rate goes up. Output rises because employees work harder at higher wage rates, responding to both the "carrot" of being paid more for each hour of work and the "stick" of having, at higher wage rates, more to lose (the potential cost of job loss). The observed relationship between the wage rate and the workers' output per hour in our hypothetical firm is presented in Table 12.1.

TABLE 12.1. THE WAGE-OUTPUT RELATIONSHIP

WAGE RATE	OUTPUT PER HOUR
$10	20 units
$12	34 units
$14	46 units
$16	56 units
$18	64 units
$20	70 units
$22	74 units
$24	76 units

It is assumed here that nothing will be produced unless the wage is at least $10 an hour. This is the fallback wage. But as the wage rate goes up, employees will put more effort into what they are doing, and output per hour (z) will rise. At first, increasing the wage brings substantial *increases* in output per hour. But as the wage is raised further, the gains in output per hour that result from each additional dollar of wage paid shrink. Eventually, raising the wage further is not worth it for the employer, because doing so beyond a certain point reduces the profit rate.

The **labor extraction curve** describes for each wage rate the intensity of work that the worker chooses to perform.

Figure 12.1 illustrates a typical pattern of how output per hour (z) changes as the wage (w) increases. The curve is termed a **labor extraction curve**. Output per hour (z) is shown on the vertical axis, and is a function of (that is, a response to) the wage rate (w) on the horizontal axis. The most important aspect of a curve like this is the *slope* at each point along it. The slope is the change in z divided by the change in w along a line tangent to the curve. This can be thought of as the *marginal increase* in *output per dollar of wage paid.*

Suppose our hypothetical firm initially pays just the fallback wage of $10 an hour (labeled $\underline{w}$ on the horizontal axis of Figure 12.1). If the firm raises the wage to $12 an hour, output per hour will rise by 14 units, and this will be a (marginal) increase in output per dollar of wage paid (z/w) of 7 units per dollar (= 14/($12 − $10)). But is there a similar gain if the firm happens already to be paying $22 an hour and raises the wage to $24? No: in that case output per hour will rise by only 2 units, a marginal increase in output per dollar of wage paid of only 1 unit of output per dollar. As mentioned earlier, wage increases gradually become less effective in calling forth additional output per hour.

In the process described here there are two separate decisions being made. The first, made by the employer, is the choice of the wage (w) to offer. The second, made by the worker, is the choice of a particular intensity of labor (e), or, in other words, how hard to work if paid the wage the employer is offering. Employers reason that paying higher wages will enable them to extract more work from their workers, and this knowledge can be represented by a labor extraction curve such as the one in Figure 12.1.

The declining power of wage increases to bring about additional output from workers (declining marginal increases in output per dollar of wage paid) is also reflected in the labor extraction curve. As Figure 12.1 shows, the labor extraction curve rises steeply at first and then becomes flatter with each increase in the wage rate. As noted earlier, the declining slope of this curve reflects the fact that workers can increase their output more easily when they are not working too hard but less easily when they

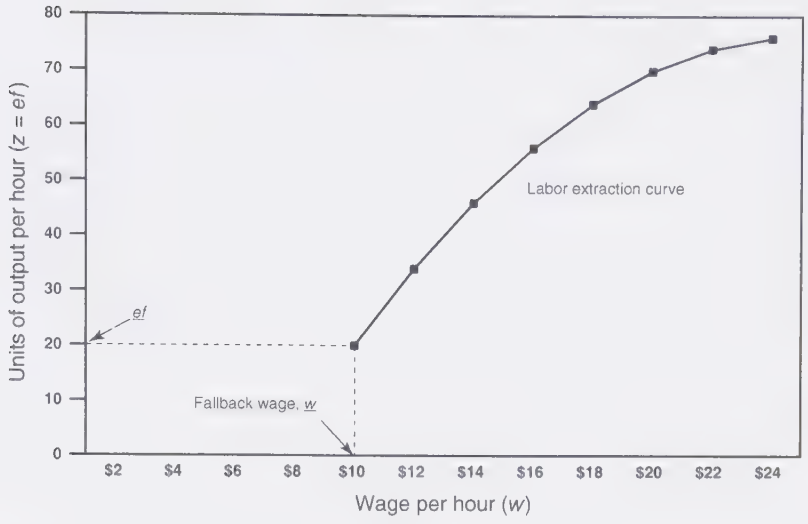


FIGURE 12.1 The wage and the intensity of labor.

The labor extraction curve in this figure shows the amount of output per hour (z) that a worker will produce at each wage rate (w) offered. Output per hour is measured on the vertical axis. Since $z = ef$, the vertical axis also measures ef. And since f (the efficiency of labor) is constant (no changes in the technology of production are possible in the short time frame of this analysis), any movements up or down the vertical axis are due entirely to changes in the intensity of labor (e).

The horizontal axis shows the hourly wage rate (w). *The fallback wage* is $\underline{w}$ (here, $10) and, at this wage rate, the worker's work effort (e) is at its minimum. This is how hard the worker will choose to work if the wage rate sinks to a level ($\underline{w}$) at which he or she is indifferent between keeping the job and losing it. When the wage is at its fallback rate and work effort is at its minimum, the corresponding rate of output per hour (z), shown on the diagram as $\underline{ef}$, will also be at a minimum (here, 20 units per hour). Thus, if the employer offers a wage equal to w, the worker's work effort will be $\underline{e}$ and the rate of output will be $\underline{ef}$

Increases in the wage rate above the fallback wage call forth more output per hour: the levels of output per hour represented by points on the labor extraction curve rise as the wage rises. But beyond a certain point (the point at which the labor extraction curve becomes horizontal), wage increases do not result in additional output per hour because the worker is already putting out as much effort as possible.

are already working hard. So the shape of the labor extraction curve will be determined by the amount of work effort (e) that is forthcoming at each wage rate (w), assuming that f (the output that can be produced with any given amount of work effort) is constant.

If a profit-maximizing firm has the labor extraction data given in Table 12.1 and plotted in Figure 12.1, what wage rate will it offer to pay? We can begin to answer this question by referring to Equation 12.5, which says that unit labor cost (ulc) is equal to w/z. But in addition we need to keep in mind the fact that the rate of profit will be maximized when unit labor cost is minimized. Putting these two ideas together, we can see that *maximizing the rate of profit* requires minimizing ulc (which is w/z). Since as w/z rises, z/w falls, maximizing profit requires *maximizing* z/w (output per hour divided by the hourly wage), or ef/w (because $z = ef$). In other words, the firm will be maximizing its rate of profit when the ratio z/w (or ef/w) is as large as possible. This makes sense intuitively because for any given level of f,

the firm's profit will be maximized when the employee's effort per dollar of wage (e/w) is at a maximum. When this is the case the firm will be getting as much output per hour (z) as it can get for a dollar of wage.

The logic of this analysis can be illustrated by adding something to the diagram in Figure 12.1. Figure 12.2 shows what we want to add. Figure 12.3 will then combine the two in one diagram. Notice that the axes in Figure 12.2 are identical to those in Figure 12.1, with the same two variables, output per hour and the wage rate, on its axes (z or ef on the vertical axis and w on the horizontal axis); the scales are also the same on both axes. Each of the two upward-sloping dashed lines (the *rays* through the origin) has a constant slope, like any straight line. That slope is z/w (or ef/w) for any point (w, z) on the ray. The economic meaning of this is that the slope (z/w) of the ray shows the average output per dollar of wage paid. It shows how much output per hour (z) the firm can get for a dollar of wage paid.

In Figure 12.2 a *steep* ray (such as line 1) indicates that for each dollar of wage paid there will be a *large* amount of effort exerted (e) and output produced per hour (z). Thus, a profit-maximizing employer will prefer to be on the steepest ray possible. But what defines "possible"? The answer is that only points on the labor extraction curve are possible, because that curve shows how much effort a worker will actually exert when faced when paid a particular wage (w). Points not on that curve represent points that will not happen. The employer must take account of the impact on worker effort when choosing the wage. So Figure 12.2 tells us that the employer prefers points on line 1, where a wage of $10 an hour elicits 15 units of output per hour, to points on line 2, where the same wage elicits only 5 units of output per hour. But the employer can only choose from among points that are *on the labor extraction curve.*

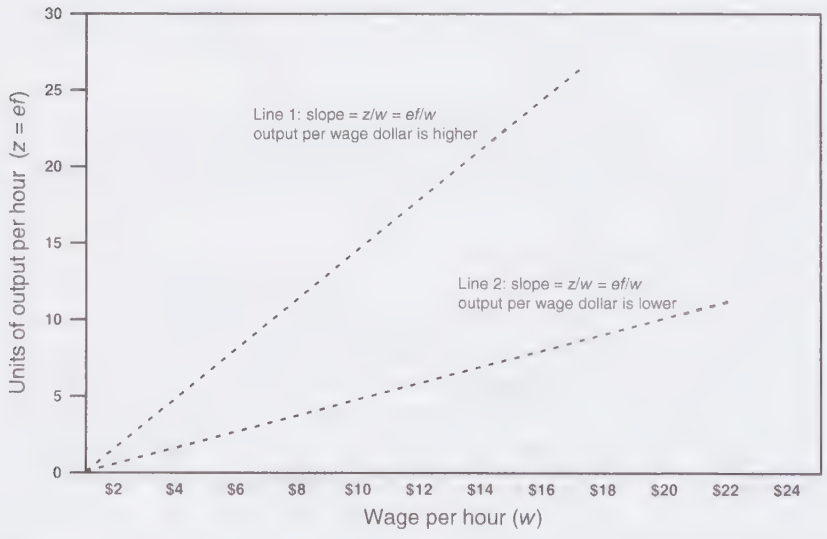


FIGURE 12.2 Output per dollar of wage paid.

This figure shows rays drawn from the origin (at 0 on the vertical axis and $0 on the horizontal one) with different slopes. A ray has the same slope at all points along it, and in this case the slope of the ray indicates how much output per hour can be achieved for each dollar of wage paid. Mathematically, the slope of a ray in this figure is z/w (output per hour divided by the wage rate). The greater the slope of the ray, the higher will be the hourly output per wage dollar paid.

We could also look at this from another angle. An employer who is maximizing *e/w* is also minimizing *w/e*, and therefore minimizing *w/ef*, the unit labor cost (remember, we are assuming that *f* is given and does not change). Recall from Equation 12.5 that the unit labor cost (*ulc*) is the wage cost per unit of output (*w/z*), and this is the *inverse* of the slope (*z/w*) of any ray, such as in Figure 12.2. This means that, the steeper the ray, the *lower* is the unit labor cost. This makes sense because when the employer seeks to *maximize profit*, it does so by *minimizing unit labor cost*.

In Figure 12.3 we see that a profit-maximizing employer will choose the point where the steepest possible ray touches (is tangent to) the labor extraction curve at a wage that we call *w**. The slope of this ray will be greater than that of any other ray drawn to a point on the curve.

Consider point A on the labor extraction curve in Figure 12.3, where 34 units of output are produced at the wage of $12 an hour. What is the output per wage dollar at this point? We can answer this question by drawing a straight line (line 1) from the origin to the 34 unit/$12 wage point on the labor extraction curve and examining the *slope* of the line from the origin. The slope is approximately 2.8 (34 divided by 12), meaning that 2.8 units of output are being produced for each dollar of wage paid.

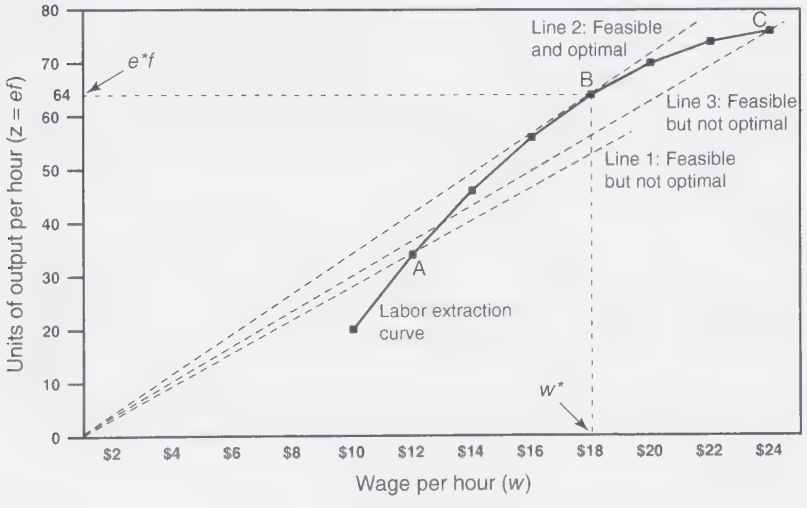


FIGURE 12.3 The labor extraction curve and profit maximization.

The labor extraction curve tells us how much output there is at each wage level (above the fallback wage), and rays show how much output per hour a dollar of wage buys. A profit-maximizing firm wants to be on a steep ray because, the steeper the ray, the more output there is per wage dollar paid. Unit labor cost is lower along a steeper ray since *ulc* is defined as *w/z*, the *inverse* of the slope of a ray (*z/w*). But no firm can be on a ray that lies *above* the labor extraction curve. This is because the labor extraction curve indicates how much effort a worker will exert per hour at each wage above the fallback wage. The profit-maximizing firm therefore wants the ray with the steepest slope (highest output per hour for a dollar of wage paid) that touches the labor extraction curve. The ray that is optimal for the firm is the one tangent to the labor extraction curve at point B, and the hourly wage corresponding to point B is *w** (here, $18). At a wage of *w**, the output per hour will be *e*f* (here, 64). Thus, *w** is the wage rate at which unit labor cost will be minimized and profit maximized, with *e*f* as the corresponding rate of output per hour. All other rays that pass through some point on the labor extraction curve will have flatter slopes and higher unit labor costs.

Hence, the particular combination of *w** and *e*f* represented by point B offers the employer the lowest possible unit labor cost and the maximum rate of profit.

What happens if the employer decides to increase the wage rate to $14 an hour? Output rises to 46 units (see Table 12.1), and the slope of the line drawn to the corresponding point on the labor extraction curve rises to about 3.3, meaning that 3.3 units of output are produced for each dollar of wage paid. The employer has clearly gained by paying his or her employees more, but the rate of profit has not yet been maximized. If the employer continues to raise the wage rate, then at an hourly wage rate of $18, the output per hour will be 64 units. The output per dollar of wage paid will then be 3.6, higher than the two previous points. The employer has increased profit by raising the wage, because each time the gain in output has more than compensated for the increased cost of labor. However, the maximum profit has been reached, and there is no gain to the employer from further raising the wage. Here, the ray from the origin to the labor extraction curve (line 2) is as steep as it can possibly be, and is tangent to the labor extraction curve. We call the wage rate here w^* because it will maximize profit, and therefore it is the wage which a profit-maximizing employer will prefer.

What will happen, though, if the employer does not know the shape of the curve, and so continues to raise the wage? Point C on the labor extraction curve represents the situation in which the wage rate is $24 an hour and the rate of output is 76 units per hour. The ray through this point (line 3 in Figure 12.3) is less steep than the line to point B (line 2). Accordingly, the profit is lower at point C than at w = $18 at point B. The profit-maximizing employer therefore does not want to pay a wage that is either too low or too high, so the employer will offer a wage of $18 an hour.

Notice an important fact about w^*: it is higher than $\underline{w}$ (the fallback wage). In other words, at a wage rate of w^* an employee would rather keep his or her job than lose it. This is fairly obvious, but there is another implication of this fact: there may be other workers outside the factory gate, on the other side of the office door, or even within the same firm, who would also like to have—but cannot get—the job that is paying w^*. Such workers might be among the unemployed, or they might be people who are currently employed in less desirable jobs who would be happy to take the job paying w^* (that would open up if the employee currently holding it were to quit or be fired).

Table 12.2 presents the same information as Table 12.1, but adds two more columns. The third column calculates the output per dollar of wage corresponding to each wage rate (w). In this column, the maximum value occurs when the employer

TABLE 12.2. THE WAGE RATE, OUTPUT PER DOLLAR OF WAGE PAID, AND UNIT LABOR COST

WAGE RATE	OUTPUT PER HOUR	OUTPUT PER DOLLAR OF WAGE PAID	UNIT LABOR COST
$10	20 units	2.0 units/$	0.500 $/unit
$12	34 units	2.8 units/$	0.353 $/unit
$14	46 units	3.3 units/$	0.304 $/unit
$16	56 units	3.5 units/$	0.286 $/unit
$18	**64 units**	**3.6 units/$**	**0.281 $/unit**
$20	70 units	3.5 units/$	0.286 $/unit
$22	74 units	3.4 units/$	0.297 $/unit
$24	76 units	3.2 units/$	0.316 $/unit

pays $18 an hour and the output per dollar of wage paid is therefore 3.6 units. The table also shows that output per dollar of wage rises as the wage rises, until this point is reached, and declines as the wage is further increased.

Notice also that the unit labor cost shown in the right hand column of Table 12.2 (corresponding to each wage level in the first column) falls and then rises as the wage rate is increased from the $10 fallback wage to $24 an hour. The minimum unit labor cost (28.1 cents per unit) is achieved, naturally, when the wage rate is $18 an hour. This is not surprising because, as we already learned, unit labor cost is minimized when the profit rate is maximized.

The data in Table 12.2 are plotted in Figure 12.4—illustrating that the unit labor cost (the lower curve) is at a minimum at that wage at which the output-per-dollar-of-wage ray is tangent to the labor extraction curve (the upper curve).

Additional Implications of the Labor Extraction Curve

This chapter has focused on the relationship between the wage rate (w) and the intensity of labor (e), assuming that *the wage rate is the only influence on the intensity of labor.* This simplifying assumption will be abandoned in the next chapter. But first there are three additional implications of the labor extraction curve analysis that should be considered.

First, if an employer can find ways to increase the supervision of his or her employees—say, by watching them more closely or by hiring others to do so—unit labor costs may be reduced. Of course, this can occur only if the costs of increased supervision are more than offset by gains resulting from greater effort on the part of

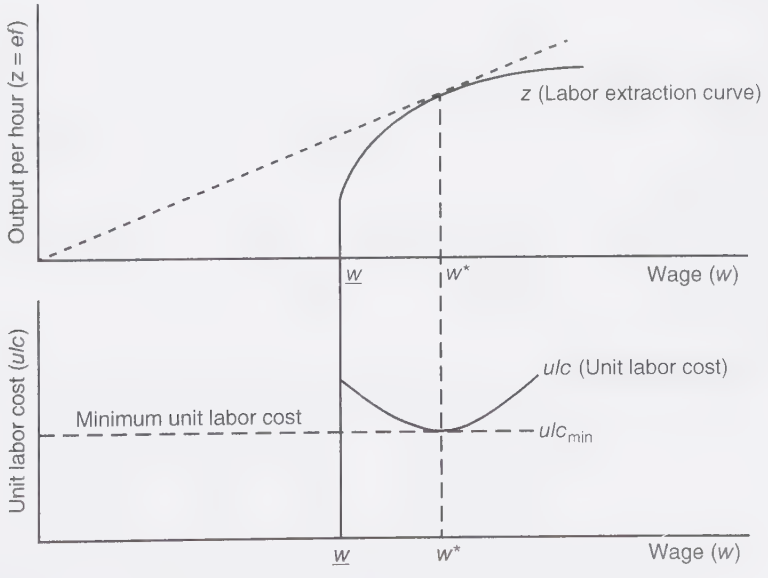


FIGURE 12.4 The wage, labor extraction, unit labor cost, and profit maximization.

This figure illustrates in a different way the idea that the point of tangency between the labor extraction curve and a ray with the slope z/w is indeed where unit labor cost is minimized. The figure makes the same point graphically as was made with numbers in Table 12.2. Both Table 12.2 and this figure show that *ulc* is minimized when the wage rate is w^* ($18 in the table and the example set forth in the text).

the employees. In recent years it has become common for employers to use surveillance cameras or other technology to monitor the performance of their employees on the job (see box "On the Road Again" in Chapter 13). An increase in supervision may bring about an *upward shift* of the labor extraction curve. Such a shift would show that employees, being supervised more closely, put out more work effort (and output) at each level of the wage rate. As noted, however, there are many labor processes, especially those requiring intellectual skills or entailing significant quality considerations, for which increased supervision may not raise e and may even lower it.

A second implication of the labor extraction curve analysis can be seen in Figure 12.5. Any increase in the worker's fallback wage will shift the entire labor extraction curve *to the right*, meaning that for any given wage, work effort and therefore output will be lower. This might happen, for example, if it became easier to find a new job at a higher wage or if there were to be an increase in the level or availability of unemployment insurance benefits.

Figure 12.5 shows what happens when there is an increase in unemployment insurance of $2 per hour. An increase in unemployment insurance by this amount shifts the entire labor extraction curve to the right by $2 because the employer has to pay an hourly wage $2 higher than before to extract any given amount of work. The fallback wage rises from $10 an hour ("old $\underline{w}$" in the figure) to $12 an hour ("new $\underline{w}$") because workers are now willing to be fired (or quit) at $12 an hour, the lost $2 being made up for by the extra $2 of unemployment insurance. And now the most

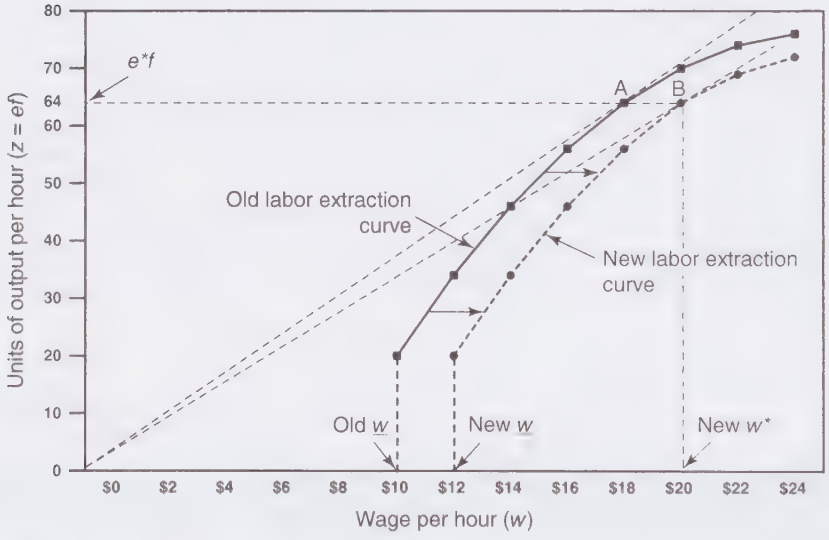


FIGURE 12.5 Effects of more unemployment insurance.

With an increase in the level or availability of unemployment insurance benefits, the worker's fallback wage rises (in this case by $2 per hour) and the entire labor extraction curve *shifts to the right*. This is because the cost of job loss at every wage rate is reduced by $2 per hour. Employers must now offer higher wages to extract the amount of work that they had previously been able to extract at lower wage rates. In this case, to get the same rate of output (e^*f) as before, the firm must pay a wage of close to $20 an hour in place of the $18 an hour that it used to pay. The slope of the line drawn to point B is lower than that of the line drawn to point A, which means that the unit labor cost at the new profit-maximizing wage offer (new w^*) will be higher than before.

profitable wage from the employer's standpoint is higher (around $20 rather than $18), as indicated by "new w^*" in the figure. Moreover, the new profit-maximizing wage involves a higher unit labor cost, since the ray to point B has a lower slope than the ray to point A. (In all of the figures in this chapter a ray with a lower slope represents a higher unit labor cost.) This example shows why U.S. employers, working through their lobbyists in Washington, usually oppose any increase in unemployment insurance.

The third implication of the analysis has already been briefly mentioned: wages can be too *low* even from the standpoint of the employer. Had the employer offered a wage less than w^*, unit labor costs would have been *higher*, not lower, and this would have meant reduced profits. This is because the savings in wages would have been more than offset by the lower level of work effort (e). For this reason unemployed workers cannot succeed in getting jobs by going to employers and offering to work for less than w^*. Such offers will be refused—and the unemployed workers turned away—because employers will see that the lower wage level will mean higher, not lower, unit labor costs. This conclusion is important because it supports the view of economist John Maynard Keynes, the father of macroeconomic theory (to be explored in Chapter 16), that in a capitalist economy there will generally be a certain amount of *involuntary unemployment*—that is, unemployment that cannot be reduced even if the unemployed workers are willing to lower their wage expectations.

The fact that unemployed workers cannot get the same kind of job by offering to work for less than the going wage rate means that they will remain unemployed. It also means that the wage will not fall even if there is excess supply in the labor market: the unemployed workers will find no takers for their labor. Indeed, neither the worker nor the employer, acting alone, can alter this state of affairs.

Once the wage rate (w^*) and the corresponding level of effort (e^*) have been reached, they will exist in what many economists call a state of equilibrium. Recall that in Chapter 8 an equilibrium was defined as a situation in which there are no forces tending to change the situation other than accidental ones or ones coming from outside the economic system ("exogenous" forces). In Chapter 8 the equilibrium involved the price and quantity of beer; here it is the relationship between a level of work intensity and a wage rate, but the concept is the same in both cases.

On the one hand, the employer controls the wage rate (w), but once w^* has been found, he or she will have no interest in changing it. This, of course, is because w^* is the wage that maximizes the employer's profit rate. The worker, on the other hand, is the one who chooses how hard to work, thereby determining e^*. But here again, once the work effort e^* is being put out, there will be no incentive for the worker to change his or her work pace, for e^* is the level of work effort associated with the wage rate w^*, and this is the equilibrium wage rate. Most likely, the worker would prefer to work less hard and get paid more, but this option is not available. Similarly, the employer would be happy to get more work for less pay, but this is not an option, either. Thus, given a particular labor extraction curve, neither the worker nor the employer will have an incentive to move away from the w^*/e^* equilibrium point. And not even an unemployed person offering to work for a wage lower than w^* can change the situation: as we have seen, such a person will not be offered a job. These, then, are the factors that make the particular situation (in which the wage rate is w^* and the work effort level is e^*) an equilibrium.

There is, however, an important difference between the beer market of Chapter 8 and the labor market analyzed here. In equilibrium the beer market clears: there is

neither excess supply nor excess demand. But in the labor market the supply of labor is greater than the demand for it—there is a positive rate of unemployment—even when the market is in equilibrium (there are no forces in the market pushing it to change). Thus, we have equilibrium *without* market clearing.

The two markets are also different, obviously, because people are not beer. But the difference in this regard is more complex than the simple statement of it suggests. When a person pays a dollar for a bottle of beer, he or she gets what is paid for and has the right to use it in any way that is legal. The buyer can send the beer down his or her throat, pour it down a drain, or splash it on someone else's face.

When an employer pays $10 or $15 to purchase an hour of someone else's labor, however, he or she does not end up with something that is totally under his or her control. The reason is that the employment contract by itself will not get the worker to work hard. It is, as we have said, incomplete. Rather, what the employer possesses is authority over the labor time of another human being, and that other human being may or may not want to work as hard or as carefully as the employer desires. An employer pays a wage that is above what an unemployed person would be willing to work for because the higher wage is a way of gaining enough control over the employee to ensure that the expected amount and quality of work will be forthcoming. Nobody pays more for a beer than he or she needs to. The beer is not going to resist being consumed; the labor, however, must be extracted.

While the labor extraction curve illustrates some important aspects of the conflict over work and pay, no simple model can capture the complexity of relationships between employers and employees. To understand the work process more fully, we must look at the firm as a social organization, one made up of flesh-and-blood people whose behavior cannot easily be summarized in a graph. This is the subject of the next chapter.

Suggested Readings

George A. Akerlof and Robert J. Shiller, *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism* (Princeton: Princeton University Press, 2010), esp. Chapter 8: "Why Are There People Who Cannot Find a Job?"

Harry Braverman, *Labor and Monopoly Capital* (New York: Monthly Review Press, 1974).

Jeremy Brecher, *Strike* (San Francisco: Straight Arrow Books, 1972).

Michael Burawoy, *Manufacturing Consent* (Chicago: University of Chicago Press, 1981).

Richard Edwards, *Contested Terrain: The Transformation of the Workplace in the 20th Century* (New York: Basic Books, 1979).

Louis Putterman, *The Economic Nature of the Firm: A Reader*, 2nd ed. (Cambridge: Cambridge University Press, 1996).

Technology, Control, and Conflict in the Workplace

13

In 2015 a U.S. coal miner working underground in Appalachia dug around three tons of coal in an average work hour. For this work the miner was paid $30.59 per hour.

Coal miners in the United States often work more than forty hours a week, working additional hours "overtime" to increase their pay and put money in the bank to live on during periods of unemployment. If a miner worked forty hours a week in a particular year, however, his or her annual income before taxes would be about $63,600. After taxes (which would vary with family size and income earned by other members of the family), he or she might be left with $50,000 to $60,000 to live on. And as everyone knows, a coal miner's job is dirty, demanding, and dangerous. Besides the danger of mining accidents, those with decades in the mines face the danger of debilitating lung disease. Of course, coal miners would like to be paid higher wages, and would certainly like their jobs to be safer.

The coal companies, however, have a different viewpoint. What they notice is that making a miner's job safer would require additional investment in safety equipment and, most likely, a slower pace of work. With existing machinery, making coal mining safer would mean slowing the pace of work to allow additional safety procedures to be instituted, and this would result in less output, which in turn would mean less revenue coming in and lower profit. Paying higher wages would also reduce profits by adding directly to costs.

If miners were paid $30 to dig three tons of coal each hour, the unit labor cost of a ton of coal would be $30/3, or $10 per ton. The market price of a ton of coal in 2015 varied widely depending on the type, but was roughly $38. At this price and labor cost per ton, the coal companies, on average, were left with more than $28 per ton for replacement of worn out machinery (depreciation), costs of supervision, new investment, and profit.

As we have seen, however, making the work of a coal miner safer requires either additional investment or a reduction in the pace of work. Either way, if other variables remain unchanged, making the work of a coal miner safer or more remunerative will raise unit labor costs or unit materials costs and reduce profit rates. For this reason the interests of the miners and those of the coal companies are often directly opposed.

The fundamental clash of interests between miners and companies has turned the nation's mining regions into centers of intense, bitter, and continuing conflict. Indeed, the conflict has often turned violent. In the 1870s, for instance, the state of Pennsylvania hanged nineteen "Molly Maguires"—militant miners seeking to form a then-illegal union. In 1914 the Colorado militia broke up a strike at a Rockefeller-owned coal mine: militiamen killed twenty-one people, including eleven children and two women, in what came to be called the Ludlow Massacre. In Kentucky, "Bloody" Harlan County has witnessed decades-long violence between coal companies and miners. Even today union organizers face frequent harassment, intimidation, and threats of violence.

Violence is the most extreme, but not the most common, expression of the continuing conflict between employers and their employees. Short of violence, how are the conflicting interests of capitalists and workers resolved? How do employers go about extracting work from employees? What do workers do to defend their interests? This chapter investigates these questions.

The main idea of this chapter is that *every capitalist labor process necessarily combines both a social organization and a technology of production; these two elements shape, and are shaped by, the conflict between employers and employees over wages and the intensity and conditions of work.*

This idea is expressed in six key points:

1. Employers organize workplaces in ways they believe will maximize the profits of their firms. The *social organization of the labor process* derives from the employers' search for profits and, in particular, from their effort to extract work from their employees. Employers establish *systems of control* in the workplace to enhance their ability to extract work.

2. Known and available *technologies of production* impose certain constraints on how employers can organize production. Decisions about which of the available technologies are actually used and what kinds of new technologies are developed are shaped by the conflict between workers and employers.

3. Employers and employees seek to advance their *conflicting interests over wages and over the intensity and conditions of work* within the context of the social organization and technology of the labor process. Workers defend their interests in a variety of ways—for example, by attempting to organize themselves in unions—and employers likewise develop strategies for maintaining or improving their profit rates—occasionally, for instance, by exploiting racial and gender-based tensions among workers.

4. Employers seek to organize labor processes in the most *profitable* way, but this need not be the most *efficient* way. Efficiency is different from profitability, and when there is a conflict between the two, the ongoing competition among capitalists pushes them to seek out profitability, not efficiency.

5. *Markets and hierarchies* are mutually reinforcing elements in the capitalist profit-making process. Although sometimes thought of as alternate and opposing methods for organizing social relations, in fact they are both necessary for the capitalist firm.

6. An alternative to the capitalist firm is the *democratic firm*, owned by its employees and managed by those elected by the employees. Despite many advantages,

democratic firms often find it difficult to survive because of the risks involved and the difficulty of raising the necessary capital.

As in previous chapters, the process of profit making is central to understanding the topics discussed here, for it is profit making that gives rise to the conflict between employers and employees. Sometimes this conflict appears to be a conflict between a single employee and his or her boss; sometimes the conflict is evident in the bargaining or struggle between a firm's owner and groups of employees or all the employees of that firm; and in other cases the conflict becomes a more general one between important groups in both the working class and the capitalist class.

The Social Organization of the Workplace

Employers try to organize workplaces in ways that will produce the highest profit rates. Their goal (profits) is the same as in their other activities, such as the ones discussed in Chapters 10 and 11. How employers organize the workplace and their immediate relations with their employees affects their profit rates most directly with respect to the intensity of work (e), the efficiency of work (f), and the wage rate (w). As we saw in previous chapters, an employer that can raise e or f and/or reduce w, other things remaining unchanged, will achieve a higher profit rate.

The most fundamental way that employers and their supervisors, foremen, and managers exercise power in the workplace, as we have seen in the previous chapter, is through their right to hire and fire. The replacement worker standing outside the plant gate keeps the employee inside the factory working hard.

If most of the employees in a workplace do not perform up to the level demanded by the employer—for example, suppose they belong to a union and they resist speedup in their work—then the employer has a bigger problem. Still, the employer has several options. For example, an employer can take measures that will provoke a strike and then lock the employees out. In such a situation the employer can close down the plant, lock the gate, wait for the employees to feel the financial pinch, and hope that they will eventually agree to work on the employer's terms. This is referred to as a **lockout**.

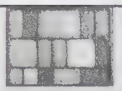
A **lockout** occurs when an employer locks the workers out of the workplace and closes down production in order to force workers to accept the employer's terms for wages, work pace, or other working conditions.

Alternatively, a business owner can move his or her plant to a place where employees will be more compliant. For example, the legal and political climate has historically been much more hostile to unions in the U.S. southern states, New Hampshire, and the Rocky Mountain states than it has been in the industrial Midwest and most of the Northeast. Similarly, a company facing a strong union can move—or threaten to move—to a foreign country where wages are lower and unions are weak or even banned. The point is that capitalists can *run away* to a more business-friendly location than the one they happen to be in currently. Because they own the capital goods they are free to invest (build new plants) or disinvest (let existing plants run down) as they please and where they please. The **runaway shop** leaves its former employees without jobs, and even the threat of relocating may scare employees into accepting the employer's terms.

A **runaway shop** is a workplace that an employer has moved from an area where workers are strong to an area where workers are weak in order to escape having to meet workers' demands.

A third strategy big firms can use in organizing the workplace is **parallel plants**, plants located in different parts of the country or the world making the same product. With parallel plants the employer can tell employees in one plant that unless they agree to low wages and/or high work effort, the company will switch production to the other plant. Then it tells the same thing to the employees at the other plant.

Parallel plants are plants owned by the same employer and producing the same product but located in different geographical regions in order to weaken the workers' ability to bargain collectively.



Making Steel: Three Views

Worker

Somebody built the pyramids . . . pyramids, Empire State building—these things just don't happen. There's hard work behind it. I would like to see a building, say the Empire State, I would like to see on one side a foot wide strip from top to bottom with the name of every bricklayer, the name of every electrician, with all the names. So when a guy walked by, he could take his son and say, See, that's me over there on the forty-fifth floor. I put the steel beam in. Picasso can point to a painting. What can I point to? . . . Everybody should have something to point to.

We handle forty to fifty thousand pounds of steel a day. . . . You can't take pride anymore. It's hard to take pride in a bridge you're never gonna cross, in a door you're never gonna open. You're mass producing things and you never see the end result of it.

—Mike Lefevre, steelworker, quoted in Studs Terkel, *Working* (New York: Pantheon Books, 1972).

Personnel Management Consultant

I can say, without the slightest hesitation, that the science of handling pig iron is so great that the man who is fit to handle pig iron as his daily work cannot possibly understand that science. . . . The man who is fit to work at any particular trade is unable to understand the science of that trade without the help and cooperation of men of a totally different kind of education.

—Frederick Winslow Taylor, founder of "scientific management," 1912 testimony before a congressional committee as reported in F. W. Taylor, *Scientific Management* (New York: Harper, 1947), 49.

Capitalist

The duty of management is to make money. Our primary objective is not to make steel.

—David Roderick, former chairman of the board, U.S. Steel (now USX), as interviewed in the film *The Business of America* (1984).

Unless the employees in these widely separated plants can somehow act jointly (a very difficult task), they will be left in a weak bargaining position.

Employers' methods of maintaining the upper hand all derive from an owner's right to say who works and who does not work in his or her workplace. Private property, as we have seen, includes the right to exclude others from its use, and private ownership of the capital goods used in production brings with it the right to hire and fire. Of course, lockouts, runaway shops, and parallel plants impose costs on business owners as well, so they may choose not to use these means of driving workers. Then, too, each employer must offer job conditions that allow him or her to fill the number of jobs needing to be filled. If other employers offer workers a slower pace of work or better pay, this may limit what the employer can do unless there is a large pool of unemployed workers. Finally, some workplaces, especially those providing services, may not be movable (a Boston hotel owner with striking employees cannot move the hotel to Hong Kong and still service the Boston customers). These considerations constrain an employer's actions.

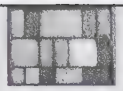
The **social organization of the workplace** refers to the way in which jobs are defined, work tasks are assigned, supervisory power is delegated, and other social aspects of the workplace are organized.

Employers' powers are further enhanced by the **social organization of the workplace**—the way in which jobs are organized. Generally, a firm's organization is hierarchical. **Hierarchy** in the workplace builds on the right to hire and fire, but it also allows for more finely tuned control. Employers cannot effectively threaten to fire workers every day and every hour for trivial or routine problems as well as important matters. Such threats, if used too often, no longer remain believable, and even if, to make the threats believable, employers do frequently fire employees, the

A **hierarchy** is an organization of power in which superiors have command over subordinates.

high turnover on the job may well disrupt a firm's operations. Successful employers have developed a series of other incentives—rewards and penalties—to supplement and reinforce their ultimate power of dismissal.

Employers organize the workplace by imposing a social organization on the labor process. They define the different jobs and responsibilities, the work rules and workers' rights (besides those defined by law), the powers of supervisors and foremen, and the rewards for good work and the penalties for the lack of it. If a union represents the workers, it negotiates with the employer about these matters.



Hardball: Owners, Workers, and Taxpayers in the Global Economy

Before deciding to expand its operations in Dublin, Virginia, Volvo Trucks North America asked the state of Virginia to share some of the costs. Since Volvo has plants worldwide, Virginia needed Volvo more than Volvo needed Virginia. So the state offered a package of tax breaks and other incentives that would cost taxpayers $54.2 million, in exchange for Volvo's promise to hire 1,277 more workers over six years (a cost of over $42,000 per job created). But Volvo proposed that the new workers get 30 percent less than the former starting wage, and wait ten months before getting health insurance. On January 13, 1999, United Auto Workers (UAW) members rejected the proposed contract.

Volvo gave an ultimatum that it would hit the road unless workers approved the contract in a new vote, and its manager visited two plants in Mexico to explore alternative locations. Dublin workers well knew that manufacturing workers in Mexico were earning less than one-fifth of the $12 per hour average production wage at the Dublin Volvo plant. The workers were given two days to vote the ultimatum up or down. On January 29 they swallowed their pride and agreed to Volvo's terms.

The same story, with a different ending, unfolded for the 548 workers in a plant in Dover, Delaware, owned by Dresser Industries, a division of Halliburton, the global oil field services corporation. The workers made gas station pumps, and most were represented by the UAW. They went on strike in July 2000, when their contract expired and contract negotiations broke down. At issue were pensions and health insurance. Halliburton and Dresser were already considering relocating (possibly to Mexico), and the strike strengthened their inclination to move. After offers and counter-offers were rejected, Halliburton announced the closing of the Dover plant on October 13, 2000.

In bargaining, the one who can walk away without suffering costs is in the driver's seat. When union members bargain with their employers, it matters greatly if they are up against a footloose manufacturer on the one hand, or an employer that cannot easily relocate on the other, such as a hotel, restaurant, hospital, or public school system.

As firms become more footloose, unions are less effective in raising wages, as a 2005 study found. The author defined an industry as more "global," the greater the fraction of its total global employment was outside the United States. By this measure, the more global industries included drugs, tobacco, and cars and trucks, as well as the soaps, cleaners, and toiletries industry. The study found that the more global an industry is, the less unions are able to raise wages and sustain union bargaining power. For example, in the global industry of soaps, cleaners, and toiletries, wages are about a fifth lower than wages in a hypothetical industry that is otherwise identical but has no global employees.

Sources: Jeffrey P. Carpenter and McAndrew Rudisill, "Fairness, Escalation, Deference and Spite: Strategies Used in Labor-Management Bargaining Experiments with Outside Options," *Labour Economics* 309 (2003): 1–16; Minsik Choi, "The Threat Effect of Capital Mobility on Wage Bargaining," in Pranab Bardhan, Samuel Bowles, and Michael Wallerstein, eds., *Globalization and Egalitarian Redistribution* (Princeton, NJ: Princeton University Press, 2005); U.S. Department of Labor, Bureau of Labor Statistics, "International Comparisons of Hourly Compensation Costs for Production Workers in Manufacturing, 2002," Release 03–507, September 26, 2003.

A **system of control** is an employer's strategy or method for governing the workplace to facilitate the extraction of work from the workers.

Such a social organization can be thought of as a **system of control**. It is the means by which the employer governs the workplace. A capitalist workplace is almost never organized democratically; it is run from the top down. And since the reason the employer organizes work is to produce output commodities in order to make a profit, the system of control is designed to enhance the employer's ability to extract work from workers.

Employers have developed several different systems of control within the firm. Each one reflects a distinct strategy within the workplace for extracting work or for reducing unit labor costs. Among the many ways of controlling workers, the following are widely used.

Simple Control

Simple control is a system of control that focuses on the supervisors' personal exercise of workplace rewards and sanctions to maintain the work pace.

One employer strategy for achieving low unit labor costs is to pay wages as low as possible, barely above the fallback wage $\underline{w}$, and to use supervisors and foremen to bully, charm, cajole, motivate, or drive employees to work hard. This strategy, **simple control**, or the *drive system*, aims at keeping wages as low as possible while forcing employees to give high levels of effort. To the degree that such a strategy is successful, it will minimize unit labor cost *ulc* (see Equation 12.6).

A small textile mill or a McDonald's restaurant illustrates simple control. The firm pays wages very near the legal minimum, and the wages by themselves provide little incentive for employees to do more than whatever minimal effort will keep them from being fired.

How do such employers get their employees to work hard? The employer (or the hired manager) personally directs the work, oversees the employees, evaluates their work, and rewards or punishes them accordingly. For instance, the employer may reward a diligent worker by assigning him or her to a better shift or a more pleasant job. Alternatively, the employer may punish an employee deemed to be underperforming by cutting back on his or her hours or giving him or her a dirty or unpleasant job to do. The unspoken threat of being fired is continually kept in the employees' minds.

Bosses may be petty and tyrannical or they may motivate workers by charm and the force of their personalities. Either way, the result is to reduce unit labor costs by driving employees to put out greater work effort. Fast work and low wages make for a low *ulc*.

Technical Control

Technical control is a system of control that incorporates a work pace designed into the machinery of production.

An alternative strategy aims at achieving the same result by different means. Given that $ulc = w/(ef)$, as Equation 12.6 says, paying the lowest possible wages (w) will keep the numerator in this equation down. With **technical control**, however, the work effort (e) part of the denominator is increased not by ever-present supervisors prodding employees, but rather by the pace of the machinery in the production process. The machinery itself impels the employees to work rapidly. Equally important, with a machine-paced production process it is easier to keep tabs on who is or is not working up to par. Those who consistently fall behind can be singled out and disciplined or fired.

For example, on a car manufacturing assembly line it is the speed of the line itself that sets the pace of work. The typical automobile assembly line in the United States turns out sixty cars per hour, so in this case each worker must do his or her job in the

sixty seconds before the next chassis comes along. Here the boss need not stand over the worker to instruct and exhort, since the line itself sets the pace. Similarly, in the call centers described in Chapter 1, the computer software that handles the routing of incoming calls can also be set up to record how much time it takes employees to dispose of their calls. Using this technology, supervisors can monitor the speed with which the operators deal with their callers and then tell the employees to work faster or be fired.

Technical control does not eliminate the need for supervision. What happens if an employee refuses to "obey" the line? In assembly line or other machine-paced factories or offices, bosses are still required for performing the tasks of evaluating employees and disciplining those who fail to meet the expected standards. Still, the pace of work is controlled, in the first instance, by the physical technology of production.

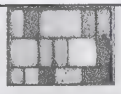
Neither does technical control resolve the issue of who will determine the pace of the assembly line (or any other machine-paced system of production). When employees are weak and not unionized, the boss may determine the line or machine speed without consulting the employees. (This type of situation was humorously but accurately portrayed in the Charlie Chaplin film *Modern Times*.) When employees are highly organized, on the other hand, bargaining between workers and employers often sets the pace of the line. Even with the pace set by bargaining, however, the line imposes the collectively agreed-upon speed of work on each individual employee.

Thus, with technical control the employer's power to design and implement the physical technology of production provides another form of command. The employee is harnessed to a line or machine that runs at a given pace, and he or she must keep up with that pace.

Bureaucratic Control

Bureaucratic control is a system of control that uses job ladders, seniority rewards, and other organizational incentives in order to extract work from workers.

Bureaucratic control is yet another strategy for achieving low unit costs. In a bureaucratic firm the employer pays relatively high wages and, more significantly, pays wages that increase in a predictable way with an employee's seniority. Workers who remain employed by the firm can expect to earn higher wages in the future and perhaps receive other benefits as well.



Technical Control: The Meatpacking Industry

Half a century ago, when the meatpacking industry was concentrated in the Midwest and West, near its historical center of Chicago, meatpacking workers were known as skilled, relatively high-paid, unionized workers with a low turnover rate. All that has changed.

From the 1960s through the 1980s, a firm called Iowa Beef Packers (IBP) led what became known as the "IBP revolution," a massive overhaul of the meatpacking production process. IBP located slaughterhouses in rural areas, away from urban union strongholds, and in the late 1960s eliminated its labor union entirely. It put livestock on an assembly line where they were slaughtered and dismembered according to a preset system, and so it made workers'

continues

tasks simple, regimented, and repetitive. The skilled work of planning and carrying out the slaughter, carving, and packing of an animal was broken down into smaller chores requiring little skill, jobs with names like knocker, sticker, shackler, rumper, first legger, knuckle dropper, and navel boner. As an IBP executive said, "We've tried to take the skill out of every step."

The assembly line allowed the firm direct control of the pace of work (e). In the 1960s a meatpacking plant could pack fifty animals an hour. Today some giant rural plants can pack four hundred animals or more per hour. At peak speeds, one cow is slaughtered every six seconds.

The meatpacking industry nowadays is both profitable and fairly concentrated; for instance, the four biggest chicken processors control about 60 percent of the domestic market. The turnover rate is high; on average, workers last three months before quitting or being fired. Many are unskilled immigrants with few other options, and some speak little English. Wages of around $11 an hour are more than a third lower now than 50 years ago, after adjusting for inflation.

But the worst is that the frenetic pace of the assembly line makes the work hard and painful and leaves many workers with health problems, especially musculoskeletal injuries like hand and arm pain from carpal tunnel syndrome. This usually comes from doing the same hand movements over and over—in the poultry processing industry, as many as 20,000 times a day—and such injuries are sometimes permanent. The cold temperatures required to keep the chicken fresh exacerbate the problem.

Officially reported injury rates have fallen since 2004, but workers who were interviewed said reporting injuries increases their risk of being fired. The Occupational Safety and Health Administration head wrote that poultry workers suffer repetitive strain from microtasks at a rate ten times higher than other U.S. workers. Other government agencies (including the National Institute for Occupational Safety and Health and the Government Accountability Office) also found that meatpacking continues to be a dangerous place to work compared to other industries.

Source: Eric Schlosser, *Fast Food Nation* (New York: Harper Perennial, 2002); Oxfam America, "Lives on the Line: The Human Cost of Cheap Chicken," Oxfam Research Report, 2015, available at https://www.oxfamamerica.org/publications/lives-on-the-line/; Southern Poverty Law Center and Appleseed, *Unsafe at These Speeds: Alabama's Poultry Industry and Its Disposable Workers*, 2013, available at https://www.splcenter.org/sites/default/files/Unsafe_at_These_Speeds_web.pdf; U.S. Government Accountability Office, "Workplace Safety and Health: Additional Data Needed to Address Continued Hazards in the Meat and Poultry Industry," GAO-16–337, April 2016, available at http://www.gao.gov.

How can high and rising wages *reduce* unit labor costs? They do so if they can bring forth additional work effort (e) at a greater rate than the rate at which wages (w) are increased. In this case the ratio between w and ef that determines ulc will fall (f being constant), driving down unit labor costs and, if nothing else changes, raising the rate of profit. (See Figures 12.1 to 12.4 and the discussion surrounding them.) Thus, rising wages can lower ulc if there is an even greater increase in work effort. This is the "carrot" approach: with high and rising wages employees will like their jobs enough, will want to keep them enough, or will feel positive enough toward their employers to put out increasing amounts of work effort.

However, there is also an important "stick" in bureaucratic control. The only way an employee can obtain a better job in a firm is by gaining seniority. But as the employee gains seniority, the greater will be his or her "investment" in the particular job. Moreover, along with the additional seniority and the elevated wage comes a higher cost of job loss in the event that the employee is fired. Thus, the longer he or she works in a particular firm, the bigger the stick becomes!

TABLE 13.1. EMPLOYERS' SYSTEMS OF CONTROL

SYSTEM OF CONTROL	WAGES	SUPERVISION
Simple control	Low, with few rewards for long service	Direct observation by bosses
Technical control	Low, with few rewards for long service	Pace of work machine-controlled: laggards machine-detected; less direct supervision
Bureaucratic control	High, rising with longer service	Promotion and firing by a "rule of law" within the firm

Because there is a conflict of interest between employers and employees, a firm's ability to make profits depends on its having an effective system of control over the labor process. The three systems of control represented in this table are used in varying degrees in different industries. Over the years they have been used to a greater or lesser extent in the economy as a whole.

Usually firms that rely on this strategy establish elaborate work rules and procedures to set up sophisticated incentives. Instead of being simply an employee, a worker is hired for a particular job, which has a job title and for which a formal job description exists. The duties and tasks of the job are outlined, and the employee's performance is measured against the prescribed duties.

Job ladders link together a series of related jobs, in which a worker over the years climbs from one job to another and gains access to jobs higher on the ladder only by first succeeding in the lower jobs.

A firm will use **job ladders** to organize a system of bureaucratic control. Job ladders link together a series of jobs, each job being one "rung" on the ladder. For instance, the jobs of file clerk, general clerk typist, secretary, private secretary, and executive secretary may be linked together on one job ladder. The employer hires job applicants from outside the firm for the bottom rung (say, file clerk), and employees are then encouraged to work their way into higher jobs on the ladder by superior performance in the lower jobs. Thus, the way an employee obtains a better job is by working hard in his or her present job. This is another carrot (positive incentive) to bring forth more *e*.

However, merely setting up job ladders and writing job descriptions does not guarantee that they will be effective. Hence, neither bureaucratic control nor technical control eliminates the need for bosses. Bosses still supervise work, direct workers, evaluate their performance, and determine who gets rewarded, promoted, disciplined, or fired, But in firms with bureaucratic control, bosses rule by "applying company policy." Since they are merely enforcing company rules, the power relations between them and their underlings are embedded in the organizational structure of the firm and, as a result, are more hidden.

The varieties of control discussed here—simple control, technical control, and bureaucratic control—are but different institutional forms of authority, or command, relations within firms. They are summarized in Table 13.1. Each workplace and firm will have its own blend of forms of control. Yet despite the multitude of control systems—including many that we have not discussed—they all exist to achieve a single goal for the employer: to reduce unit labor costs.

Technology and the Labor Process

A *technology*, as we defined it in Chapter 4, is a relationship between the inputs and outputs of a labor process. Technical change is a change in the relation between inputs and outputs. An employer will implement technical change—the introduction of new types of machinery or changes in the method of production, for example—whenever a new technology is available, affordable, and likely to raise the profit rate.

Technical change can raise the profit rate by reducing unit labor costs (*ulc*). In particular, technical change may raise the efficiency of labor (*f*) or make possible, as we will see, a greater intensity of work (*e*) or lower wages (*w*). Each of these would reduce unit labor costs and raise the profit rate, other things being equal. Employers, then, are likely to see technical change as a potentially fruitful source of higher profits.

At any particular time, known and available technologies impose limits, or constraints, on what employers can do. For instance, existing technology may dictate that at least three tons of iron ore are needed to produce one ton of iron; no matter how much the employer desires to reduce the use of this input, the current technology will not permit it.

Similarly, existing technologies place limits on the social organization of the workplace. For example, some production processes, such as an assembly line, may require many people to work together, whereas other jobs, such as those of telephone operators, may involve mostly individual tasks. And while some jobs may require few skills and little experience, the current technology may dictate that other jobs require employees with extensive skills and much experience. In these ways current technology limits how an employer can organize work.

The limits set by known and available technologies, however, are usually quite wide. The same product (for example, basic steel) is often produced by different firms using different systems of control, different proportions of skilled and unskilled workers, different relations between bosses and workers, and different wage structures. Most automobile factories involve highly fragmented jobs on assembly lines, while some car manufacturers use "teamwork" methods. Clerical workers' jobs may be insecure, dead end, poorly paid, and characterized by heavy-handed supervision (simple control), or they may be entry-level positions on a job ladder leading to more secure and highly paid positions (bureaucratic control).

So while technology places some limits on the organization of work, many different ways of organizing the workplace are often compatible with existing technologies. And, of course, the existing technologies are constantly changing. Moreover, the constraints imposed by any given technology are likely to become less important as time goes on because the employer's need to impose a system of control shapes the way technology develops.

Employers who seek technical change that will raise their profit rates will naturally be more interested in some types of technical change than others. For instance, suppose a business owner is about to invest in research to develop a new technology. The firm's research staff proposes two projects. The first project would develop a new technology that promises greatly to reduce input materials used per unit of output (*M/Z*) but would simultaneously reduce the employer's ability to extract work (*e*) from employees because it would require that workers work on their own, making supervision difficult. The second project promises to reduce *M/Z* somewhat less but does not reduce the employer's control of the work pace (*e*) at all. The business owner will be more likely to fund the second project.

Thus, the employer's need to extract work from employees can influence the direction of technical change. Indeed, the process of technical change is "biased": employers, trying to maximize their profits, encourage some types of technical change (those consistent with maximum extraction of work) while discouraging others (those that threaten their power).

The technologies that are known and available at any particular time are the products of a continuing process of technical change. But if this process is biased,

the technologies existing at any particular time will reflect this bias, in part. Not only has existing technology been shaped, but future technical change will also be shaped, by the social organization of production and by employers' need for power in the workplace in order to maximize their profits.

Conflict in the Workplace

As we have seen, employers and employees bargain over the work pace (e), wages (w), and the conditions of work. The workplace is the principal arena for their struggle, and the weapons are both social organization and technology. (They may also contest these matters elsewhere, for instance by attempting to get the government to regulate the safety or other conditions of work.)

In their conflicts employers and employees use quite different means to defend and promote their interests, or, stated differently, they exercise quite different kinds of power. First, the employers have established the battle's terrain, for it is they who hire the employees for—and organize the production process in—a given workplace. While employees, individually or collectively, may choose to withdraw from this terrain (the workplace) by quitting their jobs or going on strike, they never, or rarely, have the opportunity to organize production themselves.

Because employers, not employees, organize production, employers have the power to initiate action or change circumstances. Workers, on the other hand, are inherently in a defensive position, defending their interests in reaction to the initiatives of employers. This difference in position is evident in such employer strategies as moving a plant to another state or country, or establishing a system of control. It is also evident in the way that technical change affects workplace conflict.

Technical Change and Workplace Conflict

Technical change may affect workers in three principal ways. First, it may raise the productivity of their labor without increasing their workload. In Equation 12.6 in the previous chapter, f is used to signify the "efficiency of labor," in other words, the amount of output produced per unit of work effort. The particular value of f reflects the technology currently in use, because it is this technology that determines how much output can be produced with any given amount of effort. As can be seen from Equation 12.6, this first type of technical change lowers unit labor cost by raising f, and it does this even if the amount of work effort (e) being put out remains unchanged.

RAISING THE EFFICIENCY OF LABOR. Consider the following example. A worker in a shoe factory, using a simple electrical stitching machine and working at a comfortable pace, can produce two pairs of shoes in an hour. Hence, output per hour $z = 2$. If we assume that the amount of effort being put out by the worker is normal for this type of job, we can say that $e = 1$. Then, since we know that $z = ef$ and $z = 2$, we find that f, the efficiency of the shoe stitcher's labor, is 2. Now suppose that the owner of the shoe factory makes an investment and provides a computer-controlled stitching machine for the employee to work with (in place of the simple electrical machine). This technical change means that the way of making shoes has changed, so that even if the employee works no harder (e is still 1), he or she can now produce five pairs of shoes per hour. Accordingly, this technical change—a change in the way of making shoes—raises f from 2 to 5.

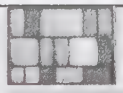
Technical change that raises f can reduce unit labor costs without requiring either that employees work harder or that they take pay cuts. Hence, when this type of technical change is a possibility, there is in some sense no inherent conflict between workers who remain on the job and their employers. For these workers, both w and e are unchanged. However, unless total sales rise to 2.5 times their former level, some workers are likely to be laid off, since the same output can now be produced with 2/5 of the workers. If there is solidarity among workers, they may well resist the change.

So in this and related situations, employer and employee may still fight over f. For example, inadequately paid employees in a particular firm may see that one way of pressing for higher wages is intentionally to obstruct production, thus reducing f. Although this type of action results in no direct benefit for the employees, it does impose costs on their employer. If these costs are greater than the wage increase being sought by the employees, the tactic may induce the employer to agree to the wage increase.

To take a specific example, if workers in an auto assembly plant feel aggrieved, they may occasionally drop a soda bottle into the door of a car as it passes by them on the assembly line. Since this creates a rattle that must be eliminated before the car can be sent to a dealer, additional employees (more work effort) must be assigned to disassemble the doors and remove the bottles. This kind of "sabotage" reduces the efficiency of labor (f) because more work is now required to produce the same number of cars. Similarly, when employees "accidentally" spoil materials or break tools during production, this, too, slows down production and reduces f. The Industrial Workers of the World—the "Wobblies"—a militant labor union whose organizing heyday was before and during World War I, captured the idea in a famous slogan: "Good pay or bum work!"

Speedup is an effort by an employer to increase the pace of work.

SPEEDUP. Technical change can affect workplace conflict in a second way. It can provide a means for employers to **speed up** work (raise e), thereby reducing unit labor costs. The most obvious way that employers can use technology to control employees is with technical control and machine pacing of work. When auto makers introduced the assembly line, it gave the employers a new way to control and speed up work. Today, technology enables employers to read employees' e-mail, to listen to their phone conversations, to record the time it takes operators to deal with callers, to keep track of how rapidly typists are typing, and, in the case of drivers (see box "On the Road Again"), to monitor the speed and location of company vehicles. All these capabilities enable employers to get more or better work from their employees.



Blowout: "Good Pay or Bum Work!"

Firestone recalled 14.4 million of its tires in August 2000 due to manufacturing flaws resulting in "tread separations" likely to cause blowouts. The National Highway Traffic and Safety Administration placed Firestone tires under investigation in cases involving 271 fatalities and more than 800 injuries. This public relations debacle moved the world's largest tire manufacturer—Bridgestone (the giant Japanese tire maker that had bought Firestone in 1988)—to think of dropping the Firestone name. But it was a mystery: How had Firestone put on the road so many blowouts waiting to happen?

Two economists (associated with the industrial relations section of Princeton University's Firestone

Library) unearthed a clue. Of three Firestone plants—in Wilson, North Carolina, in Joliette, Quebec, and in Decatur, Illinois—making the type of tires that were recalled, it was the Decatur plant's tires during 1994–1996 that were much more likely to have had the bad tread separations than those made at the other two plants, or those made at the Decatur plant in other years. Firestone engineers' lab tests confirmed the finding.

What was special about Decatur during the mid-1990s? *Labor strife.* Early in 1994 the company proposed increasing shifts from eight to twelve hours and running the plant twenty-four hours a day, with workers alternating night and day shifts. Firestone also wanted to pay new workers 30 percent less and reduce pension and other benefits. In April 1994 the 4,200 employees went on strike. Firestone replaced the striking workers at much lower pay, and later announced that the replacements would be permanent and that the strikers could seek reemployment at reduced pay when the need for additional work arose.

Over the next year many took up the offer, but under highly difficult conditions. According to a union account: "Forced to work alongside scabs who had taken their jobs . . . the strikers were assigned to the hardest jobs on the worst machines, rather than the jobs they had held for ten, twenty, and even thirty years. The company supervisors had a field day harassing, intimidating, and firing union members for the smallest infractions." Building quality tires may not have been the workers' first priority.

The economists conclude their study: "Unless another factor can be found that explains the sudden rises in defects in tires produced when Bridgestone Firestone demanded contract concessions . . . and again when replacement workers and recalled strikers worked side by side, we think that the weight of the evidence points to labor strife as being at the root of many of the defective tires." They estimate that faulty tires produced at the Decatur plant during the years of strife accounted for at least forty deaths, a number that would have doubled if there had been no recall.

A century ago the International Workers of the World (IWW), a radical American labor group, demanded, "A fair day's pay for a fair day's work" and pointedly threatened, "Good pay or bum work!" Firestone would have saved a lot of money (and lives) had they recognized the force of these demands.

Source: Alan Krueger and Alexandre Mas, "Strikes, Scabs, and Tread Separation: Labor Strife and the Production of Defective Bridgestone/Firestone Tires," *Journal of Political Economy* 112, no. 2 (2004): 253–89.

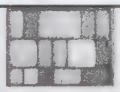
The assembly line radically altered the work process. In manufacturing in the past, the work (the thing being manufactured) had been stationary, with the workers moving around to get tools or parts or to pass on finished work to other workers. But the advent of the assembly line reversed the positions: it made the work mobile and the workers stationary. No longer did workers have much chance to move around and get to know other workers in the plant; they spent their entire workday working at one "workstation." So workers became more isolated from one another, and it grew harder to organize collective resistance to any speedup. Indeed, for employers, the fact that the assembly line enhanced their control over both the pace of work and the movements of their employees was a key advantage. Later on, employers would find that the assembly line was a mixed blessing because it allowed a fairly small group of workers to bring the entire line to a halt. Indeed, this happened in the late 1930s, and the famous "sit-down strikes" of that time led to the formation, and recognition by employers, of the United Automobile Workers union.

Deskilling means changing a production process in such a way as to make it possible to employ workers with fewer skills.

DESKILLING A third way that technical change may be used to control workers is through *deskilling*. Deskilling means changing production methods so that they require fewer skilled workers and rely more on unskilled workers.

Employers often favor deskilling because skilled workers are the ones who are generally in the strongest position to resist speedup of work and to bargain successfully for higher wages. Skilled workers are those who have some special training, knowledge, or credential. Skilled workers tend to be in shorter supply than unskilled workers, since skilled workers can usually do unskilled work, whereas unskilled workers typically cannot—or are not allowed (by their unions) to—do skilled work. Moreover, acquiring a special skill or credential may be difficult, expensive, or time-consuming. And because there are fewer skilled workers than unskilled workers, it is usually more difficult to replace skilled workers. For all these reasons skilled workers are generally in a stronger bargaining position than unskilled workers.

When employers seek to reduce unit labor costs, they naturally focus on their skilled workers. After all, the latter are likely to be the highest paid and most powerful employees in the firm. If the production process can be changed in such a way that it no longer requires skilled workers, employers can then hire unskilled workers, and these unskilled workers will cost less and be less able to resist speedup. With deskilling an employer can easily replace currently employed skilled workers with less expensive unskilled workers because the available supply of unskilled replacement workers is greater than the supply of skilled workers in most labor markets.



On the Road Again—With a Box Looking Over My Shoulder

American culture—as expressed, for example, in Willie Nelson's country & western song, "On the Road Again"—often celebrates the freedom of the cowboy and his modern incarnation, the long-haul trucker. When U.S. trucking companies installed on-board computers during the 1980s, however, truck drivers experienced a sudden curtailment of their freedoms. The computers served as "trip recorders" that greatly improved the companies' ability to monitor the actions of their drivers. The trip recorders gave the companies verifiable information on the speed, idle time, and other details of the operation of the trucks. Their sole function was to enable the companies to monitor those aspects of the drivers' behaviors in which there was a conflict of interest between the drivers and the companies.

Before the trip recorders were installed, drivers liked to take long breaks and then make up for them by driving faster than the speed that would minimize fuel costs and wear and tear on the trucks. Since the trucks were dispersed all over the country, there was no way for the companies to force their drivers to slow down. In contrast, the owner-operators—drivers who owned their own trucks—had to pay for their fuel and repairs, so they drove more slowly. As a result, they were able to compete successfully with the trucking companies on runs where the conflict of interest between the companies and their drivers was particularly strong.

After the trip recorders were introduced, the trucking companies were able to write contracts with their drivers that specified the speed at which the trucks were to be driven, and drivers who drove faster were penalized. As a result, speeders now made less money. In the end the trip recorders had two effects. First, they induced the drivers in trucks with trip recorders to drive slower, and second, they reduced the market share of the owner-operators, whose cost advantages over the trucking companies disappeared.

Too bad horses went out of style.

Source: George P. Baker and Thomas N. Hubbard, "Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking," *Quarterly Journal of Economics* 119, no. 4 (November 2004): 1443–79.

Technical change can increase the skills required in production, but it often reduces the need for skills. When it reduces the need for skills, technical change offers an important means by which employers can deskill their labor processes.

One way that employers can deskill work is to break up the whole labor process into many small tasks and then assign each employee only one, or possibly a few, tasks to perform over and over again. Since each employee now needs to know how to do only a few tasks, new workers can be trained quickly for the particular jobs.

The **separation of conception from execution** is one method for deskilling work in which the workers who plan production are different from those who carry it out.

One of the most common ways of fragmenting and deskilling jobs is called the **separation of conception from execution** in work. *Conception* means planning out the work, preparing in one's mind or in written plans how the output should be produced. *Execution* is carrying out the planned work according to the conception. Highly skilled workers such as machinists have the opportunity and ability both to plan their work and to carry it out, and part of the enjoyment that comes from their jobs is planning the work and then seeing it materialize in their own hands. Small farmers, surgeons, and artists, as well as self-employed builders, electricians, and other tradespeople continue to do both parts, integrating conception and execution.

In contrast to more traditional processes of production, modern labor processes often separate conception from execution. Employees in planning, drafting, and engineering departments determine the organization of production, but they rarely produce anything except plans or designs. On the other hand, execution occurs in a factory where the employees who actually make the product do only the physical tasks. They rarely participate in planning or improving the product they are making. The result is that planning ("white-collar work") takes place in offices, while execution ("blue-collar work") takes place in factories that are often located in entirely different places or even in different countries.

Even some white-collar work has been subjected to deskilling. Such employees as clerical workers, draftsmen, record keepers, and sales personnel have seen their jobs computerized and deskilled. They, too, have become mainly operators of computers, devices with video displays, or other office machines. They have become more like factory workers, and the need—or opportunity—for exercising mental skills in many of their jobs has declined.

In sum, technical change is a many-sided process. On the one hand, it can be introduced in such a way that it helps employers but does not damage their employees' interests; in these situations it is possible for employers and workers to share the benefits of technological progress. On the other hand, technical change can be used by employers to deskill work or introduce new controls on labor (to raise work-effort levels), with the intention in both cases of weakening workers' bargaining power. When technical changes are likely to work against the interests of workers, employers may try to hide the potential effects of a new technology from current employees since these workers will be more likely to resist the introduction of any new technology if they see that it will enable their employers to speed up their work, eliminate the need for their skills, or otherwise damage their interests.

Not surprisingly, introducing new technology often leads to contention, conflict, and bargaining between employers and their employees. When employees are weak and not unionized, they may not have the ability to resist changes that harm their interests. When employees are stronger, for instance because they are unionized, they may be able to bargain and prevent the implementation of technical changes that will not benefit them. As a result of such conflicts, the introduction of new technologies that could potentially yield benefits to both workers and employers is often delayed or blocked altogether.

Control over technical change is never equal. Whether employees are weak or strong, business owners retain the power to make or not make investments, and this enables them to decide whether, when, and where to invest in new technology. Technical change requires substantial investment, both for the research and development stage and for building and installing new equipment. Only capitalists can invest on a significant scale, so they retain the power to initiate (or withhold) technical change. Workers, even in the best of circumstances, can only hope to shape or block technical change. Only rarely will they be in a position to initiate technical change that would promote their interests.

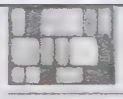
Unions

A **union** is an organization of workers established with the intention of providing a unified and stronger voice on behalf of the members' interests.

Workers organize and join **unions** to increase their strength in bargaining with employers. The principle of unions is simple. If each worker bargains as an individual with his or her employer, then each worker will be competing with every other worker. The outcome will necessarily be determined by those workers who, for whatever reasons, are willing to accept the lowest wages, the fastest work pace, and the least favorable working conditions. The employer will simply choose to make a deal with them and then replace or not hire anyone who insists on better terms of employment. If, however, all or most employees decide to bargain collectively through a union, the union may be able to achieve better results for all or most of the workers in a given firm or industry than they could obtain otherwise.

A union, then, is a means for reducing (horizontal) competition among workers. If successful, it gives its members a degree of monopoly power in the labor market. For exactly this reason, many employers (such as Delta Air Lines, among a great many others) have successfully resisted the formation of unions. The existence of unions, and the union power that remains today, are the result of more than a century of hard-fought battles between pro-union workers and anti-union employers.

RECENT TRENDS IN UNION MEMBERSHIP Although unions represented 35 percent of the U.S. nonfarm labor force in the 1950s, only 11.1 percent of all American workers belonged to unions in 2014, and in the private sector of the economy union membership amounted to only 6.7 percent of the workforce. The decline of unions in the United States since the mid-1950s has been due primarily to a seismic shift of jobs from the manufacturing sector of the economy, where unions were strong, to the private service sector, where they were relatively weak (see Figures 7.1, 7.2, and 7.7). In 1955 nearly half the private sector workforce was employed in goods-producing



What Does Labor Want?

What does labor want? It wants the earth and the fullness thereof. . . . We want more schoolhouses and less jails; more books and less arsenals; more learning and less vice; more constant work and less crime; more leisure and less greed; more justice and less revenge; in fact, more of the opportunities to cultivate our better natures.

—Samuel Gompers, first president of the American Federation of Labor, address to the International Labor League, Chicago World's Fair, August 28, 1893

sectors (mining, construction, and manufacturing), with the other half being employed in service-producing sectors (transportation and public utilities; wholesale and retail trade; and other services including finance, insurance and real estate, health services, educational services, social services, and so on). By 2015, 71 percent of the U.S. labor force was employed either in the private service sector or in wholesale and retail trade, with about half of the rest employed in goods-producing industries and the other half in government.

The decline in union membership in the United States was especially rapid during the early 1980s, when the dollar became overvalued, which widened the gap between U.S. and foreign wages and made it more attractive for U.S. firms to acquire productive assets abroad and to move operations abroad. Likewise, layoffs were widespread in the U.S. steel industry as domestic steel-using manufacturers increased their imports of lower-cost steel from abroad.

There have been two main trends in the other direction. One has been organizing public-sector workers, including police, firefighters, and many others, into unions such as AFSCME (American Federation of State, County, and Municipal Employees), as the number of government employees grew from 7 million in 1955 to more than 22 million in 2015. This effort was aided in some states by new state laws authorizing collective bargaining for certain categories of public employees. Other unions organized federal government employees, postal workers, and teachers. In 2014, 35.2 percent of all government workers belonged to a union.

The second trend has been the highly successful organizing drive by National Nurses United. This union employs an innovative organizing strategy: it not only strives for better pay and working conditions for nurses, but also advocates for better patient care and for reforming the entire healthcare system. Its broad vision and dynamic message has appealed to nurses, and its rapidly growing membership reached 185,000 in 2016.

UNION ACTIVITIES Unions engage in two different types of activities. First, they act as agents for their members in bargaining with employers. For employees in a unionized workplace, the union negotiates a contract that stipulates the wage rates and employment practices to be observed during the period of the contract. Depending on the industry and the strength of the union, the contract may specify the general work effort level (the standard or rate) required for each job, the procedures by which new technology is to be introduced and new (different) jobs created, and the method used to assign employees to different jobs. The contract also usually establishes a grievance procedure to decide cases when employees believe that the employer is not following the contract. In these ways a union contract can protect workers from an employer's otherwise unconstrained commands. Of course, the union's ability to provide effective protection depends on how strong the union is, both at contract renewal time, getting the employer to agree to a decent contract, and during the life of the contract, seeing that the employer lives up to the terms of the contract.

MANAGING THE THREAT OF REPLACEMENT One way a boss can increase his or her strength in bargaining over both e (work effort) and w (the wage rate) is to make it easy to replace each worker with someone else. The ultimate power bosses have over workers is the power to fire them. But is the boss's threat to fire workers believable? That depends on how difficult it is to find a replacement worker. Imagine two different situations.

The first is that of a skilled crane operator during a time of relatively low unemployment. The boss may bargain with this worker by threatening to find "somebody else to do the job who really wants to work." However, both the boss and the crane operator may know that there are few unemployed crane operators; that the union will protest the firing and may even get other workers to go out on strike; that it takes a long time for a new crane operator to learn the job (thus making the other workers on the job less productive while the new operator is learning); that cranes are expensive and costly accidents are more likely to happen when someone is new on the job; and that, in any case, the crane operator can easily find another job at the same or better pay. In this case, the worker is in a strong bargaining position.

In contrast, the second situation is that of an unskilled factory worker during a time of high unemployment. Under these circumstances, when the boss threatens to fire the worker if he or she does not agree to work hard for low wages (as we saw in meatpacking), the threat is entirely believable. There are many people looking for jobs who could, in fact, do this job. It does not take long to learn the job, and there is no union to protest if the worker gets fired. Here, the boss is in a very strong position.

Both bosses and workers realize that their bargaining strengths depend on how easily the workers can be replaced. The more easily the boss can replace the worker, the stronger the boss will be. This is true both for unionized and nonunion workplaces. It is also a reason why sports superstars are able to negotiate such high pay. The ultimate in replaceable workers are those whose jobs are define as temporary in the first place (see box, "Workers in the Gig Economy.")



Workers in the Gig Economy

Temp work: a respectably paid way to make some extra bucks in between the end of one job or semester and the start of another? Forty years ago, maybe; today, not so much. Still, sometimes almost any job is better than no job. But today 26 percent of temporary workers get so little pay that they qualify to receive benefits from Medicaid, food stamps, or other welfare programs. For permanent workers, that share is just 14 percent.

About 3 million workers do temporary work in the United States, and that number has been growing. In recent research by economists Lawrence Katz and Alan Krueger, temp workers were grouped with others to form a larger category called "workers with alternative work arrangements." They defined the category to include four groups: those who worked for temporary help agencies, on-call workers, contract workers, and independent contractors or freelancers. Jobs in this category grew from 10.1 percent of all employment in 2005 to 15.8 percent in 2015, an increase of just over 9 million jobs. Remarkably, this accounted for *all* of the net employment created in that decade. In other words, it appears that the number of *regular* (non-alternative) *jobs did not increase at all* during the decade. No wonder social scientist Guy Standing and others describe workers today as the "precariat."

Source: Lawrence F. Katz and Alan B. Krueger, "The Rise and Nature of Alternative Work Arrangements in the United States, 1995–2015," NBER Working Paper 22667, September 2016, available at www.nber.org/papers/w22667; Guy Standing, *The Precariat: The New Dangerous Class* (New York: Bloomsbury Academic, 2011).

PROMOTING SOCIAL CHANGES TO BENEFIT WORKERS A second activity engaged in by unions involves promoting general social changes that may benefit all workers, including not only their own members but other workers as well, whether or not they belong to a union. For instance, unions have often taken the lead in pressing the federal government to pass minimum wage legislation (raising the minimum wage), workplace safety and environmental protection laws, social security and unemployment benefits, civil rights and antidiscrimination laws, earned income tax credits (for low-wage workers), income support for people unable to work, and various other social service programs.

Discrimination in the Workplace

Unions are a method for overcoming competition among workers. By contrast, discrimination, whether practiced by businesses, unions, or individuals, tends to generate and reinforce divisions among workers.

Discrimination means treating someone differently simply because that person belongs to a certain group.

Discrimination means treating someone differently simply because that person belongs to a particular group. The most widely practiced forms of discrimination today in the United States are discrimination against women and discrimination against minorities. People with handicaps, LGBT, and older workers are also frequent targets of discrimination.

Discrimination stems from a variety of sources—historic, economic, religious, political, social, and psychological—which will not be investigated here. What is relevant to our discussion is that the capitalist profit-making process contains opposing tendencies with regard to discrimination: those that tend to perpetuate and increase it, and those that tend to erode and eliminate it.

Capitalism affects discrimination in two ways. One way weakens it and the other strengthens it. Discrimination is weakened when competing firms attempt to minimize their costs by hiring the best persons for their jobs at the lowest possible wage rates. Race and sex discrimination has generally meant that job opportunities, especially those involving high-wage jobs, have been limited for women and minorities who are qualified and able to perform the jobs. But if, facing equally qualified workers, employers hire those who cost the least, they will tend to hire women and minorities and thereby increase the demand for the labor of these groups.

Nondiscriminatory hiring will mean more job opportunities for women and minorities, reductions in their unemployment rates, and, possibly, increases in bargaining strength and pay levels. It will also mean that firms with such nondiscriminatory hiring practices will have lower labor costs than firms that practice race and/or sex discrimination. Thus nondiscriminatory firms will be able to make the same or greater profits while charging lower prices for their products. Firms seeking to hire minority or women workers and promote them into positions of authority may encounter resistance from their existing predominantly male and white workforces and sometimes from the trade unions representing them. But the firms that succeed in hiring and promoting in an unbiased way will, according to this line of reasoning, drive the biased ones out of business, and discrimination, at least in labor markets, will eventually be eliminated.

On the other hand, discrimination is perpetuated when employers try to take advantage of discriminatory attitudes in conflicts with their employees over wages and the pace of work. Just as workers try to achieve greater bargaining strength for themselves through unions, so employers may try to weaken the bargaining strength

of workers by encouraging divisions and promoting disunity among their employees. To do this, they may attempt to foster and magnify any differences and conflicts that already exist among their employees. Given the residual racial, sexual, and other forms of bigotry that have endured historically in Western societies, employers may seize upon these to divide workers. Capitalists did not invent racism, sexism, or other forms of discrimination, but they have sometimes used preexisting prejudices or biases among workers to divide and weaken them. In this they are no different from white and male workers who have sometimes appealed to the same prejudices to limit competition from women, African Americans, and recent immigrants.

Indeed, an individual employer may personally be completely free of prejudice yet be pressured by competition to discriminate in order to stay in business. This might occur if discrimination is profitable. Since firms making higher profits have a competitive advantage, the nondiscriminating firm may, over time, be forced out of business. If, on the other hand, the costs of discrimination outweigh the benefits, employers are likely to stop discriminating.

How can discrimination be profitable? To answer this question, we will compare two imaginary firms. The first firm hires both black and white workers, and it treats all its employees equally—for example, paying wages without regard to race. The workers in this firm, black and white together, form a union, and their union presses for higher wages and safer working conditions.

The second firm also hires both black and white workers, but it discriminates: it assigns blacks to lower paying jobs than whites. The employees of this second firm try to form a union, but they have a more difficult time than the employees of the first firm. Not all employees in the second firm have the same interests. Black workers think the proposed union should try to force the employer to eliminate the company's discrimination in job assignments and pay rates. The white workers, on the other hand, want higher pay but also worry that they will end up with lower wage rates, or at least slower advancement. At this point white workers may form their own union and exclude blacks in order to concentrate entirely on raising their own wages. (This actually happened for many years in the construction industry in New York City, where unions representing white construction workers managed to exclude blacks both from membership in their unions and from jobs in the industry.) However, when the predominantly white union negotiates with the employer, the employer threatens to hire more blacks to replace the white workers. Because of their inability to act collectively, both white and black workers will be stuck with lower wages than they might have negotiated had the employer not engaged in discrimination and thereby fostered divisions among the workers.

Comparing these two firms, we can see that in the first firm employers will face (black and white) workers who share a common interest in raising wages (w) and reducing the strain of work (e). To the extent that these workers are successful, the first firm's profits will decline. The second firm, in contrast, faces workers who are divided by racial conflict. These workers will have difficulty forming a union. Consequently, these workers will have less success in their bargaining for higher wages or less strenuous work.

The second firm will have a higher profit rate than the first firm as a result of its discrimination. Indeed, the first firm may be forced to begin discriminating if it wants to compete successfully in the same market with the second firm.

Neoclassical economists (for example, Milton Friedman in his book *Capitalism and Freedom*) argue that discrimination is costly, and hence competition for profits

will eliminate discrimination. But this is just one side of the coin, the only side visible when the horizontal aspect of a capitalist economy, competition, is the focus of one's attention. But when the vertical, or command, relationships of the economy are considered, a different picture emerges. When employers discriminate, it may be because discrimination is profitable. If this is the case, competition may drive even unprejudiced employers to discriminate.

If nondiscriminatory policies are more profitable than discriminatory ones, then the workings of the market will tend to eliminate discrimination. If discrimination results in higher profits, however, a capitalist economy will tend to perpetuate—or even increase—discrimination. The fact that competition among profit-seeking businesses can break down the barriers of discrimination is dramatically affirmed by the rapid integration of most professional sports after Jackie Robinson broke the color barrier and starred for the Brooklyn Dodgers seventy years ago. But it does not always work this way. More direct ways of overcoming discrimination may be more effective, as is shown by the fact that the federal government of the United States is a much more racially integrated employer—at all levels—than are the leading private corporations. The U.S. Army beats the Fortune 500 when it comes to integration.

Profitability Versus Efficiency

Profitability refers to how much profit is derived from a labor process. Profit is the excess of sales revenues over the costs of purchased inputs.

Technical efficiency refers to the relationship between the output of a production process and its inputs. (Concepts of efficiency are also discussed in Chapter 3.) A production process is said to be technically efficient if, given the existing technology, the output of it cannot be increased without using more of at least one of the inputs.

Profitability refers to how much profit is derived from a labor process.

Technical efficiency: A production process is said to be *technically efficient* if the output of it cannot be increased without using more of at least one of the inputs.

Technical change can improve technical efficiency by reducing the amount of materials and machines used per unit of output, or by increasing f—the amount of output that can be produced with any given amount of work effort (e). Any such change would reduce the amount of inputs needed to produce any specific quantity of output.

Increased technical efficiency *may* lead to greater profitability. Other things being constant, an increase in efficiency will reduce the amount (and presumably the cost) of the inputs a firm needs to produce its output without affecting its revenues from that output. Under these circumstances an increase in efficiency will increase profits. Many economists, observing this connection, presume that profitability is identical with efficiency.

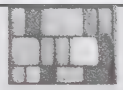
There are many situations, however, in which a less efficient technology is a more profitable one. What is profitable does not always need to be efficient, and vice versa. In fact, the adoption of a new technology may increase profitability while at the same time requiring more of some input—say, work effort, intermediate goods, or capital goods. This is often referred to as an **inefficient technical change**.

An **inefficient technical change** is one that can bring about an increase in the output of a production process only by utilizing proportionally more of at least one input.

How might a less efficient technology be more profitable? Suppose that a new method of production—say, an assembly line—allows an employer to speed up production, thus increasing output per hour of labor (z) by using technical control to increase work effort per hour (e). Assume also that the wage rate (w), the hours

of labor, and all other variables in the labor process remain the same. The new technology will be more profitable, but will it also be more efficient? Suppose that work effort per hour increases by 30 percent with the introduction of an assembly line, and output per hour goes up by 20 percent. In this case, the assembly line technology will be more profitable because the firm will be getting 20 percent more output for the same labor cost. It will not be more efficient, however, because the ratio of output per hour (z) to one of the inputs, in this case work effort per hour (e), will have *declined* (it will be 120%/130% = 92 percent of the z/e ratio before the change).

If the assembly line had increased output per hour (without requiring a more intense or faster pace of work), and everything else in the labor process had remained the same, the technical change would have been unambiguously *more efficient*. But the introduction of the assembly line in our example allows the employer



Inefficient Technologies may be Profitable

We think of modern technologies as ways of solving problems. Indeed, engineers evaluate the effectiveness of various ways of producing products, seeking those that are most efficient in the sense that the greatest amount of useful output can be achieved using given amounts of inputs. But some technologies just do not seem to make sense. The reason is that technologies are often developed to make money, not to make sense, and the two are not the same. The history of technology gives us many examples.

During the nineteenth century food canning in California was primarily the job of unskilled workers, but the task of putting the tops on the cans—capping, as it was called—required more skill. Cappers were both skilled and scarce. They could easily demand higher wages from their employers by threatening to strike just at harvest time, when the fresh food would rot if it were not quickly canned. Employers were overjoyed when James Cox invented his famous machine, called Cox's capper, which made it possible to replace the skilled workers. But the machine did not work very well, so firms that bought it did not put it into use right away. The firms just kept the machines on hand as a visible threat to the skilled workers: if the workers went on strike, the machines could take their place.

Cox himself understood the appeal of his invention: when it was invented it was not the best way of getting the cans capped, but it did put the skilled cappers in their place. Writing twenty-six years after the invention, he referred to the once-powerful position of the skilled capper and the resulting weak bargaining position of the cannery owner: "[The] helplessness of the canner . . . made him a willing advocate of every mechanical means, and made possible the working out, through frequent failures and heavy losses, the perfected mechanical means now in use."

A more recent historical example of the divergence of sensible and profitable technology is compact discs (CDs) that were used for several decades as the main way to store music, and are still in some use. Such discs may be made in such a way that they can be easily copied, thus enabling many people to enjoy the music stored on them. But the companies that produced them spent millions of dollars to develop technologies to render discs difficult, if not impossible, to copy. Such technologies made it possible for music companies to produce discs that were profitable but less useful to listeners.

Source: Martin Brown and Peter Philips, "The Historical Origin of Job Ladders in the U.S. Canning Industry," *Cambridge Journal of Economics* 10 (1986): 129–45.

to increase the pace of work, and the new production process requires using *proportionally* more of at least one input, work effort (e), in order to achieve the increase in output (z). Since e goes up more than z and nothing else changes, the introduction of the assembly line is an *inefficient technical change* even though it increases profitability.

What makes it possible for profitability to diverge from efficiency here is that the work effort, or *labor*, that actually goes into the production process is not the same as the amount of purchased *labor time*. If the employer in the assembly line example had had to pay 30 percent more for the additional work effort required by the assembly line to achieve an increase of only 20 percent in output, the introduction of the assembly line would not have been profitable. In this case, the inefficient technical change would not have occurred.

For capitalist firms, what counts is profitability, not efficiency. Hence firms will tend to introduce the methods of production that are most profitable, regardless of whether they are the most efficient ones available. When business owners invest, including when they invest in research to develop new technologies, they will invest in what is most profitable, whether or not it is more efficient as well.

Thus, the idea put forward in conventional economics that the technologies currently in use and being developed are simply the "best that modern science has to offer" is not necessarily true. Today's technologies—and current corporate research in new technologies—reflect the fact that to a large extent science has been harnessed to the criterion of profitability. The technologies that might be possible if different criteria were used—for example, to minimize the work effort required to produce necessary goods and services—might be quite different.

Markets and Hierarchies

Markets and hierarchies are sometimes thought to be alternative ways of organizing society. Markets are said to establish conditions of (horizontal) equality: voluntary exchange between buyer and seller means that either party can back out of a transaction if he or she chooses, and so a kind of equality exists between them. Hierarchy, on the other hand, implies the existence of (vertical) inequality: command relations require superiors and subordinates, and bosses have power over their underlings.

Hence, markets have sometimes been counterposed to hierarchies, as though an economic system must be organized with only one or the other. Selling gasoline retail is seen as a market-type transaction, whereas the U.S. Army and the Catholic Church are seen as being organized in hierarchies. Conventional economists define capitalism as a market system, while economic systems that use central planning are regarded as hierarchical or "command" economies. Yet in this chapter we have seen that profit-making firms, in their efforts to gain a competitive advantage over their market rivals, establish huge and sophisticated hierarchies.

It is true that when markets began to replace slavery and feudal hierarchies they offered some relief from traditional command relations. However, the larger truth is that markets brought along with them a complex system of bosses and workers, of superiors and subordinates. The capitalist system's need for hierarchy has called forth whole new fields of study (with their associated experts): industrial psychology, personnel management, occupational testing, and industrial engineering.

Corporations invest substantial portions of the surplus product to develop and refine techniques for maintaining hierarchies. Capitalism did abolish precapitalist hierarchies, freeing individuals to enter into voluntary contracts. But we now find that most people in a capitalist economy spend virtually their entire work lives in hierarchically organized jobs. Finding work is not usually a matter of entering into a contract with an equal. Rather, it is most often a matter of choosing which hierarchy to work in or which boss to work under. It may also evolve into an effort to become a boss rather than a subordinate.

In a capitalist economy, then, markets and hierarchies are not alternative and competing methods of organization, much less different types of economic systems. Its horizontal and vertical dimensions, competition and command, are mutually reinforcing and complementary elements of the system as a whole.

Democratic Firms

Must firms be run from the top down? Or might firms be run democratically, with the employees together owning the firm, electing the management, and participating in major decisions concerning technology, work safety, product design, hours, wages, and the like? Just about any firm will have a structure of authority, with some people giving orders and others taking them. What distinguishes a firm as democratic, then, is that the employees, through an election in which each worker has an equal vote, have chosen the people with authority. (See the box "Own Your Job. Elect Your Boss" in Chapter 5.)

A **democratic firm** is one that is owned by its employees and run by people who are elected by the employees.

We know that **democratic firms** are possible because thousands of them exist, and many of them are very successful. Such firms may also be reasonably large as well as successful and democratic.

Many people favor democratic firms for moral or political reasons, on the grounds that firms, like other powerful institutions, should be democratically run. There are also economic grounds for favoring democratic firms.

The economic advantages of a worker-owned and democratically run firm are clear. Because its workers are its owners, their incomes depend directly on the productivity of the firm. For this reason they have an interest in adopting technologies, work methods, and forms of organization that promote the effective use of the firm's resources, including their own time and energy. Ownership of the firm by workers does not change the facts that hard work is hard or that people might like others to shoulder the more onerous tasks of production. But it does fundamentally alter the incentives of workers: a job done well benefits the owners of the firm, so when the workers are the owners, they have strong incentives to see that work is carried out in the most effective manner.

The disadvantages are equally clear. Most workers do not have sufficient wealth to purchase "their" share of the firm, so it is difficult to achieve complete worker ownership. Banks are often reluctant to lend to worker-owned and worker-run institutions. Perhaps most important, for workers to own their own firm they would have to have most, if not all, of their wealth invested in their own firm. Because they would have all their eggs in one basket, they would tend to be very cautious in their decision-making: if their firm were to fail, they would lose not only their jobs but also most of their wealth. As a result, workers who own their own firms tend to be cautious about taking the risks necessary for innovation in product design and production processes.

By contrast, the owners of a capitalist firm (the stockholders) typically have their assets invested in many firms. Hence, their fortunes are not solely dependent on the success or failure of any one firm. And unlike the worker-owners of democratic firms, most of them are rich. For these reasons capitalist owners are more likely than are worker-owners to tolerate risk-taking and innovation by the managers of the firms they own. In addition, the capitalist firm is more likely than are worker-owners to use the latest technologies (as discussed above) to extract increased work effort from their employees. Thus, capitalist firms will tend to have lower unit costs and be more competitive in product markets than worker-owned firms.

Although many studies indicate that worker ownership and democratic participation in a firm's decision-making enhance both productivity and worker satisfaction, worker-owned and worker-run firms are the exception in capitalist economies. The fact is that a democratic firm's lack of access to capital, bias against risk taking, and commitment to decent working conditions put it at a disadvantage in competing against capitalist firms.

Suggested Readings

Eileen Appelbaum, Annette Bernhardt, and Richard J. Murnane, eds., *Low-Wage America: How Employers Are Reshaping Opportunity in the Workplace* (New York: Russell Sage Foundation, 2003).

Eileen Appelbaum and Rosemary Batt, "A Primer on Private Equity at Work: Management, Employment, and Sustainability" (Center for Economic and Policy Research, February 2012 available at http://cepr.net/documents/publications/private-equity-2012–02.pdf.

William J. Baumol, Alan S. Blinder, and Edward N. Wolff, *Downsizing in America: Reality, Causes, and Consequences* (New York: Russell Sage Foundation, 2003).

Alan Blinder, *Paying for Productivity* (Washington, DC: The Brookings Institution, 1990).

Edna Bonacich, "Advanced Capitalism and Black/White Race Relations in the United States: A Split Labor Market Interpretation," *American Sociological Review* 41, no. 1 (February 1976): 34–51, available at http://www.radford.edu/~junnever/articles/blackwhite.htm.

Edna Bonacich, "A Theory of Ethnic Antagonism: The Split Labor Market," *American Sociological Review* 37, no. 5 (1972): 547–59.

Harry Braverman, *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century* (New York: Monthly Review Press, 1998).

Jessica Gordon Nembhard, *Collective Courage: A History of African American Cooperative Economic Thought and Practice* (University Park, PA: Penn State University Press, 2014).

Henry Hansmann, *The Ownership of Enterprise* (Cambridge, MA: Harvard University Press, 1996).

Simon Head, *The New Ruthless Economy: Work and Power in the Digital Age* (New York: Century Foundation/Oxford University Press, 2003).

David Noble, *The Forces of Production* (New York: Knopf, 1984).

Guy Standing, *The Precariat* (New York: Bloomsbury Press, 2011).

William F. Whyte and Kathleen K. Whyte, *Making Mondragon* (Ithaca, NY: ILR Press, 1988).

The Mosaic of Inequality

14

In 1998 the Gallup public opinion research firm asked 5,001 Americans why "some people get ahead and succeed in life and some do not." Respondents were asked to consider twelve possible answers and to rank the answers from "extremely important" to "not at all important." The twelve possible answers were "good luck," "hard work," "inherited money," "connections," "education," "dishonesty," "parents and the family environment," "the talent one is born with," "willingness to take risks," "good looks," "one's race," and "being male or female." There was general agreement among the respondents that education, hard work, one's family environment, connections, and luck were important.

But there were also significant differences. Those who themselves had higher incomes thought that hard work was more important, while those with lower incomes thought that luck, connections, inheritances, dishonesty, and one's sex were more important. Far more than white men did, both women and African Americans thought that one's sex and one's race counted as significant factors in getting ahead. Lower income respondents as well as women and African Americans all ranked education higher than better-off respondents, men, and whites did. Clearly, there are different opinions as to why some people get ahead while others do not.

Interest in what it takes to get ahead has been growing because those who *are* ahead have been gaining ground over the rest. Indeed, there is concern about the fact that although the playing field has never been exactly level, it is increasingly tilted against the less well off. Even the president of the New York Federal Reserve Bank, William McDonough, warned in his commencement address to the 2003 graduating class at Johns Hopkins University School of Advanced International Studies that rampant inequality would tear at the social fabric and was "unsustainable in a democracy."

Is it *fair* that being "good looking" makes it probable that you will have a higher income than the rest of the people who are otherwise similar to you? (It does, by the way, for men as well as for women, and even for those in jobs that do not require a person to be "on display.") If you think this is acceptable, how do you feel about the fact (also true) that obese women earn less and that short men also earn less? Most people think it is a *good* thing that hard work and education pay off. But what about race, sex, or one's parents' wealth? The fact that these things do help one get ahead

(this *is* a fact) strikes many people as unfair. And if a high-quality education is a way to move up the economic ladder (it is), many people think it unfair that educational opportunities are more available to those with well-to-do parents.

Some people have the free time and the income that allow them to make real choices about such things as where to live and what interests to pursue. Others lack either the time, the income, or both. Racial insults, sexual harassment, and hurtful indignities are experienced by some, but not by others. Some lawyers charge well over $1,000 an hour for their services, while kitchen staff at restaurants (working just as hard under less pleasant conditions) typically get less than 1 percent of that amount.

Why do we call some differences "inequalities," find them unacceptable, and advocate policies to eliminate them? And why do we, at the same time, regard other differences as innocuous or possibly even good because they make for "diversity"? The simple answer is easy: unacceptable inequalities are those that are unfair. But deciding *what* is unfair is sometimes difficult.

Deciding what is unfair often requires knowing how differences come about. If a lawyer's high pay is the result of his hard work in school, while the kitchen worker is low paid because she is lazy, the pay difference would seem more acceptable than if the pay difference is the result of racial discrimination or has to do with the fact that the lawyer is a man and the kitchen worker a woman. The key to fairness here is *equality of opportunity*, as discussed in Chapter 3.

Further difficulties arise if we turn from the "inequality of what" issue to ask: inequality *between whom*? Is it fair that the minimum wage kitchen worker in the United States makes five times more in an hour than does the agricultural worker in some other country who tended the crops to produce the food that she is now preparing in the kitchen? Why do we worry about the high-paid lawyer and not about the much lower-paid picker in Mexico or South Africa?

In the previous two chapters we explained how two classes, the capitalist class and the working class, interact in labor markets and firms. The capitalist class is defined by its ownership and control of the capital goods used in production and its power to dispose of the resulting surplus product. The working class is defined by its lack of such ownership and control rights. But when viewed from the standpoint of the economy as a whole, this picture is incomplete. There are large numbers of management personnel who lack substantial wealth but nevertheless have control over the labor of others. As explained in Chapter 7, these managers constitute the new middle class. At the same time there are people who still can be thought of as being in the old middle class because they own the capital goods they need to carry on their work but neither have a boss nor are one.

The class structure, however, is not a set of cubbyholes into which the accumulation process sorts people neatly labeled into four homogeneous types: worker, capitalist, new, and old middle class. Rather, there is a continuum of inequalities of many dimensions: of ownership, of income, and of power, all overlaid with differences between men and women and among races and ethnic groups. Moreover, there are quite a few people who live outside the class system outlined above. They include those who depend on government *transfer payments* such as government social security benefits, unemployment benefits, and assistance to veterans, as well as those who depend on charity, those residing in institutions (like hospitals and prisons), and many of those living by crime.

There are also major differences within the four classes considered in this book. Among employers, there are the owners of the largest firms, employing hundreds of thousands, but there are also farmers, architects, and store owners with just a few

employees. The differences among employees are equally great. The people in the approximately 2 percent of the U.S. labor force who are paid at or below the federal minimum wage eke out livelihoods that, depending on whether they work full time all year or not, and how many dependents they have, can place them below the poverty line. (Those receiving a higher state or city minimum wage above the federal minimum wage are only a little better off.) They are more likely to fall below the poverty line if they are not employed full time all year. In contrast, a few members of the working class make more in a month than minimum wage workers do in a year.

In this chapter we discuss inequalities of income and wealth, not of health, happiness, or other things one might desire. We do this because information on income and wealth is especially detailed and comprehensive, and because the wealth and income data help us to see some inequalities that matter. Having more income and wealth affords more access to goods and services, more personal independence, and more opportunities to attain things like health and happiness. Those with less income and wealth tend to have less of all these things.

This brings us to the title of this chapter. A *mosaic* is an ancient art form in which a picture—for example, an image of a person or animal—is constructed by assembling small, differently colored pieces of tile. Viewed from a distance, the image is clearly recognizable, but up close what you see are only the pieces. Similarly, the many facets of inequality—race, wealth, sex, schooling, and so on—constitute a kind of "mosaic." In this chapter we examine not only the more significant pieces but also the larger picture.

The main ideas of the chapter are: (a) *among the determinants of economic success in the United States, one's race, sex, and parental income are very important,* and (b) *by almost any measure, income inequality rose dramatically in the nearly half-century since the early 1970s and remains high.*

These main ideas are expressed in the following five points:

1. Living standards are not simply a matter of material goods. People's well-being depends on their health, their material comfort, and their access to social and natural environments that contribute to their whole human development. The economy contributes to people's well-being when it provides the goods and services necessary to meet these objectives and when it shapes the kinds of social and natural environments essential to people's well-being.

2. Inequality of both income and wealth increased sharply during the nearly half-century from the early 1970s on, and remains high.

3. The children of families with high incomes are much more likely to have high incomes when they grow up, and the children of poor families are much more likely to have low incomes. The higher levels of schooling (and better quality of schooling, year for year) enjoyed by the children of well-off parents account for some of these differences, but only some.

4. African Americans and women continue to earn less, on average, than men and people of European descent. The significant income gains made by African Americans relative to European Americans from 1939 to 1979 have not continued to the present.

5. Jobs in the U.S. economy remain highly segmented: "women's work" tends to be less well paid, but even in the same jobs, women earn less than men.

Well-Being and Inequality

We often make statements such as "The Hernandez family is better off than the Jones family," "People live better in Sweden than in Mexico," or "My living standard is much higher now than when I was just out of college." What do we mean by *better off*, *live better*, and *living standard*? These terms refer to all the things that influence a person's well-being.

Influences on Well-Being and the Economy

A major influence on well-being is access to food, shelter, clothing, health care, and other necessities of life. Also important is access to the amenities and luxuries that make it possible to feel that we are living well, or at least not worse than other people around us. Moreover, a person's sense of well-being depends not only on having the respect of others but also on having a sense of belonging to a community, whether it be a family, a neighborhood, a work group, a religious group, or a nation. Without this sense of belonging, life can become meaningless.

Living standards also depend on having enough free time and enough energy left after finishing one's work to enjoy life. The workaholic who makes $120,000 a year but has no free time may not be better off than a person with plenty of free time who earns $40,000 a year. No less important is the ability to make important choices regarding, say, education and other means to the achievement of one's goals in life. The quality of our work experience is also an important influence on our well-being: few things can bring a person down as much as hating to go to work every morning. An additional influence on our well-being is the quality of our natural environment and the extent to which it allows us to experience good health and to enjoy the many pleasures that depend on our natural surroundings. How each of us regards the many influences on our well-being will, of course, differ according to our values. But, however we rank them, the basic components of well-being have to include not only material goods but also such things as health, freedom, respect, and belonging. Thus, living standards depend both on tangible things and on the intangibles that help to determine the quality of life. In Figure 14.1 we sketch some of the factors that can influence a person's well-being.

Obviously our well-being depends critically on the economy. It is through the labor processes that make up the economy that we get the food, clothing, shelter, amenities, and luxuries that make life possible and enjoyable. Less obvious, but no less important, is that the way the economy is organized influences the quality-of-life aspects of well-being: health, freedom, respect, and belonging. This is true for a number of reasons.

First, the organization of the economy affects the health status of the population. Some societies provide adequate health care to all without regard to ability to pay; in other societies adequate health care is available only to those who can pay for it. In any case, there are economies in which workers must work at such a pace that stress is a major health problem; in others the pace of work is more worker-friendly. In a laissez-faire capitalist economy (one allowing the unrestricted pursuit of profits by companies without government regulation) environmental pollution may impair the health of the population.

Second, the structure of the economy also influences how free we are to make or to influence the major decisions affecting our lives. A person's freedom may be curtailed by an economic system—such as that which existed in the former Soviet Union—that determines where a person will work or restricts what a person can do

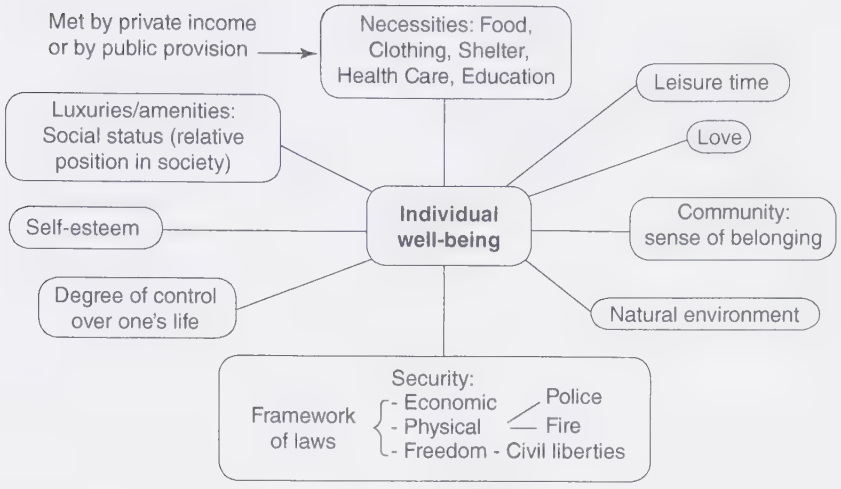


FIGURE 14.1 The determinants of well-being.

Without attempting to be comprehensive or to suggest the relative importance of each determinant, this figure displays in graphic form some of the requirements for well-being in any society.

with his or her property. A person's freedom can also be limited by lack of income, denying any real choices other than those focused on making ends meet.

Third, though the respect we feel for one another can be attributable to many different accomplishments and characteristics, if one is much poorer than others, even if one has access to the necessities of life, it can be difficult to gain respect. For example, a woman who has two children might earn $18,000 a year. In the United States she would be below the poverty line, yet the same woman earning the equivalent of the same income in Bangladesh and supporting two children would be considered reasonably successful (she would be earning twice the average income of a three-person family like her own in that country). The point is not that she could buy more if she were in Bangladesh (in fact, she would find that her dollars would buy more of some things in Bangladesh than in the United States, but less of others) but that she would *feel* better off, because she would compare herself with others less well off, and others would make the same comparison. One's sense of well-being thus depends at least in part on one's income compared to the income of others.

Fourth, closely related to the sense of respect is the sense of belonging to some group. Some economies are organized around very long-lasting neighborhoods and family units. In other economies people move frequently from one place to another in search of or because of work, with the result that neighborhoods are often made up of people who do not even know one another. Sometimes the workplace itself feels like a large community in which one cares about fellow workers and they reciprocate. In other workplaces people may be simply passing acquaintances or perhaps even hostile competitors.

Measuring Living Standards and Inequality

Measuring living standards and comparing them among people, among national averages, or between different time periods is very difficult. Many of the influences on the quality of life are hard to measure even if we could agree on how important each is.

Does a gallon of milk contribute more or less to well-being than, say, a pound of shrimp? The standard economist's answer is that the shrimp is worth more because people are willing to pay $15 for the pound of shrimp but can get a gallon of milk for less than $5.

The most common approach in economics is to measure a person's standard of living by his or her income and to use separate measures to indicate health status, income relative to others, the quality of the natural and social environment, and the like. Similarly, the standard of living in a country is often measured by the total income in the country divided by total population, that is, the per capita income. *Economic growth* is defined as an increase in the per capita income of a country. As an approximation of access to goods and services, measuring by income makes sense. But there are a number of problems with using income as a measure of well-being.

First, by measuring income and not measuring leisure, we fail to take account of one of the main determinants of well-being, free time. By the income standard the workaholic mentioned earlier is better off than someone who works half as hard and makes just a little less money. But by most people's standards the extra sleep, the free time to enjoy one's friends and family, and other such pleasures would make the leisured person with the lower income better off. In Sweden, for example, *income* per employed person increased only a little faster than that in the United States over the last half century (see Figure 15.6 in the next chapter). But the *leisure time* of the average employed Swede grew much more rapidly, as work hours in Sweden were cut back by the equivalent of two months of work time, while the reduction in work hours in the United States was less than half of this. Thus we can see that the income measure by itself understates the Swedes' improvement in well-being relative to that of Americans during this period.

Second, income measures a person's access to commodities, but many important goods and services are not commodities. Examples are home-cooked meals and all other products of household labor (house cleaning, care for one's children, and the like). Other examples of goods and services that are not commodities are public education, police protection, and other government services. Moving to a town with better schools and better police protection undoubtedly is an improvement in one's living standard, even defining this narrowly to mean simply access to goods and services. Yet this move need not be reflected in any change in measured income.

Third, the prices of goods and services often fail to measure their contributions to well-being. For example, a quarter of a pound of shrimp and a gallon of milk might cost the same, but most people would agree that the milk is in some sense more essential to well-being than the shrimp. The reason is that milk is a necessity, and shrimp is a luxury. To the person who pays $15 a pound for shrimp, doing without it would probably be no hardship. For most people, this is most likely not true of milk. Clearly, then, income is an inadequate measure of well-being because the prices that people pay for different goods—$15 for a pound of shrimp, $4 for a gallon of milk, and so on—may not properly value the contribution of each good to well-being.

Because the less well-off tend to spend more of their incomes on necessities and less on luxuries, a given amount of income is likely to contribute more to average well-being in society if more of the income goes to the less well-off. For this reason the average *standard of living* (or well-being) depends on *more than* the average amount of *income* at some point in time or in some country. We also want to know how evenly it is divided. Moreover, we are interested in the distribution of income for what it says about the degree to which economic outcomes are *fair* (see Chapter 3 on fairness).

Measuring inequality in the distribution of income, like measuring income itself, has its difficulties. A commonly used approach is to rank all households or

individuals by their income from the poorest to the richest and then divide them into fifths—or "quintiles"—of the population. We then calculate what percent of total income each fifth of the population receives. According to the most recent data from the U.S. Census Bureau, in 2014 the bottom fifth of the population received 3.1 percent of the nation's total income. The next fifth got 8.2 percent, so the poorest 40 percent received 11.3 percent of the total income, and so on up to the top fifth, which received 51.2 percent of all the income. The richest 5 percent of the population received 22.4 percent of all the income. (The source is cited in the caption for Figure 14.4.)

Growing Inequality

The recent trend toward greater income inequality is especially striking when compared with the previous trend toward greater equality. Figure 14.2 shows the fraction of total income received by the richest 1 percent of the U.S. population over the last century. The decline in the income share of the very rich from before

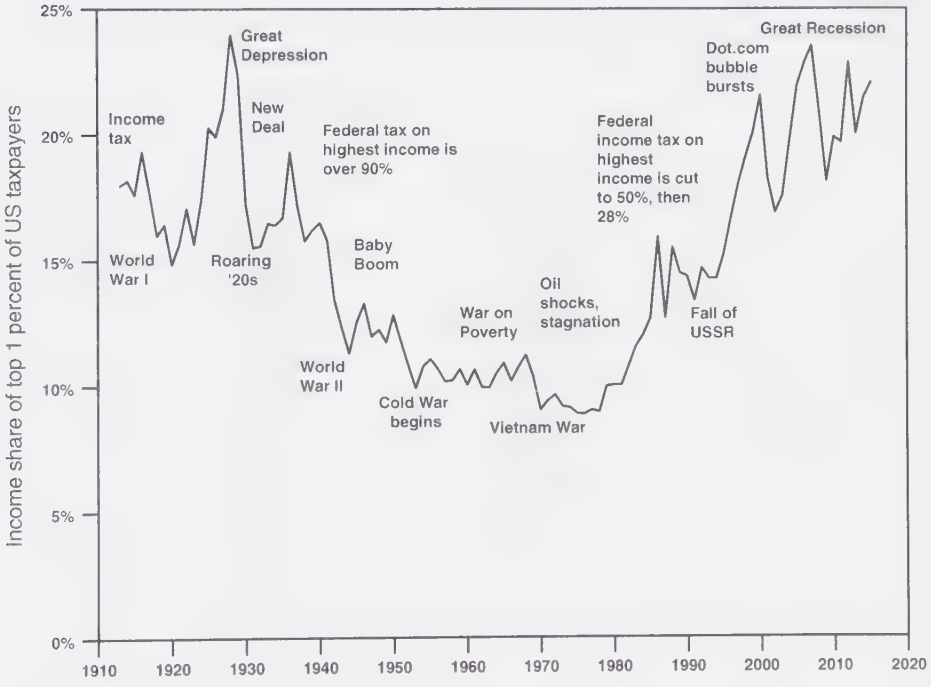


FIGURE 14.2 The lucky few: Income share of the top 1 percent of U.S. taxpayers, 1913–2015.

Share is calculated by dividing the total income of the top 1 percent of taxpayers, including capital gains, by the total income of all income earners, including capital gains. Data from the same source are also available to calculate the share if capital gains are excluded; however, the difference is not large.

Source: Data from Emmanuel Saez, "Income Inequality in the United States, 1913–1998," with Thomas Piketty, *Quarterly Journal of Economics*, 118, no. 1 (2003): 1–39 (Tables and Figures Updated to 2015 in Excel format, June 2016), Table A3: Top fractiles income shares (including capital gains) in the United States, available from https://eml.berkeley.edu/~saez/.

World War I continued with only minor reverses until the 1970s. The 1920s were especially "roaring" if you were very rich, but in the subsequent decades the top income recipients' piece of the pie shrank considerably—probably in large part because the marginal federal income tax rate on the very highest incomes was 63 percent or higher from 1932 to 1981, and was 91 percent or higher from 1944 to 1963. (This tax rate was not on the whole incomes of the very rich, but only on the income above a rather high level.) The fact that the top 1 percent received about 9 percent of the total income in 1970 means that the typical person in this group had an income about 10 times that of the average person's—so the rich person's slice was not all that skimpy. The top 1 percent's share fell from a high of 25 percent in the late 1920s to a low of 9 percent in the late 1970s. It has been rising since President Ronald Reagan took office in 1981 and began cutting taxes on high incomes and deregulating the economy.

Among "the lucky few," of course, are CEOs of large corporations. Figure 14.3 shows changes in the ratio of CEOs' compensation in the United States to an estimate of the average worker's wage in their firms, from 1965 to 2015. The figure shows an increase in this ratio from 20:1 in 1965, to 59:1 in 1989, to 299:1 in 2015.

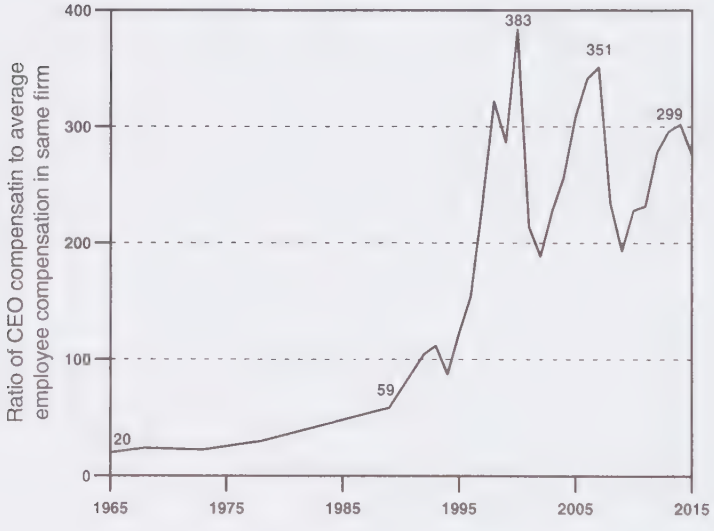


FIGURE 14.3 U.S. CEO pay in relation to the average worker's pay, 1965–2015.

For each corporate CEO, researchers used U.S. Bureau of Labor Statistics and Bureau of Economic Analysis data to estimate the total compensation (wages plus benefits) received by a typical production/nonsupervisory worker in the key industry in which the corporation operated. The CEO compensation was divided by this average worker's estimated total compensation to get the ratio for that corporation. Then for each year, these ratios for specific firms were averaged to obtain the ratio shown.

Source: Lawrence Mishel and Jessica Schieder, "CEO pay remains high relative to the pay of typical workers and high-wage earners," July 26, 2017, available at http://www.epi.org/130554. References to methodology are cited in the paper.

Figure 14.4 presents data on the recent upturn in income inequality in a different way. In this figure one can see that in the first year of the last third of the twentieth century, 1967, the top 20 percent (or top "quintile") of U.S. households received 43.8 percent of the nation's income, while the bottom quintile received 4.0 percent. By 2014, nearly half a century later, the top quintile's share had increased to 51.2 percent, while the share of the bottom quintile had shrunk to 3.1 percent. Also apparent is the fact that the shares received by all the quintiles other than the very top one fell. If "middle income" is defined as the 2nd, 3rd, and 4th quintiles, the combined share of middle-income households shrank from 52.3 to 45.7 percent of total income between 1967 and 2014. Households in the quintile exactly in the middle, the 3rd quintile, saw their income shrink from 17.3 to 14.3 percent of the total.

Figure 14.5 shows that from 1966 to 1979 all family incomes in the various quintiles grew at similar rates, but after 1979 growth rates of income diverged, with the richest enjoying much faster rates of income growth than the poorest. What the figure shows most starkly is that it is not inevitable that inequality increases; in fact, other sources show that there was essentially no increase in income inequality during the period 1947 to 1973, although Figure 14.5 only shows data from 1966 on. High income tax rates on the highest incomes likely were key in preventing inequality from rising during that period. That highest rate was reduced in 1963 to 77 percent, in 1965 to 70 percent, in 1982 to 50 percent, and later even further.

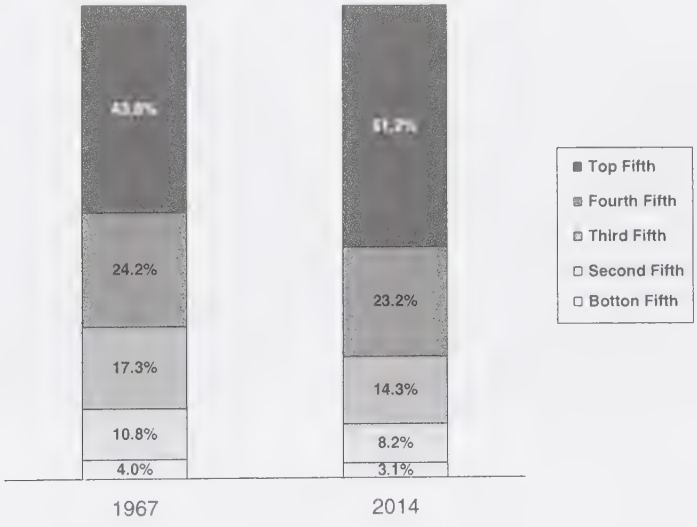


FIGURE 14.4 Income shares by quintile in the U.S., 1967 and 2014.

This figure shows increasing income inequality among U.S. households from 1967 to 2014. For each year, households were ranked by income and then divided into five equal groups (quintiles). In 2014, for example, the poorest quintile received 3.1 percent of total income (less than in 1967), while the richest quintile received 51.2 percent of all income, considerably more than in 1967.

Source: U.S. Census Bureau, *Historical Income Tables—Households. Table H-2:* Share of Aggregate Income Received by Each Fifth and Top 5 Percent of Households, All Races: 1967 to 2014, available at http://www.census.gov/hhes/www/income/data/historical/household/.

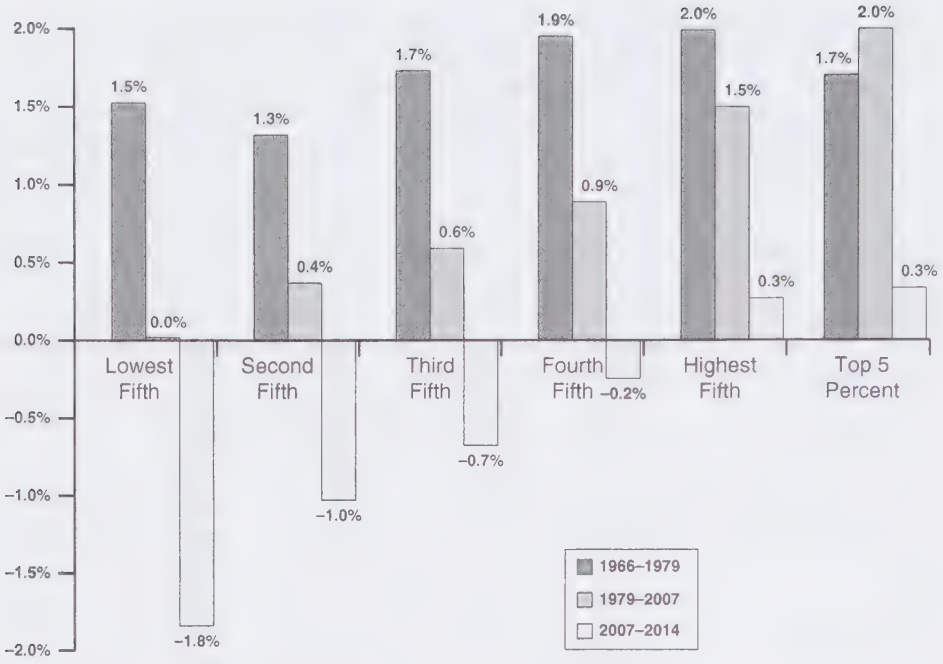


FIGURE 14.5 Growth rate of family income, by quintile, 1966–2014.

This figure shows how rapidly the average family income in each fifth (quintile) of all families grew during three distinct time periods between 1966 and 2014. Although each quintile's income grew at similar rates before 1979, growth rates diverged after that, so that inequality grew, leading to the dramatic changes shown in Figure 14.4.

Source: U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement, Table F-3: Mean Income Received by Each Fifth and Top 5 Percent of Families, All Races: 1966 to 2014, available at http://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-families.html.

Wealth Inequality

Wealth (or **net worth**) refers to the total value of a person's assets (what the person owns) minus the total value of the person's outstanding debts, or **liabilities** (what the person owes).

Income is what a person (or household, or family) receives during a certain period of time (such as a week, a month, or a year). It is a *flow* and is measured per unit of time ("I make $3,000 *per month*"). In contrast, **wealth**, or **net worth**, is measured at a point in time, and is what you own (*assets*) minus what you owe (*liabilities*, or debts):

$$\textit{Net worth} = \textit{Assets} - \textit{Liabilities}$$

So one could say, "I own assets worth $120,000 and I have total debt of $70,000, so my *net worth* is $120,000 – $70,000 = $50,000." *Wealth* is a *stock*, not a flow. As income flows into your hands, your wealth increases—until you spend it. Think of a stream emptying into a pond: The amount of water passing by someone on the bank of the stream per unit of time is a *flow* and the amount of water in the pond is a *stock*.

What kinds of assets are we talking about owning? There are two main types. The first type is *nonfinancial assets* like land, homes or other buildings (including rental properties), cars, furniture, gems, racehorses, valuable paintings, yachts, and such. The second type is *financial assets* like stocks, bonds, other securities, unincorporated businesses, and intellectual property rights (patents, copyrights,

trademarks). Economists sometimes consider one's skills and state of health to be a kind of wealth called *human capital*, but this comes up mainly in studies of education and its effects; here we always use *wealth* to refer to nonhuman capital only.

Wealth of many kinds generates either a flow of income, or else a flow of services that could be sold to get income. Owning such things as a house and a car yields housing and transportation services that, in their absence, one would have to purchase by renting a dwelling, hiring a taxi, or paying to ride a bus. Ownership of financial assets like bonds often yields income to the owner. Ownership of assets whose value rises at least as fast as inflation (such as land, or stocks during some periods) can be a fairly secure way to hold one's wealth. Ownership of assets can afford economic security, for the owner of a substantial amount of wealth can sell some assets to gain income needed for an emergency or to tide him or her over in bad times.

Figure 14.6 shows that the wealthiest typically hold a different combination of assets than the less wealthy. In 2010, fully 75.9 percent of the wealth of the wealthiest 1 percent of households in the United States was invested in corporate stock or in unincorporated business equity, while only 9.4 percent of it was tied up in residence ownership. But in the same year, households in the middle three-fifths (not the top or the bottom fifth) of the wealth distribution held most of their wealth in the form of ownership of their own residences, with corporate stocks and ownership of businesses amounting on average to only 12.0 percent of their total wealth.

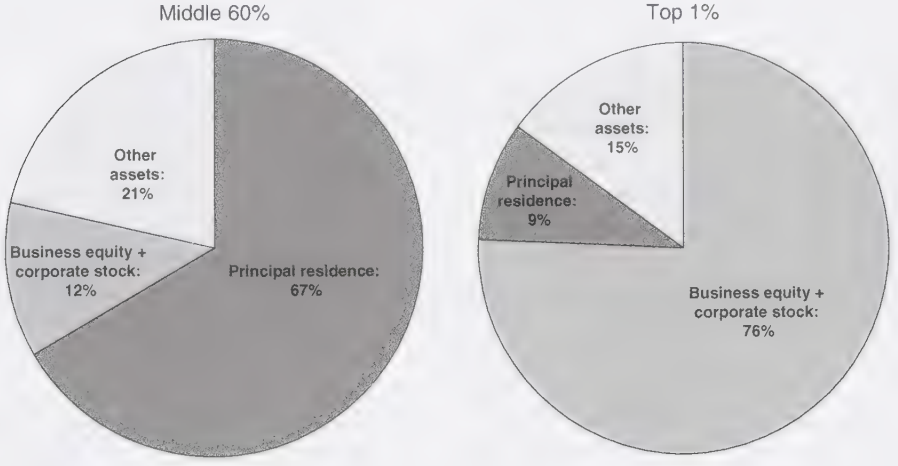


FIGURE 14.6 Composition of wealth holdings at different levels of wealth distribution in the U.S. in 2010.

This figure shows how the middle three-fifths of the wealth distribution as a group held its wealth in 2010 (left pie chart) as compared with how wealth was held by the top 1 percent of the wealth distribution in the same year (right pie chart). The three categories of wealth used in this figure are (1) ownership of a residence; (2) ownership of business equity, corporate stock, financial securities, mutual funds, and personal trusts; and (3) all other kinds of assets, including the value of pension funds, liquid assets (bank deposits, money market funds, and cash surrender value of life insurance), and personal possessions such as automobiles, household furniture, and personal items.

Source: Data from Edward N. Wolff, "The Asset Price Meltdown and the Wealth of the Middle Class," National Bureau of Economics Research (NBER) Working Paper 18559 (2012), Table 5: Composition of Household Wealth by Wealth Class, 2010, available at www.nber.org/papers/w18559.

The composition of the wealth of the rich and the not-so-rich is of considerable economic and social significance. This is due to the obvious fact that owning a home does not, by itself, enable one to be an employer of others or to work for oneself. To be an employer or to be self-employed, one needs to own assets other than one's home. An employer with a small shop employing, say, ten people would probably need to own at least a quarter of a million dollars in assets. This might be enough to buy the equipment needed to employ the workers; no less important, it would increase the employer's chances of being able to borrow additional funds from a bank or some other lending source.

Thus, while home ownership can contribute to one's personal autonomy, it is the ownership of other assets, in particular, ownership of capital goods used in production, that allows for true economic autonomy—the freedom to work for oneself, to employ others, or to choose not to work at all. One way in which a person can establish ownership of capital goods is to own stock issued by corporations and traded in stock markets. Such stock can provide not only income but also, if a sufficient amount of stock is owned, control of a corporation. As explained in Chapters 5 and 10, income from owning corporate stock takes the form of dividends, and corporations can decide whether to pay dividends. If they are paid, the corporation can determine how much will be paid per share of stock. Figure 5.2 shows that corporate stock is mainly owned by the rich: In 2011, the wealthiest 10 percent of U.S. households owned more than 80 percent of the total, while the least wealthy 90 percent of households owned less than 20 percent.

The distribution of non-home assets in the United States in 2011 (that is, counting all assets except ownership of a residence, and all liabilities except a home mortgage loan) is shown in Figure 14.7. It shows that a U.S. household is likely to have more non-home net assets the greater the educational attainment of the householder, and the higher the income of the household. In addition, white households and male households are likely to have more non-home assets.

In Figure 14.7, the lower right panel shows that U.S. households with more income tend to have more wealth. This is just what one would expect, since wealth usually provides income, and income, if saved, adds to wealth. However, what is remarkable is the *size* of the difference in asset ownership between the richest fifth of households and other ones: In 2011 the average non-home net worth of the richest fifth was *21 times* the average non-home net worth of the lowest three-fifths of households combined. Similar disparities were true with respect to educational attainment: College graduates had about eight times the non-home net worth of high school graduates. Also striking were the differences by race and ethnicity. In 2011, white households had over fifteen times the non-home net worth of black households, and over eight times the non-home net worth of Hispanic households. A household consisting of a married couple and possibly others had average non-home net worth over five times that of a single male householder (with or without other family members), and over eight times that of a single female householder (possibly with other family members such as children).

A common way of measuring the distribution of wealth is to look at how much of a country's wealth is owned by different segments of its population. As Figure 14.8 shows, the share of the top 10 percent of households in the United States increased from 68.4 percent to 76.7 percent between 1969 and 2010. During the same period the fraction of total wealth owned by the poorest 60 percent of U.S. households fell from 6.4 percent to 1.7 percent. Overall wealth inequality, in other words, has been growing hand in hand with income inequality.

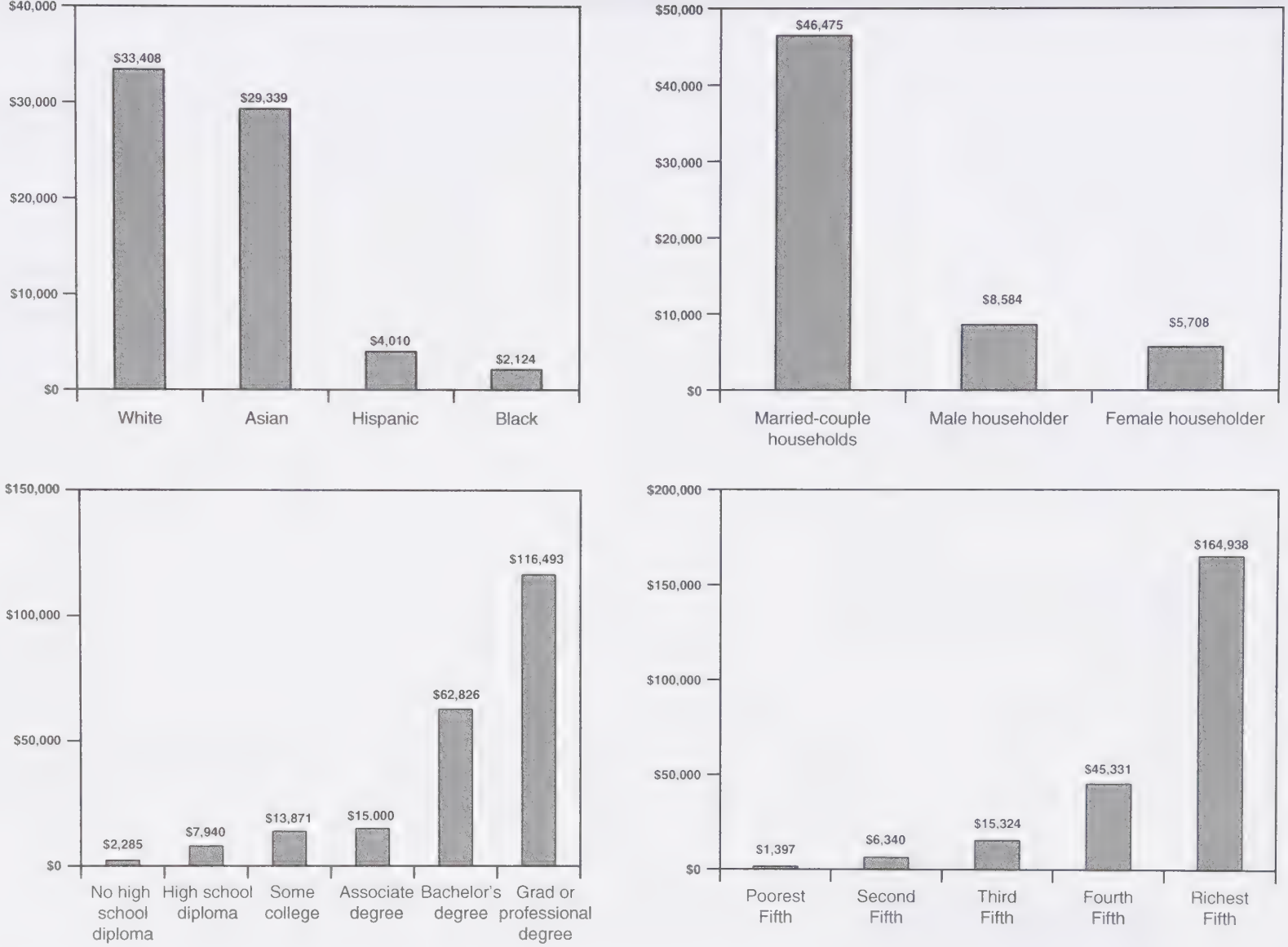


FIGURE 14.7 Distribution of median non-home net worth among U.S. households by race/ethnicity, gender, education, and income, 2011.

This figure shows how net worth, excluding ownership in a residence, was distributed among U.S. households in 2011. A household's *net worth* is the value of the assets the household owns minus the value of the debts it owes. The median is the household in the middle of a category of households that have been ranked by non-home net worth.

Source: U.S. Census Bureau, Wealth, Asset, & Debt of Households Detailed Tables: 2011, Wealth and Asset Ownership by Year, Table 1, Median Value of Assets for Households, by Type of Asset Owned and Selected Characteristics: 2011 (from Survey of Income and Program Participation, 2008 Panel, Wave 10), available at https://www.census.gov/data/tables/2011/demo/wealth/wealth-asset-ownership.html.

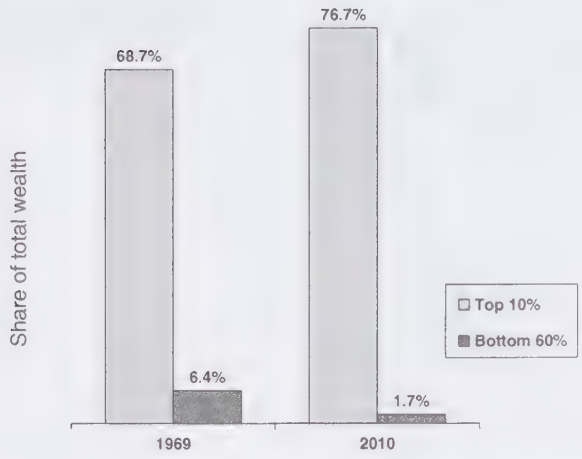


FIGURE 14.8 Increasing concentration of wealth in the U.S. between 1969 and 2010.

The ownership of wealth (net worth, or assets minus debts) among households in the U.S. became more concentrated between 1969 and 2010. The pair of columns on the left show that in 1969 the top 10 percent of American households owned 68.7 percent of the wealth in the U.S., while the bottom 60 percent of households owned 6.4 percent of it. The pair of columns on the right show that by 2010 the share owned by the top 10 percent of households had increased to 76.7 percent, while that of the bottom 60 percent of households had fallen to 1.7 percent.

Source: Data from Edward N. Wolff, "The Asset Price Meltdown and the Wealth of the Middle Class," National Bureau of Economics Research (NBER) Working Paper 18559 (2012), Table 2: The Size Distribution of Wealth and Income, 1962–2010. Available at http://www.nber.org/papers/w18559.

Unequal Chances

Those who responded to the Gallup poll that asked what it takes to get ahead in America ranked parents, good education, connections, and inherited money as a good thing to have if you are looking for economic success. They are not mistaken.

One of the enduring cultural ideals for Americans is that the United States is a "land of opportunity" where fortunes are won and lost from one generation to the next through some combination of ambition, sweat, and luck. The ideal originated during the nineteenth century, when the United States welcomed poor immigrants from the class-divided societies of Europe. Many of them found opportunities for land ownership, entrepreneurship, and schooling for their children on a scale that would have been unimaginable in the countries from which they came. The American dream means that how far you get in life is determined not by who your parents are but rather by your own abilities and by how hard you work.

Research relating the income of parents to the subsequent incomes of their grown children has shown, however, that having rich parents pays off. Figure 14.9 presents data from a study indicating, first, how likely it is that children of rich families in the United States will themselves end up being rich and, second, how likely it is that children of poor families will themselves end up being poor. The details of the study's results are given in the caption under the figure, but the most striking finding is that children whose parents are in the richest 10 percent of the U.S. income distribution have a better than 40 percent chance of ending up in the richest 20 percent of the population, while children of parents in the poorest 10 percent of the income

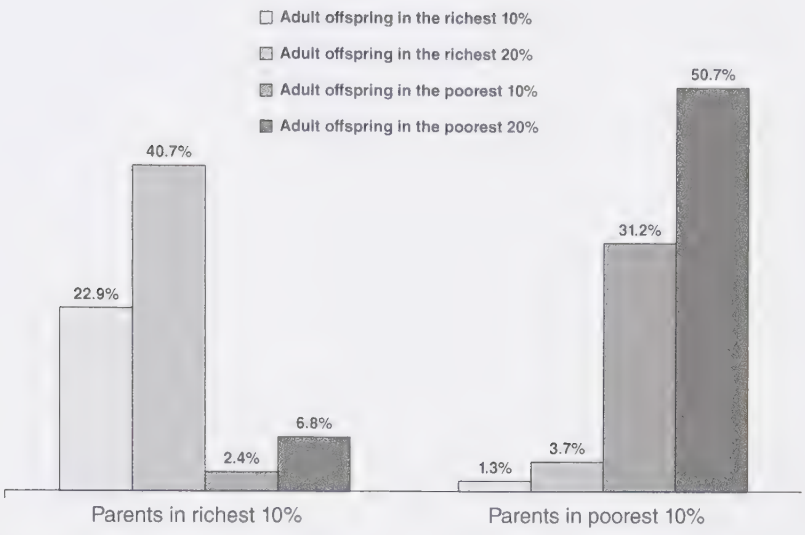


FIGURE 14.9 Unequal chances: Family background and economic success in the U.S.

What percent of the adult offspring of rich parents are poor? What percent of adult offspring of poor parents are rich? This figure uses data collected from a representative sample of individuals and families since 1968 by the University of Michigan Panel Study of Income Dynamics (PSID) to answer these questions.

The four columns grouped on the left side show what happens to the children of parents in the richest 10 percent of the U.S. population. Of these children, 40.7 percent end up, when they are adults, being in the richest 20 percent of the population, with only 2.4 percent ending up in the poorest 10 percent of the population. The four columns grouped on the right side of the figure show that of children who start off life in families in the poorest 10 percent of the population, 31.2 percent end up in the poorest 10 percent of the income distribution as adults; and only 3.7 percent of the children of poor families end up being in the richest 20 percent of the population.

Source: Data from Samuel Bowles and Herbert Gintis, "Intergenerational Inequality," *Journal of Economic Perspectives* 16, no. 3 (Summer 2002): 3–30.

distribution have a greater than 50 percent chance of ending up among the poorest 20 percent of the population.

Figure 14.9 also shows that among children whose parents are in the poorest 10 percent of the income distribution, only 1.3 percent end up as adults in the richest 10 percent of the income distribution. If the playing field were level in the sense that children's incomes were not affected by their parents' income, 10 percent of them would be in the richest 10 percent. Children from the poorest 10 percent have only a 3.7 percent chance of making it into the richest 20 percent of the income distribution. By contrast, among children whose parents are in the richest 10 percent, more than 20 percent (22.9 percent) will have incomes as adults placing them in the top 10 percent, while 40 percent of these offspring of the very rich will be in the top 20 percent of the income distribution. The figure also shows that the children of the rich are very unlikely to wind up poor, while more than half of the children of the poor wind up in the lower 20 percent of the income distribution.

Thus, children from the richest tenth of the population end up as adults in the top 10 percent of the income distribution twenty times more frequently than do children

from the poorest tenth of the population, while children from the poorest tenth of the population are fifteen times more likely than are children from the richest tenth to end up in the bottom 10 percent of the income distribution. While the cards are stacked heavily against the poor in the scramble for financial success, children of rich parents seldom experience downward mobility.

By comparison to Canada, Sweden, and many other nations on which similar research has been done, the United States is far from the "land of opportunity" that many have hoped it is. What accounts for the perpetuation of fortune and hardship from generation to generation? Two explanations are widely believed, but neither one is entirely adequate.

According to one explanation, the transmission of economic success across generations occurs because high-income parents pass on their wealth to their children (recall from Figure 14.7 that high-income families have substantially more wealth than do others). Lower-income parents lack wealth (Figure 14.7 again), so their children have to make do without a nest egg. This is true, and it explains why the grown up children of the very rich also tend to be rich. But it does not explain why the children of the somewhat rich are also very likely to be at least somewhat rich, at least by comparison with the children of the poor. Most people receive no significant inheritance beyond their parents' home. This, too, can be seen from Figure 14.7: it shows that the next-to-the-richest fifth of the population has wealth (other than a home) of only $45,000. Even this amount of wealth could well be used up covering the costs of health care and home care for an aging parent.

According to the second explanation, what counts is the "talent one is born with," in the words of the Gallup poll. It is not a nest egg that high-income parents pass on, but rather their high-income-earning genes. (Poor parents are thought to pass on their "inferior" genes). Of course, there is no such thing as a gene for high income, even though attributes such as intelligence, shrewdness, charisma, stamina, and so on may be, to some extent, passed on genetically, and each of these may have an influence on the income of both the parents and their grown children.

The most commonly suggested candidate for an income-earning trait that is passed on genetically is IQ, meaning how well one scores on an IQ test. Of course, the quality and quantity of schooling, family environment, and a host of other influences affect one's IQ, but nature as well as nurture has an influence. We know this because genetically identical twins are much more similar in IQ than are ordinary siblings (or nonidentical twins). But this explanation is even less valid than the inherited wealth account. The reason is that IQ is not a very important determinant of one's income: things such as the amount and quality of one's schooling and one's wealth are much more important.

What is it, then, that explains intergenerational inequality? The fact that children from higher-income families get more and higher-quality schooling is an important part of the story. It is also likely that successful parents teach their children, either deliberately or by example, the personality traits and behavioral patterns that contributed to their own success. Among these are such things as saving, valuing the future, ways of interacting socially with others, and believing that what one does makes a difference (the opposite of fatalism). Health is another channel: children of lower-income families frequently have health problems that often intensify in their adult years, and these bouts of illness affect incomes. Other important influences derive from the demographic and social groups to which one belongs. People whose parents live in a poor neighborhood or region are themselves likely to remain there,

and this perpetuates their low income. If they belong to a group that suffers discrimination, their children are very likely to belong to the same group. This is especially true when it comes to what is commonly called race.

Race and Inequality

Many Americans—especially white males—speak of racial discrimination in the past tense. There are indeed some selection processes for jobs, admission to educational institutions, and competitions for other valued resources in which it is a disadvantage to be white or male. But the well-publicized cases in which this is true are a misleading guide to what happens in general.



"Race": Biology or History?

"Races" do not exist in the sense that most people mean when they use the word to refer to groups of people differentiated by genetically perpetuated traits. What distinguishes people of African or East Asian or European ancestry when they are classified as "races" are physical markers such as skin color and facial characteristics. These do indeed differ markedly among these groups, and these traits are genetically transmitted.

However, a person who says "whites and blacks are different" usually has in mind something more than the obvious visible traits. They have in mind things such as culture, personality, average incomes, particular talents, and the like. But from a biological standpoint there are very few differences among groups of differing ancestry other than the superficial ones used to define the races. With respect to most of the genetic makeup of people, the members of a "race" are as different one from another as they are from members of a different "race." By a commonly accepted measure, well over 90 percent of genetic differences among people are *within* groups of similar ancestry, while less than 10 percent of the genetic differences are *between* groups.

Some genetic traits thought to be unique to a "race"—sickle cell anemia among people of African decent, for example—are, in fact, associated with particular climates. People of European ancestry from the island of Sardinia, for example, share high levels of sickle cell anemia with Africans. It has nothing to do with Africanness. It is found among people whose ancestors lived in places where malaria was common in the past, including not only Sardinia and West Africa, but parts of India, too.

What makes races distinctive, other than these physical markers, is history: over long periods of time people of different ancestries have lived under different conditions. In the case of African Americans this includes the experience of many of their ancestors having been brought to America in chains and exploited as slaves.

The conclusion we draw from this is not that race does not matter; unfortunately, it most certainly does. It is that race is not a biological fact. It is, rather, a historical outcome of how people of different ancestries have lived and have treated one another. That is why we do not consider tall people a race. Height, like skin color, has an important genetic component and is highly visible. But while the exploitation of the short by the tall may occur in the dating game and on the basketball court, it is not one of the main story lines of history.

Sources: L. Cavalli-Sforza, *Genes, Peoples and Languages* (Berkeley: University of California Press, 2000); Noah Rosenberg, Jonathan Pritchard, James Weber, Howard Cann, Kenneth Kidd, Lev A. Zhivotovsky, and Marcus Feldman, "Genetic Structure of Human Populations," *Science* 298 (2002): 2381–85; Marcus Feldman, R. C. Lewontin, and Mary-Claire King, "Race: A Genetic Melting Pot," *Nature* 424 (July 24, 2003): 374.

Discrimination in Hiring

During 2001 and 2002 an experiment carried out in Chicago and Boston by faculty members at the University of Chicago and the Massachusetts Institute of Technology showed that racial discrimination continues to exist in the labor market. The experimenters began by first downloading a number of resumes from the Internet and then changing some of them—adding experience, certification, and other qualifications, so that some resumes would be of higher quality than others; anything in them that might have identified a specific person was removed. Following this, "white-sounding" and "black-sounding" names were randomly assigned to the resumes. These names were obtained from historical birth records and were chosen according to the frequency with which they occurred in black and white households. The resumes were then sent out to 1,300 potential employers in the Boston and Chicago areas.

Each employer was sent four resumes: one high-quality one with a "white-sounding" name, one high-quality one with a "black-sounding" name, one low-quality one with a "white-sounding" name, and one low-quality one with a "black-sounding" name. It turned out, as shown in Figure 14.10, that one factor alone—whether a resume had a "white-sounding" or a "black-sounding" name on it—accounted for most of the differences in the frequency with which the businesses called back the people who had (supposedly) submitted the resumes, inviting them to come in for interviews. "White-sounding" names, whether male or female, were more likely to get called back for interviews than "black-sounding" names. A "Brad" was five times more likely to be called back for an interview than was a "Rasheed" with comparable qualifications, and a "Kristen" was six times more likely to be called back for an interview than was an "Aisha."

The study whose results are shown in Figure 14.10 was exceptionally well designed. But other studies of race discrimination have been conducted in which otherwise identical white and African-American car buyers, apartment seekers, and loan applicants have been treated differently. A disturbing aspect of the "race-sounding names" experiment is that while qualifications do matter when it comes to opportunities in the labor market, how much they matter is significantly affected by race. Figure 14.11 shows that high-quality resumes with "white-sounding" names are called back 30 percent more frequently than are good resumes with "black-sounding" names. Moreover, high-quality resumes with "black-sounding" names did not elicit many more callbacks than did low-quality resumes with "black-sounding" names. The improvement in the callback ratio due to the quality of the resume was so small that it could have occurred by chance.

Problem of Differential Pay

On top of discrimination in hiring, there is also the very persistent problem of differential pay. Figure 14.12 shows that differences in pay by race and sex have been discouragingly persistent over nearly half a century. Data also exist for earlier periods, and may indicate that there was progress from an earlier period of even greater pay gaps—but the data are not sufficiently comparable for us to be sure of this. That said, there are several historical events that should have narrowed the pay gap between black and white workers and between female and male workers. One was World War

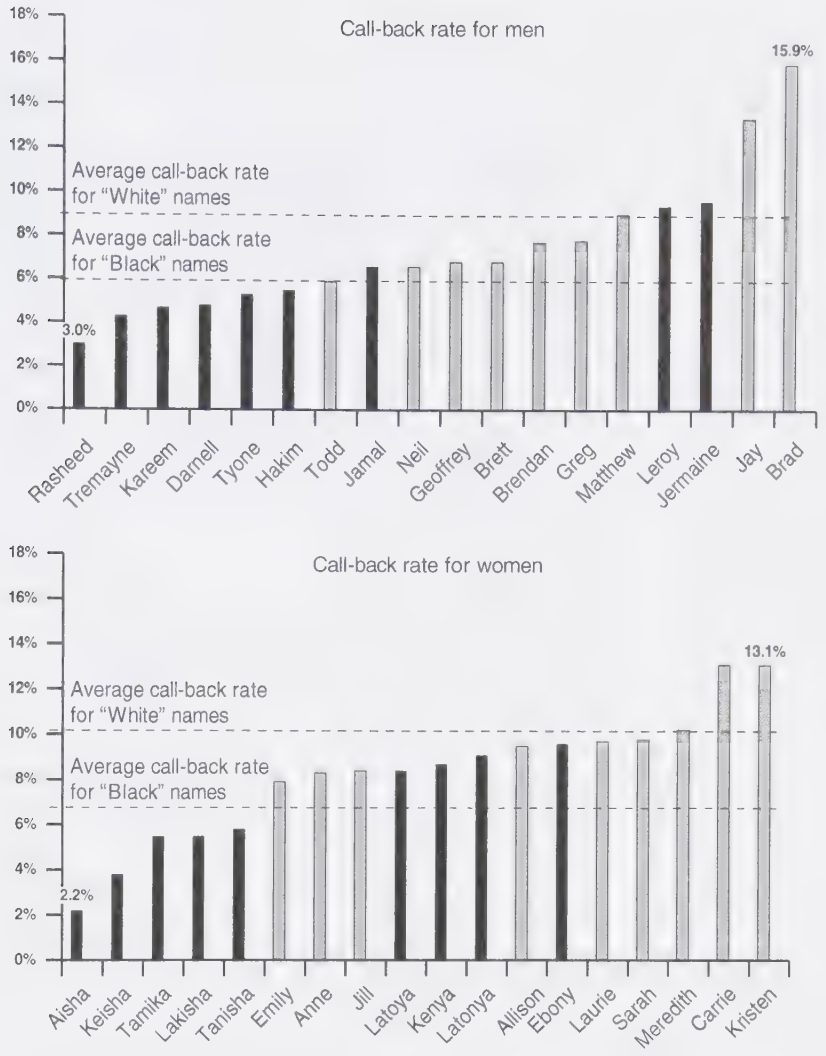


FIGURE 14.10 Racism by any other name: Labor market discrimination in the U.S., 2001–2002.

This figure displays the results of an experiment that was carried out in Chicago and Boston in 2001 and 2002. The two charts here show that for both men and women, a resume with a "white-sounding" name had a much better chance of eliciting a positive response than did a comparable resume with a "black-sounding" name on it. (Please see the text for a fuller description of this experiment.)

Source: Marianne Bertrand and Sendhil Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review* 94, no. 4 (2003): 991–1013. Available at http://www.jstor.org/stable/3592802.

II, when the war created a shortage of civilian labor, drawing both African Americans and women into the labor force and giving them new skills. Then, in 1964, the Civil Rights Act outlawed discrimination on the basis of race, color, religion, sex, and national origin.

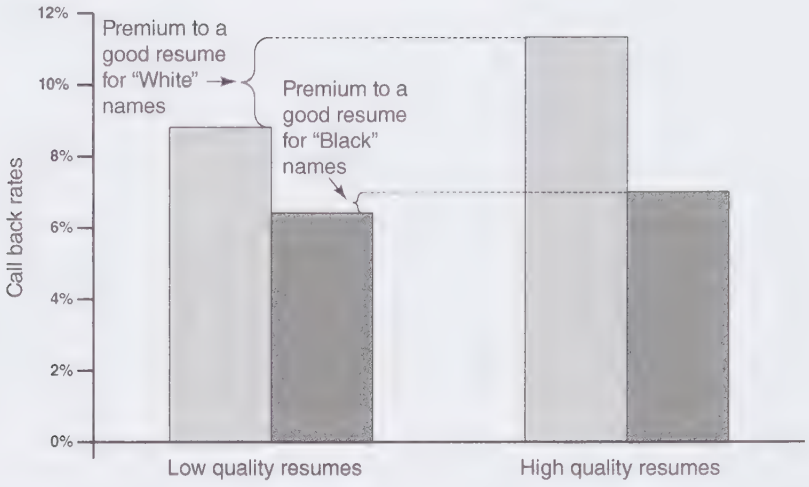


FIGURE 14.11 Good resumes pay off, if your name is OK: Labor market discrimination in the U.S., 2001–2002.

The Bertrand/Mullainathan experiment also showed that while qualifications affect one's opportunities in the labor market, such opportunities are significantly influenced by race. Resumes with "white-sounding" names had callback rates higher by 30 percent if the resumes were of higher quality. For resumes with a "black-sounding" name, having a good resume rather than a bad one did not result in a statistically significant improvement in callback rates.

Source: Marianne Bertrand and Sendhil Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic* Review 94, no. 4 (2003): 991–1013. Available at http://www.jstor.org/stable/3592802.

Figure 14.12 tracks progress in eliminating race and sex discrimination from 1971 on. Ratios shown in all four panels are based on median money income of people who worked full time throughout the year, so, for example, it does not reflect the fact that African Americans were and are more likely to be out of work than are whites. (*Earnings* refers to income from work, that is wages, salaries, and other compensation for work.) If progress for lower-paid African-American or women workers happened to take the form of moving from part-week or part-year employment to full-year employment, that progress would not be reflected in Figure 14.12.

The left-hand panels show that the white-black earnings gap for full-time, year-round workers was about the same in 2010–2012 as in 1970–1972 for both men and women, although there were temporary periods of improvement in between. Black women in those two periods earned 85 percent of what white women earned, while the ratio of black men's to white men's earnings was 68 percent in 1970–1972, and just slightly higher at 71 percent in 2010–2012.

In contrast, the right-hand panels show that the male-female earnings gap did narrow substantially between 1970 and 2012 among both African Americans and whites. Among black workers, women earned 72 percent of what men earned in 1970–1972, and this rose to 87 percent by 2010–2012; however, most of this change had already occurred by the early 1990s. Among white workers, women earned only 58 percent of what men earned in 1970–1972, and this rose to 75 percent by

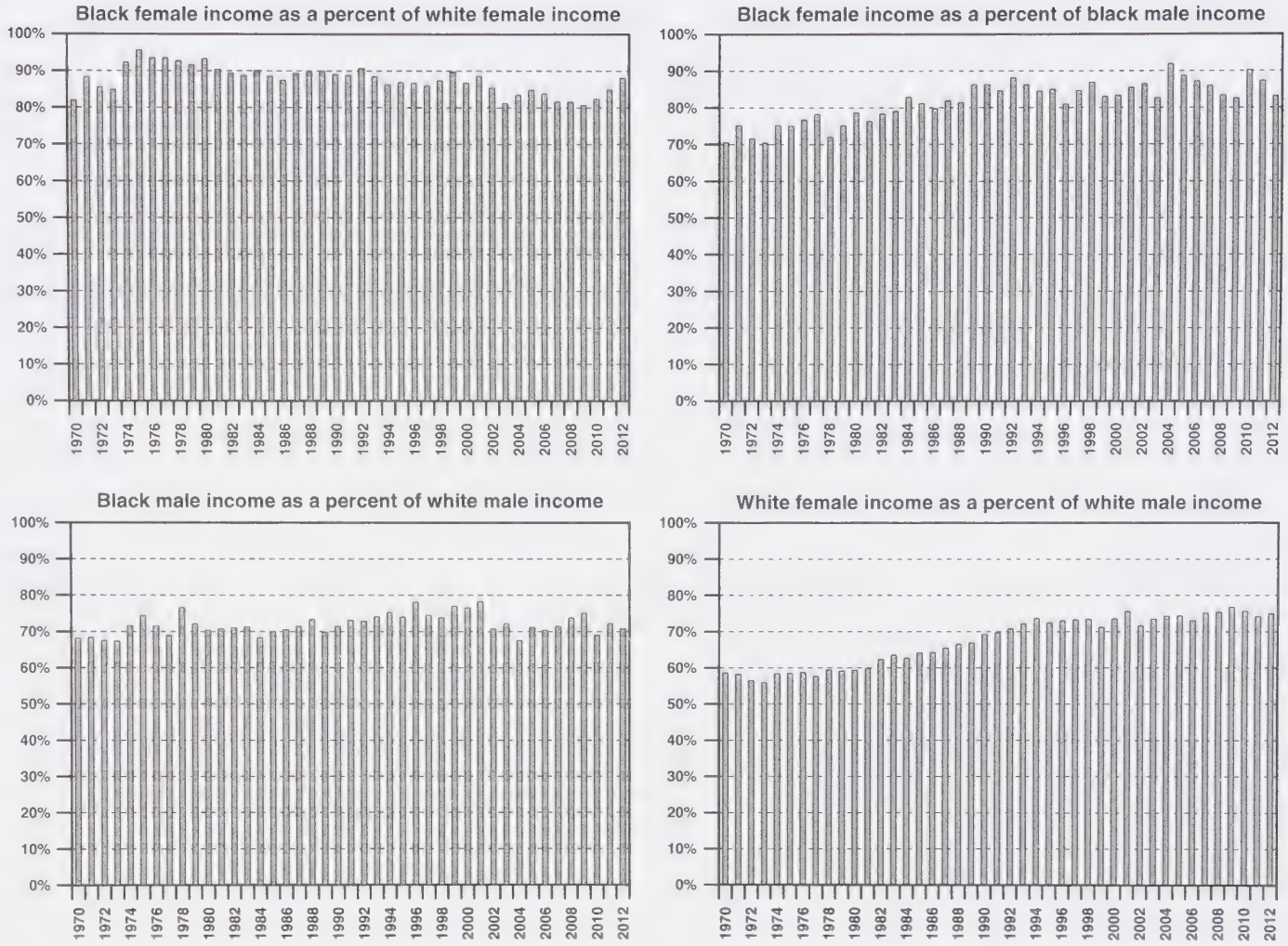


FIGURE 14.12 Slowed progress in the U.S. toward a color-blind and gender-neutral economy, 1971–2011.

This figure shows how the incomes of various groups changed in relation to those of other groups over four decades. All incomes on which ratios are based are the median annual incomes of full-time year-round employees. This helps to correct for any differences in the extent to which the different groups are engaged in full-time or part-time work. However, if the lower income group in any pair actually works on average fewer hours per year, because of working less than thirty-five hours a week or less than fifty weeks a year (or both), then the actual income gaps will be larger than those shown here. Comparable data were evidently not gathered before 1970.

Source: U.S. President's Council of Economic Advisors, *Economic Report of the President (ERP)*, various issues; for 1971: *ERP* 1993, p. 381: Table B-28, available at https://fraser.stlouisfed.org/docs/publications/ERP/1984/ERP_1993.pdf, and available at http://www.gpo.gov/fdsys/browse/collection.action?collectionCode=ERP&browsePath=2013&isCollapsed=false&leafLevelBrowse=false&isDocumentResults=true&ycord=0; for 1981 and 1991: *ERP 1997*, Table B-31, and for 2001: *ERP 2003*, Table B-33, and for 2011: *ERP 2014*, Table B-9: Median money income (in 2012 dollars) and poverty status of families and people, by race, 2003–2012, all available at https://www.gpo.gov/fdsys/browse/collection.action?collectionCode=ERP&browsePath=2014&isCollapsed=false&leafLevelBrowse=false&isDocumentResults=true&ycord=195

2010–2012, also a considerable increase. Again, however, almost all of this change had already taken place by 1994, and there was only minimal progress after that.

While racial discrimination in the labor market is part of the explanation for the racial earnings gaps documented in Figure 14.12, it is far from the entire story. Educational differences also matter. While the average number of years of schooling attained by white and African-American people are similar, the quality of schooling—as measured by expenditures and quality of teachers, for example—differs between the races. Finally, we have seen that having high-income parents contributes to having a higher income oneself, and few African Americans have high-income parents.

Women's Work, Women's Wages

Why do we see such persistent earnings gaps by gender? One reason women earn less than men is *job segregation*. Women tend to work in different kinds of jobs than men do, and women's jobs pay less than men's jobs, on average. Secretaries, elementary school teachers, and nurses, for instance, are usually women, whereas 99 percent of carpenters, 93 percent of mechanical engineers, and 91 percent of aircraft pilots and flight engineers were men in 2015. Mechanical engineers make over $1,500 a week, and pilots and flight engineers make over $1,800 a week, but carpenters—a much broader category with a wide variety of skill levels—have median earnings somewhat less than $700 a week. Figure 14.13 gives other examples of the pay in jobs that mostly women do, and the pay in jobs that mostly men do. Again, we are looking only at full-time year-round workers.

It turns out that in the various kinds of jobs that mostly women hold, there seems to be a lower ceiling on median weekly earnings than is true for jobs a substantial fraction of which are held by men. For both men and women, taking a job of the sort that is held mainly by women tends to guarantee relatively low pay. However, the reverse is not true: Taking a job of the sort that is held mainly by *men* does *not* guarantee relatively *high* pay. Men dominate a number of occupations with quite low pay, such as jobs that are physically demanding, but require little education or training.

Job segregation occurs even within occupations and industries. For example, there are industries in which some firms hire mostly men and other firms hire, in those *same* occupations, mostly women. Even the same firm, especially if it has plants located in different regions of the country, may hire mostly men at one plant and mostly women at another. Pay differences, with men earning more than women, usually accompany such segregation. Some of the differences are illustrated in Figure 14.14.

Why women are paid less than men is a matter of dispute. On average, women do not experience many of the economic disadvantages experienced by African Americans: school quality does not differ among men and women, nor (for obvious reasons) do women have poorer parents than men. One reason is that experience on the job is rewarded with higher pay, and in many jobs women have less experience than men. This is in part due to the time women take off of paid work to raise children or for other family responsibilities less likely to fall on men.

Some people attribute pay differences in the same job to women's lesser physical strength or some other skill. However, jobs requiring physical strength—such as farm workers and stock handlers for example—show relatively little difference in pay between men and women, especially by comparison to lawyers, physicians, and insurance adjusters, jobs in which physical strength is not rewarded. For many,

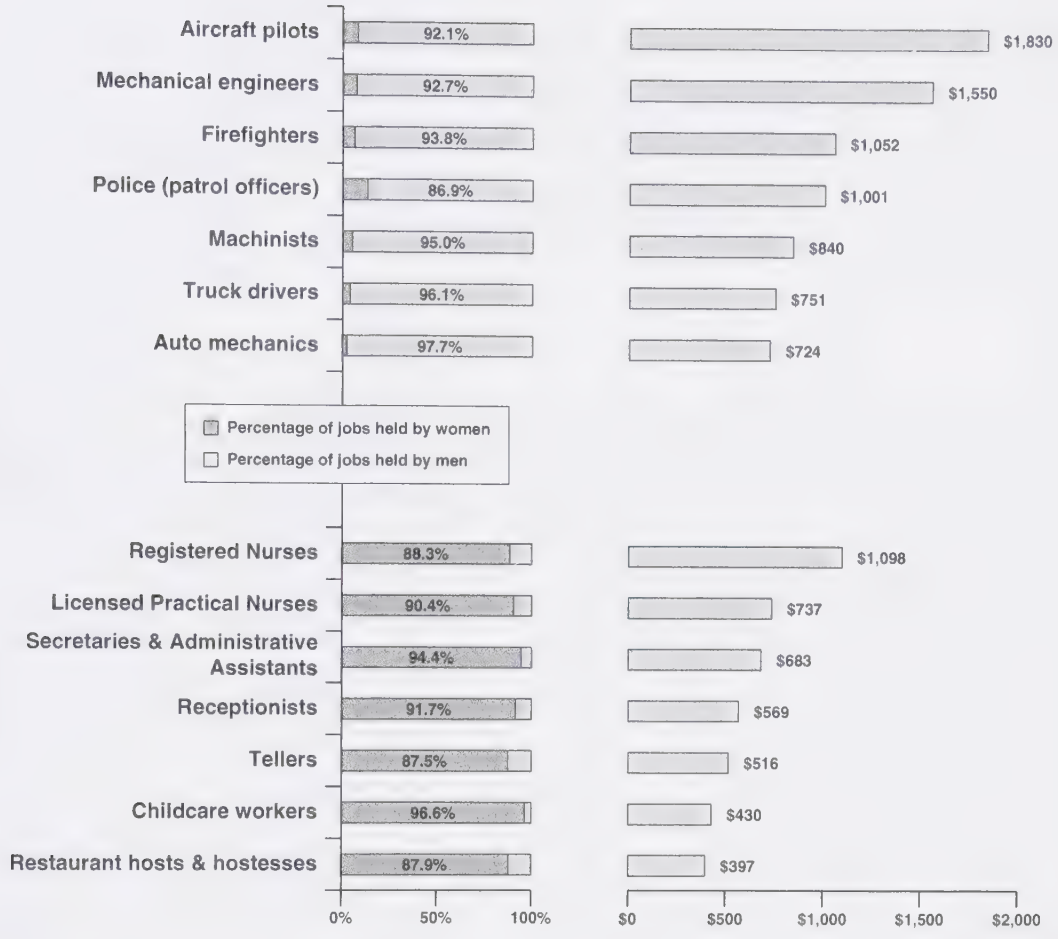


FIGURE 14.13 Women's work, women's wages: Gendered occupations and unequal earnings in the U.S., 2015.

In certain occupations almost all the workers are women, while in certain others almost all are men. The occupations displayed here include some of the most "gendered" occupations. Median weekly earnings are listed for each occupation, for full-time wage and salary workers, for the gender that dominates the occupation. (Data for the other gender were in many cases not reported in the data source.)

Source: U.S. Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey," Household Data, Annual Averages, Table 39: Median weekly earnings of full-time wage and salary workers by detailed occupation and sex, available at http://www.bls.gov/cps/cpsaat39.htm.

job discrimination increases job segregation by sex and thereby increases the differences in average wages and other disparities between male and female workers.

Social norms about "appropriate" work for women also make a difference. Some economists, including Nancy Folbre of the University of Massachusetts, have argued that there is a "care penalty"—relatively lower pay for jobs that involve taking care of others.

But it is also true that women who take "male" jobs such as truck driving or auto repair are sometimes seen as sexually unattractive. Another "callback" study by Folbre and Lee Badgett (also of the University of Massachusetts) confirmed this.[1]

[1]Nancy Folbre and Lee Badgett, "Job Gendering: Occupational Choice and the Marriage Market," *Industrial Relations* 42(2): 270-298 (2003).

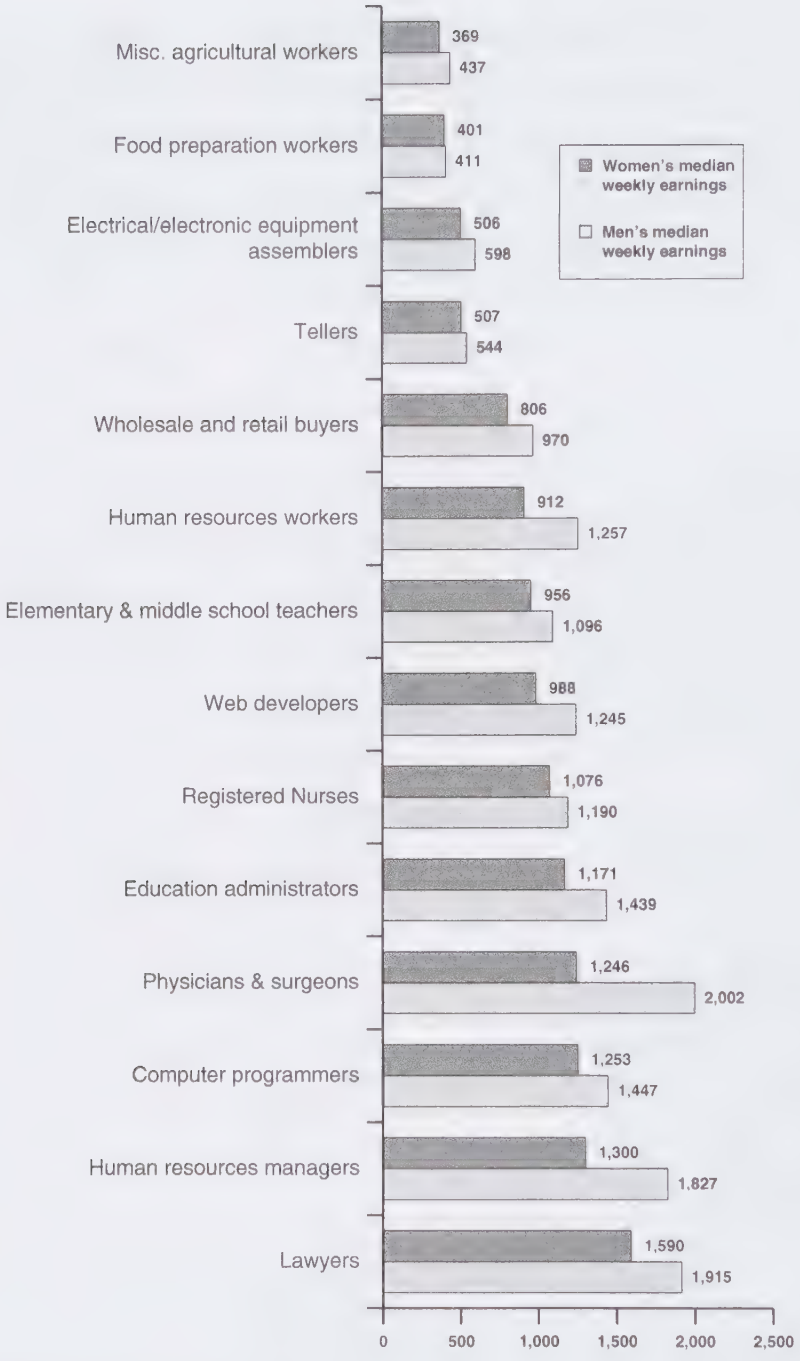


FIGURE 14.14 Men's and women's median weekly earnings in selected occupations in the U.S., 2015.

This figure illustrates the fact that with very few exceptions, in nearly all occupations the median weekly earnings for men exceed those for women. This includes jobs overwhelmingly held by women, such as registered nurses. Note the large gender pay gap for human resource personnel.

Source: U.S. Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey," Household Data, Annual Averages, Table 39: Median weekly earnings of full-time wage and salary workers by detailed occupation and sex, available at http://www.bls.gov/cps/cpsaat39.htm.

They asked survey respondents to rank fictitious personal ads from women and men seeking dates according to the likely number of positive calls they would receive. Ads that portrayed either women or men in atypical occupations, such as a female electrician or a male nurse, were rated lower than others in more typical jobs who had otherwise similar hobbies, relationship preferences, and physical attributes. Women in atypical occupations without much education paid an especially high price in the "dating market." Such dynamics could discourage some women from entering male-dominated occupations, and some men from entering female-dominated occupations.



Is *She* Well Served if *He's* in Charge?

Women are underrepresented in politics around the world. For example, in 2015 women were just 22 percent of members of the world's national parliamentary bodies like the U.S. Congress. But does the underrepresentation of women in the legislatures of the world result in policies that favor men's interests?

India, the world's most populous democracy, became a laboratory for studying this question after its constitution was amended in 1993 to require that women be the heads of not less than a third of each state's local government councils. In many states the villages that were to be required to elect a woman as *pradhan* (council head) were selected simply by choosing the first village, the fourth, the seventh, and every third village thereafter from a list of all villages in the state.

A detailed study of 261 villages in the states of West Bengal and Rajasthan investigated how policies were affected, if at all, in poor villages with newly elected female *pradhans*, compared to other poor villages. All the villages studied lacked public services. Tap water was available in only one in ten of the Rajasthan villages and one in twenty of the West Bengal villages. Public health facilities were available in less than a tenth of the villages in West Bengal and fewer than half of the villages in Rajasthan.

Researchers studied what issues were raised by men and women in the council meetings. In both states women complained more often than men about the lack of tap water. This is not surprising: Where there is no tap water, women have to carry water in pots on their heads, often for long distances. In West Bengal, where women do most of the paid work of road maintenance, women complained more often than men about the condition of the roads. In Rajasthan, where road maintenance is shared between men and women, men often travel in search of work. Since men there could count on at least half of the road maintenance jobs, they disproportionately favored road improvements.

The councils with newly elected female *pradhans* adopted policies in line with the interests of women. In both West Bengal and Rajasthan they invested more in providing water than did villages not selected to have a woman *pradhan*. In West Bengal, councils headed by female *pradhans* invested more in roads, while in Rajasthan less was invested in road improvements.

The study brings good news about democracy: who gets elected does make a difference (but see the box "Democracy Disconnect" in Chapter 19). So as women achieve representation in legislatures worldwide in proportion to their numbers, so will women's interests—at least—be better served.

Source: Raghabendra Chattopadhyay and Esther Duflo, "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India," *Econometrica* 72, no. 5 (September 2004): 1409–43; United Nations, Millenium Development Goals Chart 2015, available at http://www2.unwomen.org/~/media/headquarters/attachments/sections/library/publications/2016/mdg-gender-chart-2015-for-web.pdf?v=1&d=20160222T174956

Conclusion

Why do some families in the United States have much more income than others? And why do the income differences among families change over time, increasing as they have been for the last three decades of the twentieth century or decreasing as they did for the four decades prior to that?

One way to answer this is to make an analogy between a family and a farmer. The farmer's income will depend on how much of each crop he is able to market and the price he gets for each. The family's income is determined in the same way. Like the farmer and his crops, the family has a set of assets that potentially can earn income: the skills and time of its members and perhaps some land or other capital goods either directly owned or through the ownership of stocks. During a given year the family will put some of these assets on the market: renting the land, putting some of their skills and time at the disposition of an employer in return for wages, and so on. Like the farmer, the family's income depends on what they have to put on the market, how much of each they can sell, and the price that each of their income-earning assets fetches.

SUGGESTED READINGS

Lee Badgett and Nancy Folbre, "Job Gendering: Occupational Choice and the Labor Market," *Industrial Relations* 42, no. 2 (2003): 270–98.

Samuel Bowles, *The New Economics of Inequality and Redistribution* (Cambridge, UK: Cambridge University Press, 2012).

Samuel Bowles and Herbert Gintis, "The Inheritance of Inequality," *Journal of Economic Perspectives* 16, no. 3 (2002): 3–30.

Samuel Bowles, Herbert Gintis, and Melissa Osborne, "The Determinants of Earnings: A Behavioral Approach," *Journal of Economic Literature* 39 (December 2001): 1137–76.

Samuel Bowles, Herbert Gintis, and Melissa Osborne, eds. *Unequal Chances: Family Background and Economic Success* (Princeton, NJ: Princeton University Press, 2004).

Angus Deaton, *The Great Escape: Health, Wealth, and the Origins of Inequality* (Princeton: Princeton University Press, 2013).

Len Doyal and Ian Gough, *A Theory of Human Need* (New York: Guilford, 1991).

Barbara Ehrenreich, *Nickel and Dimed: On (Not) Getting By in America* (New York: Henry Holt, Metropolitan Books, 2001).

Robert Frank, *Luxury Fever: Why Money Fails to Satisfy in an Era of Excess* (New York: Free Press, 1999).

Katherine Newman, *No Shame in My Game: The Working Poor in the Inner City* (New York: Vintage, 2000).

Michael Omi and Howard Winant, *Racial Formation in the United States*, 3rd edition (New York: Routledge, 2014).

Kate Pickett and Richard Wilkinson, *The Spirit Level: Why Greater Equality Makes Societies Stronger* (New York: Bloomsbury Press, 2011).

Thomas Piketty, *Capital in the 21st Century* (Cambridge: Belknap Press, 2014).

Robert Pollin, Mark Brenner, Jeannette Wicks-Lim, and Stephanie Luce, *A Measure of Fairness: The Economics of Living Wages and Minimum Wages in the United States* (Ithaca, NY: Cornell University Press, 2008)

Joseph E. Stiglitz, *The Price of Inequality: How Today's Divided Society Endangers Our Future* (New York: W. W. Norton, 2013).

Edward N. Wolff, "Changes in Wealth Inequality in the 1980s and 1990s in the U.S." in Edward N. Wolff, ed., *International Perspectives on Household Wealth* (Northampton, MA: Edward Elgar, 2006).

15 Progress and Poverty on a World Scale

Six centuries ago the Arabic explorer and geographer Ibn Battuta traveled throughout Asia, Africa, the Middle East, Russia, and Spain. He also visited the land now called Bangladesh, where today a sixth of the people are undernourished, and the country's Gross National Income per person (at purchasing power parity, or PPP) defines it as a lower-middle income country—just above the thirty-one poorest countries classified as low-income. He described traveling along its waterways, "passing between villages and garden lands." "It was as if we had been going through a market," he observed. "This is a country in which rice is extremely abundant."[1]

In the fourteenth century Bangladesh (then Bengal) was relatively affluent in relation to the other parts of the world that Ibn Battuta had visited. Europe, by contrast, was reeling under the impact of a deadly bubonic plague that in many areas took the lives of a quarter of its population. Hunger and the desperate struggle for survival were the daylong and lifelong lot of all but a handful of rich and lucky people. The plight of Europe was not unusual: poverty has been the global norm for millennia. In the fourteenth century, the relative affluence of Bengal made it the exception, not the rule.

Along with the United States, Japan, and a few other countries, Europe is now the exception. It has one of the highest standards of living anywhere. The amount of goods and services available surpasses anything imaginable even a few centuries ago. And in some countries people maintain a high standard of living while working fewer hours than before. In Denmark, for example, the average work week in 2016 was thirty-two hours, compared to over thirty-eight in the United States.

However, in Europe today, and even more so in the United States, some people are still desperately poor. In the United States in 2015 more than one child in five (22 percent) was living in poverty. In 2010 (the latest data available), 32 percent of people in Bangladesh were poor, yet a few—a very few—lived in a condition of affluence comparable to that of the rich in Europe and the United States.

[1]Quoted in Henry Yule, *Cathay and the Way Thither* (London: Haklyut Society, 1916), 80, 91.



Cheap Bananas and Runaway Shops: Are *We* Rich Because *They're* Poor?

"Are *we* rich because *they* are poor?" This is a haunting question for people in countries where the income levels are many times greater than in poorer countries.

One way to tackle the question is to ask another. If the people of the poor nations of the world were to succeed in raising their incomes by a substantial amount, would this force a reduction in the income of the typical rich-country resident? The answer to this question has at least two sides.

On one hand, the fact that wages are low in poor countries (see Figure 15.8) means that the cost of production in those countries is low, and this means that people in Europe, the United States, and other rich countries get cheap shoes, cheap clothes, and cheap bananas. In this sense, rich-country residents benefit from the poverty in the rest of the world.

The other side of the coin is that when these cheaper goods are imported by rich countries, workers in the rich countries who were producing those goods often lose their jobs. And when firms in rich countries such as the United States expand operations in low-wage countries while closing plants back home (see box "Globalization: U.S. Winners and Losers" in Chapter 7) or "outsource" things such as software development or medical records processing to low-wage countries (see box "Hardball: Owners, Workers, and Taxpayers in the Global Economy" in Chapter 13), low wages in large parts of Africa, Asia, and Latin America can undermine living standards of some groups in the rich countries.

In the United States and other rich countries, some benefit from the *cheap banana*, while others' wages are eroded by the threat of the *runaway shop.* Many auto workers and software engineers would find, if they looked closely at the labels of the clothes they are wearing, that they both benefit and lose. Big questions rarely have simple answers.

Throughout the world today, poverty is not the sole barrier to a decent standard of living. Even those with adequate incomes now find the quality of life and even life itself threatened by HIV/AIDS and other epidemics, the poisonous environmental consequences of rapid, careless economic growth, the failure to regulate new technologies, and in many countries the danger of violent conflict.

How do these massive improvements and striking inequalities in living standards come about? Part of the answer, in a word, is capitalism. We have seen already in Chapter 1 the dramatic changes in output per capita and increases in real wages that have occurred since capitalism took hold in Europe and North America. But wages or per capita output levels, as we have seen in the previous chapter, are just part of what make up our economic well-being. We need to understand the entire process by which some countries and some people get rich while others remain in poverty. We need to comprehend what determines a person's and a nation's living standards. (See box "Cheap Bananas and Runaway Shops: Are *We* Rich Because *They're* Poor?")

In this chapter we focus on the world economy. Already we have seen that capitalism is an internationally oriented economic system: firms seeking profits do not let national boundaries get in their way when there is money to be made. In Chapter 1 we saw that inequalities in the world have grown since the early days of capitalism. In later chapters, including Chapter 13, we saw that as firms have moved production operations abroad, often their foreign affiliates have been able both to pay lower wages than in the United States, and at the same time to threaten to run away from

their U.S. plants, in order to pressure U.S. workers to accept lower wages and worse working conditions. Here we build on these ideas by looking at the experiences of other countries and by studying the processes of investment from one country to another.

The main idea of this chapter is that *dramatic differences in income levels in the world reflect both the massive productive potential of the capitalist economic system and the uneven distribution of the results of the capitalist economic growth process. Growth and distribution in turn depend on the rules and institutions that regulate international investment, production, taxation, and other activities that affect development.*

This central idea is expressed in seven key points:

1. Around the world, people's incomes and living standards differ widely both within and among countries, and have evolved in diverse ways over time. In other words, there is *uneven development* among countries, nearly all of whose economic systems are predominantly capitalist. For a country to be capitalist does not guarantee its economic development.

2. There are several kinds of cause-effect relationships that tend to create both *virtuous circles*—in which prosperity leads to further prosperity—and *vicious circles*—in which poverty leads to further poverty. Two consequences are that both technology gaps and power gaps between rich and poor countries tend to persist.

3. Nations that started the process of economic development early tend to have an advantage over latecomers, partly due to economies of scale, partly to the technology gap (knowledge and skills that are not easily available to people in poor countries), and partly to the *power gap* (disproportionate influence of powerful countries and industries on evolving rules and institutions at all levels).

4. Key differences among national economies have much to do with the *rules and institutions* that permit, shape, or restrict various economic activities. This includes rules made both by national governments and in international agreements and organizations. Such rules can either aid or inhibit growth in productivity, and may either break up vicious circles of poverty or reinforce them.

5. In poor countries, major increases in productivity can often be achieved by learning and using new technologies. Foreign investment has the potential to help by transferring technology to local managers and workers; but to realize this potential often requires suitable government policies. The World Trade Organization agreement, and other recent agreements, reduce the scope for such government policies.

6. Those countries most successful in raising productivity have often been those in which the government has played an important role in the economy.

7. Investment by transnational corporations (TNCs) sometimes can foster development, but its effects in practice are often mixed. Half of U.S. corporations' foreign investment is in tax havens, and half of their foreign profits are reported in these tax havens, although there is typically no production and almost no employment there.

Poverty and Progress

Today one of the most striking aspects of the world economy is the extent of inequality. In the United States, the richest fifth of households receives more than twenty-eight times as much income as the bottom fifth does. Even more shocking, though, is the fact that average income in the richest 5 percent of households in the United States is roughly *seventy-five times* the average income in Bangladesh, taking into account the differing buying power of the incomes in both countries. And Bangladesh, while very poor, is not the poorest country in the world. This dubious honor goes to the Democratic Republic of the Congo, whose average income in 2013 (taking account of differences in buying power among countries) was about one-fifth that of Bangladesh and one-eighty-fourth that of the United States.

In 1700 Mexico's per capita income was about the same as that of the British colonies that would become the first thirteen states of the United States. Cuba and Barbados were significantly richer. Around 1800, Cuba had a slightly higher per capita income than did the United States, and Haiti was probably the richest society in the world (on a per capita basis, counting slaves). Since 1820, however, the share of income received by the richest 10 percent of the world's population—almost all of them living in North America or Europe—has risen, and the share of the poorest 60 percent of the world's population—almost all of them in Asia, Africa, and Latin America—has fallen. By the year 2000, the per capita income of Mexico (at Purchasing Power Parity—see box, "What's a Hamburger Worth?" in this chapter) was less than a third of the U.S. level, and Haiti's was even lower; both are still true today.

In recent years this process of income divergence worldwide between the haves and the have-not countries has slowed and perhaps even reversed. The reason for this is that China, the world's most populous nation, has experienced extraordinarily rapid income growth over the past two decades, growing far more rapidly even than England did as the leader of the Industrial Revolution in the nineteenth century. Inequality is also increasing rapidly within China, but the overall effect of China's dramatic catching up is to reduce world inequality. India, the world's second-most-populous nation, is also catching up, though more slowly. As in China, inequality among Indians is growing rapidly. Worldwide, however, the very wealthiest *people* have become wealthier relative to the rest of the world's population. But income and wealth differences tell only part of the story. Figure 15.1 presents comparative data on incomes and some other aspects of living standards.

It is clear from the upper left panel in Figure 15.1 that the buying power of the average income in the United States is much greater than the buying power of the average income in China, India, and Kenya. In fact, in 2011 the buying power of the average income in the United States was also greater than the buying power of the average income in Sweden and Japan. (Bangladesh, not shown, has income a little above that of Kenya.)

The top right panel of Figure 15.1 presents a standard measure of health in a population—life expectancy at birth, that is, the average number of years a newborn infant can expect to live. Notice that although Japan and Sweden are less rich than the United States, their populations both live several years longer than people in the United States. Differences in health between these relatively high-income countries (as a group) and the others may be greater or smaller than the differences in their income levels, depending on the specific measure. Life expectancy in Kenya, for example, is more than three-fourths the U.S. figure, while its average income is barely

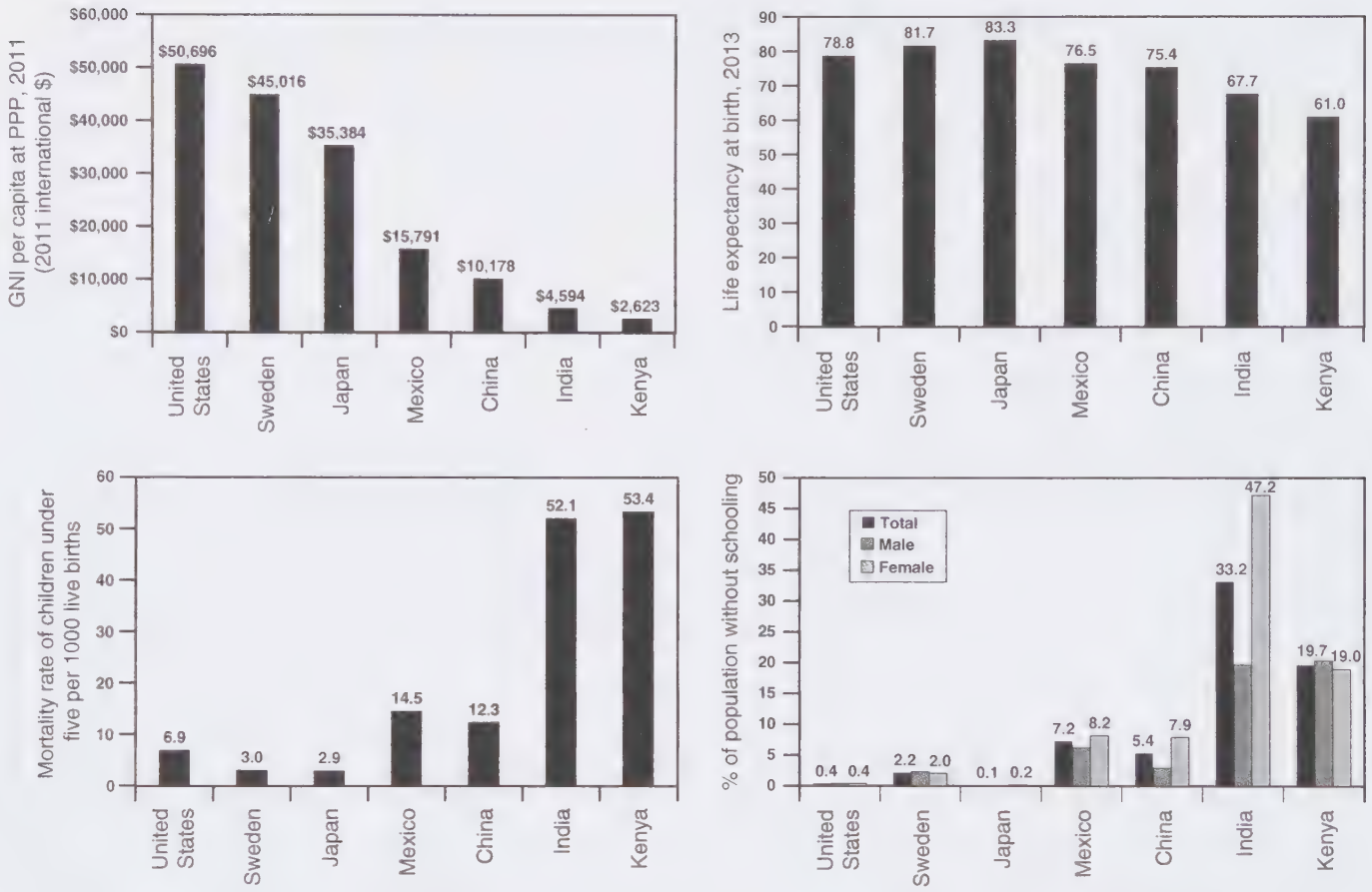


FIGURE 15.1 Income, life expectancy, child mortality, and schooling in seven countries.

The four panels in this figure compare countries with a wide range of income levels, life expectancy, child health, and schooling, using data from years ranging from 2010 to 2013. Notice that there is no exact correspondence between income per capita and the social indicators. For example, India had nearly twice the Gross National Income per capita (at purchasing power parity) that Kenya had in 2011, yet in 2010 nearly half of Indian women fifteen or older had no schooling, compared to Kenya, where less than 20 percent of such women had never been to school. The same two countries are nearly equal in child mortality, which is the number of children under five who die during a year out of every one thousand who were born alive.

Sources: World Bank, World Development Indicators 2015; Robert J. Barro and Jong Wha Lee, "A New Data Set of Educational Attainment in the World, 1950–2010," *Journal of Development Economics* 104 (September 2013): 184–98, available at www.barrolee.com

one-nineteenth that of the United States. However, disparities are much greater for *child mortality rates* shown in the bottom left panel of Figure 15.1 (the number of children each year, out of every one thousand live births, who die before reaching age five). Sweden and Japan are healthier than the United States by this measure, but about one in nineteen babies born in Kenya die before the age of five (that is, fifty-three per thousand).

The fraction of the population with no schooling is shown in the lower right panel, by gender. Note that nearly half the women in India have not been to school at all, though most girls today in India do attend primary school at least for a few years.

Not surprisingly, countries with higher Gross National Income (GNI) per capita, adjusted for Purchasing Power Parity (see box "What's a Hamburger Worth?" in this chapter), tend to have longer life expectancy and lower child mortality rates. Richer countries have long provided more of their people with educational opportunities, and hence a smaller percent of their adult population is without schooling.

The living conditions typical of China, India, and Kenya are much more common in the world today than the living conditions of the United States, Sweden, and Japan. (Well over a third of the world's people live in India and China; Americans make up less than one-twentieth of the world's population.) Figure 15.2 shows how

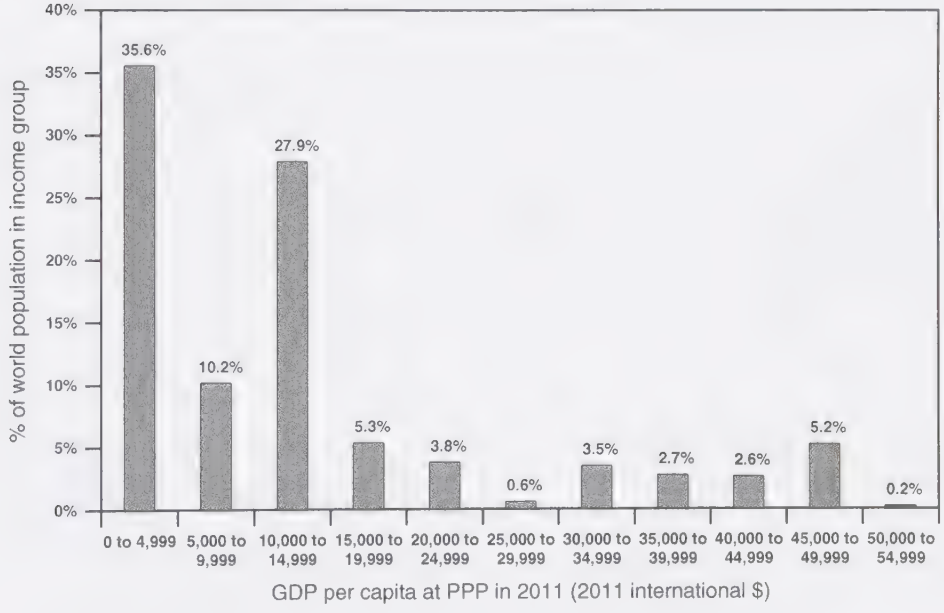


FIGURE 15.2 The income distribution of the world, 2011.

This figure uses data on the average income of each country (GDP per capita) in 2011. The numbers are adjusted for differences in purchasing power across countries (see box "What's a Hamburger Worth?" in this chapter). The height of each bar shows the percent of the world's population that had the annual income (in 2011 international dollars) in the range indicated on the horizontal axis. The fact that India's average income was between $0 and $4,999, and that China's was between $10,000 and $14,999, account for the tall columns for both these income ranges. In 2011, China and India each had close to 1.3 billion people, or nearly one-fifth of the world's population each. In this figure, all residents of a country are treated as if their income were the average income of that country.

Source: World Bank: World Development Indicators 2015.

the income of the world is distributed among families at various income levels. Along the horizontal axis are levels of family income measured in units of comparable buying power, translated into dollar amounts. The vertical axis indicates the fraction of the world's population that receives the level of income shown on the horizontal axis. Most of the people represented in the left-hand bars are in India ($0 to $4,999), China ($10,000 to $14,999), and the other large, relatively poor countries such as Bangladesh, Nigeria, Pakistan, and Ethiopia. Note that in this diagram, all the people in a country are treated as having the average income of that country; inequality within countries is not reflected in the diagram. (Figure 15.3 gives a different perspective.)

By most measures, about three-quarters of the income inequality in the world is between nations, not within them. This means that even if income were to be distributed equally within every country, the world's income would still be distributed very unequally. Nevertheless, within-country inequality is substantial in many countries—for example, the United States and Brazil. Figure 15.3 displays the income levels of each fifth of the populations of India, Brazil, and the United States in 2011. As the figure shows, the average income of the poorest fifth of the U.S. population (the shortest white bar) is higher than that of the richest fifth of Indians (the tallest black bar).

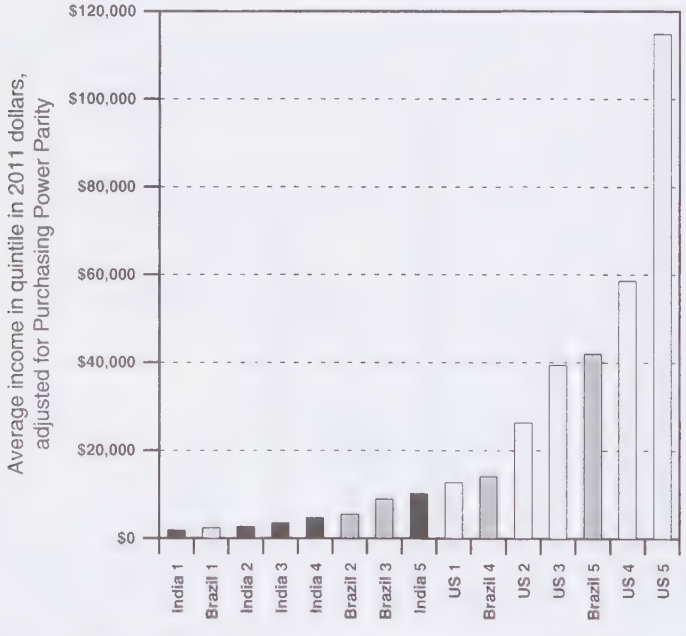


FIGURE 15.3 Comparative inequality: Incomes by quintile in India, Brazil, and the U.S., 2011.

This figure compares the income levels of three countries, showing the income by quintile (fifth) of each country's income distribution and placing the various quintiles—no matter from which country—on the horizontal axis in order of each quintile's income. The labels refer to quintiles: "US 5" refers to the highest-income quintile of the U.S. population, and "Brazil 1" refers to the poorest quintile of the Brazilian population. The height of each bar reflects *real* income (corrected for inflation) adjusted additionally (by the purchasing power parity method) for the relative buying power of a unit of each nation's currency (see box "What's a Hamburger Worth?" in this chapter). Data for the U.S. are for 2010 rather than 2011.

Source: World Bank: World Development Indicators 2015.

But the degree of income inequality in Brazil allows the richest fifth of that country's citizens (the tallest gray bar) to enjoy an average income above that of the middle quintile in the United States (the "Brazil 5" gray bar is a bit taller than the "US 3" white bar). Meanwhile, the poorest fifth of Brazilians ("Brazil 1") has average income only a little above that of the poorest fifth in India ("India 1", the shortest black bar). Of course, if we divided each country into income deciles (tenths), then each country's bars would likely be scattered more from one end to the other of the diagram.

Are living standards improving throughout the world? It depends on where you look and what you look at. By almost any measure, living standards have been rising *on average*. Access to medical services, basic nutrition, and education has improved in most countries. Figure 15.4 summarizes some of the progress made since 1960. Using the same measures of well-being as in Figure 15.1 but reflecting world averages, the figure shows that income and longevity are both rising, while child mortality and the fraction of people who have not been to school are both falling. The United Nations and its member countries targeted Millenium Development Goals to improve health, education, gender equity, and a host of other indicators by 2015—with considerable, but not universal, success. Recently the UN and its member countries set a new round of more ambitious Sustainable Development Goals, with the aim of promoting greater efforts to bring everyone in the world closer to a flourishing life. Yet widespread drought, violent conflict, and resulting forced migration make it challenging in many parts of the world to achieve these goals.

Productivity and Income

The facts of economic progress and inequality among nations point to an important question: Why have some countries made such striking economic progress while others have done so poorly? Why is there such *uneven development* among countries?

In some ways, unevenness in development is self-reinforcing. There are ways in which "them that have, get" and those that have little, get little. In other words, there are both *virtuous circles* and *vicious circles* in development processes.

Vicious Circles of Low Productivity

An example of a vicious circle is the presence of extremely debilitating diseases in a poor country, as described in the box, "Sleeping Sickness: It Really Is a Failure of Capitalism" in Chapter 9. In some countries, there are widespread endemic diseases borne by parasites, part of whose life cycle is outside the human body. Thus a person who eradicates the parasite from his or her body with oral medicine will likely get the parasite, and the disease, again from the environment. This is a special kind of vicious circle whose solution is comprehensive, carefully designed eradication programs that treat a whole community and the whole environment. Such programs are currently being implemented under the auspices of the World Health Organization, in cooperation with other organizations. Private drug companies have donated drugs to some such efforts.

Eradicating such endemic diseases has enormous benefits for productivity. When someone is very ill in a poor family, not only can that person not work nor contribute to the family income, but another family member is also taken away from work by the need to care for the sick person. Further, the cost of doctor visits (and

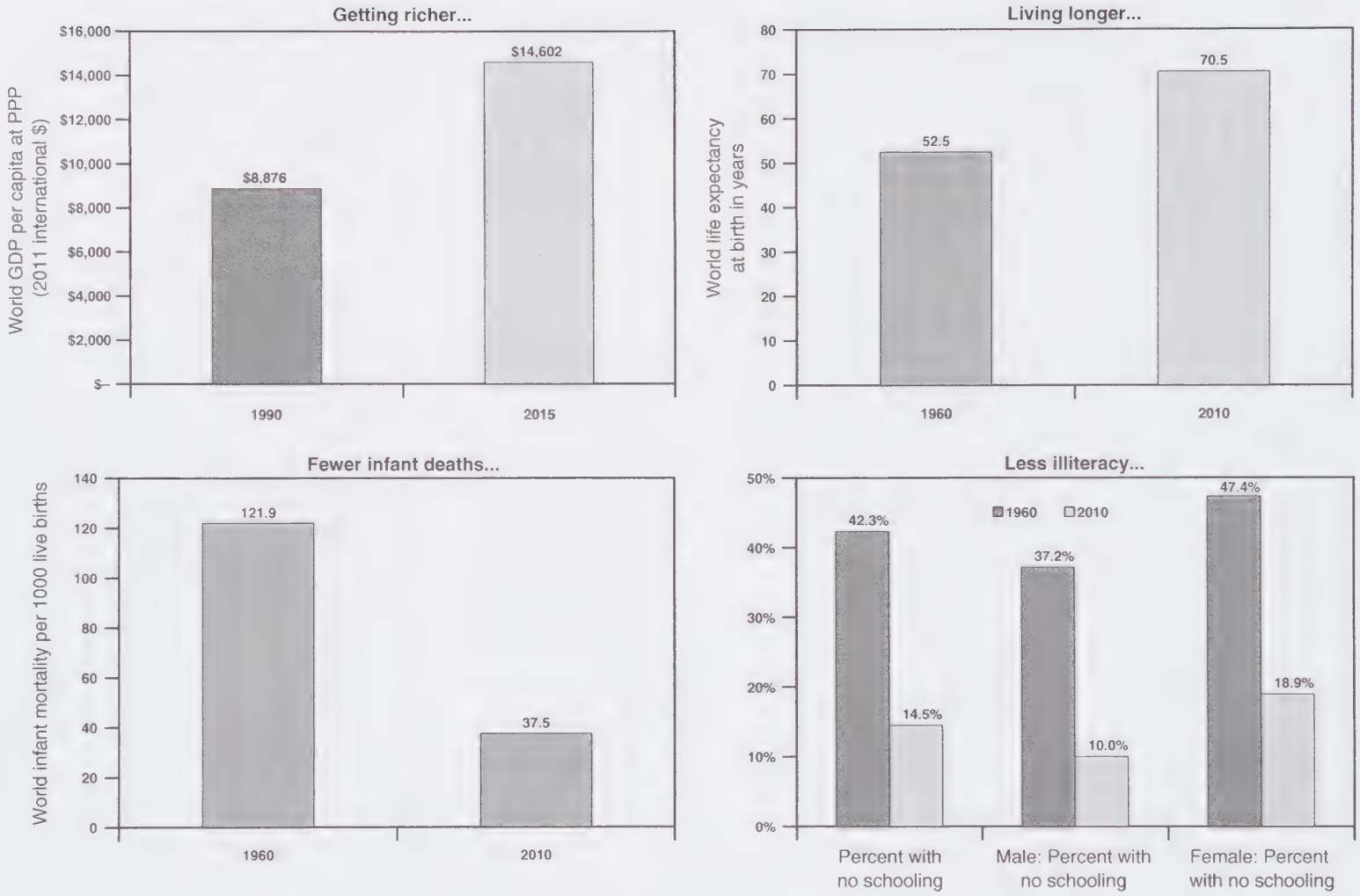


FIGURE 15.4 Progress on a global scale.

The numbers in this figure are averages for all the countries reporting the relevant information, and these countries contain the vast majority of the world's people. As the figure shows, dramatic gains in health and education have been made in the last fifty years.

Sources: World Bank: World Development Indicators 2015; Robert J. Barro and Jong Wha Lee, "A New Data Set of Educational Attainment in the World, 1950–2010," *Journal of Development Economics* 104 (September 2013): 184–98, www.barrolee.com.

the transportation to them) and medicines is a hardship on families that are poor already and live in countries without a national health system. But the Neglected Tropical Diseases program of the World Health Organization (with help from other organizations) has begun to tackle the vicious circle of endemic disease with significant success. This illustrates the fact that some vicious circles can be broken up, and problems solved, by an initiative from government or other institutions such as well-run aid organizations.

There are also virtuous and vicious circles in the competitive process. The process is cumulative: The winners of each round of competition start the next round with an advantage over their competitors. The game of competition in a capitalist economy—as we saw in Chapter 7—is not like tennis, in which each player starts with a score of 0 at the beginning of a new set. It is more like boxing, in which the fighter who has taken some punishing blows in the first round starts the second less able to fight and proves an easy target for a competitor. The world economy is one of cumulative advantages and disadvantages. But what exactly is the nature of the advantages or disadvantages, leaving latecomers trapped in low productivity growth?

ECONOMIES OF SCALE. As Chapters 6 and 7 have already explained, transnational corporations (TNCs) have affiliates in multiple countries, and their operations are located wherever they contribute best to the firm's global after-tax profitability. Countries that are already developed, and transnational corporations based in these countries, can enjoy virtuous circles relative to the rest of the world based on their previous successes. For example, if their industries are already large enough to have achieved economies of scale (see Chapter 11), so that their average cost per unit produced is low, then they can sell their product everywhere in the world at a price that smaller producers would find unprofitable. In this way they can maintain their dominance in the industry; but this makes it hard for poor countries to advance.

In poor countries, producers that do not export face a very limited market, particularly if the country is small; and this makes it hard to achieve economies of scale. Of course in principle poor countries could export their products, and thereby achieve scale economies by increasing the number of units of output they produce. The difficulty is that other countries may themselves be trying to increase the scale of their own production, and so may have protectionist policies, such as tariffs, to limit imports and shelter their own infant industries. Or if rival producers have already achieved scale economies, they may sell at a price with which the producer in a poor country cannot compete.

Establishing regional free trade agreements among developing countries, and so creating larger markets, is one strategy for escaping this vicious circle. In the 1970s, several countries in the northwest part of South America formed the Andean Pact to promote regional trade and development. Among other things, the member countries made a group decision about which country would first establish a tractor plant (it was Venezuela). This gave that plant a chance to achieve scale economies by selling to the whole region. Thus planning within regional agreements is one possible solution.

The auto industry in the United States was in a dominant position in the world auto market for some time because it had achieved large economies of scale. However, competition in the industry has now become fairly intense as the auto industries in some lower-wage countries (like South Korea) have become as efficient, or more efficient, at producing cars on a large scale (remember Hyundai's enormous auto

assembly plant mentioned in Chapter 11). The availability of relatively cheap U.S. cars in most places in the world—precisely because of scale economies—makes it harder for fledgling small scale auto companies in developing countries to produce and sell profitably. For developed countries, this is a *virtuous circle* of development; for developing countries it is a *vicious circle*.

South Korea faced this obstacle in the early 1980s, and managed to break out of that vicious circle and establish Hyundai as a competent and growing automaker—but only with considerable government encouragement, pressure, and support. The story of how it did this is told in Linsu Kim's book *Imitation to Innovation*; a briefer version is in Alice Amsden's *Asia's Next Giant* (see Suggested Readings at the end of this chapter).

Uneven development in auto and truck production is shown in Figure 15.5: initially, production was established in the United States, and production grew in Japan; then American and German automakers established significant production in lower-wage countries; in the 1960s and 1970s, Mexico and especially Brazil (both included in "Other countries") developed their vehicle production capacity but almost entirely in the form of subsidiaries of transnational firms. In the 1980s, South Korea developed its national production, and by the twenty-first century, all these national firms had become transnational, still with a parent firm in their source country, but producing in several host countries (most of them included in "Other"). Most recently, China has suddenly become a major vehicle producer.

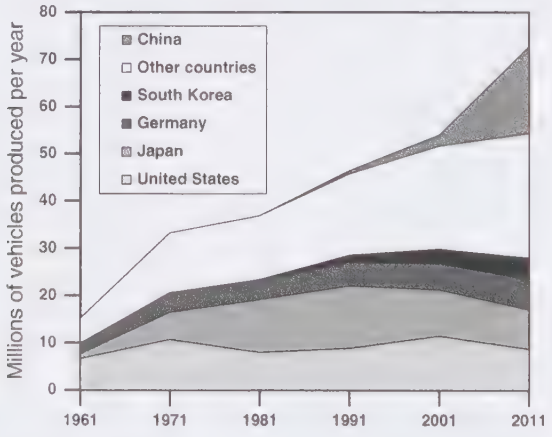


FIGURE 15.5 Global vehicle production, 1961–2011.

U.S. vehicle production, the bottom light-gray region, has been roughly level since about 1971, as U.S.-based auto makers have shifted some of their production to other countries. Japan's production increased initially, but was declining by the 1990s, when it also shifted some of its production abroad, including to the United States. Other auto makers did the same thing. South Korea's production grew rapidly until the twenty-first century, and it too began increasingly producing in other countries, including the United States. China's production, and production in other countries (Brazil, Mexico, India, and others) have grown rapidly for similar reasons.

Source: U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology, Bureau of Transportation Statistics, Table 1–23: World Motor Vehicle Production, Selected Countries, available at http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_01_23.html_mfd; data from WardsAuto.com, Motor Vehicle Facts & Figures (Southfield, MI: Annual Issues), 14 and similar pages in earlier editions.

TECHNOLOGY GAP. Another reason why vicious circles of poverty exist is the technology gap. A firm generally has little motive to share its technological secrets, fearing that a rival firm will learn these secrets, build on them, innovate further, and so become the new industry leader. Closely guarding technological advantages allows transnational corporations (TNCs) from rich countries that are already technologically advanced to enjoy a virtuous circle of technological superiority. By the same token, firms that are technologically far behind get stuck in a vicious circle: Because they do not have access to cutting-edge technology, they find it difficult to innovate and gain access to the information they need to learn new technology; and lacking the labor skills needed to operate the technology, they have a hard time attracting manufacturing plants that use the latest technologies. However, suitable government policies can help a country break out of such a technology gap. Several illustrations follow.

Governments have often taken a leading role in technology transfer, which is one of the few ways in which poor countries can break out of the vicious circle of poverty and catch up with the more advanced countries. It is difficult for the private sector to take this on, because the process of adapting and introducing new technologies, while at the same time developing a labor force able to use them, often requires committing substantial funds over a long period of time, and it entails taking great risks with few immediate opportunities for profit.

In the late seventeenth century, when naval architecture and shipbuilding were on the cutting edge of modern technology (akin to computer design today), Czar Peter the Great of Russia, traveling in disguise, took a group of Russian shipbuilders to England and Holland to learn the most advanced methods of ship design and construction. In the late nineteenth century the Japanese government sent countless missions to Europe to learn about new techniques. Government expenditures for health and education to develop and maintain the productive capacities of the labor force have also figured prominently in latecomer strategies for catching up with the rich countries.

In South Korea in the 1960s and 1970s the government developed an effective way to elicit technology transfer from TNCs, without paying too high a price. One or more local Korean firms would try to persuade a foreign patent holder to license the Korean firm to produce the product. The licensing agreement would specify that the Korean firm had to meet strict quality standards, but with considerable help from engineers employed by the patent holders; and the product would be stamped with the brand name of the patent holder and sold in specified markets so as not to compete with licensed producers in other countries, or with the patent holder itself. Once licensing agreements were signed, the education, training, and technical assistance from foreign engineers in building and outfitting the plant were invaluable to Korean firms.

There was an important issue, however, which the government helped to resolve. If different Korean firms competed against one another for a licensing agreement from the same foreign firm, they might bargain away all the potential gains, for example by agreeing to pay a high royalty to the patent holder. So the South Korean government took over negotiation of such agreements, a strategy that allowed it to strike a better bargain than the rival Korean firms likely would have done. In this case the central government's practice of taking on the negotiation of licensing agreements greatly affected the ability of the South Korean economy to develop. This offers one example of how much difference rules and institutions can make in development.

China also succeeded in getting technology transfer from foreign investors. China's size and relatively high level of education and skill made it attractive enough as a site for foreign investment to allow it to impose a strong technology transfer policy on such investors. Firms typically are reluctant to share details of technology because they fear those who learn these details will develop a superior technology and surge past their former teachers in market competition. Yet China was a key site for foreign investors, and so they complied with the requirement to make an explicit plan for teaching Chinese managers and workers the principles and practice of their technology, and then to implement that plan.

After the establishment of the World Trade Organization in 1995, however, WTO member countries were no longer allowed to regulate foreign investment as they pleased, because they were bound by the new investment provisions of the WTO Agreement. Thus China was required, as a condition of joining the WTO in 2001, to officially junk its technology transfer policy. However, fifteen years later, researchers report that China still uses various means (but less openly) to lean fairly hard on foreign investors to transfer technology to Chinese firms. In the Chinese case, too, the original technology transfer rule allowed China to learn much more than if it had relied solely on voluntary transfer of technology. Once again, rules and institutions matter greatly. By the same token, the creation of the WTO with all its associated agreements also made a significant difference, narrowing the scope of policy decisions allowed to developing countries.

Vicious and virtuous circles exist, but it is often possible in principle to break out of them with the right kind of policies and programs established by government or by aid organizations. Sometimes, however, in TNCs' pursuit of new rules and institutions that favor their profitability over the long term, at the same time they may weaken the prospects for poor countries to develop. A good illustration comes from relatively new rules under the World Trade Organization agreements.

Intellectual Property Rights

One determinant of how rapidly a country develops is the price of its exports relative to the price of its imports. Many goods imported from rich countries are patented or copyrighted, and this means they can be sold at monopoly prices. The collective term for patents, copyrights, trademarks, and the like is *intellectual property rights*. A government may grant a monopoly right in the form of a patent to an inventor of a distinctly new and useful product or process. The purpose is to reward the inventor and create an incentive for more innovation. For the duration of the patent, the patent holder has the sole exclusive right to produce and sell the product. This means the patent holder is able to charge a price for the good that is higher than would prevail in a competitive market with many firms producing it and competing to sell it. The patent holder sometimes licenses the right to others (as in the South Korea story above), allowing them to produce the good, typically for a fee. Under the WTO Agreement, the length of a patent is fixed at twenty years and made uniform worldwide.

Two kinds of patents have been particularly troublesome for developing countries. One is patenting of seeds. Before 1980, there were no life forms that were patented—not seeds, not genetic material of any other kind, not living creatures. In 1980 a U.S. Supreme Court decision made it possible for the U.S. Patent Office to grant a patent for a living organism, and this was followed by extension of patents to seeds and other life forms. This allows giant agrochemical firms like Monsanto

to sell seeds to poor farmers, and then prohibit them from saving the seed from their crop to plant during the next season, unless they pay Monsanto (or the holder of the patent on the seeds) an additional fee. In some cases farmers have been charged the fee although they did not even consciously cultivate the crop, but only failed to actively eradicate it from growing on its own in their fields. U.S. farmers have challenged such charges in court but have often lost.

For developing countries, the other big problem with intellectual property rights has been that urgently needed patented drugs have often been priced far out of reach of their populations. Devastating diseases like HIV/AIDS impose great suffering and large costs in lost work time and lost productivity, not only for those who are ill, but also for those taking care of them. HIV/AIDS has been, and continues to be, a public health crisis in some countries of sub-Saharan Africa, with a million dying of AIDS each year, and 1.5 million new cases arising. Around 2001, the retroviral drug combination needed to treat an HIV/AIDS patient for a year cost $10,000, a cost that has now fallen to $100 per person per year. The decline occurred for several reasons. One was that in 2001, in the city of Doha, Qatar, at a WTO Ministerial meeting to negotiate various development issues, the ministers issued a declaration modifying the rules on intellectual property. They said that member countries had the right to declare a public health emergency and issue *compulsory licenses* requiring pharmaceutical companies to allow production of such urgently needed drugs in the country with the public health crisis, at a price close to cost. In some cases the drug could also be imported from other developing countries producing the needed drugs partly for their own public health emergency.

The overall trend of expanding intellectual property rights and enforcing them more globally that came with the establishment of the WTO has been another "virtuous circle" for drug companies based in rich countries, and a "vicious circle" for people in developing countries. In developing countries, people cannot afford high-priced brand-name medicines (expensive even for people in rich countries without a national health system); and the developing countries themselves (since they signed the WTO agreement in the early 1990s) in most cases can only establish local pharmaceutical manufacturing of drugs whose patents have been allowed to expire, or drugs for which their government declares a public health emergency under the 2001 Doha declaration mentioned above.

Even so, drug companies have developed a number of strategies to extend their drug patents beyond the standard twenty-year period, either officially or by imposing burdens on potential competitors, such as forcing them to separately conduct expensive clinical trials, even though a chemically identical drug has already passed such tests.

When the intellectual property rights agreement that was part of the overall WTO agreement extended these rights beyond their former reach, economists estimated that the United States would be the country to gain the most from the rule change, and there would be a few other winning countries, as well as many countries that would be small losers, and a few rather large losers. The United States was projected to gain the equivalent today of $9.1 billion, and Brazil to lose about $2.7 billion. The losses were due to paying more for patented and copyrighted goods, and the gains were in the countries where the patents and copyrights were held.[2]

---

[2]Keith E. Maskus, *Intellectual Property Rights in the Global Economy* (Washington, DC: Institute for International Economics, 2000), 181–86.

Other Government Interventions to Promote Development

In most countries where the disadvantages of initial poverty have been overcome, the government has played a major role in providing the conditions for rapid productivity growth. The earliest example was in the United States right after the War of Independence. At the time many of the founding fathers, including Alexander Hamilton, reasoned that U.S. manufacturing industries could never compete with their larger and more experienced British counterparts unless the American "infant industries" were sheltered from competition and given some time to get established. Hamilton advocated a heavy tax—a tariff—on imported British manufactured goods, and this policy was then implemented. Tariff protection of industry was also a prominent feature of economic policy in nineteenth-century Germany, helping it to overtake Britain as an industrial power.

In other countries governments have pursued different strategies. In addition to the technology transfer policies already described in the previous section, in South Korea beginning in the 1960s the government used its police and armed forces to help employers in their conflicts with workers. This kept wages low, thus assuring high profits for companies (bringing to mind the warning, "Don't try this at home.") The government also prevented most companies from investing the profits anywhere but in South Korea. Sending wealth out of the country—a common practice by the rich in most poor countries—was deterred by capital controls designed to curb such "capital flight." Under pressure from the government, the major banks in South Korea channeled large sums of money into infant industries. And in sectors in which these strategies failed to bring about sufficient private capitalist investment, the government simply did the investing itself, as it did, for example, in building one of the world's most efficient steel plants under government ownership (POSCO, which after years as a state-owned enterprise was privatized by the year 2000).

Requisites for Economic Development

What does economic development require? Scholars have identified eight key conditions that appear necessary in order for economic development to occur. In addition to these, there are some other factors that may play a role in development, including the existence of natural resources. We discuss the essential elements below.

Eight Essential Elements

- **stability and peace:** political stability, an absence of civil conflict, and a currency whose value is fairly stable, along with an inflation rate that is not too high
- **financing** for development initiatives
- a **well-functioning government** that mostly does not waste or embezzle public funds, nor corruptly extort payments from the public for performing services
- **skills** needed in industry, agriculture, and government—or some means of acquiring skills (including those that can only be acquired on the job)
- **education**, such as for engineers, managers, and others overseeing production activities and developing new products (and an absence of major brain drain in the form of migration of such educated persons to richer countries)

Infrastructure, of an economy consists of its roads, railways, airports, harbor facilities, bridges and tunnels, water and sewage systems, public utilities and electricity grids, and communications networks such as telephone lines and the internet.

- **infrastructure**, such as adequate electrical power supply, transportation, and communication
- **absence of major public health problems** involving disabling diseases like malaria, sleeping sickness, Ebola, HIV/AIDS, and others; presence of a public health service that prevents, monitors, and controls outbreaks of disease
- **access to foreign markets** (absence of prohibitive trade barriers)

STABILITY AND PEACE. Both political stability and stability in the value of the national currency are important to reduce the risks of initiating new production operations. They allow potential investors to plan ahead several years into the future, which is key in encouraging investment.

Peace is also necessary to reduce risk for potential investors. If civil war is imminent or already in progress, it is likely to shut down investment because of the danger that what is built may be destroyed, or that economic activity will be curtailed by the need to stay away from public roads where there may be robbery or violence.

Peace is important for another reason. Civil conflict has its own vicious circles, described in Paul Collier et al., *Breaking the Conflict Trap*—see Suggested Readings (Collier refers to vicious circles as *traps*, and defines *conflict traps* that tend to perpetuate violent civil conflict once it has begun). Military spending tends to beget more military spending. Rebel groups tend to behave more and more like businesses, and then are reluctant to lay down arms because it means giving up making a rather good living. In addition, military spending tends to persist for years after civil conflict ceases. In other words, it requires prolonged peace to escape from elevated government military spending, and in the interim, this cuts deeply into development spending.

As for the goal of a stable currency, the cause of fluctuating currency values is often rapid inflation (generalized increases in most prices); so those who control the monetary system need to show determination to act in ways that keep inflation within reasonable bounds.

FINANCING. Other items on the list, such as infrastructure and education, will require government spending, although some of the cost might be financed by foreign aid. Foreign aid, however, has its problems. Foreigners, including outside "experts," sometimes make poor decisions based on inadequate knowledge of the local situation. If a project is not based on a need perceived by the local community, development projects are unlikely to work well. And corruption in distributing and spending funds is a persistent problem.

Financing often comes partly from government tax revenue, but it is often insufficient. Tax collection in many countries used to be mainly from import tariffs, but as trade barriers have been dismantled and tariffs reduced, less money has flowed into government coffers from that source. In poor countries, systems for collecting other kinds of taxes often are not very efficient. This can force partial reliance on borrowing, foreign aid, or other sources of financing.

Another source of financing, though not through government, is remittances—money—sent back home by migrants working in other countries. In Mexico, for example, many migrants to the United States join "hometown associations," which pool contributions from their members and send funds to build schools, clinics, or other projects in their hometown. Unlike government-to-government development aid, this form of financing usually avoids leakage of funds into the pockets of corrupt

officials. However, it will not finance mega-projects like dams or airports. Remittances can also be invested in private businesses (or just spent on consumption or home repairs).

Public borrowing to implement major development projects is another method of financing. But borrowed money must be paid back, and that eventually requires collecting taxes, whose problems have already been mentioned. Corruption can also be a problem. Private borrowing from banks or even from relatives to invest in business is also common.

In the last several decades, a problem has arisen with tax revenues: tax competition among developing countries to attract foreign investment has often meant offering tax holidays to TNCs, which can cut deeply into government revenues. The same problem exists in the United States, with states competing with one another for auto plants. But efforts to restrain tax competition have been limited and largely ineffective.

Another large problem is the enormous illicit financial flows out of developing countries. Public officials who embezzle from public funds have learned to use tax havens to put their wealth beyond the reach of tax and criminal enforcement authorities; and those already holding private wealth have learned to do the same, and avoid having that wealth taxed. Of the total wealth of the rich in Africa, for example, $500 billion, or 30 percent, is estimated to be held in accounts in offshore tax havens, implying tax losses to African countries of $15 billion.[3]

A WELL-FUNCTIONING GOVERNMENT. A well-functioning government is essential to the economy of any country. In developing countries, though, corrupt government hurts more, because when there is little income to begin with, officials who embezzle or steal it leave people with too little to eat, and with inadequate health care and schools. Corruption of public officials also undermines confidence in the government, discouraging inflow of aid, loans, or investment. For example, donors may deny aid to a government that they think will misuse or divert the funds to persons, and for purposes, that the donors have no intention of funding. In these ways, corruption can retard development. Petty corruption, and simply poor management of public spending, are often problems as well.

SKILLS AND EDUCATION. While education teaches knowledge, skills must be learned mostly on the job, so there is a chicken-and-egg problem: Industrial enterprises are really the only places in which the full range of industrial skills can be learned, but enterprises are unlikely to locate in places where those skills are lacking.

Education is also important, and can be publicly funded, but frequently there is a mismatch between what knowledge is really needed and what is taught. In many developing countries, new college graduates have a hard time finding work. This can set in motion brain drain, in which those with college or postgraduate degrees leave the country in order to find a suitable job. This produces another virtuous circle for developed countries (a high concentration of educated people, and jobs that employ them, attracts additional educated people), but a vicious circle for developing nations (a low concentration of the highly educated means they seek work elsewhere)."

---

[3]Gabriel Zucman, "Taxing across Borders: Tracking Personal Wealth and Corporate Profits," *Journal of Economic Perspectives* 28, no. 4 (Fall 2014): 121–28, Table 1.

INFRASTRUCTURE. Studies show that expanding and improving infrastructure is strongly associated with accelerating economic growth. Roads in poor condition will prematurely destroy vehicles and slow down shipment of inputs and output. Too few transport vehicles means that existing buses and trains will be overloaded and break down more often. Inadequate electrical power supply causes frequent, sudden power outages, which interrupt production and can damage both work in progress and the machinery used to create it. Good infrastructure is thus essential to development.

PUBLIC HEALTH. Epidemic and endemic disease are both problems. Effective public health measures can in principle do wonders to increase the productivity of the labor force, both by making the sick well, allowing them to work, and by freeing the time, energy, and money of caregivers. (See "Guinea Worm Eradication Campaign" in the Suggested Readings.)

ACCESS TO FOREIGN MARKETS. If an enterprise has access to foreign markets without paying large tariffs, this increases the chance that the enterprise can achieve economies of scale, reduce the average cost of a unit of output, and expand further. This can also aid in development.

Other Factors in Development

One commonly held view of what causes the wealth and poverty of nations is that the rich countries are those with abundant natural resources. The wealth of the United States, for example, is often attributed to its vast resource base. An abundance of natural resources is a source of income, of course, and lack of access to coal, iron, and sources of power can be a barrier to growth. (See box "Why Did the English Economy Take Off? Competition, Coal, or Colonies?" in Chapter 1.)

But an abundant resource base is hardly a guarantee of affluence. The tiny country called the United Arab Emirates is oil-abundant, and its per capita income is among the highest in the world, but the mineral-rich Democratic Republic of the Congo has the lowest per capita income in the world. A country with higher per capita income than the United Arab Emirates is resource-poor Denmark. And Japan—which during 1950–1990 was one of the fastest growing of the world's major economies—is practically without natural resources. Japan has no iron ore, no oil, and very little coal, yet it produces more steel than the United States. Researchers have found evidence that resource-rich countries that are poor tend to be undemocratic, corrupt, and somewhat more prone to civil conflict. One difficulty is that if the government is supported entirely by resource revenues rather than taxation of the population, government officials may not pay much attention to the people's needs and views.

Indeed, having abundant resources that are being extracted and sold at a profit is not enough. What a country needs is to gain a substantial share of the revenue from those resources, and then to spend it on well-designed development initiatives. Both are problems. From about the 1940s to the 1960s, oil-rich developing countries began adamantly demanding that transnational oil companies pay higher royalties on the oil they extracted. Finally some of them simply nationalized the oil companies.

Iran's popular prime minister Mohammed Mossadegh and the Iranian parliament did nationalize the Anglo-Iranian Oil Company in 1951, but they paid dearly for it by a U.S.-backed coup that ousted Mossadegh in 1953, and oil returned to foreign control. However, since the Shah—who was restored to the throne after a brief period outside the country—spent much of the oil revenues on weapons,

the majority of ordinary Iranians remained fairly poor. This eventually helped to ignite the 1979 Iranian revolution, which among other things wrested control of the oil back into Iranian hands, so that 100 percent of the proceeds of oil sales went to the government.

In other words, the point is not only how much surplus is produced in a country, but who gets that surplus. Resources may result in significant production, but if that production is initiated and controlled by a foreign TNC, only a fraction of the surplus may remain in the country and so its contribution to development may not be large.

Production that is owned and controlled by nationals of the country will at least result in the surplus going to those nationals. Even then, however, corruption may cause that surplus to leave the country and be spent on European real estate or shopping trips to Paris, perhaps after first being hoarded in an offshore account. So a relatively honest, well-functioning government is also important for development.

Yet despite a long list of pitfalls, development does take place, though with considerable unevenness among countries. Figure 15.6 shows differences in the long-term rate of growth of productivity among selected countries.

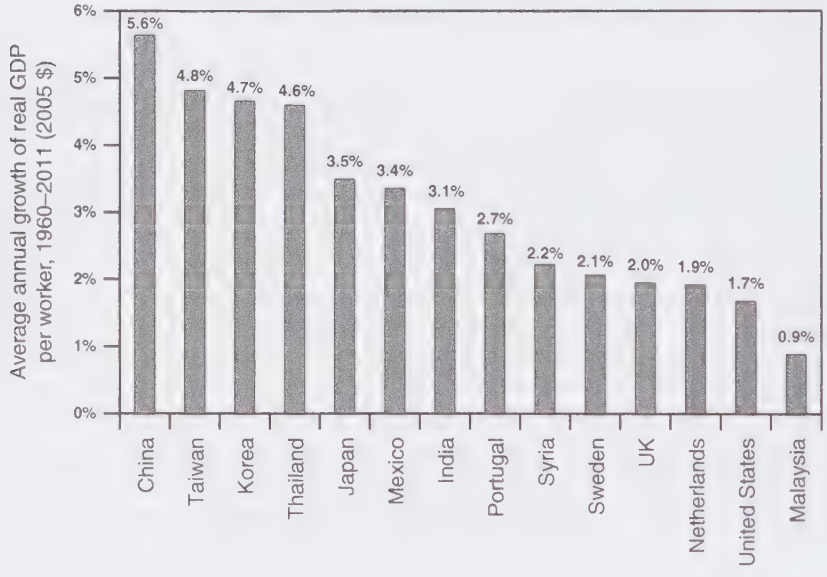


FIGURE 15.6 Average annual growth of real GDP per worker, 1960–2011.

This figure shows how rapidly productivity grew, in *real* terms, in various capitalist economies in about the last half century. "Real" means that "nominal" values, that is, values in "current" dollars, have been adjusted for inflation, in this case by calculating each country's GDP as if it had been priced each year in "constant" 2005 dollars. Productivity is usually defined as the amount of output produced per hour of labor, but data on hours of work are not available for all nations. An alternative measure is the total output of an economy (GDP) divided by the number of people employed in a given year. This is the measure used here, so the productivity growth of each country shown in the figure is the percent change in this variable from year to year, averaged for the years from 1960 to 2011.

Source: Penn World Table, Version 8.1, University of Groningen, April 13, 2015, available at http://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt8.1 (DOI: 10.15141/S5NP4S); Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table," *American Economic Review* 105(10): 3150-3182, available for download at www.ggdc.net/pwt.



What's a Hamburger Worth?

How can we compare living standards in two economies, like those of the United States and China? Each country, of course, first measures its output (income) in its own currency. Although many European nations now have a common currency (the euro) and can measure and compare their various economies directly, we still must solve the multiple currency problem to be able to compare most of the world's economies.

China's GDP, for example, is measured in *yuan*, while that of the United States is measured in dollars. An Internet search for "USD to yuan" in mid-2016 revealed that a dollar could buy 6.64 Chinese *yuan* at that time. This is also called "the *exchange rate* of the dollar against the *yuan*." In mid-2016 a Big Mac hamburger cost $5.04 in the United States, on average. But it only took $2.79, converted into *yuan*, to buy a Big Mac in China. As pointed out in a July 2016 issue of *The Economist*, this indicates that the value of the *yuan*, expressed in dollars, was less than it should have been (the *yuan* was *undervalued*). Shouldn't a Big Mac cost the same everywhere? So comparing the size of Chinese and U.S. average incomes based on the dollar-*yuan* exchange rate was misleading at that time.

Although one cannot make valid international comparisons using only the prices of Big Macs in different countries, the Big Mac story contains more than a grain of truth. The World Bank and the International Monetary Fund have for years used the *Purchasing Power Parity* (PPP) method of making international economic comparisons.

Economists are usually interested in comparing the living standard in one country with that in another, not comparing the prices of individual goods. They might ask, "How many dollars would a person need in the United States to have a standard of living similar to the average living standard in China?" This comparison needs to consider not only that prices of goods and services differ in the two countries, but also that their people in eat different foods, wear different kinds of clothes, and generally buy different baskets of commodities. The *Purchasing Power Parity* (PPP) method of comparison tries to take all these things into account. Exchange rate comparisons do not.

Use of the PPP approach often greatly affects results. In 2015, for example, China's per capita income was one-seventh that of the United States when incomes were compared using the exchange rate between *yuan* and dollars. Using the PPP approach, however, China's per capita income was almost one-fourth that of the United States. Since the PPP measure is well designed for comparing living standards, we use it whenever possible in this book. (See Figures 15.1, 15.2, 15.3 and 15.7 in this chapter.)

Sources: "The Big Mac Index: Exchange Rates, to Go," *The Economist*, July 21, 2016, available at www.economist.com/content/big-mac-index; World Bank, World Development Indicators.

Variations in Rate of Growth of Productivity

The rate of growth of productivity differs greatly even among advanced capitalist nations. From Figure 15.6 we can see that since 1960 the growth of productivity (output per worker) in the United States has lagged behind that of other capitalist nations. Productivity growth in South Korea and Japan was exceptionally rapid because they were able, early in the period, to borrow advanced technology from the leading nations. Most of the European countries have also enjoyed productivity growth higher than that of the United States.

As the figure shows, in China, over half a century, average annual growth of productivity (measured by *output per worker*) has been 5.6 percent, while in Sweden and the UK it has been about 2 percent, and in the United States it has been even lower, at 1.7 percent. Sweden's relatively low rate of growth of productivity has been largely due to the fact that over this period Swedish workers greatly reduced the number of hours they worked per year, thus trading off faster growth of output for significant increases in leisure time such as vacations.

The governments of the rich countries often adopt policies that impede development in the poor countries. Intellectual property rights have already been mentioned. Development in poor countries is also impeded when rich countries' governments—through financial assistance, military aid, direct military intervention, and other means—protect the political position and economic privileges of the backward-looking traditional elites in poor countries. Because these elites often resist economic and social change, external support of them effectively blocks economic development.

Agricultural subsidies in rich countries have long been a problem for the development of poor countries. When U.S.-subsidized corn, rice, wheat, and cotton



Does Inequality Grease the Wheels of Progress?

For decades economists believed (and taught their students) that rapid growth in productivity and other measures of good economic performance require high levels of income inequality. A "trade-off," they called it: if you want greater equality, you must pay for it with inefficiency and slower economic growth. They reasoned that heightened inequality gives stronger incentives to work hard, to manage well, and to take the risks necessary for innovation. The idea was even enshrined in the title of a famous book, *Equality and Efficiency: The Big Tradeoff*, by the Brookings Institution economist Arthur Okun, who wrote, "The conflict between equality and economic efficiency is inescapable." The alleged trade-off became a powerful justification of inequality: a more just society would be nice, but simply too expensive.

Far from being inescapable, however, the tradeoff turned out to be hard to find in practice. Suspicion about its existence was prompted by the fact that in the last quarter of the twentieth century the income distributions in Japan, Korea, and Taiwan were among the most equal in the world, yet productivity growth in these countries was far higher than in highly unequal countries like South Africa and Brazil, as well as in the United States, the UK, and other more unequal countries.

Many economists now think that economic inequality impedes rather than promotes efficiency and the growth of productivity. Inequality often breeds conflicts such as strikes and fosters hostile relationships between employers and employees, wasting output and effort. Inequality of opportunity, including racial and other forms of discrimination, deprives many of adequate education, with a resulting loss of talent for the economy as a whole. And when sharp inequalities of wealth leave most people without the assets they would need to go into business on their own, the economy loses their capacities for management and innovation as well as the benefit of the incentives inherent in owning the fruits of one's own labor. In these cases inequality throws sand in the wheels of progress.

Sources: Luxembourg Income Study, available at: http://www.lisproject.org; Penn World Tables—International Comparisons of Production, Income, and Prices 9.0, Robert C. Feenstra, Robert Inklaar, and Marcel P. Timmer, "The Next Generation of the Penn World Table," *American Economic Review* 105, no. 10 (2015): 3150–82, available for download at http://www.ggdc.net/pwt; Samuel Bowles and Herbert Gintis, *Recasting Egalitarianism: New Rules for Markets, States, and Communities* (London: Verso, 1999).

(and subsidized crops from Europe and Japan as well) flood world markets, small farmers in many developed countries suffer—an issue that has been contentious for well over a decade in the World Trade Organization. Yet rich countries have used this as a bargaining chip to get concessions in other areas that are sometimes costly for poor countries.

In sum, a country with a high level of income, a large market, decades of experience with modern production techniques, and a large surplus to invest in new equipment, education, research, and infrastructure has a huge competitive advantage over a less well-endowed country. And the advantage may grow over the years as the rich country uses its gains to maintain its productivity growth, while the poor country cannot get productivity off the ground.

The fact that development is uneven does not mean, of course, that latecomers can never catch up. Germany and the United States were both latecomers—after the United Kingdom—in the nineteenth century, but they rapidly overcame their disadvantages and now have per capita incomes much higher than the UK. To cite another example, South Korea was a latecomer that took off during the 1960s and 1970s using policies described earlier in this chapter. And leading nations do not always stay ahead. The United States overtook the UK, and China is now growing much more rapidly than Japan. But as already noted, some of the policy doors that were open in that earlier period have been closed by the WTO and other agreements, leaving less policy space for the governments of developing countries, and so fewer avenues for escape from low productivity growth.

Figure 15.7 illustrates the dramatic contrasts of success and failure in the process of uneven development over half a century. In the two decades 1960–1980, per capita income in the small southern African country of Botswana grew by a factor of eight (buoyed by diamond discoveries), while South Korea's per capita income grew by almost a factor of six. During the fifty-year period, however, Ghana's per capita income slightly fell. A contrast affecting far more people is China coming from behind and then pulling away from India after 1980: during 1980–2000, China's average income grew by nearly a factor of four. The vertical axis is a logarithmic scale, so the distance from $400 per capita to $4,000 per capita is the same as the distance from $4,000 to $40,000 per capita. The prolonged and repeated civil wars in the Democratic Republic of the Congo, with literally millions killed, were devastating to the economy and set back the country by decades, reducing its real average income over the fifty years shown to less than one-fourth of its 1960 level.

Foreign Investment and Development

When a developing country hosts foreign investment, does this always aid its development? In the two decades after World War II, many colonies of European countries achieved independence, often after bitter and costly nationalist struggles. In newly independent nations of Africa and Asia, leaders sought to meet the long-suppressed needs of their people for decent work, income, and dignity. Many leaders sought to bargain with TNCs that had been extracting their oil or mineral resources, but paying only small royalties to the government. Some insisted that henceforth, foreign investment would be welcomed only if it contributed to national development. New laws were passed to regulate foreign investment to this end.

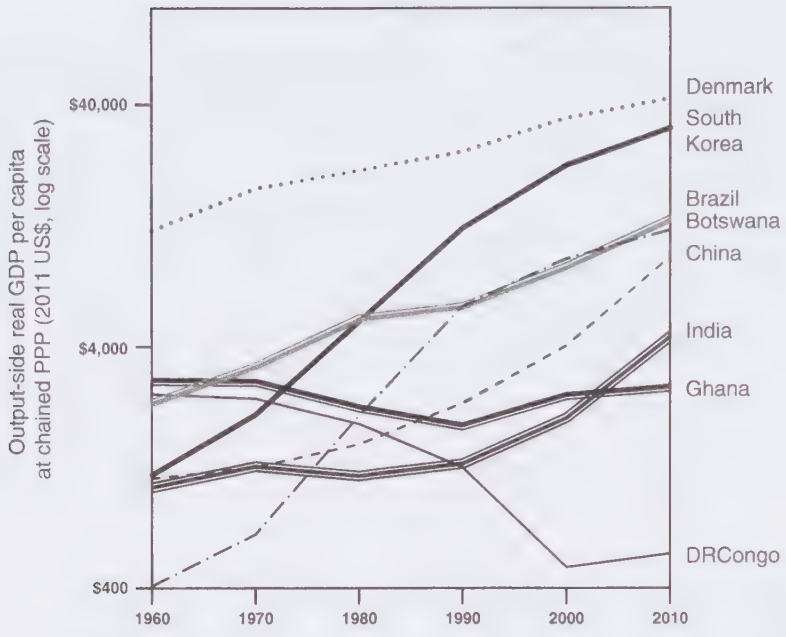


FIGURE 15.7 Uneven development, 1960–2010.

Three-fourths of the eight countries shown here experienced increasing average output per person over the half century shown. DRCongo's sharply declining growth was largely the result of fierce civil conflict.

Source: Penn World Tables—International Comparisons of Production, Income, and Prices 9.0, Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer, "The Next Generation of the Penn World Table," *American Economic Review* 105, no. 10 (2015): 3150–82, available for download at http://www.ggdc.net/pwt. Author's calculations from Table CGDPo (2011): Output-side real GDP at current PPPs (in mil. 2011 US$), divided by population. See also http://www.rug.nl/research/ggdc/data/pwt/.

But while leaders continued through the 1970s to assert their governments' right to impose conditions on foreign investment, in many countries they also actively sought to attract such investment. Brazil's successful attempt to create an auto industry in the 1950s and later was one example; Mexico made similar efforts in the 1960s. After all, TNCs had modern technology, and developing countries wanted both managers and workers to learn that technology through on-the-job training in factories that used it.

So governments behaved as if they thought investment by TNCs could bring greater benefits than costs, if adequately regulated. By the 1980s, they began seeking to attract investment even without regulating it much. The benefits they saw included (and still include) new knowledge and skills related to modern technology, employee compensation, and the ripple effects of re-spending of that money in the local economy. It sometimes also included the increased appeal to other firms, such as suppliers, to locate nearby, creating more jobs and incomes.

There were also costs, which included any government subsidy, any displacement of local businesses producing similar products, and any negative effects such as environmental damage that was not cleaned up or compensated for. An example of the latter is environmental damage to native lands in Ecuador by Texaco's careless disposal of wastes from oil and natural gas extraction during 1964–1990. Decades of legal efforts to get ChevronTexaco to pay for cleanup led to a $9.5 billion

judgment against Chevron, and legal action continued to try to compel Chevron to actually pay this amount. In other places in the world, toxic runoff from mining operations has contaminated water systems in several countries, generating widespread protest.

Another problem is that if several jurisdictions compete to attract a TNC's investment project (like an auto plant), these jurisdictions will compete against one another, bidding up the subsidies they offer to the foreign investor, even to the point where the subsidy eats up much of the expected benefit. And if officials rely on information provided by the TNC that is more likely to overestimate than to underestimate the benefits, then the community can even bid too high, giving up more than all the benefits. Unfortunately, the precise impact of a project is hard to measure, even after the fact, so we never actually know for sure.

Nevertheless, because TNCs are virtually the only way a developing country can get access to both recent technology—the knowledge, the equipment, and the skills—and world markets, TNCs are much sought after and regarded as central to development.

Investment Decisions by Transnational Corporations

As noted above, a TNC makes global profits by locating its affiliates in various countries so that each affiliate contributes to its profitability in some way. But decisions about where to establish an affiliate do not all rest on the same considerations. (Subsidies offered do matter, of course, but we now focus on other factors in the decision.) The main advantage of one site may be abundant oil, minerals, or other natural resources; the attraction of another may be closeness to large markets, or a workforce with a high level of skill, or exceptionally good infrastructure for transportation, communication and energy supply, or a government eager to please foreign investors—or several of the above. The major U.S. TNCs typically operate in dozens of countries, and some, such as the Ford Motor Company, operate in more than one hundred countries. Profits from international operations, as we saw in Figure 7.3, have increased in importance relative to profits made in the United States.

What appears to determine the location of investments in production operations? The following factors are some of the key ones, not necessarily in order of importance, and not necessarily the same in every industry. Notice that many of them coincide with the requisites for development from the standpoint of developing countries:

- A labor force with the necessary skills
- For extractive industries (mining, oil, gas), substantial resource deposits
- A government that provides a stable, favorable business environment, with little risk of political upheaval or of erratic changes in the value of its currency
- Good infrastructure: a well-functioning transportation and communication system and reliable, continuous electrical power supply
- Absence of violent civil or interstate conflict
- Location close to major markets
- Low wages relative to the skill level of workers

For firms looking for sites for production operations, low wages are a strong attraction only if workers already have the required skills, or the work to be done is low-skilled, or if workers can be trained relatively quickly, at relatively low cost, to do it, building on the knowledge and skills they already have. Of course, an adequate infrastructure, political and currency stability, and low materials costs are also needed. To paraphrase a common saying, "The bitterness of low skill or poor infrastructure lingers long after the sweetness of low wages is forgotten." Figure 15.8 shows that wage differences around the world are indeed great. As Chapter 13 made clear, the relative importance of skilled workers also differs by industry and may change over time (recall the poultry processing factory); some factory work requires relatively little training, while other work demands much more skill and experience.

Where do U.S. firms invest? The U.S. government collects data each year on the value of investment abroad by U.S. corporations, by country. The data are for "direct investment," a term defined as holding at least 10 percent ownership (which is usually enough for a controlling share) of the stock in the corporate entity owned in another country. Much of this investment is in rich countries, which may be surprising in view of what earlier chapters say about TNCs moving abroad to pay low wages. However, many of these rich countries are of a special type—those that use TNCs to reduce the amount of taxes they owe, and the amount they pay.

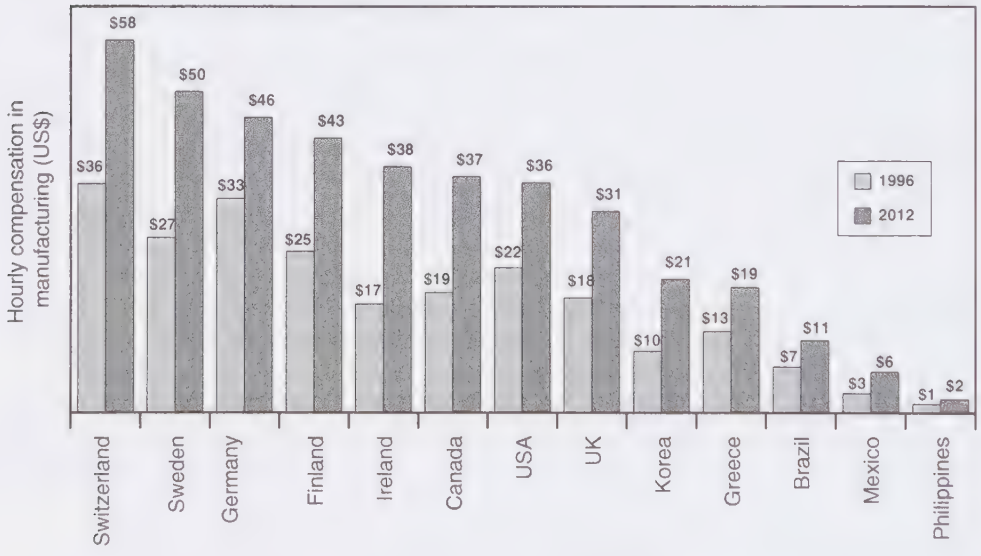


FIGURE 15.8 Hourly wages and benefits in the manufacturing sector in selected countries, 1996 and 2012.

This figure compares the total hourly compensation of manufacturing sector workers in sixteen countries in 1996 and 2012. Compensation in each country is first measured in units of the country's own currency and then converted into U.S. dollars at market exchange rates in each year. This measure is good for comparing the hourly labor cost to employers in different countries. It is not appropriate for comparing the relative purchasing power of labor earnings in manufacturing in different countries, since conversion to dollar values uses market exchange rates, not purchasing power parity.

Source: U.S. Bureau of Labor Statistics, *International Labor Comparisons*, August 2013. Hourly Compensation Costs (Wages and Benefits), in the Manufacturing Sector, Report through 2012, Table 1: Hourly Compensation Costs in Manufacturing, U.S. Dollars, and as a Percent of Costs in the United States, ichcc.xls, available at: http://www.bls.gov/fls/#compensation.

Transnational Investment and Tax Havens

A **tax haven**, or offshore financial center, is defined here as a jurisdiction that offers strict banking secrecy laws and low or zero tax rates on foreign-owned business activities.

Companies locate their affiliates in places that they judge will contribute to their profitability. But in recent decades this has often meant locating a sizeable share of their affiliates in *tax havens*, also known as offshore financial centers. A tax haven has two distinctive features. First, its government imposes a low or zero tax rate on the incomes of foreign-owned corporate entities located in the tax haven. Second, strict secrecy laws protect information about these entities from the prying eyes of the tax authorities or crime investigation authorities of other countries.

Such tax havens include a number of Caribbean islands, the best known of which are perhaps the Cayman Islands. And there are many other islands in the world that are, to one degree or another, tax havens—some of them in the Pacific, some just off the coast of the United Kingdom, some in other locations. Key and longstanding tax havens also include some countries of Europe: Switzerland, the Netherlands, and Luxembourg, for instance—as well as some jurisdictions that are neither countries nor islands, including, according to some researchers, the state of Delaware.

One main tool accountants can use to reduce a company's income tax liability to the IRS is *transfer pricing*. Within a transnational firm that has a parent firm in its home country and many affiliates in other countries, one affiliate often sells goods (inputs or outputs) to another in a different country. That sale within the global firm is not a "market" or "arm's length" transaction (the global firm is selling to itself), and the price set for the transaction may not be the price the goods would fetch on the open market. There are reasons why the corporation might not want the sale price to be the market price of the good. In fact, the price that one affiliate charges another for such transfers can in principle be almost anything. So the global firm's accountants choose a price that gives the lowest tax liability. They do have to be a little bit careful, because under IRS rules, if the price departs too much from the price on the open market (the "arm's-length" price), then if the IRS finds the firm in violation of the tax laws, the firm can be penalized. Nevertheless, it is expensive and difficult for the IRS to investigate and prosecute such violations, and the IRS does not always win. In practice, corporations evidently do a lot of transfer pricing in an effort to reduce the taxes they owe.

But how does this work? The trick is to make the global firm's profits show up in places (tax havens) where those profits will face a low or zero tax, and where the global firm's affiliate is shrouded in secrecy by local law, while reporting low profits in countries with higher tax rates. A hypothetical example will illustrate. Suppose a TNC has an affiliate in the Cayman Islands, in the Caribbean, a well-known tax haven. Suppose then that the TNC's Mexican affiliate that produces finished goods sells this output to the tax haven affiliate at a price just above the Mexican firm's cost of production (so that almost no profit is recorded in Mexico). The Cayman Islands affiliate then turns around and resells the output to the U.S. parent firm at a much higher price, producing a large profit in the Caymans. The U.S. firm slightly marks up the price and sells the good wholesale or retail, recording only a small profit. This transfer pricing has, as if by magic, shifted profit away from Mexico, where corporate profits are taxed at 30 percent, and the United States, where they are taxed at 35 percent, to the Cayman Islands, where they are taxed at 0 percent. Notice that "profit shifting" in this way *reduces the profit reported in the United States.* Thus the downward trend in Figure 10.2, showing declining U.S. corporate profit rates, is in part an artifact of accounting practices rather than an actual change in economic activity. By the same token, the upward trend in Figure 7.3, showing the share of foreign source income in

all U.S. corporate profits, arises from "profit shifting" accounting practices that simultaneously reduce reported U.S. profits and increase reported profits from abroad, both of which reduce tax liability.

There is more to the story. The global corporation based in the United States still would have to pay income tax on the Cayman Islands profits *if and when they are brought back to the United States*. But as long as these earnings are reinvested outside the United States, the tax payment can be deferred indefinitely.

This is not a small matter. The watchdog group Citizens for Tax Justice has estimated that in 2016, $2.4 trillion (that's $2,400,000,000,000) of unrepatriated profits were being held offshore by just the U.S. transnational firms who were in the Fortune 500. CTJ estimated the unpaid taxes on that amount at up to $695 billion. Even these taxes may never get paid, for two reasons. One is that, under pressure from TNCs, in 2004 Congress declared a tax holiday for firms holding profits abroad, and about $300 billion was repatriated and firms (legally, under the tax holiday) paid about one-seventh of the taxes they owed. Chapter 19 will take this story a little further to explain how this actually worsened the problem, because it motivated *more* firms to hold *more* money abroad for years or decades at a time.

It is not surprising, then, that a sizeable share of U.S. investment abroad is in tax havens. "Investment" in this case is a word that should be used cautiously. What actually happens in places like the Cayman Islands is that a firm establishes an entity that exists in cyberspace, has a local address but no labor force, and no other visible signs of economic activity except a single building in the Cayman Islands (Ugland House) with small office shared by over 18,000 corporations (a minority of which are U.S.-based). The TNC's shell entity reports exceptionally high profits on its "activities."

Thus when U.S. firms establish foreign affiliates, they do so with a variety of purposes. Some affiliates extract or manufacture output abroad and sell it; their contribution to the firm's global profits may be from taking advantage of economies of scale, or from exploiting deposits of minerals or fossil fuels, or both. Some affiliates are located in, or close to, large markets, saving on transportation costs and delivery times, And some contribute to minimizing the tax bite, often by carrying on accounting practices in tax havens that shift reported profits to places in which they will be taxed lightly or not at all.

How does the existence and corporate use of tax havens affect development? First, TNCs can use tax haven entities in bribing persons in developing countries. Second, this is especially effective because it introduces such officials to a way to hide any ill-gotten funds. Third, wealthy persons in developing countries who hide their funds in tax havens can easily avoid being taxed on those funds. Fourth, they can even establish a tax haven entity that appears, to their home country government, to be a "foreign investor," and is therefore eligible for the government's subsidy program for foreign investors. All these strategies undermine government finance, perpetuate corruption, and harm development.

To be clear: most nonfinancial TNCs invest abroad both in production operations and tax havens. In the United States, over 70 percent of Fortune 500 firms have tax haven subsidiaries, where skilled labor, markets, and infrastructure are irrelevant. There, the key considerations are the tax rate on foreign firms' activities (sometimes zero, but not always), the absolute commitment to secrecy regarding these activities, and the stability of this arrangement. Figure 15.9 shows the share of several tax havens, or groups of tax havens, in the total stock of U.S. direct investment abroad, as reported to the U.S. Bureau of Economic Analysis. It shows that

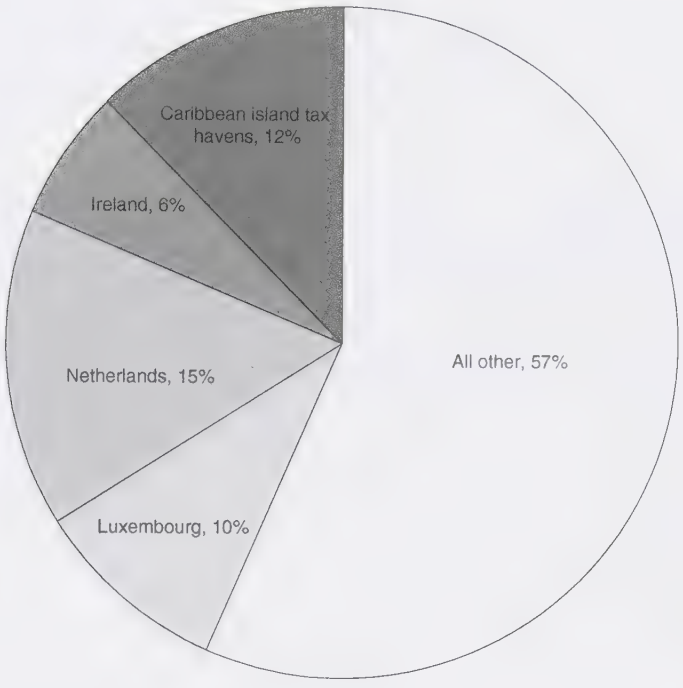


FIGURE 15.9 U.S. Direct Investment Abroad, 2014.

This figure emphasizes the importance of tax havens as sites for U.S. direct investment outside the United States. "Caribbean island tax havens" include Barbados, Bermuda, the Bahamas, the Cayman Islands, the British Virgin Islands, and others. Ireland, Luxembourg, and the Netherlands are also tax havens, as is Switzerland. A few additional tax havens are not shown, but account for only a very small percent of the total. The term "direct investment" refers to ownership of at least 10 percent of the stock in an affiliate, which usually gives full control.

Source: U.S. Bureau of Economic Analysis, "U.S. Direct Investment Abroad for 2012–2014: Detailed Historical-Cost Positions and Related Financial Transactions and Income Flows," *Survey of Current Business*, September 2015, 1-36, Table 14: U.S. Direct Investment Abroad: Selected Items by Detailed Country, 2010–2014.

about half of the value of those holdings lies in tax havens. It also happens to be true that slightly over half of all the profits reported by U.S. TNCs abroad is reported as having been made in tax havens.[4]

Thus there are two distinct reasons why U.S. firms (and the transnational firms of other rich countries) often invest in rich countries rather than in poor ones. The first reason applies to firms whose investment is in production operations. For these firms, the key considerations are closeness to large markets, availability of highly skilled and educated labor, excellent infrastructure (power supply, transportation, communication, health system), and both political and currency stability. Political stability includes security of property rights. Extractive industries like mining and fossil fuel extraction are, of course, exceptions, since they must locate wherever the resource deposits are found, and transport skilled personnel there or train local workers. The second reason is that some rich countries are tax havens.

[4]Gabriel Zucman, "Taxing Across Borders: Tracking Personal Wealth and Corporate Profits," *Journal of Economic Perspectives* 28, no. 4 (Fall 2014): 121–28, Figure 2, 128.

The way in which TNCs conduct their production operations, or the way in which the firms to which they outsource some of these operations, also affects development, as the box "The High Cost of Low-Priced Clothing" explains.



The High Cost of Low-Priced Clothing

Garment industry competition is fierce. A few pennies' difference in cost can make or break apparel producers—so every penny counts. This fosters dangerous working conditions: Workers face not only pressure to work faster, but excessive risk of injury or death because many factory owners are reluctant to spend money on safety measures.

In 1911, the famous Triangle Shirtwaist Factory fire in New York killed 146, some of them crushed because locked doors prevented their escape. This avoidable tragedy spurred garment workers to unionize and raised awareness of workplace safety issues. A century later history repeated itself in Dhaka, the capital of Bangladesh, on eight times the scale. On April 24, 2013, the eight-story Rana Plaza building, housing garment factories, collapsed, killing 1,135 out of about 5,000 workers and injuring many more. Five months earlier, the nearby Tazreen Fashions factory fire, killing 113, had provoked enormous outrage: 100,000 people were said to have attended the burial of about half the victims. Rana Plaza was much worse.

About 80 percent of Bangladesh's exports are clothing, and it is the world's second largest clothing producer after China. Pressure to respond to domestic and international concerns was intense. Rights groups (the Clean Clothes Campaign, the International Labor Rights Forum, and others) pressed global retailers to whom Tazreen and Rana Plaza firms were key suppliers to donate to a $30 million fund to compensate families of the dead and injured. Some gave, some did not. Wal-Mart gave just $0.25 million, though owned by the richest family in the United States, with a $130 billion fortune. Nevertheless, $2.2 million was distributed to the Tazreen workers' families by mid-2016, and compensation for Rana Plaza had also been initiated.

But what the 5 million garment workers of Bangladesh need is safe working conditions. These disasters were preventable—and there have been other smaller fires as well. According to the police report charging forty-one people, including the owner, with murder in the Rana Plaza collapse, one day before the disaster an engineer inspected the building and deemed it unsafe: A heavy generator on an illegally constructed upper floor was shaking and weakening the structure, as numerous cracks showed; managers told workers to work anyway. At Tazreen, the fire started where flammable yarn and fabric were piled near electrical generators—not in fireproof storage rooms as required by law.

Leading in inspection and remediation of safety violations have been the multi-stakeholder Accord on Fire and Building Safety and the Alliance for Bangladesh Worker Safety, covering a fourth of garment factories and half of the nation's garment workers. With government and International Labor Organization help, over half of garment factories were inspected by mid-2016. Fixing of physical safety violations was slow, but the first thirty-one factories that completed remediation all were suppliers to the Alliance and the Accord. Years after Rana Plaza, most of the fixes remain to be done, and garment workers continue to risk their lives for a job whose minimum wage is just $68 a month.

Sources: Jim Yardley, "Horrific Fire Revealed a Gap in Safety for Local Brands," *New York Times*, December 6, 2012; Julfikar Ali Manik and Nida Najar, "Bangladesh Police Charge 41 With Murder Over Rana Plaza Collapse," *New York Times*, June 1, 2015; Alliance for Bangladesh Workers' Safety, *Protecting and Empowering Bangladesh's Garment Workers*, Second Annual Report, September 2015, available at http://www.bangladeshworkersafety.org; Ibrahim Hossain Ovi, "Only 31 RMG Factories Fully Remediated 3 Years after Rana Plaza Disaster," *Dhaka Tribune*, April 23, 2016, available at http://archive.dhakatribune.com/rmg-factories/2016/apr/23/only-31-rmg-factories-fully-remediated-3-years-after-rana-plaza-disaster; "Rana Plaza Building Collapse," https://failures.wikispaces.com/Rana+Plaza+Building+Collapse.

Conclusion

In this chapter we have posed some of the most difficult questions in economics, those concerning economic growth and world inequalities. The question of long-term economic progress—the wealth of nations—was one of the most important for the founders of political economy. The question remains an important one today.

We have seen that both the gap in technology and the gap in power between poor and rich countries have a tendency to persist because there are vicious circles from which it can be difficult to escape. As Chapter 6 began to explain, the rules and institutions that govern economic activity (the constitution, laws, regulations, policies, and programs)—both those within countries and those written into international agreements—greatly affect what is produced, how it is produced, and who gets the net output. They also affect how much revenue a nation's government is able to collect, and they sometimes affect how highly motivated government officials of developing countries are to act in the public interest and choose actions that help break out of such vicious circles. While at first glance one might think foreign investment in poor countries would tend to close the productivity and income gap, the drawbacks (poor infrastructure, low skill level) often more than offset the advantages, and a large portion of TNC holdings abroad are actually in tax havens, not in poor countries.

We will continue to examine the government's role in the economy in Chapter 19. In the next three chapters we lay the foundation for our discussion of the government's role by looking at how the economy works from a macroeconomic perspective.

SUGGESTED READINGS

Alice Amsden, *Asia's Next Giant* (New York: Oxford University Press, 1992, paperback).

Samuel Bowles, *Microeconomics: Behavior, Institutions and Evolution* (Princeton, NJ: Princeton University Press, 2004).

Samuel Bowles, Steven N. Durlauf, and Karla Hoff (eds), *Poverty Traps*, reprint edition (Princeton, NJ: Princeton University Press, 2016).

James Boyce and Betsy Hartmann, *A Quiet Violence: Life in a Bangladesh Village* (London: Zed Press, 1986).

James Boyce and Leonce Ndikumana, *Africa's Odious Debts: How Foreign Loans and Capital Flight Bled a Continent* (London: Zed Books, 2011).

Christopher Chase-Dunn, *Global Formation: Structures of the World Economy* (Oxford: Basil Blackwell, 1989).

Paul Collier, *The Bottom Billion: Why the Poorest Countries are Failing and What Can Be Done About It* (NY: Oxford University Press, 2008).

Paul Collier, V. L. Elliott, Håvard Hegre, Anke Hoeffler, Marta Reynol-Querol, and Nicolas Sambanis, *Breaking the Conflict Trap: Civil War and Development Policy* (New York: The World Bank and Oxford University Press, 2003).

Herman Daly and John Cobb, *For the Common Good: Redirecting the Economy toward Community, the Environment, and a Sustainable Future* (Boston: Beacon Press, 1989).

Mostafa Elm, *Oil, Power, and Principle: Iran's Oil Nationalization and Its Aftermath* (Syracuse, NY: Syracuse University Publications in Continuing Education, Revised edition, paperback, 1994).

Alexander Gerschenkron, *Economic Backwardness in Historical Perspective* (Cambridge, MA: Harvard University Press, 1962).

"Guinea Worm Eradication Campaign," available at https://web.stanford.edu/class/humbio103/ParaSites2005/Dracunculiasis/eradication.htm

James Heintz and Robert Pollin, *The Question of Sweatshops: Globalization and the Struggle for Decent Work* (New York: New Press, 2004).

Kenneth P. Jameson and Charles K. Wilber, eds., *The Political Economy of Development and Underdevelopment*, 6th ed. (New York: McGraw Hill, 1995).

Linsu Kim, *Imitation to Innovation: The Dynamics of Korea's Technological Learning* (Cambridge, MA: Harvard Business Review Press, 1997).

Peter Lanjouw and Nicholas Stern, eds., *Economic Development in Palanpur over Five Decades* (Delhi: Oxford University Press, 1998).

Branko Milanovic, *Global Inequality: A New Approach for the Age of Globalization* (Cambridge, MA: Belknap Press, an imprint of Harvard University Press, 2016).

Erik Reinert, *How Rich Countries Got Rich . . . and Why Poor Countries Stay Poor* (New York: Public Affairs, 2008).

Helen Shapiro, *Engines of Growth: The State and Transnational Auto Companies in Brazil* (Cambridge, UK: Cambridge University Press, 2006).

The World Bank, *World Development Report* (Washington, DC, and New York: The World Bank and Oxford University Press), available at http://www.worldbank.org/en/publication/wdr/wdr-archive.

Gabriel Zucman, *The Hidden Wealth of Nations: The Scourge of Tax Havens* (Chicago: University of Chicago Press, 2015).

Aggregate Demand, Employment, and Unemployment

16

One of the hardest things to understand about our economy is the fact that there are often (a) factories lying idle while at the same time (b) a significant number of people are looking for work, and (c) many human needs are not being met. Any two of these could be explained, but why all three are typically true in a capitalist economy is puzzling. People need the products that the unemployed workers could produce if they were put to work in the idle factories. Yet they are not being put to work.

For example, 2009 was a year in which the American economy shrank by about 3 percent in real terms. In midyear, one-third of the nation's industrial capacity lay idle (the capacity utilization rate was 67 percent). Even so, about $4.5 trillion worth of goods were produced in the United States that year (goods only, not including services)—but this was a drop of about $1 trillion from the year before, and at least $2 trillion less than what could have been produced at full capacity. Two trillion dollars of additional output, if divided up equally among the 307 million men, women, and children in the United States that year, would have given each an additional $6,500 or so worth of goods. The output was not lost because of a shortage of workers: on an average day in 2009 there were 13.9 million unemployed people looking for work—about 7 million more than in 2007. This has been called the Great Recession.

The fact that there can be idle tools, unemployed people, and unsatisfied needs all at the same time is hard to understand because it is inconsistent with our own personal experience. If we need a bookshelf and have the tools, lumber, and time, we can set to work and make ourselves a bookshelf. When we organize our own production—in families, in small groups, or on our own—we do not end up with idle tools, idle hands, and unmet needs. The reason is that when we produce for ourselves, our purpose is to make something that we need (or want). If we have unmet needs and the tools, time, and materials (or money to buy them) to produce what we need, we just do it.

The capitalist economy, by contrast, is organized quite differently. A business owner's purpose in producing commodities, as we saw in Chapter 5, is not to meet a need; it is to make a profit. If the employers who direct the production process do not believe it is profitable to do so, they will not engage in production; they will simply leave idle the tools and factories they own and cease hiring workers. The fact that someone may need the goods or services that could potentially be produced is

irrelevant. The system of capitalist competition forces owners of for-profit corporations to make their decisions solely with regard to profitability—and in the United States, at least, courts have ruled that if corporate leaders make their decisions on any other basis, they leave themselves vulnerable to lawsuits by shareholders. The one exception is if the corporation is organized as a "benefit corporation," a concept to be discussed briefly in Chapter 19.

One of the results of this profit-driven process is that the economy alternates over long periods between good times and hard times. In Chapter 7 we explained the long-term alternation of boom and crisis in the U.S. economy by chronicling its consecutive social structures of accumulation. In Figure 16.1 we present these long swings in concrete terms, showing the rates of growth of total output in the successive periods of consolidation and decay. The figure caption describes the four long swings pictured. Of course, during each of the long swings there are the less pronounced ups and downs associated with business cycles (see Chapter 7). While the long term booms and crises occur over several decades, the expansion and recession phases of a particular business cycle are generally completed in a period of less than ten years. Thus there can be four or even more business cycles within a long swing.

Each long swing begins with a long boom, during which the SSA (in its "consolidation" phase) makes possible relatively rapid growth, followed by a period of stagnation (the "decay" phase of the SSA). Business cycles occur within both the booms and the hard times of long swings, with the rate of output fluctuating in the

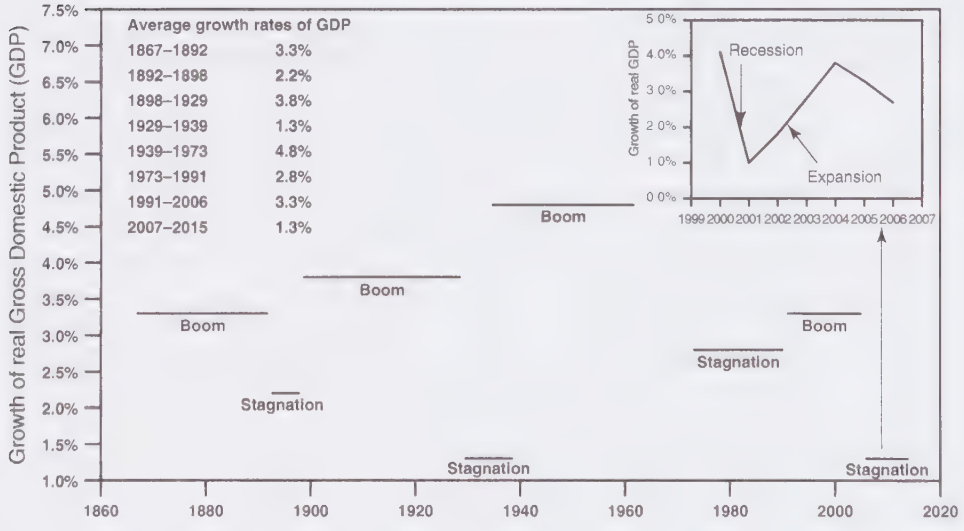


FIGURE 16.1 Long swings and business cycles.

A long swing is a boom followed by a period of stagnation; it typically extends over a period of thirty to fifty years. Business cycles are shorter ups and downs of the economy, normally over less than ten years (from peak to peak). This figure shows that the U.S. economy passed through three long swings from 1867 to 1991, and seems to be near the end of a fourth. The first (the stage of *competitive capitalism*) lasted from the 1860s to 1898, with a boom period from 1867 to 1892 and stagnation from 1892 to 1898. The second (*corporate capitalism*) was 1898–1939, with good times until 1929 and then the Great Depression of the 1930s. The third long swing (*regulated capitalism*) covered about 1939–1991. Around 1991 we entered into a fourth long swing (which we refer to as *transnational capitalism*), with a boom until 2007 and stagnation afterward.

Sources: U.S. Bureau of Economic Analysis, National Income and Product Accounts, Table 1.1.1: Percent Change from Preceding Period in Real Gross Domestic Product, available at http://www.bea.gov/iTable/iTable.cfm?ReqID=9&step=1#reqid=9&step=1&isuri=1

short term without altering the long-term trend. The inset graph in this figure depicts a fairly typical business cycle: a high rate of growth in 1999 turned into a falling and then a negative quarterly growth rate (with declining output for two quarters) until a trough was reached in 2001, followed by accelerating growth during the boom of 2002–2006. (The annual growth rate for 2001 did not fall below 0 percent because this particular recession was short, and there were just two quarters of negative growth and two quarters of greater positive growth during the year.)

When the economy expands over a considerable length of time, employers tend to hire more workers, so the level of unemployment generally falls during long booms. Conversely, a long period of decay is usually characterized by rising unemployment. The unemployment rate in the U.S. economy between 1890 and 2014 is shown in Figure 16.2. Short-term ups and downs in employment mark business cycles of up to ten years in length. While the unemployment rate tends to go up in periods of economic stagnation or decline during peacetime, it generally falls with the onset of a war as workers are hired to produce military goods and as people who would otherwise be in the civilian labor force are drawn into military service. For instance, the very high unemployment rates that had plagued the United States

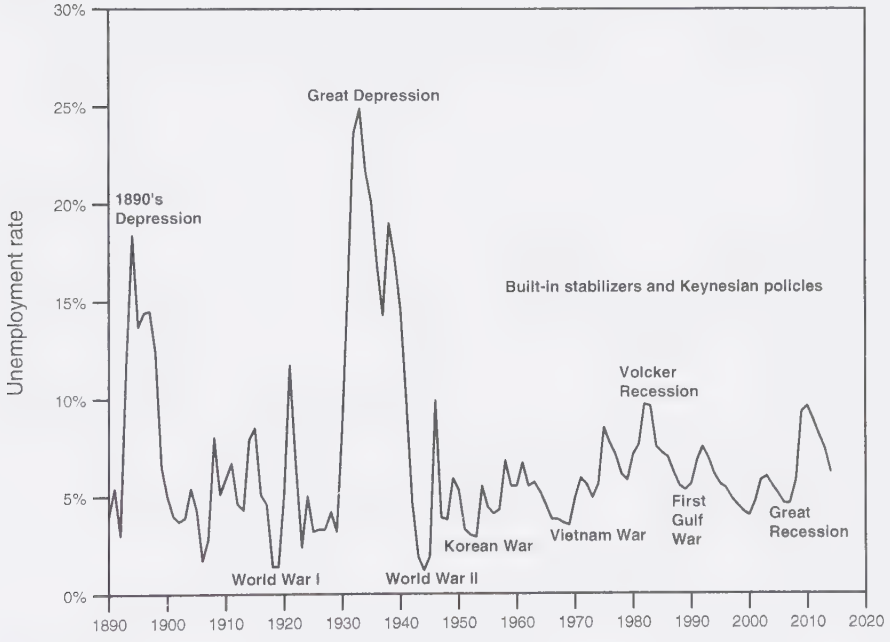


FIGURE 16.2 The U.S. unemployment rate, 1890 to 2014.

The unemployment rate presented in the figure is the ratio of the number of people officially counted as unemployed to the number officially counted in the civilian labor force. The unemployment rate reached about 18 percent in the 1890s and peaked at 25 percent during the Great Depression of the 1930s. Note that the rate of unemployment was held to less than 10 percent in the half-century after World War II by means of Keynesian policies and built-in stabilizers such as unemployment insurance and other social insurance programs established as part of the "New Deal" in the 1930s.

Sources: U.S. Bureau of Labor Statistics: Databases, Calculators, and Subjects, Labor Force Statistics from the Current Population Survey, series LNU04000000, available at http://data.bls.gov/timeseries/LNU04000000?years_option=all_years&periods_option=specific_periods&periods=Annual+Data.

throughout the 1930s came down dramatically with the onset of World War II. Especially in recent decades, unemployment rates have typically fallen rather slowly after a recession ends and the economy once again begins to grow.

Two main ideas are central to this chapter. The first is that *levels of employment and unemployment depend on the total demand for goods and services*, and the second is that *government policies can reduce the level of unemployment and smooth out fluctuations in the level of total demand*. These main ideas are expressed in seven key points:

1. The total value of goods and services demanded need not equal the total value produced in a year: total demand may exceed or fall short of supply.
2. When demand for goods and services exceeds supply, firms will hire more workers; when demand falls short of supply, firms will lay off workers; and when demand equals supply, firms will not alter the size of their workforce.
3. When demand for goods and services equals supply (and firms are neither expanding employment nor cutting back), there will generally be some unemployment.
4. Both a government's demand for goods and services and its policies regulating the rate of interest affect the level of employment; government deficits and low interest rates generally bring about higher levels of demand for goods and services and hence lower levels of unemployment.
5. An important influence on the level of demand for goods and services is the distribution of income between profits and wages. This is because the wage rate influences the demand for consumer goods, and the profit rate influences the demand for investment goods.
6. However, raising the wage may result in either more jobs, fewer jobs, or no change in the number of jobs, for a wage increase may raise, lower, or leave unchanged the level of total demand for goods and services.
7. The economy typically experiences fluctuations in employment and in output called business cycles. During the expansion phase of a business cycle, output and employment increase rapidly, and unemployment is low. During the recession phase of a cycle, output falls, and the level of unemployment rises. Government policies can dampen the fluctuations of the economy to maintain steadier levels of income and employment.

Counting the Unemployed

According to an old saying, a person with one watch always knows what time it is, but a person with two watches is never sure. In May 2016 the U.S. unemployment rate (not seasonally adjusted) was 4.5 percent—or maybe 9.4 percent—because the Bureau of Labor Statistics (BLS) has more than one measure of what it calls "underutilization of labor." In fact, it has six (U-1 through U-6), but we look just at the two most commonly cited.

Each month the BLS announces the unemployment rate using the U-3 measure. In some press releases, it refers to the U-3 as "the" unemployment rate, although it

also releases monthly the new values of the other five measures. But some economists think the U-6, a more comprehensive measure of unemployment, merits greater attention than it gets. In the 2007–2009 economic crisis the (seasonally adjusted) U-3 reached 10 percent, while the U-6, the highest one, topped 17 percent. This is because 10 percent of the labor force was unemployed by the U-3 definition, but many others were either marginally attached to the labor force or working part time but seeking full-time work. The U-6 counts these others as unemployed; the U-3 does not. In May 2016 the U-3 was 4.5 percent, while the U-6 was 9.4 percent. In measurement, the nature of the tool matters, so it is worth examining these two tools more closely.

Monthly professional surveys of about 60,000 representative U.S. households provide the BLS with data to compute unemployment rates and other measures. According to their answers to a detailed series of questions, members of the civilian non-institutionalized population age sixteen or older are classified as *employed*, *unemployed*, or *not in the labor force*. If we denote the number employed by E and the number unemployed by U, the unemployment rate is $U/(U + E)$. Since the labor force is defined as the number employed plus the number unemployed, an equivalent definition is *U/*(the number of persons in the labor force). Both the U-3 and the U-6 define the unemployment rate in this way, and both draw from the same data from the monthly household survey.

But the devil is in the details: the U-3 measure classifies people as employed or unemployed (or neither) in a different way from the U-6 measure, which is why they produce different results.

One difference is that the U-3 classifies people as "not in the labor force" if they are not working, but have not actively sought work in the last four weeks, even though they have done so in the last twelve months, and say they would accept a job if offered one. Although many people would probably regard such "discouraged workers" as unemployed, the U3 does not; but the U6 does count them as unemployed—and therefore also in the labor force. Imagine someone who lives outside a major urban area, and has long been unemployed: he or she may well find it hard to repeatedly approach the few existing local employers every few months and repeatedly ask for work, only to be repeatedly rebuffed. The U-6 "unemployed" category includes not only such discouraged workers, but also all others "marginally attached to the labor force," who may lack childcare or reliable transportation to search frequently for work—but who do want to work, would accept a job if offered it, and could then afford both childcare and transportation. In May 2016, this difference between the U-6 and the U-3 explained about one percentage point of the nearly five percentage point gap between the U-6 unemployment rate (9.4 percent) and the U-3 (4.5 percent).

The rest of the gap was due to the difference in the way the two measures classify people who work part time but want to work full time (where "full time" means 35+ hours per week). Such people working "part time for economic reasons" are counted as *employed* by the U-3, but *unemployed* by the U-6 measure. In May 2016 this difference in definitions accounted for most (about four percentage points) of the gap between the U-6 and the U-3. Table 16.1 shows how the Bureau of Labor Statistics data for May 2016 are combined to calculate both measures. In May 2016, the U-6 unemployment rate was 9.4 percent, while the more frequently cited U-3 unemployment rate was 4.5 percent. (Both figures were "not seasonally adjusted"; the seasonally adjusted figures differed from each of these by less than a third of a percentage point.)

TABLE 16.1 THE OFFICIAL U-3 AND U-6 MEASURES OF UNEMPLOYMENT, MAY 2016

MAY 2016		MILLIONS OF PERSONS
	Employed (by U-3 definition) (E)	151.6
	Part-time for economic reasons (PT)	6.2
	Unemployed (by U-3 definition) (U)	7.2
	Marginally attached to the labor force (MA)	1.7
U-3	$\frac{U}{U+E}$	4.5%
U-6	$\frac{U+MA+PT}{U+MA+E}$	9.4%

Sources: U.S. Bureau of Labor Statistics, Department of Labor, Data Retrieval: Labor Force Statistics (CPS), data for "Employed" and "Unemployed" from Table A-1: Employment Status of the Civilian Population by Sex and Age (Not seasonally adjusted); data for "Part-time for economic reasons" from Table A-8: Employed persons by class of worker and part-time status"; data for "Marginally attached to the labor force" from Table A-16: Persons not in the labor force and multiple jobholders by sex, not seasonally adjusted; and data for U-3 and U-6 unemployment rates from Table A-15: "Alternative measures of labor underutilization"; all tables available at http://www.bls.gov/news.release/empsit.toc.htm.

What Determines Employment and Output?

The key to understanding the ups and downs of unemployment and output is the total demand in the economy for goods and services. In order to understand changes in overall economic performance, we need to focus on this total demand and examine the reasons why it rises or falls.

One important aspect of the cyclical process is the fact that the amount of money spent by business owners on investment goods varies from year to year, and these changes, in turn, drive accumulation of capital at a faster or slower pace. To what extent does a particular set of social institutions provide a favorable climate for the accumulation process? The successive stages of U.S. capitalism and the corresponding long swings in economic performance are defined with reference to distinct social structures of accumulation and their supportive or dampening effects on investment.

To simplify the analysis in this chapter, we ignore the fact that a part of total output must be devoted to replacing the materials used up in production and to repairing the wear and tear on capital goods. We focus here on *net output*, and we assume that net output per hour of labor (y) is constant. Net output Y is defined as in Chapter 10, and is divided between owners of capital and employees. Net output *per hour* (y) equals the hourly compensation to labor (wages plus benefits) plus the remainder, which is profit per hour. (We leave out interest, rent, and taxes in order to keep matters simple.) In this chapter we are interested in the macroeconomy, so we use N to represent the total number of hours worked in the economy. (In previous chapters H was the total number of hours worked in a *firm*, but we will not be using that notation here.) Thus:

(16.1)
$$y = \frac{Y}{N} = w + \frac{R}{N}$$

where Y/N is value of total net output divided by the total number of hours worked in the economy (N), w is still the hourly pay that workers receive, and R/N is the profit per hour.

Net output will depend on how many hours of labor are employed. To keep things simple we also assume (in this chapter, but not in later chapters) that the total

number of people seeking work is given (and called *LS* for "labor supply"), that prices are constant, and that wages do not vary with the level of employment. All this allows us to focus on the main points of this chapter.

Aggregate Supply and Aggregate Demand

Macroeconomics is about total supply of goods and services in the economy ("aggregate supply"), and total demand for these goods and services ("aggregate demand"), both expressed in dollars. To be concise we will often talk about "the product market" or "the goods market" when what we really mean is "the market for goods and services."

A fundamental question in macroeconomics is this: Will the aggregate (total) demand for goods and services in the economy in a given year tend to equal the aggregate (total) supply in that year? That is, will demand be enough to buy all that is produced, and no more? Or will there be excess supply or excess demand in the goods market?

A **macroeconomic equilibrium** exists when aggregate demand equals aggregate supply (no excess aggregate demand or excess aggregate supply exists) and there are no forces tending to change the situation other than accidental ones or ones coming from external sources.

A situation when no excess demand or excess supply exists in the goods market is termed a **macroeconomic equilibrium**. If $20 trillion worth of goods and services are produced, will $20 trillion of goods and services be demanded? If $20 trillion in goods and services are produced, but $21 trillion are demanded, we say there is *excess demand*. If instead, $19 trillion are demanded, some goods are unsold and we say there is *excess supply*. If $20 trillion are demanded, then aggregate demand equals aggregate supply, and we say that *macroeconomic equilibrium* exists. This means that the economy will continue producing at its present level for the near future, as long as there is no change coming from external sources.

A second, equally fundamental question is: If demand and supply for goods and services are equal, will the level of aggregate supply be large enough that everyone willing to work at the going wage can find a job? We can ask this same question a different way: If a macroeconomic equilibrium exists, will it be characterized by *full employment*? By **full employment** we mean a situation in which nearly everyone seeking work can find it.

Full employment is a situation in which almost everyone seeking work readily finds it.

The first question—concerning macroeconomic equilibrium—involves markets for products, while the second question—concerning full employment—focuses on markets for labor. Even when (as sometimes happens), the answer to the first question is yes (aggregate demand equals aggregate supply), the answer to the second question will generally be no (not everyone willing to work at the going wage will be able to find a job). In other words, labor markets will not tend to clear. Rather, there will generally be an excess supply of labor—also known as *unemployment*.

Recall that the equilibrium of the Iowa City beer market studied in Chapter 8 was one in which the quantity of beer supplied was equal to the quantity of beer demanded. In equilibrium the beer market clears, as do most product markets. By contrast, in Chapter 12 we found that even when the labor market is in equilibrium, it typically will not clear. There will generally be workers out of work, with nothing they can do to change the situation, and there will also be nothing that prospective employers would choose to do that would result in more employment.

In this chapter we examine the determinants of how much will be produced, how many jobs there will be, and how many people will be unable to find work. We study the equilibrium levels of output, employment, and unemployment, keeping in mind

The **equilibrium level of employment** is the amount of employment that corresponds to a particular macroeconomic equilibrium (when there is no excess demand or excess supply in product markets, all taken together).

that labor markets generally do not clear. The amount of employment associated with a macroeconomic equilibrium is termed the **equilibrium level of employment**.

Consider the following hypothetical example. A worker produces $25 worth of goods in an hour. If $25 of output is produced, then $25 of income will be generated for somebody as a result. If we ignore the fact that some of the output will go to the government for provision of public goods and services, we can say that in this example $25 worth of goods ends up on the shelf of a retail outlet. Assume now that for each $25 worth of goods produced the worker gets $15 in wages and his or her employer gets $10 in profits. Whether there is enough demand to buy the supply on the shelf depends on whether the worker and the employer spend what they have received in income. If the worker spends all of the $15 in labor income and the employer spends only $5 out of the $10 in profit income (saving the rest), there will be $25 worth of supply on the shelf but only $20 worth of demand. The result would be excess supply in the product market.

The opposite could also be true. The employer might want to spend $15, using his or her $10 of profits and an additional $5 borrowed from someone else, while the worker continues to spend his or her $15 of wages. As a result there would be $30 worth of demand and only $25 worth of goods. The result would be excess demand in the product market.

Next, consider what happens if the value of products that are produced within a year is greater than the value of products sold. In this situation, on balance some goods remain unsold. If so, we say there is excess supply in the product market. In this case the seller might respond to the situation in two ways. First, the seller might lower the price, seeking to attract additional customers. Second, the seller might not reorder as much of the good (or maybe temporarily not reorder at all) to prevent the goods from piling up on the shelves. With fewer orders coming in, the managers of firms that produce the good will have to lay off workers, so unemployment will increase.

Excess demand produces just the opposite result: the shelves are bare before a new shipment comes in. The seller then places a larger order. As a result, employers hire more workers to produce the larger amounts now demanded.

We thus reach an important conclusion: when excess supply exists in product markets firms will employ fewer workers (or cut back the hours of existing workers), and when excess demand exists firms will employ more (or increase working hours of existing workers).

There are two effects of employing fewer workers when excess supply exists. First, there will be less produced and hence a lower level of supply. Second, there will be fewer workers receiving wages, and since those laid off will have less money to buy consumer goods, there will be a lower level of aggregate demand. In the example above, if one worker is hired for one hour less, output will be $25 lower, and demand will also be lower. If we continue to assume that workers receive $15 of the net output they produce and that they spend all of their income, total demand will be lower by the $15 that the worker would have spent. Notice that the fall in supply ($25) will be greater than the fall in demand ($15), simply because workers receive in compensation only a fraction of the value of what they produce. If it were otherwise, there would be no profit and capitalist enterprises would go bankrupt at once.

These effects raise a troubling question: will the reduction in employment prompted by excess supply in the product market tend to eliminate the excess supply (by reducing supply)? Or will the reduction in employment make the excess supply problem worse (by its effect on demand)? The answer is simple, and has just been given: Whenever there is excess supply and workers are therefore laid off, this reduces the dollar value of aggregate supply more than it reduces the dollar value of

demand. That is, aggregate supply in our example falls by $25 for each hour a worker is laid off, but aggregate demand falls by only $15 per hour, so that excess supply falls by the difference, namely $10 per hour. In this very simple model, laying off workers should move the macroeconomy closer to equilibrium.

A similar problem occurs when excess demand exists: firms will hire more people, increasing supply, and this will tend to reduce the excess demand. Yet the additional hiring will give more income to workers, thus increasing demand. It should be clear from the excess supply example that, in this situation, hiring more workers will add more to aggregate supply than to aggregate demand, and so reduce excess demand.

Measuring Total Output

Imagine someone asking you how much stuff you have in your suitcase, and if it is more than he has in his. How would you determine your answer? You could list everything in yours—a book, two shirts, a pair of shoes, a hair dryer, and on and on. But this would not really answer the question because it would not allow you to figure a specific quantity to compare with the specific quantity of his suitcase contents. If you wanted a single number total, you might weigh the contents, or you could indicate the dollar value of each item and total their worth.

When it is so difficult to quantify what is in a person's suitcase, imagine how hard it is to calculate the total output of a nation. The total output of a country in any particular year is usually considered to be the market value of all the goods and services produced. In some cases, the good or service need not be sold in order to be counted as output: agricultural products grown and consumed at home are typically counted, even though their value is only estimated.

In measuring output, then, goods and services are valued at their market price (or if they are not sold, roughly at what their price would have been if they had been sold). Services that are not offered for sale—such as home cooking or care given to parents or children by family members—are not included at all. Since such work continues to be done mostly by women, the officially measured value of a nation's output fails to take into account much of the labor performed by women. (See Figures 4.1 and 4.2 on the importance to the economy of *both* production *and* reproduction.)

Terms for Measuring the Macroeconomy

Economists use one or another of the following terms to put a value on the total net output of an economy.

- The *Gross Domestic Product* of a nation (its GDP) is the total output produced within its territory (including what is produced by nonresidents' labor or the capital they own).
- *Gross National Income* (GNI) is another measure of total income, which conceptually is the incomes received by nationals of a country from labor or property they own, whether the labor or the property was used inside or outside the country. It does not include incomes received by foreign nationals who are resident in the country. A country's GNI is therefore obtained by subtracting from its GDP the incomes received by foreigners resident in the country, and adding the incomes received from abroad by nationals of the country (with some small adjustments). A closely related concept, *Gross National Product*, is rarely used these days, but was commonly used decades ago, and is sometimes the main form in which historical data are available.

- *Net Domestic Product* (NDP) and *Net National Income* (NNI) refer to the above measures (GDP and GNI) *minus depreciation* (the amount of wear and tear on machinery and other capital goods resulting from production during the relevant time period, usually a year). However, the depreciation of natural resources and the environment that occurs when we erode the soil or use up clean air or water, for example, is generally *not* taken into account in the usual indicators of national output. In this and other ways, an increase in GDP or GNI, or even in NDP or NNI—may not accurately measure a change in our well-being.

Note that all of the above concepts are calculated using *value added* or what we have been calling *net output*. However, to fully implement this concept, the statistical authorities would have to subtract from the total value of each firm's output *both* the inputs used up in production *and* the depreciation of machinery. The "gross" measures, however, subtract only the inputs used up, but not the depreciation (which *uses*, but does not *use up*, machinery). Only the "net" concepts (NDP, NNP, and NNI) subtract both the value of inputs used up and depreciation of capital equipment, and thus fully reflect the "net output" concept, except, again, that they fail to deduct the "using up" of environmental resources.

When comparing total outputs for a given country at different points in time, we generally make two modifications in our measures of national output, one to correct for inflation and the other to take into account population growth. First, imagine an economy in which inflation (rising prices) occurs but the quantities of goods and services actually produced do not change. In this case the market value of total output will rise even though nothing real changes. Since there is nearly always some inflation occurring in an economy, changes in total output measured on the basis of *current prices* will tend to overstate the amount of growth in output.

Comparisons of total output between one year and another are more meaningful if they are made on the basis of *constant prices*. This means correcting for changes in the general level of prices, and usually this is done by referring to the price level in a "base" year so that comparisons can be made over a number of years using the same benchmark. When total output measures are corrected for inflation in this way, the total output is termed *real*, or *inflation adjusted*. (See box "Money Wages and Real Wages" in Chapter 17 for an explanation of how this type of inflation adjustment is done for wages.)

Second, we sometimes express total output measures on a per person basis. We do this to indicate roughly how well-off the typical individual is in a particular country and also to take into account any growth (or decline) in its population. To calculate such a figure we simply divide total output by the number of people in the country at any given time. The result is called *per capita* output (or income). Adjusting a country's gross domestic product by correcting it for inflation and dividing it by the population, we arrive at what many economists regard as the best measure of an economy's health: *real per capita GDP*. Real per capita NDP might be a better measure conceptually, but depreciation is hard to measure accurately, so many countries do not measure it at all.

Analyzing Aggregate Supply and Aggregate Demand

Aggregate supply is the total supply of goods and services produced during some period, say, a year.

Aggregate supply (AS) is the total supply of goods and services produced during some period, say a year. We continue to consider only net output. Then aggregate supply is equal to the number of hours of labor employed in a year (N) multiplied by net output per hour (y). Thus:

(16.2) $AS = (\text{net output per hour})(\text{total hours of employment}) = yN = Y$

For example, if $y = \$25$ and $N = 2$ million hours, $AS = \$50$ million. Because we have assumed that y does not change, we can graph AS in relation to N as a straight line, the slope of which is y (see Figure 16.3).

For every dollar's worth of AS that is produced, a dollar in income will be received by someone—by workers as wages, by capitalists as profits, or by some combination of the two. Thus, total income (the sum of all persons' incomes) before taxes equals AS. In other words, *net output in dollars equals income in dollars.* That is why economists often talk about aggregate output and aggregate income interchangeably. In principle they have the same dollar value.

Aggregate demand is the total demand for goods and services during some period, say, a year.

Aggregate demand (AD) is the total demand for goods and services during a year. The first component of aggregate demand is C, the total amount of money spent by households buying food, clothing, electricity, and other consumer goods and services. Consumption spending (C) usually amounts to about two-thirds of aggregate demand.

The second component of aggregate demand consists of *investment* (I), that is, the amount of money spent by businesses to buy capital goods to maintain or expand the production process. Governments also spend money on goods and services—everything from schools to bombs. Total government spending (demand for goods and services) is labeled G. Lastly, foreigners buy products from us (exports), and we buy products from them (imports).

In adding up the components of aggregate demand, we are interested in how many more dollars people outside the United States spend on goods produced in the

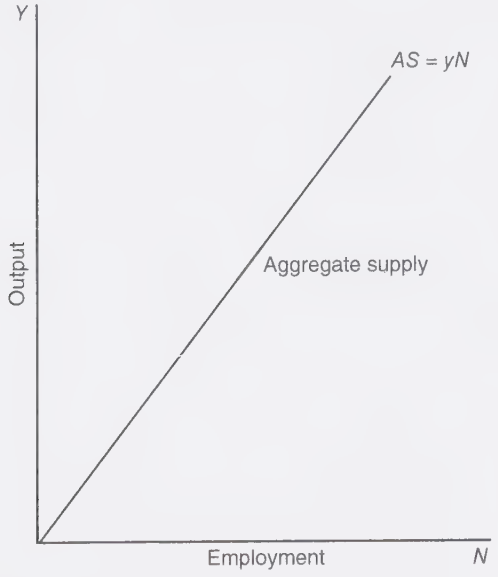


FIGURE 16.3 Aggregate supply.

The aggregate supply curve (AS) shows the total amount of goods and services (at their current market value) that can be produced with any given number of labor hours (N). The measure of output Y is *net output* or *value added*; in fact, the label "Output" on the vertical axis (here and in the rest of figures in this chapter) refers to *net* output. The value of AS is equal to yN, which is the net output of goods and services per labor hour employed (y) multiplied by N, the total number of labor hours employed. For simplicity, we assume that each hour of labor produces the same number of dollars worth of net output y. Since y is constant as N changes, this makes AS a straight line with slope y. AS is the same as Y.

United States than the dollars people in the United States spend on goods produced outside the United States. When they buy more of our products than we buy of theirs, this net demand for U.S. goods adds to aggregate demand. We refer to the dollar value of this excess as **net exports** (X), which here means exports minus imports. When we buy more of their products than they buy of ours, exports minus imports, or net exports X, is a negative number and detracts from aggregate demand. In this chapter we ignore net exports, and we begin by also ignoring government spending. So assuming that $G = 0$ and $X = 0$, aggregate demand, AD, is simply consumption plus investment:

Net exports are equal to the monetary value of total exports of goods and services minus the monetary value of total imports of goods and services.

(16.3) $$AD = C + I$$

What determines C, the amount of spending on consumer goods, also known as the demand for consumer goods? C depends on two things: first, the *total amount* of income and, second, the *distribution* of income between profits and wages. The total amount of income matters because the more income people have, the more they will spend on consumer goods. The distribution of income matters because people who earn wages tend to spend a larger fraction of their income than those who receive profits do. Why? There are two reasons.

First, people whose income mainly takes the form of wages tend to be less rich than those whose income mainly takes the form of profits. And while the average rich person consumes far more than the average person who is not rich, the rich spend a smaller *fraction* of their income on consumption, saving the rest. Thus, a shift in the distribution of income from profits to wages would raise the share of the nation's total income that is spent on consumer goods (C). This also works in reverse: Shifting income from wages to profits, as has happened in recent years in the United States, can weaken aggregate demand and slow economic growth.

Saving is income minus consumption spending.

Saving is defined as income minus consumption spending ("saving" essentially means "not spending" on consumption goods; the money could go into a bank, or be hidden under a mattress, or buried in the back yard, or go into purchasing securities or stocks). From a macroeconomic standpoint there are three components of a nation's total domestic saving (we continue to assume that there is no trade). The first is the saving that is done by families and other households. The second is the saving that is done by businesses when they do not spend all their sales revenue on costs, or dividend payments to their stockholders, or stock buybacks; in such situations, as described in Chapter 10, the money that is saved is referred to as "retained earnings." The third form of saving occurs when governments take in more money from taxes than they spend; when this happens a government is said to have a *budget surplus* (the opposite of which is a *budget deficit*).

A family has a flow of *saving* during a year if it spends less than 100 percent of its income by the end of the year on goods and services. If nothing is left after consumption, then its saving is zero that year. If the family spends more than all its income (such as by taking money out of savings from previous years' income, or by running up a credit card debt), then its *net worth* will decline. This will happen either because it sells some of its assets or because it incurs new debt. In this case in which a family consumes more than its income, it is said to have engaged in *dissaving*.

The typical family supports itself on wage and salary income during its working years and does not have much income from property (see Table 5.2, "Sources of Income for U.S. Taxpayers"). The vast majority of such families save very little over

the course of a lifetime because they break even in some years, save in some years (typically during middle age), and dissave in other years (typically during their twenties and in old age). This pattern is not true of the rich, who usually save a substantial fraction of their income throughout their lives because they generally have more income than they need for consumption.

A Basic Macroeconomic Model

To explain basic macroeconomic concepts, we now introduce a model with a number of simplifying assumptions. Imagine a hypothetical economy in which everyone either works or is unemployed and even the owners of firms (employers) receive wages (for the work they perform as managers). To keep things simple, we assume that all employed persons receive the same wage (w), whether or not they also receive a share of profits. The total wage income in the economy, then, is the wage per hour (w) times the hours of employment (N), or wN. The other form of income is profit income, which is simply net output (yN) minus total wages (wN), which can be expressed as $(y - w)N$. It is easy to see that total wages plus total profits equal net output yN (which is also total income).

To reflect the fact that the rich save more than the typical family, we make two additional assumptions. The first is that all profits are saved (the rich pay for their consumption entirely out of their wages). The second is that a large fraction of wages—received by both employers and workers—is consumed.

If people spend a large fraction (c), say 95 percent, of their wage income buying consumer goods and save the rest (including all of any profits they receive), then the demand for consumer goods is the fraction c multiplied by total wage income (wN). Thus:

(16.4) $$C = cwN$$

For example, if in our model $c = .95$, $w = \$15$, and $N = 2$ million hours, then $C = \$28.5$ million. The letter c represents the fraction of wage income that is spent on consumer goods; the rest is saved.

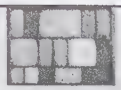
We will consider what determines investment shortly, but for now let us say that it is simply a given amount, denoted by $\underline{I}$. Then, aggregate demand is

(16.5) $$AD = C + I = cwN + \underline{I}$$

Aggregate demand (AD) is shown in Figure 16.4 as a relationship between *employment* on the horizontal axis (N in the term cwN) and total demand on the vertical axis ($C + \underline{I}$) or ($cwN + \underline{I}$).

Aggregate demand rises with the level of employment because as more people have jobs, there is more wage income to pay for the purchase of consumer goods (so cwN or C rises). From Equation 16.4 we can see that for every additional hour of labor hired (which increases N by one hour), the demand for consumer goods (C) rises by cw (the fraction of the wage spent on consumer goods times the wage rate)—and hence aggregate demand rises by the same amount. In mathematical terms, then, the slopes of both the C and the AD curves are equal to cw.

In order for all goods produced in a year to be sold in that year (that is, for product markets to clear and for there to be neither excess supply nor excess demand),



Income, Consumption, and Saving

Total income is the sum of the incomes (wages plus profits) received by all income earners. Since we are assuming that all output is sold, and hence that each dollar of aggregate supply generates a dollar of income, we already know that we can write total income as yN, which is net output per hour (y) times the number of hours of labor employed (N). In our simplified analysis, all consumption is paid for out of wage income (wN), and c is the fraction of wage income spent on consumer goods. Thus

$$\begin{aligned}\text{Total income} = \text{Aggregate supply} &= yN \\ \text{Total consumption} &= cwN \\ \text{Total saving} = \text{income} - \text{consumption} &= yN - cwN\end{aligned}$$

We could arrive at the same result by noticing that total saving consists of all profit income $(y - w)N$ plus the fraction of wages that is not spent on consumption: $wN - cwN$, which can be rewritten as $(1 - c)wN$. Thus

$$\begin{aligned}\text{Total saving} &= \text{wages saved} + \text{all profit income} \\ &= (1 - c)wN + (y - w)N \\ &= wN - cwN + yN - wN \\ &= yN - cwN\end{aligned}$$

For example, if employment is 90 hours and the net output per hour of labor (y) is \$25, then net output (yN) is (\$25/hour)(90 hours) = \$2,250. If the hourly wage (w) is \$15 and the fraction of income that is spent is 0.95, then total consumption (C) will be cwN = (0.95)(\$15/hour)(90 hours) = \$1,282.50. Total saving is net output (\$2,250) minus consumption (\$1,282.50), yielding \$967.50 as the amount of total saving.

We have assumed (to keep matters simple) that people save only out of wage income. This implies that people whose income consists mainly of profits consume a smaller fraction of their income than do workers. For example, if $c = 0.95$, people living wholly on wages will consume 95 percent of their income and save 5 percent of it. By contrast, a person whose income is 80 percent from profit and 20 percent from wages will consume 95 percent of the 20 percent wages, or (95%)(20%) = 19 percent of his or her total income, leaving 81 percent to be saved.

aggregate demand must be equal to aggregate supply. We can combine Equations 16.1 and 16.4 to explain what is needed to make aggregate demand equal aggregate supply:

(16.6) $$AS = yN = cwN + \underline{I} = AD$$

When aggregate supply and aggregate demand are equal, we have a macroeconomic equilibrium in which everything that is being produced is being sold. In Figure 16.5 we graph on the same axes the equations for aggregate supply and

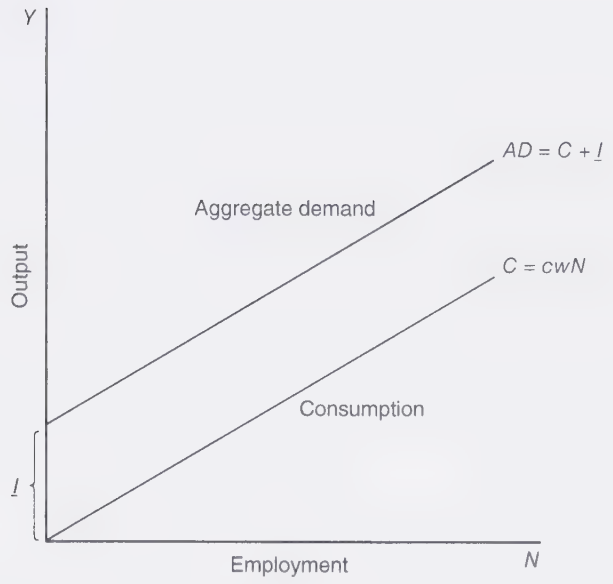


FIGURE 16.4 Aggregate demand with investment constant.

The total demand for goods and services (aggregate demand, AD) is the demand for consumer goods (C) plus the demand for investment goods (I). For simplicity, we assume that the demand for consumer goods is a constant fraction (c) of wage and salary income. If the wage rate (w) does not vary, total wage income will be wN (with N being the level of employment). The total demand for consumer goods at any level of employment (N) (measured in hours) will be cwN, the fraction of wage and salary income spent on consumer goods (c) times the wage rate per hour (w) times the total number of hours employed (N). The total demand for consumer goods at any level of N can then be represented by a straight line whose slope is cw. Finally, since we assume here that investment ($\underline{I}$) is constant, aggregate demand ($AD = C + \underline{I}$) appears as a line parallel to the C line and a distance $\underline{I}$ above it at all levels of employment (N). The slope of the AD line is also cw.

aggregate demand. Their intersection tells us the level of employment N such that aggregate supply AS equals aggregate demand AD. At this level of employment (labeled N^*) there will be just enough demand to buy the amount that N^* hours of work produces. The intersection of AS and AD is the level of employment and output that makes yN equal to $cwN + \underline{I}$. At this point on the graph, the equilibrium condition stated in Equation 16.5 is satisfied.

But what happens at other levels of employment? First consider $N+$, a level of employment greater than N^*. In Figure 16.5 we can see that at the $N+$ level of employment, the AS line lies above the AD line. So if $N+$ hours of labor are hired, aggregate supply exceeds aggregate demand. This means that the workers will produce more goods than can be sold. When the suppliers notice this, they will reduce employment, laying off workers. As long as employment is greater than N^*, excess supply will persist, and layoffs will continue until the employment level is reduced to N^*.

Similarly, if the firms employ $N-$ hours of labor, more goods will be demanded than are being produced, and excess demand will exist, since (see Figure 16.5) the AS line lies below the AD line. The workers are not producing enough goods to keep up with incoming orders. In this case more workers will be hired, and the employment level will rise to N^*, and then stop rising because all goods produced are now being sold.

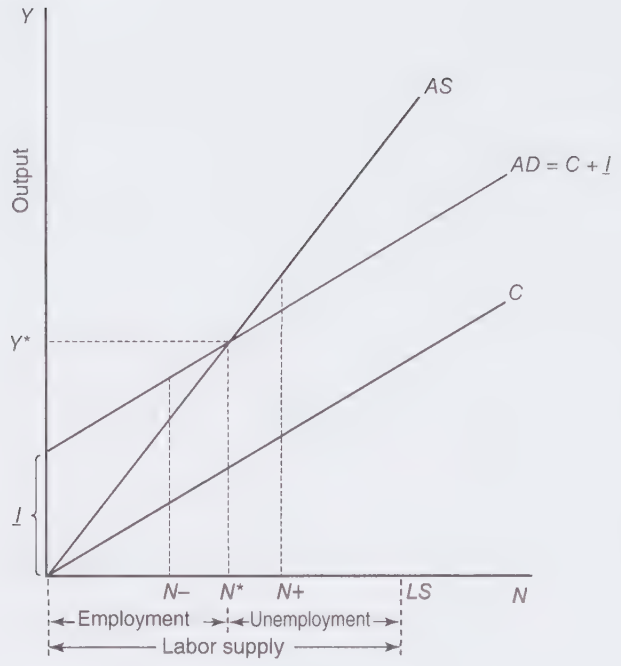


FIGURE 16.5 Macroeconomic equilibrium.

This figure shows the interaction of aggregate demand (*AD*) with aggregate supply (*AS*). It uses exactly the same *AS*, *AD*, and *C* curves as in Figures 16.3 and 16.4 and it is again based on the assumption that investment *I* is the same at all levels of employment (*N*), in fact the same level as in Figures 16.3 and 16.4. Aggregate supply equals aggregate demand where the *AS* and *AD* curves intersect, at the *N** level of employment and the corresponding *Y** amount of output. The intersection occurs where $AS = yN = cwN + \underline{I} = AD$ (see Equation 16.5). The result is that *N** and *Y** are the *equilibrium* levels of employment and output. When employment is *N+* some goods that are produced are not sold: $AS > AD$, which means there is excess supply in the product market. When employment is *N–* more goods are demanded than are being produced: $AS < AD$, which means there is *excess demand* in product markets. *LS* is the total labor supply: the total number of hours people would work if they had a chance. The difference between *N** and *LS* is unemployment, and is drawn larger on the diagram (for clarity) than it would normally be in reality.

Thus at neither at the *N+* nor at the *N–* level of employment will the macroeconomy be in equilibrium. At *N+* employment will tend to fall (move to the left) until it reaches *N**; and at *N–* employment will tend to rise (move to the right) until it reaches N*.

If the total labor supply is *LS*, the difference between *N** and *LS* is *unemployment*. An economy can be at a macroeconomic equilibrium such as the one indicated by *N** and *Y** (total supply equal to total demand in the product market) while at the same time there is an excess supply of labor (unemployment) in the labor market. Note, by the way, that the distance between *N** and *LS* on the horizontal axis in the figure is as great as it is only to allow room for labeling the diagram, not to reflect a situation that might actually prevail in an economy (with approximately half the labor force unemployed).

We can now answer the "troubling" question posed earlier in this chapter: When excess supply in the product market causes workers to be laid off, will this tend to eliminate the excess supply (by reducing supply)? Or will layoffs make the excess supply level worse (by reducing demand)? Given what we have assumed (that both *c*

and investment spending I are fixed), the answer to the first question is "yes", and the answer to the second question is "no." When more goods are being produced than are being sold, and employers let some workers go, this will narrow the gap between aggregate supply and aggregate demand, nudging the economy toward equilibrium, at which that gap disappears. The economy will then be operating at the N^* level of employment. The reason, quite simply, is that workers are paid only part of the net output. Suppose (as before) that they produce $25 an hour worth of net output, but are paid only $15 an hour, and they spend 100 percent of their wages. Then reducing total hours worked by exactly one hour will reduce aggregate supply by $25, but will reduce aggregate demand by only $15. The result will be a $10 reduction in the difference between the two, that is, in the level of excess supply.

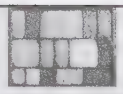
Notice that this whole line of reasoning implies that when the macroeconomy is not at equilibrium, it tends to move toward equilibrium, in which the dollar value of what is produced is the same as the dollar value of what is sold. If we changed our assumptions that $\underline{I}$ and c are both fixed, however, the result might sometimes be different.

The level of employment N^* is the *equilibrium employment level*, and the corresponding level of output, AS^* (which is equal to yN^*), is termed the *equilibrium level of output*. The fact that AS^* is the equilibrium output level does not mean that AS^* is the amount of output that will actually be produced in any given time period, or that N^* hours of labor will actually be hired. The actual levels of output and employment may be above or below their equilibrium levels. But, as we have just seen, the economy will tend to gravitate toward the equilibrium levels of output and employment. (Exceptions will be discussed briefly later in this chapter, in the section "The Business Cycle and the Built-in Stabilizers.")

The word *equilibrium* has a reassuring ring to it. When things are "in equilibrium" one tends to think of harmony and peace, but this impression is misleading when the equilibrium is a macroeconomic equilibrium. The reason is that, as we will explain, even when there is a macroeconomic equilibrium there will generally be people out of work. For those who are out of work, their lives—and particularly their household finances—will *not* be in equilibrium.

It is important to note here that the N^* equilibrium level of employment is generally not a *full employment* level. Instead, unemployment exists in macroeconomic equilibrium. We can see this if we compare the equilibrium employment level N^* with the total labor supply level LS, indicated on the horizontal axis in Figure 16.5. When N^* is lower than (to the left of) the total labor supply level LS, there will be unemployment. In this model the level of unemployment, if there is any, is shown as the hours of labor that workers are offering to supply in the labor market but for which there is no demand, or $LS - N^*$.

Stepping back from our simplified model, recall the importance of unemployment in a capitalist economy as a basis for the power of employers over workers. As discussed in Chapter 12, the equilibrium level of employment will normally involve some amount of unemployment. Although the unemployed workers would like to get jobs and would even be willing to work for a wage lower than the one currently being paid, they cannot get work. The reason for this is that given any particular level of unemployment, employers will be paying a particular wage, and they will not choose to lower the wage just because an unemployed worker shows up at the hiring office offering to work for less. (Remember—from Chapter 12—that even from the employer's standpoint, wages can be too low. Employers have no desire to lower the wage to such a point that the quality and amount of work effort, e, will be reduced



Unemployment Hurts

People who are out of work have to deal with a lot more than the loss of a paycheck. For many people, having a job is essential to self-respect, the respect of others, and even the feeling of being a part of a society.

Just how devastating a job loss can be is indicated by surveys that ask people about the quality of their lives. A study of Americans found that people who are widowed, divorced, or separated are much unhappier than those who have not suffered these separations. But the unhappiness associated with being without a job is as great as that associated with having lost a spouse.

Could it be that unhappy people tend to be unemployed rather than the other way around, and that is what the surveys are picking up? Not according to one study that looked at the same persons when employed and when unemployed. People reported being much unhappier when they were interviewed while jobless than in interviews when they had a job. In this study, as in the one referred to above, the psychological costs of joblessness exceeded the effect of the loss of income on one's happiness.

Sources: David G. Blanchflower and Andrew J. Oswald, "Well-Being over Time in Britain and the USA," *Journal of Public Economics* 88, nos. 7–8 (2004): 1359–86; Liliana Winkelmann and Rainier Winkelmann, "Why Are the Unemployed So Unhappy? Evidence From Panel Data," *Economica* 65, no. 257 (1998): 1–15.

so much that unit labor costs will rise.) So, in the real world, unemployment is likely to persist, even when unemployed workers would be willing to work at or below the going wage (see box "Unemployment Hurts").

Unemployment and Government Fiscal Policy

Is there anything that can reduce unemployment? Yes, there is. This message of hope was the main lesson offered in the economic theory of John Maynard Keynes, the early twentieth-century British economist who was the father of macroeconomic theory.

Deficit spending occurs when the government finances its purchases by borrowing from the public.

Government **fiscal policy** uses taxes and spending to regulate the level of total output and employment.

The basic idea of Keynesian macroeconomic policy is that unemployment is the result of insufficient spending by consumers and firms. In this situation a government can increase aggregate demand (*AD*) and reduce unemployment by spending not only the revenue it receives from taxes but also what can be raised by borrowing from the public, say, by selling government bonds. When a government spends more money than it receives from taxes, it is said to be engaging in **deficit spending**. By engaging in greater or smaller levels of deficit spending a government can regulate the level of total output and employment. When a government consciously regulates the rate of taxation and its level of spending we say that it is pursuing a **fiscal policy** (*fiscus* is the Latin word for "treasury").

Governments might spend to build schools or to build weapons. What determines the impact of government spending on the level of employment (at least in the short run) is not so much the nature of the spending as how it is financed. To generate more jobs, a government needs to spend more without taxing more. If the government increases spending but at the same time raises taxes to pay for it, taxpaying

consumers and firms will have less money at their disposal, and the resulting decrease in their spending will at least partially offset the increased spending by the government.

If, on the other hand, government decides to spend more without raising taxes, this will raise aggregate demand (AD) and increase the level of employment. As noted above, the government can spend more without raising taxes by borrowing money from the public. In order to do this, the government will normally offer new treasury bonds for sale in the bond market. People will then use their savings to buy treasury bonds from the government if the rate of interest on them is high enough to attract buyers. (A treasury bond is a government promise to repay, at a later date, the face value of the bond, and to pay interest on the face value in the meantime.) Funds that are saved by consumers or firms and used to buy treasury bonds from the government enable an increase in government spending. Such spending directly boosts total demand.

The government can also increase its spending without raising taxes simply by printing money and spending it. When this happens in the United States, the government—through the Department of the Treasury (the agency that prints the money)—is technically borrowing from the Federal Reserve System, the nation's central bank.

Effects of Deficit Spending on Employment

To see how deficit spending will increase the level of employment, return to the equation for aggregate demand (Equation 16.4). The demand for goods by the government is typically denoted by the letter G, and the amount of taxes is represented by T, so the level of government borrowing, which we call B, is simply $G - T$. Because we are focusing here on the *net* effect of government taxing and spending on the level of employment, we simply make the highly unrealistic assumption that there are no taxes and that all government spending is financed by borrowing. So G and B are the same. We will use the letter B (instead of G) to remind us that we are considering government deficit spending rather than spending financed by taxes.

To see the effects of deficit spending, assume that B is a specified amount, and call it $\underline{B}$. Then, adding $\underline{B}$ to Equation 16.4, aggregate demand becomes:

(16.7) $$AD = C + \underline{I} + B = cwN + \underline{I} + B$$

The result, as can be seen from Figure 16.6, is to shift the aggregate demand line upward by the amount $\underline{B}$. Adding $\underline{B}$ generates a higher level of equilibrium employment N^*. The conclusion is that deficit spending can generate higher levels of employment, thereby reducing unemployment.

The relationship between deficit spending and employment can be clarified by rewriting Equation 16.5 (the equation that defines macroeconomic equilibrium), adding government borrowing (B).

(16.8) $$AS = yN = cwN + I + B = AD$$

As before (in Equation 16.5), this new equation for macroeconomic equilibrium just states the situation in which aggregate supply (AS) is equal to aggregate demand (AD) in the product market.

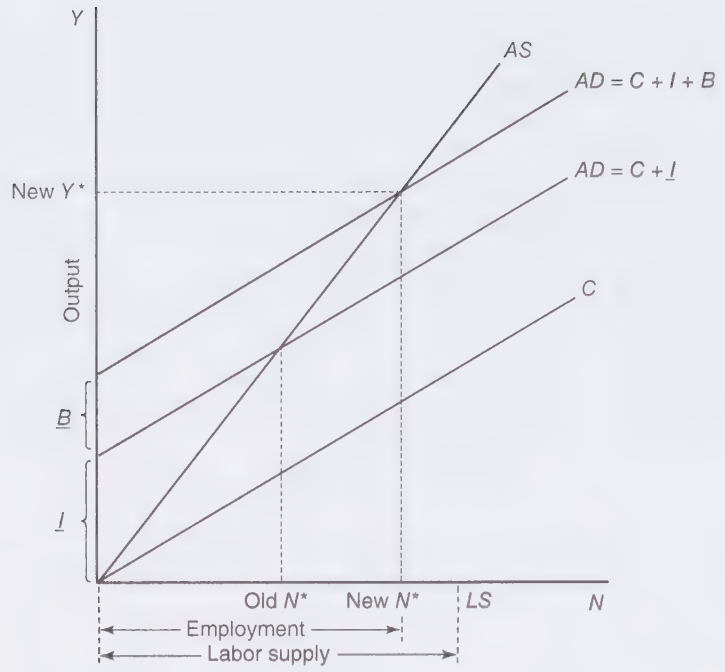


FIGURE 16.6 Macroeconomic equilibrium with deficit spending.

The only difference between this figure and Figure 16.5 is that it adds government deficit spending (B) to aggregate demand (AD). With this addition, $AD = C + I + B$. If the level of government deficit spending is $\underline{B}$, then the new AD curve will be above the old one by this amount at every level of employment. The effect will be to shift the AD curve upward by the amount $\underline{B}$ and so to increase the equilibrium level of employment from "old N^*" to "new N^*." The figure shows that when employment is below the desired level (say, at "old N^*"), deliberate government fiscal policies involving deficit spending can raise the level of employment (say, to "new N^*"), thereby reducing the amount of unemployment.

To derive the equilibrium level of employment (N^*) that corresponds to the macroeconomic equilibrium just defined, we subtract cwN from both sides of Equation 16.7, which moves the cwN term to the left-hand side and changes its sign:

$$yN - cwN = I + B$$

This equation can then be rewritten as follows:

$$(y - cw)N = I + B$$

Then, dividing both sides by $(y - cw)$, we get the equation for N^*:

(16.9)

$$N^* = \frac{I + B}{(y - cw)}$$

We add the asterisk to N here, signifying that this equation (16.9) is the equation for the *equilibrium* level of employment (N^*), where aggregate supply and aggregate demand in the product market are equal, and, therefore, the level of employment

corresponding to this macroeconomic equilibrium will also be in equilibrium. The denominator of the fraction on the right-hand side of the equation, $(y - cw)$, is the net output per hour of labor (y) minus cw (the fraction of the wage spent on consumer goods c times the wage w). Because the wage must be less than the net output per hour (otherwise there would be no profit!), and the fraction of the wage spent on consumer goods (c) is normally less than 1 (at least on average nationwide), cw must be less than y. So $(y - cw)$ is a positive number.

From Equation 16.9 and Figure 16.6, we can see that an increase either in investment I or in deficit spending B will increase employment. How many new hours of employment will an additional million dollars of government deficit spending generate? Of course it depends on what is the value of net output (y) that is produced in each hour of labor. For instance, if y = \$25/hour, then to get the added total output of \$1,000,000 requires $y(\Delta N)$ = \$1,000,000, or in this case \$25(ΔN) = \$1,000,000, where ΔN is read "delta N" or "the change in N" and refers to the change in employment). Clearly this means that the increase in the number of hours of labor needed to produce this output is ΔN = \$1,000,000/\$25 = 40,000 hours. Thus 40,000 hours of new employment are needed to increase aggregate supply by \$1 million. If a full-time job is 2,000 hours of labor in a year, the additional \$1 million of spending will result in twenty full-time jobs.

More generally, if we refer to the change in aggregate demand as ΔAD and to the change in aggregate supply as ΔAS, and to the change in the number of hours employed as ΔN, then this calculation can be expressed as $\Delta N = \Delta AS/y = \Delta AD/y$.

The \$1 million of additional government deficit spending will increase aggregate demand (AD) by \$1 million, putting the macroeconomy initially in a state of disequilibrium. As explained earlier in this chapter, the economy will then tend to move toward a new equilibrium.

It might seem that, once the \$1 million is spent to purchase the \$1 million of newly produced goods, the macroeconomy would be back in equilibrium. In fact, the effects will not stop there. The reason is that the newly employed workers will take their wages and spend them on consumer goods, creating additional demand, which in turn will create additional jobs; and then the workers in *those* new jobs will also spend, creating yet *more* jobs, and so on.

Suppose that the \$1 million is spent by a local government to build a new gymnasium for the regional high school. The money will be used to employ architects and engineers to draw up the plans and workers to begin the construction. The money will also be used to buy the construction materials needed to build the gym. All in all, the town will spend \$1 million buying inputs—labor, engineers' plans, cement, machinery, and so on. The companies that supply the concrete, construction machinery, and engineers' plans will find that the demand for their output has increased. They will hire more people to meet the demand or ask their current employees to work overtime. Jobs will be created not only in building the gym, but also in building or producing the things needed to complete the project.

Furthermore, all the workers who are newly employed or who work overtime building the gym will now have more money to spend. They will head to clothing stores or car dealers to make purchases. Whatever they buy will mean that some company will find that the demand for its product—clothes, cars, or whatever—has risen. That company will then employ more people in order to boost production and meet the demand. The people so employed will, in turn, go out and spend more money.

Multiplier Effect of Deficit Spending

What is the effect of these many increases of employment and spending—first, the direct effect of the original increase in government deficit spending and then the effects of the successive increases in employment and spending? Like the ever-widening ripples spreading out on the surface of a pond after a stone has been thrown in, the effects of the additional expenditures will spread throughout the economy. Will the effects ever end?

The effects will eventually diminish and come to an end because some of the new money received in each successive round of the process will not be spent; people will choose to save some of it instead. Although an additional hour of employment requires an extra $25 of spending, the worker employed by this spending will receive a wage of less than $25 and will not spend all of it. So the spending on each round will become progressively smaller, with each new round adding a smaller and smaller amount to total employment.

Equation 16.8 tells us what the sum of all these effects will be. If, as in our example, government deficit spending is increased by $1 million, we add $1 million to the numerator of Equation 16.8. How much will N^* increase? The answer is ($1 million)/$(y - cw)$, because the change in N^* will be equal to the amount of the change in spending ($1 million) times $1/(y - cw)$. If, as before, y is $25, c is .95, and w is $15, then $(y - cw)$ is $10.75, and a million dollars of deficit spending will create ($1 million)/($10.75/hour), or approximately 93,000 hours of new employment. This is the equivalent of 46.5 full-time year-round jobs, each of 2,000 hours per year.

Figure 16.7 shows the upward shift in aggregate demand that could result either from an increase in investment spending or an increase in government deficit spending. In the lower left corner we see from the aggregate supply curve *AS* that 40,000 hours of labor produce $1 million of output, since each hour of labor produces $25 worth of output. The economy initially is in macroeconomic equilibrium at point A, with 200,000 hours of labor producing $5 million of output. There, the *AS* and *AD* (*original*) curves intersect, and aggregate supply equals aggregate demand.

A rise in aggregate demand of $1 million shifts the *AD* curve upward by $1 million, to the dashed line *AD* (*new*), as shown by the vertical arrows; and the vertical intercept of the *AD* curve shifts from $2.0 million to $3.0 million. Now macroeconomic equilibrium will move to point B, where $7.32 million of output is produced by 293,000 hours of labor.

Producing the additional $1 million of investment goods in response to the additional demand at first requires only an additional 40,000 hours of labor. But in the process, new incomes are generated that create still more demand, so output continues to rise. Eventually the ripple effects of the $1 million rise in spending move equilibrium to point B, where the demand for labor is now 293,000 hours—an increase of 93,000 hours from point A. The employment multiplier (the change in employment divided by the change in the value of output) here is 93,000/$1,000,000, or .093. That is, an added $1 million of spending, either by the government or by business, creates (.093)(1,000,000), or 93,000, new hours of employment. If each person works 2,000 hours a year, this creates 46.5 new jobs.

The **employment effect** is the change in the number of hours of labor employed that results from the direct and indirect effects of a change in business investment or government deficit spending.

Notice that the total **employment effect** of the additional government spending in this case (93,000 hours) is more than double the direct employment effect (40,000 hours). The reason is that additional spending and employment will have resulted from the second, third, fourth, and so on, rounds of the process. Moreover, an increase of $1 million in investment spending (I) would have had exactly the same

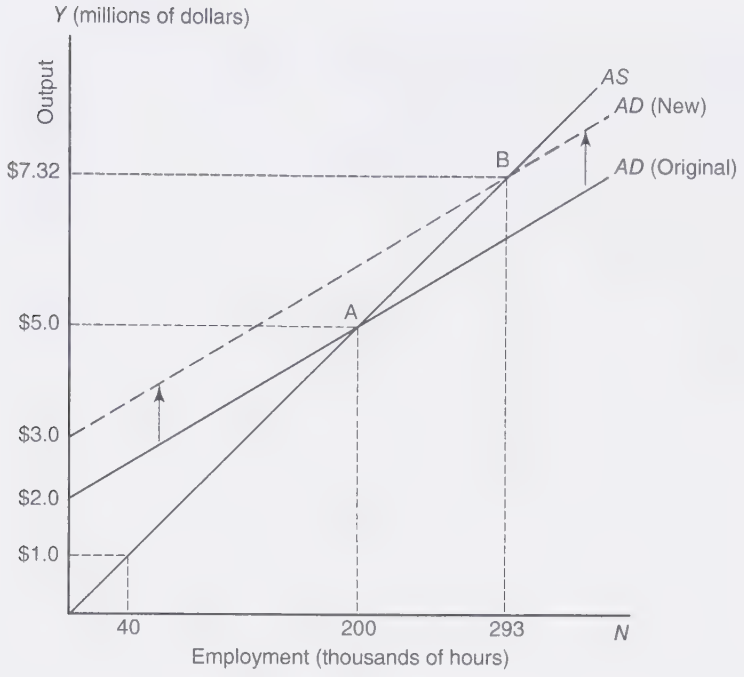


FIGURE 16.7 The employment multiplier.

How much will a $1 million increase in aggregate demand, such as from a rise in investment spending, boost employment? In this diagram, it increases equilibrium employment by 93,000 hours, as the economy moves from the equilibrium at point A before the increase to the new equilibrium at point B after the increase. Employment increases from 200,000 hours to 293,000 hours.

effect as the increase of $1 million of government deficit spending (B) did in this example (see Equation 16.8).

The **employment multiplier** is the hours of new employment directly and indirectly created by an additional dollar of investment or other spending.

How can we determine *how large* the employment effect will be? To answer this question we must introduce the concept of the **employment multiplier**, which enables us to know how large the total effect of a change in either I or B will be on N^*. The employment multiplier is defined as the amount of new employment created directly *and* indirectly by one additional dollar of deficit spending or business investment. To see how this works, take Equation 16.8 and make the numerator of it $10. If we then increase this numerator from $10 to $11 we can see how much N^* will change as a result. From Equation 16.8 we can see that the dollar amount of N^* will increase by exactly $\frac{1}{(y-cw)}$, since we are increasing the numerator by the difference between $11 and $10. And since the employment multiplier is defined as the change in N^* resulting from a $1 increase in B or I, we can see that the employment multiplier is $\frac{1}{(y-cw)}$. In our example, then, the employment multiplier is 1/10.75, or .093.

Figure 16.7 shows graphically how the employment multiplier works: the change in N^* from 200,000 to 293,000 is the effect of an increase in aggregate demand from $2 million ($AD$ original) to $3 million ($AD$ new)—and it does not matter whether this increase in aggregate demand results from an increase in B, or from an increase in I, or from some combination of the two making up an increase of $1 million. If we had known the value of the employment multiplier in advance (economists often

try to estimate its value), we could have predicted that the employment effect of an additional $1 million in aggregate demand would have been 1 million times .093, or 93,000 additional hours of employment.

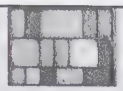
One last question: What makes the employment multiplier larger or smaller? The short answer is that it depends on the relationship between the *slope* of the *AD* line and the *slope* of the *AS* line. Recall that the slope of the *AS* line is *y* (net output per hour) and that the slope of the *AD* line is *cw* (the fraction of the wage spent on consumer goods times the wage rate). But *y* and *cw* are the two terms in the denominator of $\frac{1}{(y-cw)}$. Since the value of the employment multiplier will be large if the denominator is small (and vice versa), the employment multiplier will be relatively large if *cw* is only a little smaller than *y* (so that the *AD* curve is only a little flatter than the *AS* curve). In this situation, when the *AD* curve shifts upward, its intersection with the *AS* curve will slide a long distance up the *AS* curve, producing a large increase in equilibrium employment and output. Likewise, the employment multiplier will be relatively small when the difference between *y* and *cw* is relatively large (so that the *AD* curve is a lot flatter than the *AS* curve). The slope of the *AD* line (namely, *cw*) will always be less than the slope of the *AS* line (namely, *y*), because workers generally consume less than they earn ($c < 1$), and are paid less than they produce ($w < y$). But if *cw* is a *little* less than *y*, then the employment multiplier is large; and if *cw* is a *lot* less than *y*, the employment multiplier is small. Figures 16.5 and 16.6 in this chapter illustrate these concepts. We illustrate further with an example that contrasts with the example above.

If *y* is $25, *c* is .95, and *w* is $15 (as in our previous example), then $(y - cw)$ is $10.75, and the value of the employment multiplier is 1/10.75, or .093. In this case, as we have seen, an increase of $1 million in aggregate demand results in the creation of 93,000 additional hours of employment. In contrast, imagine the situation in which *y* is $25 (as before) but *w* is $23 (nearly all of net output is paid to workers as wages; very little is taken as profit), and *c* is 1 (workers spend all of their wages on consumer goods). In this case, the slope of the *AD* line will be quite close to that of the *AS* line, and the value of the employment multiplier, calculated as 1/(25 – 23), will be much larger (.5). Under these conditions an increase of $1 million in aggregate demand results in the creation of 500,000 additional hours of employment (1 million × .5), more than five times as much as the 93,000 hours in the previous example. The size of the employment multiplier does make a difference!

Although we have focused here on the effect that an increase (or decrease) in aggregate demand will have on the number of hours of labor employed in the economy—the *employment multiplier*—we can also calculate the effect of an increase (or decrease) in aggregate demand on the *output* of the economy. This is called the *output multiplier* (see box "The Output Multiplier").

We have reached three important conclusions. First, left to its own devices, the private economy will not generally provide jobs for all, because there is no reason to expect that when aggregate demand equals aggregate supply (macroeconomic equilibrium), the equilibrium level of employment will be such that everyone willing to work—or seeking work—at the going wage rate will be able to have a job.

Second, government spending can generate higher levels of employment. Sometimes people interpret this second conclusion to mean that government fiscal policy can achieve *full employment*. The reasoning is that if deficit spending can increase the amount of employment, and if we know how much employment is generated by each million dollars of deficit spending, the government could simply increase



The Output Multiplier

While the employment multiplier tells us how *employment* will vary with a change in aggregate demand, the *output multiplier* tells us how *output* will vary with a change in aggregate demand. The output multiplier is defined as the amount of additional output that results from a $1 increase in aggregate demand. It is equal to the employment multiplier times *net output* per hour (y), since *net output* varies with the amount of employment and the employment multiplier tells us how many hours of additional (or reduced) employment results from $1 of extra (or less) spending (aggregate demand). Thus,

$$\left(\frac{1}{y-cw}\right)(y)=\frac{y}{y-cw}=\frac{1}{\left(1-\frac{cw}{y}\right)}$$

Using the same numbers as in our example in the text, the output multiplier can be calculated as follows: (.093)(25) = 2.32. Thus, in our example an additional $1 million of spending brings about a little over $2.3 million more net output: each additional dollar spent generates .093 hours of new employment, and each hour of new employment produces an additional $25 worth of net output.

its spending (without raising taxes) until enough jobs were created that everyone seeking work could find it. In the next chapter we explain why a full-employment outcome is generally not possible.

Third, the total employment effect of additional government spending or business investment is considerably greater than the direct effect. This is because of the ripple effects that result from any particular spending decision (by either government or business), these effects being represented in the concept of the employment multiplier.

In the next section we explain why the presence of ripple effects in the economy is a mixed blessing. As in life, so with the economy: there can be too much of a good thing. When the employment multiplier is very large, the economy tends to go up and down like a roller coaster.

The Business Cycle and the Built-In Stabilizers

The **business cycle** is a periodic expansion and contraction of output and employment usually taking place over a period of three to ten years.

In the primarily agricultural economies of the precapitalist era, people's incomes fluctuated from year to year, sometimes bringing surprising and welcome affluence or devastating hard times. The main source of this instability was the weather. The right amount of rain at the right time spelled a good harvest and a bountiful year. Too much or too little rain, freezes, hail, or locusts during the growing season spelled belt tightening and a hard winter ahead.

People in capitalist economies experience similar fluctuations in fortune as a result of the **business cycle**. The term *business cycle* refers to the periodic expansion and contraction of output and employment that usually take place every few years. A typical business cycle is shown in the inset chart in Figure 16.1.

Business cycle expansions are characterized by increases in employment and income.

Business cycle contractions (or **recessions**) are periods of increased unemployment and reduction in output and income.

Business cycle expansions are characterized by rapid increases in employment and income. **Business cycle contractions**, or **recessions**, are periods of increasing unemployment, slower growth in total income, and sometimes even reductions in total income. The business cycle generally runs its course, passing from expansion to recession and back to expansion, in anywhere from three to ten years. The top of the expansion is termed the *business cycle peak*, and the bottom is called the *trough*. One can also see the business cycle in the ups and downs in the rate of unemployment (see Figure 16.2).

The good times associated with the business cycle are welcome, and the bad times are hard. But it does not all average out; most people would be happy to do without the business cycle. Its gyrations are unpredictable. For both businesses and families an unexpected downturn in the business cycle can spell financial disaster. For families it means the possibility of being laid off or having work hours reduced and unexpectedly having to meet expenses with less income. For businesses planning for the usual amount of consumer demand, it may mean that goods are produced that will remain on the shelves unsold for months or even years.

Like waves breaking on the shore, each business cycle is a little different from the last: some are mild, some severe, some peak early, others later. Recessions and booms may last only half a year, or they may extend for two or more years. As a result, nobody knows quite what to expect; the business cycle keeps everyone off balance.

In the agricultural economies discussed earlier, weather conditions caused economic uncertainties. In a capitalist economy uncertainties arise because the millions of decisions that together determine total output and employment are *all made independently*. Without any means of coordinating production and spending decisions, the economy we live in often produces too much or too little output in relation to the demand for it. When a prolonged imbalance between aggregate demand and aggregate supply occurs, the economy goes into a recession (if there is too little demand) or expansion (if demand is too great).

To understand how a business cycle works, recall our prior discussion of the size of the employment multiplier. Using the same numbers as in our earlier example (the employment multiplier is .093), imagine what would happen if, for some reason (say, the possibility of an impending war), business owners decide to reduce their annual investment (I) by $1 million. In this case the direct effect of the reduction in investment would be a loss of 40,000 hours of employment, and the direct and indirect effects together would bring about a loss of 93,000 hours of employment. The effect of this much lost employment might lead to a deterioration of the business climate, and this in turn might result in a further reduction in investment, possibly setting off a downward spiral that would cause the economy to go into a recession (see the box later in this chapter "What Determines Investment?").

What would happen, however, if the employment multiplier were .5? In this case (as explained earlier) the direct and indirect effects of a $1 million reduction in investment spending (I) would be a loss of 500,000 hours of employment, dramatically increasing the possibility of a recession. Thus, as we have said, a large employment multiplier is a mixed blessing. On the one hand, it means that any given amount of additional business or government spending will have a relatively large employment effect, creating a larger number of new jobs. On the other hand, however, a large employment multiplier will magnify the roller coaster nature of the business cycle.

Countercyclical policies are policies aimed at dampening the business cycle.

One of the objectives of macroeconomic policy is to smooth out business cycle fluctuations. Policies aimed at dampening the fluctuations of the business cycle are called **countercyclical policies**. They are of two types, conscious policy changes and automatic changes through built-in stabilizers, to be explained shortly.

Conscious Policy Changes

First, the government may attempt to offset changes in consumers' or businesses' spending by altering its spending, taxing, or (as we will see in the next section) interest rate decisions. For example, if investors decide to spend $1 million less on capital goods, the government can purchase $1 million more (remember the $1 million gymnasium). If nothing else changes, the result will be that aggregate demand is maintained at the same level, and no contraction in output or employment takes place. If the government could accurately anticipate every change in spending, and if it were free to react instantly by altering its spending level, it could smooth out (or dampen) the business cycle in this way. But these are pretty big ifs.

It generally takes months, and sometimes a year or more, to figure out how much businesses and consumers are going to spend and how their decisions will affect the economy. If even economists cannot agree on whether the economy is headed for a recession (or whether the business cycle expansion will soon peak)—frequently debated topics—the government cannot know enough to intervene with sufficient accuracy to offset the fluctuations of the business cycle. Nor can the government, in most cases, act quickly enough even when it has accurate information. New taxes and new expenditure programs must be drafted, approved by Congress, and then implemented—a process that may take years.

Automatic or Built-in Stabilizers

As an alternative to conscious policy changes intended to offset changes in private spending, a second countercyclical policy is used. Government spending and taxing may be deliberately designed so that as the economy contracts (and unemployment rises), the government will *automatically* spend more and tax less. The idea is to give the economy something like its own thermostat so that it will regulate itself automatically and avoid severe business cycles.

The best example of an automatic economic regulating mechanism is unemployment insurance (UI), by which unemployed workers who qualify can receive payments that replace at least a portion of their lost wages. As unemployment rises, the total amount of unemployment benefits paid out also rises, allowing unemployed people to spend even though they are receiving no wages. As a result, the level of demand for consumer goods will fall less than it would have if there had been no unemployment insurance. For example, if business demand for investment goods declines and the producers of such goods have to lay off workers, there will be a smaller reduction in output and employment than there would have been otherwise. The effect of unemployment insurance, then, is to reduce business cycle fluctuations.

Similarly, if government spending is financed primarily through income taxes, there will be comparable effects on the business cycle. The reason is that as incomes fall during a recession, so does the amount of income taxes that are collected. Then, if government spending continues at the same level while tax revenues decline, the government spending will sustain part of aggregate demand, and the declining income tax bite will leave a larger share of income dollars in households, which will help sustain their spending.

In fact, if the income tax is progressive—meaning that people with higher incomes pay a higher tax rate (a higher percentage of their incomes)—then as incomes fall, taxpayers will fall into lower tax brackets. This will reduce total income taxes collected by an even larger amount than if the tax system were not progressive. This

allows households to keep and spend more of their before-tax incomes, helping even more to sustain consumer spending.

Such policies as unemployment insurance and income tax financing of the government help to stabilize the economy. They reduce the severity of both upturns and downturns that take place as a result of shifts in spending by consumers and businesses. These policies are called **built-in stabilizers** because their stabilizing effects do not require particular government decisions to counteract the cycle. Rather, the effects are automatic because they are built into the structure of government spending and taxation.

Built-in stabilizers are automatic changes in government spending and taxation that dampen the cycle without conscious policy decisions directed toward this end. Examples are unemployment insurance and income taxation, especially progressive income taxation.

Both built-in stabilizers and consciously tailored changes in government spending and taxation—the two kinds of countercyclical fiscal policy—were used very effectively in the United States in the post-World War II period. As Figure 16.2 shows, business cycle fluctuations were much less severe in the United States for about half a century after World War II, after these policies were adopted. The level of unemployment is now at least somewhat more predictable. However, the built-in stabilizers appear to be working less well than in the past. Currently less than one in four of the unemployed actually collect unemployment benefits, due in large part to changes in state regulations about who is eligible and how long a person can collect benefits. Until 2011, all states paid at least twenty-six weeks of unemployment benefits. But after 2011, nine states reduced the duration of benefits; North Carolina, for example, reduced the period during which one could collect benefits from twenty-six weeks to only fourteen weeks. One reason for this is that the state unemployment funds were partially depleted by very high unemployment in the Great Recession, and until those funds are replenished or else supplemented by the federal government, some states lack the money to pay the level of benefits they paid in the past.[1] Similarly, the income tax rate structure is now far less progressive than in the past, because rich taxpayers pay a tax rate on their taxable income that is little different from that paid by middle income and low income taxpayers, weakening its contribution as a built-in stabilizer.

In the next chapter we will see that countercyclical policies cannot completely eliminate unemployment.

Many of the fluctuations that make up the business cycle come from changes in investment spending by firms. To understand this source of instability we need to take a closer look at the investment process.

Investment, Aggregate Demand, and Monetary Policy

Monetary policy refers to government influences on the rate of interest intended to regulate the level of investment, output, employment, and other macroeconomic outcomes.

At the outset of this chapter we assumed that the level of investment was given and did not vary with wages or the level of employment. This assumption was made for the sake of simplicity, but it was not realistic. Investment not only varies, it can also be influenced by government policies regulating the rate of interest. Government policy that influences the rate of interest in order to regulate the level of investment, output, employment, and other macroeconomic variables is termed **monetary policy**.

[1]Will Kim and Rick McHugh, "How Low Can We Go? State Unemployment Insurance Programs Exclude Record Numbers of Jobless," Economic Policy Institute Briefing Paper, March 9, 2015, available at http://www.epi.org/files/2015/how-low-can-we-go-state-unemployment-r3.pdf.

The demand for investment goods includes the purchase of such things as machinery and software as well as the building of factories, offices, and the like. To the economist the word *investment* does not mean buying and selling stocks or bonds. It means maintaining and adding to productive capacity (see Chapter 4). Firms purchase investment goods both because they wish to repair the wear and tear on their existing capital goods (this is called *depreciation*) and because they want to buy new equipment in order to expand their production (this is called *net investment*). Thus:

Total demand for investment goods = depreciation + net investment

Investment is, as we have seen, an important part of aggregate demand for goods and services, and hence an important source of jobs. It is essential to the economy for other reasons, also. Most important, investment generally increases productivity and thus allows for a higher standard of living. It accomplishes this by acquiring more or better equipment to be used in production, thereby making possible either more output or more leisure, or both. The long-term effects of net investment on net output per hour of labor (y) will be considered later.

Here we focus only on the short-term effects of changes in the level of investment (I) on aggregate demand (AD) and employment (N). Thus, in this chapter we treat investments in (that is, purchases by businesses of) such things as buildings, computers, and software only as a part of the total demand for goods and services; we do not examine the impact that additional amounts of these capital goods will have on net output once they are put to use.

Investment may be made by governments or by families, but the vast majority of investment in the United States is made by businesses. Businesses invest for one main reason: to make money—except, for example, when they are compelled by law to invest in equipment to comply with legal standards, such as emission limits. As explained in Chapter 11, investment is essential to enable a business to keep up in the competitive race, for only by renewing and expanding the capital goods used in production can the firm protect and enlarge its share of a market.

However, business owners are not *required* to invest. They can do many things with their money. Building new factories and purchasing equipment are only two possibilities. Rather than investing, they might choose to spend their money on such things as luxury consumption, lobbying the government for better treatment of their company, or advertising. Some of the alternative ways of spending corporate profits (that might otherwise be used for investment) are listed in the box "Investment and Other Uses of Corporate Profits."

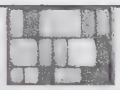
A few of the alternative uses of funds that would otherwise be used for investment, such as luxury consumption, contribute to aggregate demand. Lending and paying off debt, however, do not. If business owners are to spend their money on investment, it must be because they expect to benefit more from this activity than from the other possible uses of their funds.

Business owners usually decide how much they want to invest based on the expected profit rate and the rate of interest. As explained in Chapter 11, the expected amount of profit per dollar of investment (expected r) is one of the major determinants of the amount of investment undertaken. Of course, nobody can know in advance what the profit rate will turn out to be on any particular investment project. Much less can anyone know what the profit rate will be over the lifetime of the capital goods such as equipment, buildings, and software that have been purchased (their lifetimes may be very short or may extend for decades into the future). Thus,

business owners have to make investment decisions based on estimates, guesses, and hunches. To emphasize the unpredictable quality of the future and the significance of this unpredictability for the investment process, we say that investment is determined in part by the expected profit rate minus the interest cost of funds used for investment.

To determine the expected profit rate, a potential investor has to estimate the values of two main variables: the demand for the product and the cost of its production. The investor must guess whether there will be enough demand to absorb the quantity of the product he or she is preparing to produce and whether all the projected new productive capacity will be used.

Also, the investor will have to calculate whether wages (w), work intensity (e), work efficiency (f), and unit materials costs (umc), will be such that the cost of production per unit (uc) is low enough to make the project profitable, given the product's expected selling price (P_z). An investor may look at the current and the recent past profit rates and then come up with an estimate of the profit rate in the future. Thus,



Investment and Other Uses of Corporate Profits

When a firm makes a profit, investing in the United States is only one of the many things it can do with its surplus funds. Firms can use these funds in various ways, including the following:

Uses of Profit by U.S. Firms	Example
1. Invest in United States	IBM builds a new plant in Kentucky
2. Invest elsewhere	GE builds an assembly plant in Asia
3. Lend (at interest)	ExxonMobil lends money in Eurodollar market
4. Repay existing debt	AIG repays bailout money
5. Pay dividends to stockholders	GM declares a quarterly dividend of $1.25 per share
6. Hold funds for future investment	Ford adds to its "retained earnings" account
7. Buy back own firm's stock from other shareholders	Pfizer uses an amount equal to over half its profits for stock buybacks
8. Political expenditures	Lobby to influence political leaders; buy TV commercials

Political expenditures may take the form of election campaign contributions, payment of bribes to public officials, or advertising to influence public opinion; the goal is to get more favorable treatment from the government (such as lower taxes or relief from pollution-control laws). Economists do not consider any of these expenditures to achieve political objectives to be "investments," because they do not increase or maintain productive capacity. Political scientists like Thomas Ferguson do call them investments, however, referring to the fact that they are made with the expectation of a return.

the present profit rate will be a major influence on the expected profit rate and hence on the amount investors decide to invest (I).

Another major determinant of the level of investment, in addition to the expected profit rate, is the rate of interest (i). As we saw in Chapter 11, the interest rate is important for two reasons. First, most investments are made with at least some borrowed money, and a higher interest rate will increase the cost of the investment and reduce the profit available to the investor. Second, any business considering making an investment must determine whether the firm would be better off lending its available funds to other businesses (perhaps through financial intermediaries such as banks) rather than building new productive capacity. If the interest rate is high, the cost of making an investment will be high, and the rate of return on lending out funds will also be high. In this situation the high interest rate, combined with the incentive for lending out funds, will tend to discourage the making of investments.

Given the importance of the expected profit rate and the rate of interest as the two main determinants of the level of investment, we can now reconsider the problem of aggregate demand and employment using our more developed understanding of investment. For any given interest rate, the level of demand for investment goods depends on the expected profit rate, which, in turn (as explained above), depends on the current and recent past profit rates. The influence of the profit rate on investment—that is, how much an increase of \$1 in profits will increase investment—is called the **profit effect on investment**. We use the letter j as the symbol for this effect. For example, if profits increase by \$1, and as a result investment increases by 50 cents, then the profit effect on investment is \$0.50/\$1.00 = .5, so $j = .5$. The profit effect on investment could also be greater than one: a \$1 increase in profits could so excite investors that it stimulates an increase in investment of \$1.50 (the investors might finance their purchases of investment goods by borrowing); in this case the profit effect on investment (j) is 1.5.

The **profit effect on investment** is the amount by which investment will increase for each \$1 increase in total profits.

We can then say that total investment (I) is composed of two parts. The first is influenced by the interest rate and factors other than profits, and it is represented by $\underline{I}$ in Equation 16.10. The second part is influenced by profits. It can be represented as j multiplied by total profit (R).

Then, since $R = (y - w)N$, we can say that total investment is determined as follows:

(16.10) $$I = \underline{I} + jR = \underline{I} + j(y - w)N$$

An increase in the interest rate will lower I (by lowering $\underline{I}$), and a decrease in the interest rate will have the opposite effect. And because higher levels of employment bring about higher rates of capacity utilization, they will increase the rate of profit, which, in turn, may also stimulate more investment. These relationships are shown in Figure 16.8. Investment is shown as increasing with the level of employment (for the reason just stated), and changes in investment due to changes in the interest rate are shown by the dotted lines in the figure, with higher investment at every level of employment being associated with lower interest rates, and vice versa.

There is another factor that can influence the expected profit rate. If businesspeople become more optimistic about the future—perhaps because a pro-business Congress has been elected and will probably enact legislation favorable to business—the expected profit rate will rise (even if the current profit rate remains unchanged). A heightened optimism will have the same effect as a fall in the interest rate would

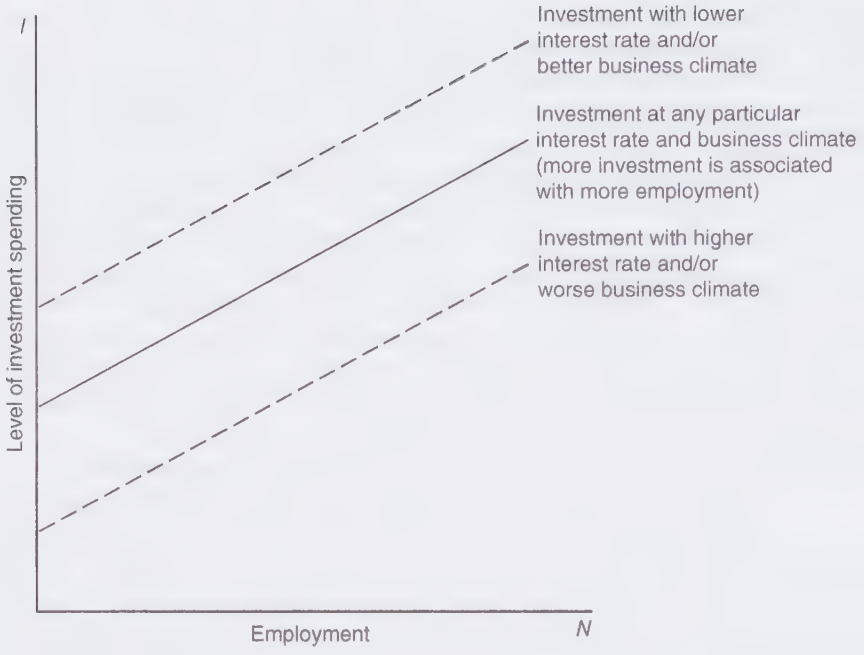


FIGURE 16.8 How investment responds to changes in employment, the interest rate, and the business climate.

This diagram shows that investment spending actually tends to increase as employment N increases. In earlier diagrams, to keep matters simple, we assumed that investment spending would stay at the same level regardless of employment N, and this assumption would make investment a horizontal line as a function of employment. But investment actually tends to increase as employment N increases (see box "What Determines Investment?"). This is because rising employment uses more of existing production capacity, and profits rise, and these changes motivate a good manager or owner to invest in expanding production capacity in order to be able to meet growing demand. The shape in the diagram reflects Equation 16.10. The investment spending function also shifts in response to changes in the interest rate or the business climate. A fall in the interest rate or an improvement in the business climate will encourage investment spending, shifting the investment function upward to the upper dashed line. A rise in the interest rate or a worsening of the business climate will have the opposite effect, shifting the curve downward to the lower dashed line and lowering investment spending at every level of employment.

have, but it works through the term jR on the right-hand side of Equation 16.10 (which shows the effect of profit expectations on investment). When increased optimism raises the expected rate of profit, it will shift the investment curve upward. When this happens we can say that the upward shift in the investment curve is the result of an improvement in the **business climate**. The influences on investment of the expected profit rate, the interest rate, and the business climate are depicted in the accompanying box (see box "What Determines Investment?").

The **business climate** refers to political, ideological, psychological, and other influences on the expected profit rate. It is sometimes termed the investment climate.

To focus on the ideas introduced here, we assume that the government does not engage in deficit spending, so that B (for borrowing) is zero in the following discussion. In Figure 16.9 we show how a change in the business climate or in the interest rate will change the equilibrium level of employment (N^*). As before, a macroeconomic equilibrium—in which aggregate supply is equal to aggregate demand in the product market—will determine the equilibrium level of employment in the labor market.

Also as before, the equilibrium level of employment falls short of the total labor supply (LS), and so unemployment continues to exist. In the previous section we

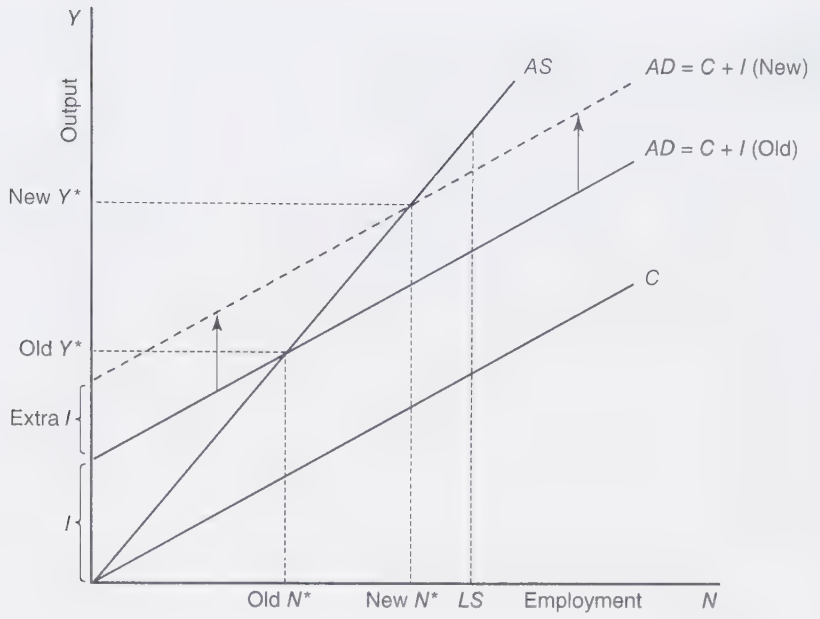
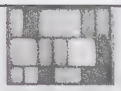


FIGURE 16.9 Effect of an increase in investment due either to a decline in the interest rate or an improvement in the business climate.

A decline in the interest rate or an improvement in the business climate (or both), will increase investment and therefore both aggregate demand (AD) and the number of labor hours employed (N). We ignore deficit spending here just for simplicity. Then since $AD = C + I$, an increase in I shifts the AD curve upward and raises the equilibrium level of employment (N^*). The new AD curve is the dashed line, and the equilibrium level of employment moves from "old N^*" to "new N^*." Thus lowering the interest rate or improving the business climate can reduce the unemployment rate.

showed how the government could use fiscal policy to generate more jobs. Now we look at how monetary policy (mentioned earlier in this section of the chapter) can be used to boost the equilibrium level of employment.

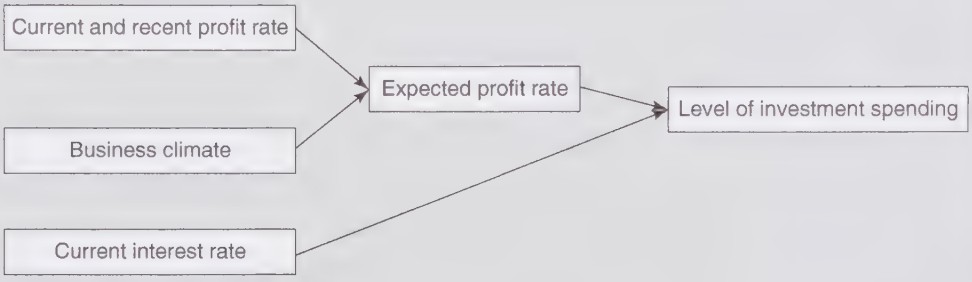
All governments exercise some degree of control over the monetary systems of their economies. In the United States the Federal Reserve System regulates private banks in various ways, establishing rules and regulations (supplementing those provided explicitly by the legal system). The Federal Reserve System (known as the Fed) also lends money to private banks. Through its activities the Fed can usually influence interest rates, making it easier or more difficult for private banks to make loans to their customers. The Fed cannot literally dictate what the banks' interest rate (or rates) will be, but it can influence these rates, pushing them up or down. For example, the Fed might purchase some of the treasury bonds that banks or individuals hold and pay them with money that the Fed (and only the Fed) has the authority to create simply by writing a check. This would have the effect of pumping more money into the economy (increasing the supply of money), and this, in turn, would push interest rates down. Conversely, if the Fed were to sell treasury bonds (that the Treasury had issued and sold at some time in the past), it would push interest rates up. The reason is that an interest rate is simply the price of (borrowing) money. Therefore, an increase in the supply of money will have the same effect that an outward shift of a supply curve would have in a product market, lowering



What Determines Investment?

One of the hottest debates in economics is about what determines the level of investment spending, and this debate influences the choice of policies to promote investment. Most economists agree that the expected profit rate is an important factor and that profit expectations are based in part on the current (or recent past) profit rate. The arrows in the figure below indicate causal effects. Thus, the top arrow indicates that current and recent profit rates influence or determine the expected profit rate.

What determines the level of investment spending?



Most economists think that the business climate is important, but there is little agreement about what constitutes a *good* business climate. Some emphasize pro-business policies such as low taxes on profits or less emphasis on environmental regulation; others emphasize the importance of stable market conditions so that businesses can count on an adequate demand for their products in the future.

The importance of the interest rate is also hotly debated. Conservative economists tend to give it a larger role than do liberals because influencing the rate of investment by changing the interest rate can be done without enlarging the size of government or having it engage in deficit spending. The factual evidence about the importance of the interest rate is mixed, but it is certainly true that very high interest rates discourage investment.

(or raising, in the case of a leftward shift) the price of the product (see Figure 17.5 in the next chapter).

Thus, when there is a high rate of unemployment, the Fed could push interest rates down, bring about a higher level of investment, increase aggregate demand (see Figure 16.9), and thereby raise the level of employment. Lower interest rates will also lead to an increase in consumer spending. What is significant for the analysis of how monetary policy can affect employment is that the government—or, more precisely, the Fed, an agency that is independent of the president and the Congress—can influence the rate of interest.

Wages, Aggregate Demand, and Unemployment

Some economists think that if the problem is unemployment, a lower wage is the solution. Economists who think this way make unemployment part of the solution, too, because, they argue, the existence of unemployment will tend to force wages down until the market clears. But this approach to reducing unemployment is wrong on both counts: first, the existence of unemployment need not cause wages to fall, and, second, if wages did fall, the result might be more, rather than less, unemployment.

The existence of unemployment (excess supply in the labor market) does not mean that wages will fall. Wages can remain unchanged even if there is unemployment. When there is excess supply in the milk market, the price of milk typically falls, supply is reduced, demand increases, and the market eventually clears. But as we saw in Chapter 12, the labor market is different. To the milk *buyer* there cannot be such a thing as too low a price of milk: a price of zero would be fine, thanks. But to an employer there can be too low a price of labor. As we saw in analyzing the labor extraction curve in Chapter 12, the wage has to be high enough not only to attract workers but also to motivate them to work hard and well. A firm that pays its workers so poorly that they do not mind losing their jobs is unlikely to be a profitable company. As explained in Chapter 12, the profit-maximizing wage will tend to be higher than what an unemployed person would be willing to work for because this higher wage will minimize unit labor cost and thus maximize the rate of profit. But given some particular amount of unemployment, a wage will be chosen by the employer—or be settled upon between employer and employees—and the fact that there is unemployment will not cause the wage to change. So unemployment itself will not cause wages to fall.

The second problem with the conventional approach is that even if wages did fall, the result might be more, not less, unemployment. This point is hard to understand because it is inconsistent with our personal experience. If a person is looking for work and is asking $20 an hour, he or she might have to look for quite a while (or forever) to find a job. But if this person offered his or her services at $10 an hour, he or she would might find work sooner. For an *individual*, asking for a lower wage sometimes improves a person's chances of finding employment. From this, one might conclude that if *everyone* were willing to take a pay cut, then fewer people would be unemployed. But this conclusion would, as a general rule, be wrong.

If everyone were to take a wage cut, the result could be (a) lower unemployment, (b) higher unemployment, or (c) no change in unemployment. The reason the outcome is uncertain can be understood using the macroeconomic analysis presented in this chapter.

An employer's demand for labor is a *derived demand*—that is, it is derived from (caused by) the demand for the commodities produced by the particular firm. But the demand for any one firm's output depends on the level of demand in the economy as a whole, or aggregate demand. Higher wages will increase employed workers' incomes, and this will tend to result in an increase in the demand for consumer goods (C). The result may be an increase in aggregate demand ($C + I$), and hence in employment. But if the higher wages reduce profits, the amount of investment (I) may fall. And if investment falls more than consumption rises, the result will be a decline in aggregate demand and less employment (more unemployment).

The question is: how will an increase in the wage (w) affect total employment (N^*)? Whether a rise in the wage leads to a growth or a reduction in total employment

depends on how firms respond to the rise in w. In a situation in which employees insist on being paid a higher wage, firms may anticipate lower profits and cut back on their investment spending. If this happens broadly across the economy, the result will be a smaller total amount of investment (I), and possibly a lower level of aggregate demand ($C + I$). But if firms anticipate that the rise in wages will lead to more aggregate demand and rising sales (through the influence of w on consumption), they may decide to keep I at its present level—or even increase it—and the result will be an increase in employment.

Figure 16.10 shows how an increase in the wage (w) can lead to an increase in employment. It shows what happens when firms do not change their level of investment even when w is rising. An increase in w will increase the amount of demand for consumer goods (C) at each level of employment. This is represented on the graph as a rotation upward of the consumption line from "C (old)" to "C (new)." The slope of the C line rises because its slope is cw, and, assuming that c (the fraction of the wage spent on consumer goods) does not change significantly, an increase in w raises cw. At the same time, because we are assuming that the level of investment ($\underline{I}$)

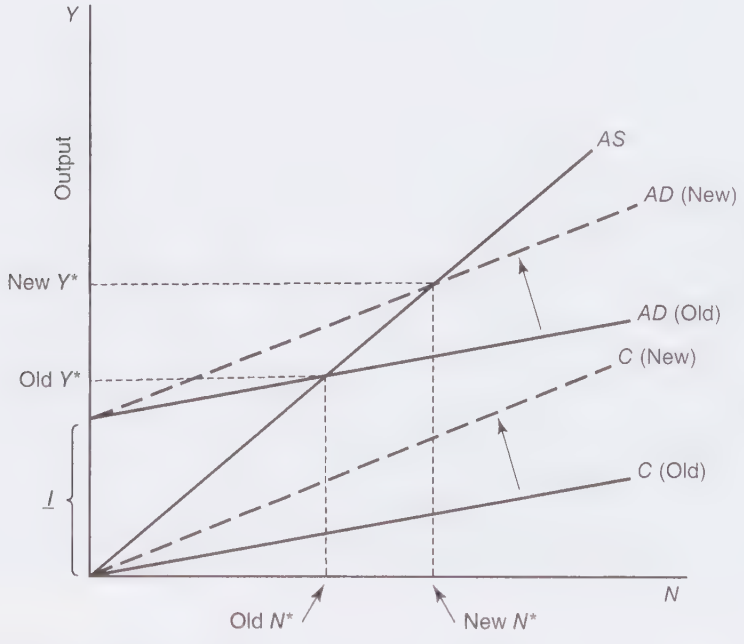


FIGURE 16.10 A wage-led employment situation with investment constant.

This diagram shows how an increase in the wage can increase equilibrium employment under some conditions. The amount of investment is assumed constant here, meaning it is unaffected by changes in profits. This means that it is the vertical intercept of the AD curve (both "old" and "new"). An increase in the wage (w) increases the slope of the C curve from cw (*old*) to cw (*new*), but the AD curve continues to have the same vertical intercept as before—so the AD curve simply rotates counterclockwise from AD (*old*), the solid line, to AD (*new*), the dashed line. Its intersection with the AS curve travels up to the new intersection of the AD (*new*) and AS curves, and the new equilibrium employment level N^* (*new*) is higher than the old. The level of output and income also increases from *old* Y^* to *new* Y^*.

remains constant, the *AD* line rotates upwards—from "*AD* (old)" to "*AD* (new)"—its slope always remaining the same as the slope of the consumption line. As the *AD* line rotates upward, the equilibrium level of employment increases, from "old N^*" to "new N^*." Under these conditions, then, an increase in the wage (w) leads to a rise in the amount of labor employed.

When is an Employment Situation Wage-Led?

From the earlier discussion, however, we know that we cannot jump to the conclusion that an increase in the wage (w) will *always* lead to an increase in employment. To determine when a rise in w will have a positive effect on employment and when a rise in w will have a negative effect on employment, we must consider the effect of a change in w on investment (I) as well as on consumption (C). (As we know, aggregate demand is the sum $C + I$.)

Since $C = cwN$, we can see that an increase in w of \$1 will raise consumption (C) by the dollar amount cN. Similarly, if w is reduced by \$1, consumption will fall by cN dollars. To determine how a dollar's change in w will affect investment (I), we refer back to Equation 16.10. In it, w is preceded by a minus sign in the term wjN on the right-hand side. So, holding all the other variables constant, we can see that increasing w by \$1 will reduce investment by jN dollars. Similarly, a reduction in w of \$1 will lead to an increase in investment of jN dollars.

We can now see how a change in the wage (w) will affect aggregate demand. Since a rise in w of \$1 will lead to an increase in consumption (C) of cN dollars, and since the same \$1 increase in w will lead to a decrease in investment (I) of jN dollars, the *net* effect of a \$1 wage increase on aggregate demand will be $(c - j)N$ dollars. Whether the wage increase ends up expanding or shrinking aggregate demand will depend on which is larger, the fraction of a dollar in wages that is spent on consumption (c) or the profit effect on investment (j). For example, if $c = .95$ and $j = .5$, a \$1 increase in the wage will *expand* aggregate demand by $(c - j)N$, or $(.95 - .5)N$ dollars. Thus, a wage increase expands aggregate demand when c is greater than j, and unemployment falls as a result. Alternatively, if $c = .95$ and $j = 1.5$, a \$1 increase in the wage will *shrink* aggregate demand by $(c - j)N$, or $(.95 - 1.5)N$ dollars. In this case the wage increase leads to a reduction of aggregate demand (because j is greater than c), and unemployment rises as a result.

A **wage-led employment situation** exists when rising wages increase employment.

A **profit-led employment situation** exists when rising wages reduce employment.

A situation in which a wage increase affects employment positively—c is greater than j—is termed a **wage-led employment situation**. A situation in which a wage increase affects employment negatively—j is greater than c—is termed a **profit-led employment situation**. The Appendix found on the book website gives an example of a wage-led employment situation that is more realistic, though slightly more complicated, than the one shown in Figure 16.10.

What types of circumstances determine whether a wage increase will have a positive or a negative effect on employment? The answer, as might be expected, is that "it depends." In some countries and at some times, a wage-led employment situation exists; in other places and at other times, a profit-led employment situation exists. Where most wage or salary recipients are poor, the fraction of wage income consumed (c) is likely to be high, and there will probably be a wage-led employment situation. Where businesses are highly mobile and able to relocate anywhere in the world in response to differing profit rates, the profit effect on investment (j) is likely to be large, so there will probably be a profit-led situation.

(Yet if part of the reason for a U.S.-based TNC to relocate more of its production abroad is as part of a tax avoidance strategy, even an increase in the profit rate in the U.S. might not inspire greater investment in the U.S., since increasing U.S. profits also means increasing the profit income subject to relatively high U.S. corporate income tax rates.)

Where a nation is in relation to the business cycle may also be important. At low levels of employment and wages—at the trough of a recession, for example—a wage increase is likely to generate jobs (a wage-led situation). And at high levels of wages and employment—toward the end of a business cycle expansion, for example—a wage increase is likely to reduce investment, shrink aggregate demand, and reduce employment (a profit-led situation).

Why do circumstances like these make a difference? At low levels of employment and at low wages, the fraction of wages spent on consumption is likely to be high: when most people have a member of their family unemployed or are themselves unemployed, saving is impossible and, as a result, all available income is spent on necessities. In this case c will be high, possibly as high as 1 or even greater than 1 (as families borrow to maintain their living standards).

Also, at low levels of employment and low wages, the profit effect on investment (j) is likely to be relatively low. The reason is that at low employment levels, businesses are likely to have unused productive capacity: they will have machinery and perhaps entire factories idle for lack of demand. Under these conditions they are unlikely to make investments, even if profits are rising, so the profit effect on investment (j) will be low. Thus, depressed economic conditions are likely to be situations in which c exceeds j, a wage-led situation exists, and a wage increase will tend to generate additional employment.

By contrast, at high levels of employment and wages—during a boom, for example—people will be better off and will be making more than they had in recent years, perhaps even more than they had expected to earn. In this case they will find it easier to save, so c will be relatively low. At the same time, businesses will find most or all of their existing productive capacity in use, so they will be ready to invest if the expected profit rate is right. Thus, the profit effect on investment (j) will be high. With c low and j high, there will be a profit-led situation, and a wage increase is likely to lead to more unemployment.

Implications for Economic Policy

The analysis of wage-led and profit-led situations has important implications for economic policy. For example, when a wage-led situation prevails, there will be greater justification for an increase in the legally mandated minimum wage as an employment-boosting policy. Raising the minimum wage usually brings about rises in wages at every level. With the rise in the average wage (w) there will also be an increase in total employment and hence a reduction in unemployment. In such a situation, then, a policy that improves the living standards of low-wage workers will be beneficial to virtually everyone in the labor force. On the other hand, when a profit-led employment situation exists, there will be a case for cutting taxes on corporate profits on employment-generating grounds. Such a tax cut will, under these circumstances, bring about an increase in after-tax profits, raise the expected profit rate, stimulate investment, increase aggregate demand, and increase total employment and reduce unemployment.

Conclusion

Increasing wages may actually increase the number of jobs. With two options for creating jobs—monetary and fiscal policy—and with the fact that under certain conditions wage increases might help rather than hurt, one might think that the government could make unemployment a thing of the past, as Keynes hoped.

If, to Keynes, the *economics* of full employment seemed promising, the *politics* of full employment must have seemed even more hopeful. Policies to increase the number of jobs are certain to be politically popular. What could be closer to a politician's dream platform than promoting full employment by raising both wages and profits, allowing the government to spend money without raising taxes (deficit spending)?

There can be no doubt that the policies suggested by Keynes have taken some of the roller coaster out of the economy (see Figure 16.2). But the path to full employment turns out to be strewn with thorns, not roses. The main policies we have reviewed—deficit spending, reducing interest rates, and increasing the minimum wage—often work at cross-purposes. Also, as we have seen, not everyone benefits just because (nearly) everyone else has a job.

And, as the next chapter also makes clear, Keynes's optimism neglected to take into account some of the realities of the capitalist economy. In particular, he overlooked the fact that sustained full employment had never been achieved, either in his native England or in the United States. Indeed, few countries have achieved it, and those that have, have done so only temporarily during wartime or more permanently by changing some of the basic rules of the game of the capitalist economy.

SUGGESTED READINGS

George A. Akerlof and Robert J. Shiller, *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism* (Princeton: Princeton University Press, 2010).

Jared Bernstein and Dean Baker, *The Benefits of Full Employment: When Markets Work for People* (Washington, DC: Economic Policy Institute, 2003).

Robert L. Heilbroner, *Beyond Boom and Crash* (New York: Norton, 1978).

John Maynard Keynes, *The General Theory of Employment, Interest, and Money* (New York: Harcourt Brace Jovanovich, 1972 [1936]).

Paul Krugman, *The Return of Depression Economics and the Crisis of 2008* (New York: W. W. Norton, Reprint edition, 2008).

Alejandro Reuss, Luis Rosero, Bryan Snyder, Chris Sturr, and the Dollars & Sense collective, eds., *Real World Macro*, 34th edition (Cambridge, MA: Dollars & Sense, 2017).

Thomas I. Palley, *Plenty of Nothing: The Downsizing of the American Dream and the Case for Structural Keynesianism* (Princeton, NJ: Princeton University Press, 1998).

Studs Terkel, *Hard Times: An Oral History of the Great Depression* (New York: Pantheon, 1970).

17 The Dilemmas of Macroeconomic Policy

"A big jump last month in [U.S. workers'] hourly earnings suggests that the wage stagnation [that has been] plaguing so many Americans may be lifting," one newspaper reported in July 1996. On the same day, however, the same newspaper carried a second story with the headline: "Signs of Unexpected Growth Send Markets Tumbling." This article reported that the people on Wall Street who buy and sell stocks and bonds "reacted with alarm to the disclosure that a lot more Americans were working than had been anticipated." Bond traders began selling their clients' bonds as soon as the employment news came out, and in the stock market as well, prices began falling when the market opened (shortly after the jobs report was released), and they "tumbled" 2 percent by the end of the day.[1]

Just the opposite chain of events occurred in February 1995. Newspapers reported then that job growth had been unexpectedly weak in January, and, as a result, the unemployment rate had risen to 5.7 percent—the first such rise in two and a half years. Most people regard a rise in unemployment as bad news. Not Wall Street. On the day newspapers reported that the unemployment rate had gone up, one headline read: "Big Rally for Stocks and Bonds." The article under this headline began as follows: "Wall Street rejoiced yesterday as Main Street showed signs that its economy was slowing and unemployment was rising."[2] People often find such reports hard to understand.

In this chapter we explain why a capitalist economy cannot tolerate too much employment (too little unemployment). In Chapter 18 we go on to discuss an additional obstacle to the achievement of full employment, the fact that many economists and policy makers believe high employment will cause inflation, that is, a general rise in prices. (Expectations about inflation were a factor in both of the historical situations just mentioned.) We begin here with unemployment.

Unemployment imposes two types of cost on society. The first type of cost is the waste that results from it—the loss of goods and services that are needed and could

[1] *New York Times*, July 6, 1996.
[2] *New York Times*, Feb. 4, 1995.

have been produced but were not. The second type of cost is the personal insecurity and hardship that unemployment creates. Workers never know when unemployment will hit, or where; its occurrence is impossible to predict or to plan for.

It is not just those actually unemployed who suffer. Even employed workers who have never been unemployed have to worry that they *might* be laid off at some time in the future. The most vulnerable workers are, of course, those recently hired, those with few skills, and those who are young or old or members of minority groups. But many others—auto workers with seniority, computer programmers in fast-growing companies, salespeople of all kinds, secretaries with seemingly safe jobs, long-term employees, professionals, technicians, and managers—must also worry about being laid off.

Why is unemployment such a big problem? After all, as we saw in Chapter 16, fiscal and monetary policy can generate additional jobs and reduce the amount of unemployment. So, if only nine out of every ten workers are employed, why not devise a policy that would result in the tenth worker being employed? Why not shorten the length of the work week so that firms would have to employ the tenth worker if they wanted to produce the same amount of output as they had been producing before? The French government passed a law in 1998 limiting the length of the work week to thirty-five hours, in the hope that employers would hire more workers and reduce the unemployment rate. There is still contention over whether the law should remain as is or be made more flexible, or whether the work week should be lengthened again. A proposal to change it, however, provoked angry protests by French workers and students in 2016.

Alternatively, there could be a system that insured people against unemployment, just as they are insured against other calamities—fires, storms, explosions, accidents, and so on. When your home burns down, your insurance pays to replace it. Why not provide all workers with an insurance program so that if a person is laid off, the insurance would make up for his or her lost wages? In fact, the governments of many countries do provide insurance systems of this kind.

Unemployment compensation consists of regular payments made to some unemployed workers from a government insurance fund to which employers contribute.

In the United States, the states collect small fees from companies in proportion to the size of their workforces and then pay **unemployment compensation** (*unemployment insurance*, or *ui*) to unemployed workers. However, the unemployment compensation that is paid to American workers amounts, on average, to only one-half of their lost wages, and fewer than one-fourth of the unemployed receive it, although this varies considerably by state. This is because a worker has to qualify for this benefit in order to receive it, and eligibility rules differ from state to state. And even when a person qualifies for it, the benefit is only paid for at most 26 weeks, so a person who is unemployed for more than that does not receive it after that unless there is a high unemployment rate nationally, *and* Congress mandates extended unemployment benefits. Since 2010, nine states have tightened eligibility rules, and some have reduced the duration of benefits; North Carolina shortened the duration to only fourteen weeks of unemployment benefits. In mid-2016 there were 8 million people unemployed in the United States (by the most commonly used U-3 definition; see the section "Counting the Unemployed" in Chapter 16). Barely 2 million of these were actually receiving unemployment compensation, however—some because they exhausted their benefits long ago, some because they had not worked during the period during which their wages would have counted, and some for other reasons.

Unemployment compensation payments help to stave off acute hardship for as long as they last, but this hardly reduces the fear of unemployment. A person's fear of the consequences of fire would hardly be allayed by fire insurance that was not easily available and that paid to repair only half the damage.

If there are ways of eliminating the insecurity associated with unemployment, why are they not implemented? The answer is simple. A capitalist economy *needs* to have a certain level of unemployment and the insecurity it generates. We explain in this chapter that if there is a sustained period of low unemployment for some reason (war, for instance), and as a result workers lose their fear of being laid off, the economy works in such a way as to restore enough unemployment to bring back the fear of job loss.

In the previous chapter we saw that unemployment does not occur because all of a society's needs have been met. If they were, there might not be enough work to keep everyone employed. But as we saw, unemployment and unused productive capacity can exist side by side with unmet human needs. In this chapter we show that unemployment and its attendant insecurity exist not because we cannot think of ways to employ everyone, but because U.S.-style capitalism requires insecurity in order to function, and unemployment is one of the major components of insecurity.

Is unemployment, then, like death and taxes, unavoidable? Not really. Figure 17.1 shows that unemployment rates vary greatly among advanced capitalist nations, with some countries having had higher average rates and others lower ones over the past four decades. How do some countries achieve lower unemployment rates? The six countries represented in Figure 17.1 are similar in many ways: they face the challenges of rapid technological change and intense international competition that are often seen as obstacles to the achievement of full employment. Yet some of these nations do provide jobs for most of their workers in most years, while others do not.

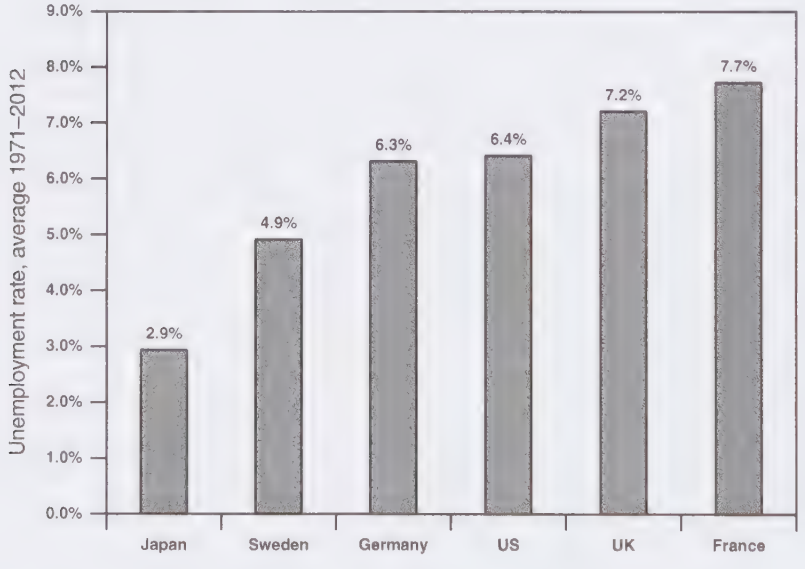


FIGURE 17.1 Average unemployment rates for six countries, 1971–2012.

This figure shows that there are significant and persistent differences among the unemployment rates of advanced capitalist nations. Each country defines its labor force, employment, and unemployment statistics in its own way, but the U.S. Bureau of Labor Statistics has recalculated the unemployment rates for these countries, using U.S. definitions.

Source: U.S. Department of Labor, Bureau of Labor Statistics, ILC Tables: Labor Force, Employment, and Unemployment, Annual labor force statistics, 1970–2012, Full series and underlying levels, By indicator (XLS), "International Comparisons of Annual Labor Force Statistics, 1970–2012," June 7, 2013: Table 1-2: Unemployment Rates, available at http://www.bls.gov/ilc/#laborforce.

The key to achieving low unemployment is replacing the whip of job insecurity with other ways of regulating work and pay. Capitalist nations that have low rates of joblessness often achieve them with mutually beneficial understandings worked out by representatives of business, labor, and government.

In this chapter we explain why unemployment continues to be a major problem in most capitalist countries today even though the theoretical knowledge outlined in the previous chapter, knowledge of how macroeconomic policy can be used to achieve full employment, is widely available. The main idea of this chapter is that *macroeconomic policy cannot achieve full employment in a durable way without fundamental changes in the institutions that coordinate labor processes and regulate labor markets.* This main idea is expressed in four key points:

1. High employment (low unemployment) sustained over a period of years tends to reduce profit, a phenomenon termed the *high-employment profit squeeze.* Profits decline because low unemployment shifts the balance of power toward workers and away from their employers, driving up wages and unit labor costs.
2. The availability of imports and the existence of foreign markets for any country's goods place additional limits on the effectiveness of macroeconomic policies aimed at high employment. Policies designed to stimulate demand for domestically produced goods may lose their effectiveness if consumers and businesses choose to spend their additional income on goods produced in other countries (creating jobs abroad). Similarly, it may not be possible to get foreigners to purchase domestically produced products (generating additional income and jobs at home) without pursuing policies that keep unemployment high and business costs low to enable a nation's firms to sell their products at competitive prices in the global marketplace.
3. Monetary policy and fiscal policy approaches to job creation are both effective, but they may work at cross purposes. Their effectiveness is limited because the government borrowing required by deficit spending often works against the lowering of interest rates that is the goal of full employment monetary policy.
4. Sustained high-employment levels and a further dampening of the business cycle's gyrations are possible, but not without new policies and fundamental changes in institutions.

In the previous chapter we assumed, for the sake of simplicity, that changes in the level of employment do not bring about changes in the level of net output per worker hour, y, or the wage, w. But y and w do, in fact, vary with the level of employment, and in order to identify the difficulties that confront policy makers seeking to promote high employment, we have to understand how they vary. As in Chapter 16, however, we continue to assume that current spending on investment goods, a component of aggregate demand, does not contribute to improvements in productivity (z). Although we close this chapter by listing some institutional changes that might enable a country such as the United States to reduce its rate of unemployment, we assume here that the overall institutional framework of the economy—what we call a social structure of accumulation—continues to be that of "transnational capitalism." (The concept of a social structure of accumulation was developed in Chapter 7.)

The High-Employment Profit Squeeze

When macroeconomic policies and other influences *do* reduce unemployment, problems develop. The source of the difficulty is this. Just as changes in the distribution of income between wages and profits affect the level of aggregate demand and employment (see Chapter 16), the process also works the other way: Changes in aggregate demand and employment affect the distribution of income between wages and profits. Wages tend to rise with high levels of employment (low unemployment), lowering the profit rate, and, conversely, wages tend to fall with high levels of unemployment. One reason for this is that when there is only a small pool of unemployed people, the employed workers will have more bargaining power and will thus be in a better position to push for higher wages. If wages rise, unit labor cost (*ulc*), the wage cost per unit of output, will also rise, and this will tend to reduce the profit rate (see Chapter 12). A reduction in the profit rate will tend to discourage investment, and the process may then come full circle: if investors hold back, aggregate demand will be diminished, output will decline, and unemployment will rise.

Profits will tend to fall when employment is high (unemployment low) for another reason. High employment occurs when there is strong aggregate demand. But if aggregate demand is strong, the output of the economy will most likely be growing rapidly, and the demand for materials used in production—raw materials, energy, and the like—will be rising. Rising demand for input materials (given limited supplies) tends to push up their prices, so another consequence of high employment is that unit materials cost (*umc*) rises.

The **Consumer Price Index (CPI)** is a measure of the average prices a typical family pays for the goods and services it buys, relative to the prices of the same or similar items in a reference year, called the base year.

As labor and materials costs rise, profits will fall—unless businesses are able to raise their prices enough to cover the cost increases. But businesses will generally not be able to pass along all cost increases to their customers in the form of higher prices because competition with other businesses in all parts of the world will prevent them from doing so. Any individual business that raises its price too much will find that its share of the market, its sales, and most likely its profits, will also have fallen. If all businesses raise their prices at the same time, the result will be inflation, a general rise in prices. In the following two subsections we discuss the two types of cost increases that can push down the rate of profit.

In order to fully understand the following account of the high-employment wage push, it helps to understand how the money wage is adjusted to take account of inflation so that we can compare wages at different points in time. Economists use the term **real wage, or real income**, to refer to wages or income adjusted to take account of the effects of inflation. When we talk of an increase in the *real wage*, we mean an increase in the actual (money) wage large enough so that, even if prices are rising, the new money wage buys more than the old money wage. To understand better what we mean by the *real wage*, see the box "Money Wages and Real Wages."

The **real wage, or real income,** refers to wages or income corrected to take account of the effects of inflation.

The High-Employment Wage Push

When employment is low and unemployment high, workers feel fortunate to have their jobs. They may not like their work, but the alternative for them is no work at



Money Wages and Real Wages

If a worker's hourly wage rose from \$10.00 in 2016 to \$10.50 in 2017, was she better off in 2017? It depends on how much prices rose. Could she buy more goods and services in 2017 with her 2017 wage than she could in 2016 with her 2016 wage? If so, then her *real wage* (the *purchasing power* of her wage) did rise from 2016 to 2017.

In this case, her wage rose by 5 percent (\$10.50 – \$10.00)/\$10.00 = 5%). So if the overall price level rose by less than 5 percent, then in 2017 she could buy more goods with her new, higher wage than before. Suppose in both years she worked 100 hours per month, earning \$1000/month in 2016, and \$1050/month in 2017. If in 2016 she spent \$1000/month on goods, and their total price rose in 2017 to, say, \$1025, then in 2017 she could buy the same things monthly and still have \$25 left over (\$1050 – \$1025) to buy more. Since the price of her typical purchases rose by 2.5 percent and her money wage rose 5 percent, her real wage was higher in 2017.

To compare wages over time, we convert *money wages* ("nominal" or "current-dollar" wages) to *real wages* by adjusting for inflation using the Consumer Price Index (CPI). To find the CPI, government statisticians first find the price of a fixed basket of typical consumer goods each year. Suppose the standard basket cost \$10,000 in 1983 and \$15,000 in 1995, and suppose 1983 is named the base year in which the CPI is set to 100. By definition, $CPI_{1995}/CPI_{1983} = P_{1995\ basket}/P_{1983\ basket}$. In this case,

$$CPI_{1995} = (CPI_{1983})(P_{1995\ basket} / P_{1983\ basket}) = 100(\$15{,}000/\$10{,}000) = 150$$

The table below shows the money hourly wage for private sector production workers in selected years, as well as the CPI-W (CPI for urban wage earners and clerical workers), and the resulting real wage. For instance, \$4.14(100)/44.5 = \$9.30.

	Average money wage	CPI-W (July 1983 = 100)	Real wage (in 1983 \$)
1973	\$4.14	44.5	\$9.30
1983	\$8.20	100.0	\$8.20
1993	\$11.05	142.1	\$7.78
2003	\$15.36	179.6	\$8.55
2013	\$20.14	230.1	\$8.75

In 1973 the average *real* hourly wage, expressed in 1983 dollars, was \$9.30 per hour, though the average money wage was only \$4.14 in 1973. From 1973 to 1983, prices rose faster than wages, so the real wage fell. A \$4.14 money wage in 1973 could buy more than a money wage of \$8.20 in 1983. The real wage slid further by 1993-5, but later rose, and by 2013 was nearing its 1973 level.

Sources: U.S. Bureau of Labor Statistics, Employment, Hours, and Earnings from the Current Employment Statistics survey (National), Average hourly earnings of production and nonsupervisory employees, total private, seasonally adjusted, available at http://data.bls.gov/timeseries/CES0500000008; and CPI – Urban Wage Earners and Clerical Workers, available at http://data.bls.gov/timeseries/CWUR0000SA0. In the data, the actual base "year" in the data is given as "1982–84." In the table, the CPI-W shown is for July of each year, since the BLS now reports July 1983 as the month in which the CPI-W was 100.1 (its closest approach to 100.0 in any month in 1983).

all—or at least a prolonged spell of unemployment. In this situation workers will be willing to work hard and well for relatively low pay. They will not be inclined to demand improvements in working conditions, nor would their employers be likely to accept such demands. Furthermore, when unemployment is high, labor unions are not likely to resort to strikes as a way of pushing up wages: with high unemployment there will be many people out of work and eager to take the jobs of striking workers.

As unemployment falls, however, the picture changes. Workers begin to feel more confident about finding new jobs. Employers know that it will be harder to replace a worker who quits or goes on strike. The shoe is now on the other foot: workers can demand higher wages, call for safer working conditions, and insist on a pace of work that is less exhausting. In order to hold on to their current employees—especially those considered essential to a company's ability to make a profit—employers must be willing to accept at least some of their workers' demands, even if doing so will result in a lower rate of profit.

Why do workers have more bargaining power when employment rises? As the level of unemployment falls, workers' alternative income-earning opportunities improve, and this lifts what was defined in Chapter 12 as the fallback wage. Recall that the fallback wage is based on the wage available to a worker in some other job or the amount of income the worker can expect to receive while unemployed. If an employer is willing to pay only the fallback wage, the worker will not care whether he or she keeps the job; in this case the worker is unlikely to be responsive to any demand by the employer for hard work.

If we assume that there is a legally defined rate of unemployment compensation (*ui*) and that a worker can anticipate earning a particular wage in another job, the worker's fallback wage will depend on how long he or she expects to be unemployed (see Chapter 12). The important point here is that the expected duration of unemployment will vary with the rate of unemployment. The more unemployment there is, the longer the worker must expect to remain without work, and, hence, the lower will be the fallback wage. Conversely, as unemployment falls, the expected period of unemployment will be shorter, and the fallback wage will rise.

When the fallback wage rises, a process called the *high-employment wage push* comes into play. Figure 17.2 shows how such a process occurred during the long economic boom of the 1990s in the United States. The first dot on the left side of the graph refers to the second quarter (April–June) of 1992. (In the rest of this chapter we will refer to quarters by roman numerals; in this case we are referring to 1992:II). This point shows that in 1992:II the employment rate was 92.4 percent, the unemployment rate 7.6 percent, and the real wage $7.79. The term *real* means that the nominal wage then being paid would in 1992:II have purchased a combination of goods and services that in 1982 would have cost $7.79 (in the data used for Figure 17.2, the base year being used by government statisticians to make real wage calculations was 1982).

Of course, the nominal wage in 1992 was higher than this—it was more than $11.10 per hour—but prices had also risen since 1982, so the purchasing power of $11.10 in "real" 1982 dollars was considerably less than $11.10. See the box "Money Wages and Real Wages.")

To convert the nominal wage of $11.10/hour in 1992:II to real wages in 1982 dollars first requires knowing that the appropriate Consumer Price Index (CPI) is the

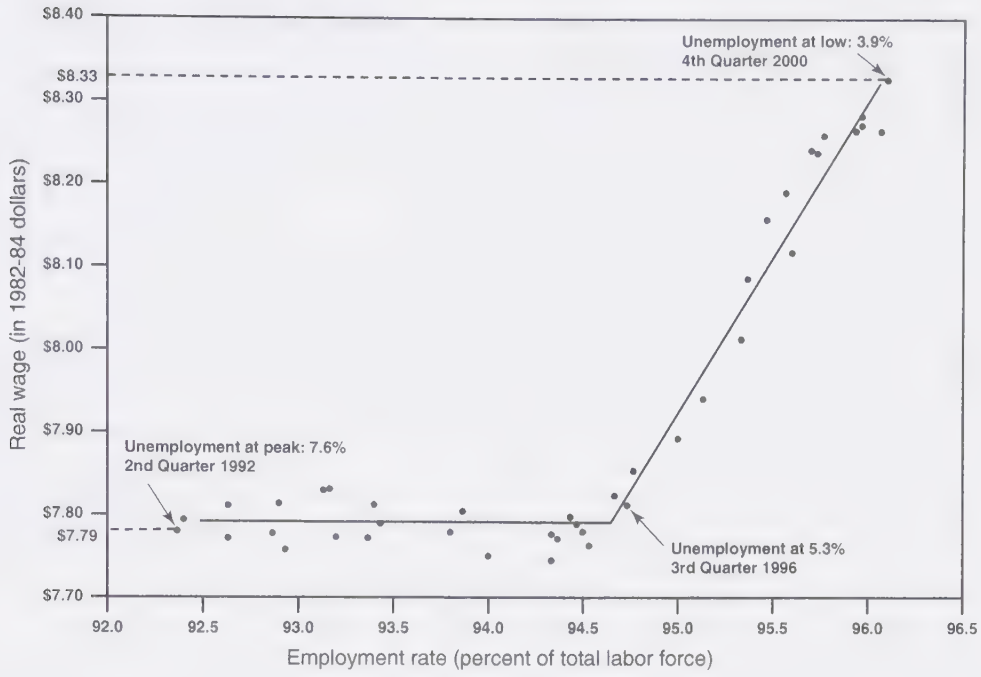


FIGURE 17.2 Employment and real wages during the 1990s boom in the U.S. economy.

This figure illustrates that a "high-employment wage push" took place in the last decade of the twentieth century. The data show that when unemployment fell below about 5.3 percent after the third quarter of 1996, the real wage began to rise steadily.

The figure plots the average real wage of U.S. production workers (in 1982 dollars) on the vertical axis from the second quarter of 1992 (when unemployment peaked at 7.6 percent and the average real wage was $7.79), to the fourth quarter of 2000 (when unemployment reached its cyclical low of 3.9 percent and the real wage was $8.33). The 7.6 percent unemployment rate in the second quarter of 1992 implies that the employment rate then was 100% – 7.6% = 92.4%, and similar calculations apply to the other statements in the graph about the unemployment rates in particular quarters.

Source: U.S. Department of Labor, Bureau of Labor Statistics, detailed statistics on "Employment & Unemployment," "CPI-Urban Wage Earners and Clerical Workers (CPI-W)," and "Compensation & Working Conditions," available at http://www.bls.gov/data/home.htm. Calculations by Arjun Jayadev, University of Massachusetts, Boston.

one for urban wage earners and clerical workers (the CPI-W). We also need to know that the CPI-W in 1992:II was 142.5. The real wage is found as follows:

$$\frac{\text{nominal wage in 1992:II}}{\text{(CPI-W in 1992:II)/100}} = \frac{\$\ 11.10}{142.5/100}$$

which is $7.79, the real wage shown in Figure 17.2. To say that the CPI-W was 142.5 in 1992:II means that the overall price level had, by that quarter, increased by 42.5 percent over its average level in 1982. In 2000:IV, production workers earned an average of $14.50 per hour, but when converted to real 1982 dollars this was just $8.33. From this, if you think it through, you may be able to figure out the CPI-W in 2000:IV.

The U.S. economy expanded steadily between the beginning of 1991 and the end of 2000, and as a result the employment rate rose from 92.4 to 96.1 percent, bringing down the unemployment rate from 7.6 to 3.9 percent. What Figure 17.2 shows is that after a while, unemployment was reduced enough to give workers the power to push up wages. Thus, by 2000:IV the real wage had risen from $7.79 to $8.33. This is indicated by the point on the graph that is higher and farther to the right than all the others.

Notice in Figure 17.2, however, that the real wage remained virtually unchanged, hovering around $7.80, until the second half of 1996. Only in 1996:III did the real wage come unstuck and start to go up. The fact is that U.S. workers, on average, saw no improvement in their living standards during the first five years of the economic boom that began early in 1991. Moreover, if the expansion of the economy and the corresponding fall in the unemployment rate had stopped in 1996, there probably would not have been any improvement in their living standards. It was only when the unemployment rate fell below 5.5 percent (to 5.3 percent) in 1996:III that workers' wages headed up (see the sharp upturn in the line in Figure 17.2). At this point, the real wage began to rise—to $7.81 in 1996:III—and it continued to rise steadily until it reached its peak (for this business cycle) of $8.33 in 2000:IV. The steady rise in the real wage between 1996:III and 2000:IV can be attributed to the fact that during this period the unemployment rate fell from 5.5 to 3.9 percent.

Wage increases will drive up unit labor cost unless labor productivity (output per hour) goes up just as fast as (or faster than) the wage. But labor productivity generally is unlikely to rise because at high rates of employment (low rates of unemployment) workers will be less afraid of losing their jobs. Accordingly, they will hardly be inclined to put out the higher levels of work effort that would be required to offset (with productivity increases) the rise in their wages. In addition, a low unemployment rate generally means that workers with less experience and skills have recently been hired, and they will likely be less productive until they gain some experience.

The Materials Cost Push

Just as high employment rates produce higher labor costs, the cost of materials will also go up as employment rises. And the increase in unit materials cost (*umc*) will have the same effect on the profit rate as the high-employment profit squeeze.

When output and employment increase, more materials must be used in the production process in order for output to increase. As firms demand more materials, the sellers of materials will probably be able to raise their prices. Such price increases will in turn tend to push the rate of profit down—hence the high-employment materials cost push.

To understand how the materials cost push works, consider a particular input, for example, the oil that is used in a great number of production and consumption processes. Increases in the output of the economy will not only generate more employment, they will also require the use of additional amounts of other inputs, such as oil, and as a result the demand curves for these inputs will shift to the right. As explained in Chapter 8, the effects of such shifts will be to create excess demand in input markets, and this will tend to push up the prices of input materials. The result is that just as a rise in employment ensuing from an increase in aggregate demand will tend to raise the wage and squeeze profit, an increase in the demand for material inputs will tend to have the same effect by raising unit materials cost.

Rising Costs Squeeze Profits

Would employers, then, like to have the employment rate be as low as possible (the unemployment rate as high as possible)? No, of course not. It is true that widespread unemployment strengthens the bargaining power of employers vis-à-vis both workers and materials suppliers. But to make a profit a firm has to do more than produce a product at a low cost, it has to produce and sell a high volume of its product.

Production of a high volume of output cannot be achieved without employing substantial numbers of workers. Moreover, selling large amounts of products is easier to do when many people are employed and earning and spending plenty of money. At high rates of unemployment the income of many workers will be reduced, and since workers are also consumers, the total demand for the economy's output will be commensurately lower. It is clear that unemployment has a direct effect on the income of unemployed workers; in addition, however, high unemployment weakens the bargaining power and limits the wage increases of employed workers as well. Thus, high rates of unemployment will hold down the incomes of all workers.

We can conclude, then, that employers will see that some levels of employment are too *low* to allow them to maximize their profits; in this case, consumers will not have enough money to buy their products. Similarly, employers will see that some levels of employment are too *high* to allow them to maximize their profits; in this case, wages and other costs will be too high.

When employment is below a certain level, increases in employment (reducing unemployment) will raise profits. However, as employment increases, a point will be reached at which the profit rate will be at its maximum and further increases in employment will only cause the profit rate to fall. This is because beyond the profit-maximizing rate of employment, rising wage and materials costs will squeeze profits. When increases in employment have a negative effect on the profit rate, we say that there is a **high-employment profit squeeze**.

A **high-employment profit squeeze** occurs when the high demand for labor creates both labor and materials cost increases, which in turn reduce profits.

Here is how the high-employment profit squeeze works. Intuitively one might think that with more labor hours employed more units will be produced, and if these additional units of output can be sold, total profit (R) will increase. Even if the additional units can be sold, however, there is a problem with this line of reasoning: it is based on the assumption that the cost of producing the additional units will not rise as employment and output rise.

Under certain circumstances the assumption of constant costs might be a reasonable one to make, and in this case total profit would certainly increase as employment and output expanded. This might be the situation when, at low rates of employment (high unemployment), increases in demand for labor and materials will not result in significant increases in wages or materials costs. These are circumstances when increases in employment (N) are most likely to have a positive effect on the rate of profit (r).

Under other circumstances, however, assuming labor and materials costs to be constant would not be justifiable because, in fact, these costs will increase as additional labor and materials are used to produce more output. Eventually, the increases in input costs will outweigh the positive effect of increases in employment and output on the rate of profit. In such situations the net effect of increases in employment (N) on the rate of profit (r) will be negative, hence there will be a high-employment profit squeeze.

Figure 17.3 shows how movements in the two important variables, unit cost (*uc*) and output produced (*Z*), will have varying effects on the rate of profit (*r*) at different levels of employment. The top panel of the figure shows that as employment expands (moving from left to right on the horizontal axis), unit cost (*uc*) rises at first

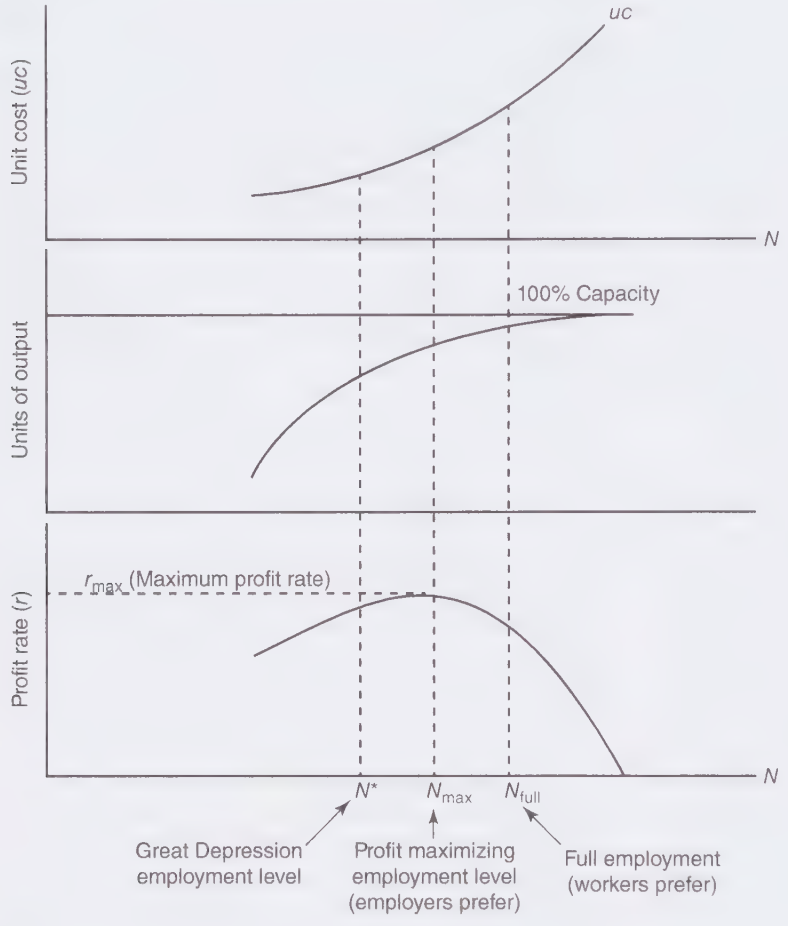


FIGURE 17.3 Unit cost, output, and the profit rate.

This figure illustrates how a rising number of labor hours employed (*N*) affects the number of units of output and the cost per unit, and how these in turn affect the profit rate of a firm. Moving from left to right in the figure, the bottom panel shows that the rate of profit rises as long as the positive contribution to profit resulting from producing and selling more output (middle panel) outweighs the negative effect on the profit rate resulting from rising unit cost of production (top panel). At the point where these positive and negative effects are just in balance, the level of employment will be the one (N_{max}) where the profit rate is at its maximum. As employment increases beyond that point, unit cost will start rising *more* rapidly than before, while output will start to rise *less* rapidly than before as the enterprise reaches its maximum productive capacity, and cannot quickly expand that capacity. Rising unit cost begins to drive down the profit rate, which is not what the employer wants. Thus the N_{max} level of employment—the one that maximizes profit—is the one the employer prefers, but it most likely is not full employment (N_{full}); some people will still be unemployed. Workers would prefer a higher level of employment, but have no easy way to reach it. The N^* level of employment shown in the bottom panel simply reminds the reader that the equilibrium level of employment varies depending on how much demand is coming from households, businesses, government, and foreigners. In principle, depending on these factors, N^* could be at almost any level of employment, although it is usually below N_{max}.

slowly and then, at higher levels of employment, goes up more and more rapidly. Thus, the unit cost curve is nearly flat at low levels of employment and then becomes steeper and steeper as employment rises. Rising unit materials cost (*umc*) is a factor driving up unit cost here, but the main contributor to the upsurge in unit cost is the high-employment wage squeeze (increasing *ulc*) discussed earlier and displayed in Figure 17.2.

The middle panel of Figure 17.3 shows that as employment increases, output (as one would expect) also increases. At first the output curve rises somewhat steeply, showing that each additional labor hour hired results in a significant increase in output produced. However, as employment goes to higher levels, the increases in output generated by additional employment become smaller and smaller, and the output curve flattens out as firms approach their maximum output (maximum productive capacity).

Finally, the bottom panel of Figure 17.3 shows how the two counteracting influences—the positive effect of increases in output, on the one hand, and the negative effect of rises in unit cost, on the other—determine what the rate of profit will be at different levels of employment. At low levels of employment (high unemployment), each increase in employment brings about a sharp increase in output (shown in the middle panel) but only a modest increase in unit cost (shown in the top panel). As a result, the profit rate rises as employment increases. In this range of employment (where the line in the lower panel is rising), the contribution that each additional unit of output makes to profit outweighs the negative effect of the rising costs associated with increases in the amount of labor employed.

However, as employment continues to increase, additional hours of labor produce smaller and smaller increments of output, so the output curve levels off and the contribution of output to profit diminishes as more labor is hired. As additional hours of labor are hired, unit costs rise at an increasing rate (see the upward curving line in the top panel). At some point the negative effect of increases in cost (with both unit labor cost and unit materials cost rising) outweighs the (declining) positive effect on profit of more output—and at this point the profit rate itself (shown in the bottom panel) begins to decline. In the bottom panel the level of employment at which the profit rate is at its maximum is labeled N_{max}. At this level of employment the two influences on the profit rate just balance out, and any increases in employment beyond this point will only reduce the rate of profit.

Figure 17.4 shows the high-employment profit squeeze during the 1990s economic boom in the United States: A rising employment rate initially led to a rising profit rate, but as the employment rate continued to rise, the profit rate eventually peaked and began to fall. We start in the lower left area of the figure with the data point for 1992. It shows that there was 7.6 percent unemployment (92.4 percent employment) in that quarter, and the profit rate was 5.7 percent. As employment rose over the next five years, the profit rate rose with it, just as the bottom panel of Figure 17.3 suggests. The profit rate peaked at 9.5 percent in 1997:III, and then, as the employment rate continued to rise and the unemployment rate fell to 3.9 percent, the profit rate declined to 5.8 percent, again following the pattern laid out in Figure 17.3.

The profit rate peak occurred a full year *after* the real wage upturn shown in Figure 17.2 (1997:III as opposed to 1996:III); this is consistent with the analysis in Figure 17.3, which explains why the profit rate may keep rising even as workers successfully bargain for higher real wages. It is only when the increases in both labor and materials costs overwhelm the positive effects of economic expansion

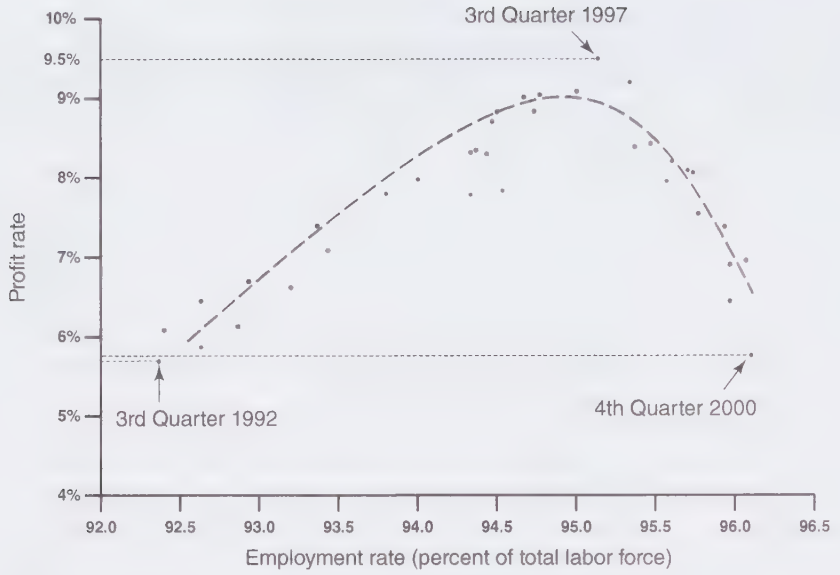


FIGURE 17.4 The employment rate and the profit rate during the 1990s boom in the U.S.

This figure charts the rise and fall of the profit rate during the economic boom of the 1990s, covering almost exactly the same period as Figure 17.2. In both figures, the data show that a high-employment profit squeeze happened in the U.S. economy during the 1990s, following the pattern described in Figure 17.3. Figure 17.2 plots real wages against the employment rate, while this one plots the profit rate against the employment rate. The profit data are for the U.S. nonfinancial business sector; this excludes banks but represents more than 85 percent of the U.S. corporate business sector.

Source: U.S. Department of Commerce, Bureau of Economic Analysis, National Income and Product Accounts (NIPA), Table 1.14: Gross Value Added of Domestic Corporate Business in Current Dollars and Gross Value Added of Nonfinancial Domestic Corporate Business in Current and Chained Dollars, available at https://www.bea.gov/iTable/iTable.cfm?ReqID=9&step=1#reqid=9&step=3&isuri=1&903=55. Calculations by Arjun Jayadev, University of Massachusetts, Boston.

that the profit rate begins to fall. In the 1990s boom, the cost increases pushed the profit rate down to 5.8 percent by 2000:IV, almost exactly the same as the 5.7 percent profit rate at the beginning of the expansion in 1992:III. By the end of 2001:I the U.S. economy had stopped growing and begun to contract, beginning the brief 2001 recession.

And so policy makers who seek to achieve full employment cannot escape the logic of a capitalist economy. Policies that succeed in increasing employment beyond a certain point will eventually drive the profit rate down. This will reduce investment and, as a result, reduce employment.

Beyond the economic logic of capitalism, there is a political obstacle to achieving full employment. People who receive most of their income from profits often have considerable political influence (money talks!), and many of them will resist the adoption or implementation of any policy that might cut into profits. (In Chapter 19 we consider how the government works in a capitalist economy, including the connection between money and politics.) If the stock market falls when the government reports higher employment, it should not be surprising that some business owners will openly state that in certain situations, at least, an increase in unemployment—even if brought on by a recession—may be a good thing.

Exports, Imports, and Aggregate Demand

The high-employment profit squeeze is not the only obstacle preventing the achievement of full employment. High-employment policies can also run into problems in the international economic arena. The main problem is that reducing domestic unemployment may conflict with the need to remain competitive in world markets. When domestic unemployment is low and wages start rising, prices will rise as employers attempt to hold on to their profit margins. As a result, buyers may find that goods produced in other countries are less expensive than domestically produced goods.

Unless all countries are pursuing full-employment policies, the higher costs and prices resulting from one nation's high-employment policies may mean that its goods and services will not be competitive in international markets. The result will be a fall-off of exports and a flood of imports as consumers and businesses shun high-priced domestically produced goods and buy cheaper imports. Some of this may occur because U.S. transnational corporations (TNCs) shift production to overseas locations in response to rising U.S. wages—but of course this will also slow down, and perhaps reverse, the rise in wages.

Two problems then arise. First, as we will see shortly, falling exports and rising imports mean that there will be less aggregate demand for goods produced domestically. And second, there is the long-term reality that buying more from, while selling less to, the rest of the world cannot go on indefinitely. In the short run an excess of imports over exports can be paid for by borrowing, as the United States did on a massive scale during the late 1990s and early 2000s. But sooner or later the loans will have to be paid back. Here, we focus on the first problem, the short-term effect of trade deficits on aggregate demand.

We are interested in the demand for goods produced domestically, no matter whether such demand comes from inside or from outside the country. Only goods production within a nation employs that nation's workers, and it is domestic employment that we are interested in. So, for example, the sale of a U.S.-built drill press to a German firm and the sale of a pound of U.S.-grown wheat to a family in Mumbai both contribute to U.S. aggregate demand and employment. By contrast, when an Atlanta business buys a Japanese-built computer or a Chicago resident buys a bar of Swiss chocolate, that will not generate jobs in the United States. Such purchases will contribute to aggregate demand, and hence increase employment, in other countries.

Once we take into account imports and exports, we can see that not all U.S. demand for consumer goods is part of aggregate demand in the United States. As noted, an American's purchase of Italian wine is not. Neither is all demand for investment goods part of aggregate demand. The purchase of a Japanese-made computer by an American business is not. So when international trade is taken into account, the demand for *imports*—demand that is not directed toward domestically produced goods—must be *subtracted* from aggregate demand. On the other hand, the demand for *exports*—the demand by foreigners for goods produced domestically—is *added* to aggregate demand.

A simple way to make the required adjustments is to use the term *net exports* to represent the difference between the total value of exports of goods and services and the total value of imports of goods and services. Net exports are positive—they increase aggregate demand—if a country exports more than it imports; conversely, net

imports are negative—they decrease aggregate demand—if imports exceed exports. We use the letter X to represent net exports:

$$X = \textit{net exports} = \textit{total exports} - \textit{total imports}$$

With this simplification, we can take into account both exports and imports by just adding X to the equation for aggregate demand (Equation 16.7) that was developed in chapter 16:

(17.1) $$\text{Aggregate demand} = AD = C + I + B + X$$

As before, C = total demand for consumption goods, I = total demand for investment goods, and B = government deficit spending. From Equation 17.1 it tells us that an increase in X, like an increase in B, C, or I, will increase aggregate demand and promote higher levels of employment (assuming nothing else changes), while a fall in net exports will have the opposite effect.

The Demand for Exports and Imports

As with investment and consumption, we need to know what determines the level of net exports. The short answer is that the level of net exports will be determined (a) by the level of incomes at home and abroad and (b) by the prices of foreign goods as compared with those of comparable domestically produced ones. To get a fuller understanding of how the value of net exports is determined, we need to look at each of these relationships more closely.

First, consider the effect of rising incomes in a particular country on that country's net exports. As the incomes of U.S. families and businesses grow, both will buy more because there is more money to spend. Among the things they will buy are imported goods, including cars, machinery, and other items. For this reason imports tend to rise when incomes rise. Exports, however, do not grow because incomes grow. If rising incomes cause imports to rise while exports remain at their existing level, net exports will fall.

Since incomes rise or fall with employment, the same conclusions can be reached by looking only at changes in the employment rate. At low levels of employment, incomes will also be low and hence spending by consumers and businesses—including spending on imports—will be limited. But increases in employment will generate more income, and, as noted, some of the additional income will be spent on imports. Hence, imports will tend to rise with increases in employment and fall with contracting employment (rising unemployment). Unless the level of exports changes, then, rises in employment will reduce net exports, and less employment will have the opposite effect.

Just now we have been assuming that a country's exports are constant. Of course, this is unrealistic. Just as changes in a country's own level of employment and income affect its imports, changes in the levels of employment and income in other countries affect its exports. Consider the situation of the United States. If in the rest of the world unemployment is high and people are poor, the demand for U.S. exports will be low. If at the same time people in the United States are prosperous and buy a lot of imports, net exports will also be low, perhaps even negative. But higher incomes in other countries will lead to greater U.S. exports and net exports.

The other main influence on the level of U.S. net exports is the prices of imports into the United States compared to the prices of goods made at home. If domestically

produced goods are cheaper than goods produced elsewhere, then imports will be low and exports will be high (foreigners will find that the goods they import from the United States are more attractive than goods produced in their own country). In this case U.S. net exports are more likely to be positive. But of course if goods produced elsewhere are relatively cheap (or equivalently, goods produced in the U.S. are relatively expensive), then the volume of imports will be high compared to the volume of exports. This will be so because buyers in the U.S. (be they individuals or firms) will prefer to spend their money on foreign-made goods, so imports will be strong and net exports may even be negative.

It is also true, however, that U.S.-based TNCs deliberately use transfer pricing to shift their profits to low- or zero-tax jurisdictions, and so reduce their tax liability. One way they can do this is if their affiliates outside the United States charge high prices to the U.S. firm for goods and services (such as the right to use intellectual property) that the affiliate holds in a tax haven. They can also do the same for the goods they produce abroad and ship to market in the United States: charge a high price, which swells the reported value of U.S. imports. The U.S. merchandise trade deficit grew from the mid-1970s to the present, which likely is largely a consequence of the growing presence of U.S. TNC affiliates abroad using such practices. In 2015 the annual U.S. trade deficit was about $500 billion.

To make clear the influence of relative prices on the value of net exports, we use P_z to refer to the price of goods produced domestically (as in earlier chapters), and P_{im} for the price of goods produced elsewhere (P_{im} for price of imports). The ratio of the price of goods produced elsewhere to the price of domestically produced goods is the real price of imports:

(17.2) $$\text{the real price of an import} = P_{im}/P_z$$

Suppose that a country called Grainland exports grain and imports plows, while Plowland exports plows and imports grain. Suppose further that one plow on the world market can be bought with 10 bushels of grain. So the real price to Grainland of importing a plow from Plowland is 10 bushels of grain. In monetary terms, a set of prices consistent with this assumption would be $20 for a bushel of grain and $200 for a plow. Then from the standpoint of Grainland P_{im}/P_z will be $200/$20, and the real price of importing a plow will be the 10 bushels of grain that Grainland must export in order to earn the $200 it needs to buy its plow.

If the real price of the imported plow rises to 11 bushels of wheat per plow (such as if the plow's price rises to $220), this will also affect the size of the surplus product in Grainland. Now Grainland will have to export one additional bushel of grain for every plow it wants to import. If nothing else changes, this will leave less and less of its total output as surplus product for those who control the production process. As Chapter 10 made clear, when the price of material inputs rises (and nothing else changes), profit falls. Thus rising real prices of a country's imported inputs will reduce the country's surplus product. That is how it works in our hypothetical Grainland. What about in a real modern economy?

The Foreign Exchange Rate

Imagine that a farmer in the United States wants to import a plow in from Germany. Buying the imported plow (a transaction likely to be carried out by a U.S. agricultural equipment dealer) will require two steps. The first step will be to sell dollars

and buy euros (the currency used in Germany), assuming that the German producer of the plow wants to be paid in euros. The second step will be for the U.S. dealer to pay the German producer the euros to buy the plow. The cost of this step in the transaction will depend on how many euros each dollar will buy on the day the transaction takes place.

Thus, the dollar price of the import depends on two things. First, it depends on the price of the import in euros, and second, it depends on how many euros a dollar will buy, known as "the value of the dollar (expressed in euros)" or "the **exchange rate** of the euro to the dollar." (It is also the number you get when you do an Internet search on "USD to euro").

A **foreign exchange rate** is how much of one currency another will buy. The exchange rate of the dollar to the euro is the number of euros a dollar will buy, and is also known as "the value of the dollar in euros").

Of course, given the fact that the exchange rates of currencies are allowed to fluctuate in global currency markets, the dollar exchanges for different amounts of the currencies of other countries, and these exchange rates change daily. Thus, on August 6, 2016, a dollar exchanged for .90 euros, 101.8 Japanese yen, or 0.77 British pounds.

The price of an import in dollars (P_{im}) is the price of the import in its own currency (P_{imf}) divided by the value of the dollar (v) expressed in the number of units of that currency that a dollar will buy ("the exchange rate of the dollar to the euro").[3] The "f" in "P_{imf}" signifies that this is the price of the import in its own "foreign currency." Thus

(17.3) $$P_{im} = P_{imf}/v$$

If the plow in our example had been imported from Germany, knowledge of the exchange rate between dollars and euros on the day the plow was imported would enable us to figure out its original cost in euros. If the plow had been imported on August 6, 2016, when the value of the dollar (v) was 0.90 euros, the price of the plow ($200 in the U.S.) would have been 180 euros in Germany: $200 = 180 €/(0.90 € per $). (For simplicity here we make the unrealistic assumptions that there are no transportation costs and there is no markup of the price by the agricultural equipment importer.) If the German price of the plow remained constant, a rise in the value of the dollar (meaning that a dollar could buy more euros than before) would reduce the price of the plow as an import to the United States. You can check that Equation 17.3 will give you this result, or just use your intuition: a higher value of the dollar means that each dollar will purchase more in another country, so fewer dollars will be required to pay for the plow. When the value of the dollar rises, then, people in the United States will tend to buy more products made in other countries (more imports), and the opposite will be true when the value of the dollar falls.

In the early 1980s, the sharp increase in the value of the dollar led to complaints by manufacturers of steel and machinery, as well as by crop exporters, that they were having a hard time competing both on the world market and at home. In the mid-1980s, when the value of the dollar was allowed to fall, these problems eased considerably.

In general, when a country's currency rises in value (relative to other currencies), the country will import more goods, because those goods now are cheaper than

[3]Note that sources published before the year 2000, roughly speaking, may use these terms differently; in an earlier period many economists used the term "the exchange rate of the dollar against the yen" to mean the *inverse* of "the value of the dollar expressed in yen" (so that the product of the two quantities was 1).

before, and will export fewer goods, because for buyers in other countries the goods have become more expensive than before.

We explain what determines the value of a particular country's currency in the following section. So far, we have established that net export demand is determined by the following factors:

- household and business incomes at home and abroad
- prices at home and abroad
- exchange rates between the domestic currency and various foreign currencies

With this information we can now consider the effect of exports and imports on employment and on high-employment policies.

International Trade and Macroeconomic Policy

How do high-employment macroeconomic policies work when they are pursued in an economy with significant exposure to the global economy? Do the availability of imports and the existence of foreign markets for domestically produced goods place limits on the effectiveness of macroeconomic policies aimed at high employment? In considering the relationships between international trade and macroeconomic policy one is confronted by two basic questions: First, can policies deliberately aimed at promoting net exports be successful in achieving high levels of domestic employment? Second, does being competitive in a global economy require pursuing macroeconomic policies that keep unemployment high in order to maintain costs of production low enough to enable businesses to price their products competitively?

Promoting Net Exports

Promoting exports and discouraging imports (promoting net exports) would seem a promising way to promote domestic employment. If net exports could be promoted without bringing about changes in any other important economic variables, the result would certainly be an increase in jobs. But in the real world, keeping variables constant is often difficult.

Imports can be discouraged by imposing taxes (tariffs) on them or by putting restrictions (quotas) on the quantity of goods that can be brought in from another country. The United States once did this successfully to hold down automobile imports from Japan. Such efforts as the "Buy American" campaigns that have been mounted in the United States may also reduce the demand for imported goods. The government of a particular country can promote exports by paying some of the costs of their production (as the United States does with farm products) or by persuading other countries to give favorable treatment to its exports (as various imperialist nations have done with their colonies).

Policies aimed at discouraging imports have worked at times in the past. In the nineteenth century both the United States and Germany used tariffs to protect domestic industry against lower-cost manufactured goods from Great Britain. By doing this they made possible the development of their own manufacturing industries while at the same time promoting employment and growth in their own economies.

In recent decades the governments of South Korea and Japan have actively promoted exports of their nations' products by subsidizing the export prices of them. Partly as a result of such policies, both of these countries achieved high employment rates and rapid economic growth.

However, export-promoting and import-restricting strategies will not work for everyone, much less for everyone at the same time. After all, more than one country can play the game, and the success of any one country will be limited by what other countries may do.

In the first place, other countries may retaliate. If the United States makes it difficult for Brazil, for example, to export shoes to the United States, Brazil is likely to place obstacles to the exportation of farm equipment to Brazil. Any country attempting to create jobs by keeping out goods from other countries may find its own goods kept out of other countries' markets. The result may be a loss in jobs, rather than a gain.

Second, it is impossible for all countries to expand their net exports at the same time. This is true for the reason that the total amount of net exports in the world must always add up to zero. Imagine the case of a two-country world with one of the two countries enjoying positive net exports. In this case the second country would have to be experiencing negative net exports. With the first country exporting more than it is importing, the second country would have to be importing more than it is exporting, and the second country's negative net exports would have to balance exactly the first country's positive net exports. The logic is the same for a world of two hundred countries, and what is involved here has nothing to do with retaliation; it is a matter of simple arithmetic.

Third, even if two countries manage to increase their exports to each other, it could happen that neither one experiences an increase in its net exports, in which case there would be no increase in aggregate demand and no increase in employment in either country. Again, there is no retaliation involved here, but in neither country will a successful export promotion strategy result in an increase in employment.

Imagine that the United States successfully promotes its exports of chemical products to Europe and that Europe successfully promotes its exports of automobiles to the United States. The exporting industries of both countries will be booming and will be hiring more workers. Chemical workers in the United States will have no trouble finding work. Auto workers in Europe will not worry about unemployment. But will there be a net gain in jobs in either country? Probably not. Europe's exports of automobiles will be automobile imports in the United States, and U.S. autoworkers will be laid off. In Europe, it will be the chemical industry workers who will suffer unemployment. The increases in U.S. imports from Europe will count as a subtraction from U.S. net exports, and the increase in Europe's imports from the United States will be a drag on Europe's net exports. If both double their exports, the *net* exports of both Europe and the United States will remain unchanged.

Note that *consumers* of automobiles and chemical products in Europe and the United States may benefit if Europe is better at producing automobiles and the United States is better at producing chemicals. In this case the increasing specialization of Europe in automobile production and the United States in chemical products will lead to greater total output of both types of products. Hence, there could be increases in total consumption of both automobiles and chemical products.

While strategies to increase net exports cannot work for everyone, there are circumstances in which they may work. Promoting exports and discouraging imports can lead to an increase in the total number of jobs in a particular country if

the country adopting such policies has one or both of two unusual advantages. First, the policies may work if the country is very powerful and can thus prevent retaliation. And second, net export promotion may succeed if a country can produce many goods much more cheaply than other countries at its prevailing wages and existing foreign exchange rate. A country that can claim either of these advantages will most likely be able to use net export promotion strategies successfully to improve its competitive position in the global economy.

Competing in Global Markets

Is it possible for a country to compete in global markets without maintaining a high rate of unemployment? To compete in global markets a country must be able to sell its products at prices that are competitive with the prices of other countries' products. But to preserve low prices, costs must be relatively low, including unit labor cost. As we know from the discussion of the high-employment wage push earlier in this chapter, however, unit labor cost rises when unemployment falls. So how does a country achieve high net exports while at the same time pursuing high-employment macroeconomic policies?

Here is the problem. With high unemployment, prices can be low, and a country's goods can be competitive on world markets. Exports will boom, and there will be little demand for imports since they cannot compete with domestically produced goods. In this situation net exports will be high. But high net exports will add to aggregate demand and thus generate higher employment.

As we have seen, an increase in employment will lead to falling unemployment and rising costs. As this happens businesses will raise their prices as much as they can to maintain their profit rates in the face of rising costs. But as prices go up, exports will fall off, imports will increase, and eventually imports will exceed exports so net exports will be negative. At this point international trade will be having the effect of reducing aggregate demand and placing a damper on further employment increases.

The impact of rising costs and prices on a country's net export position can be explained more precisely using Equation 17.2. With increasing prices the real price of imported goods will fall (as P_z rises, P_{im}/P_z falls). Responding to the more favorable price of goods produced elsewhere, buyers will buy fewer domestically produced goods and more imported goods. The result will be falling net exports, possibly even subtracting so much from aggregate demand that no further increases in employment will occur. So as imports rise and exports fall, the effect is not to create jobs at home but to create them elsewhere.

The independent macroeconomic policies of different countries have a distinct "after you" quality. Each country would like *the others* to adopt expansionary macroeconomic policies—for example, deficit spending and low interest rates—so that its exports will increase, generating more employment and income for its own people. The other countries' higher employment rates and incomes will increase the demand for the home country's exports, thus creating jobs in the home country. If all countries adopted deficit spending policies at the same time, their efforts would be mutually supportive. But any one country trying to expand employment alone could well face inflation, mounting imports, and little job creation to show for it.

The difficulties we have just identified arise from two sources: the high-employment profit squeeze and the lack of coordination among nations pursuing their own macroeconomic policies. Thus, the success of high-employment policies requires two things. First, a way must be found to reduce the cost pressures that

develop with high employment. Second, the major governments of the world must adopt macroeconomic policies that are complementary rather than competitive and self-defeating. This would involve coordinating decisions to expand aggregate demand rather than discouraging imports or waiting for other countries to expand their aggregate demand. We consider some policies to achieve these objectives in our concluding section. Before considering those options, however, we examine a last major dilemma of macroeconomic policy.

Monetary and Fiscal Policy at Odds

Both monetary policy (influence by the central bank on interest rates) and fiscal policy (changes in government spending and taxation), can be effective tools for increasing the amount of output and employment. But they often work at cross purposes, with the influence of one counteracting the influence of the other, so that the potentially beneficial effects of both are greatly reduced.

The main problem is that deficit spending designed to increase aggregate demand and generate more jobs requires the government to borrow money—often vast amounts of it—in order to make ends meet. When the government borrows large sums, it is more difficult for others to get credit (consumers wanting to finance a car purchase, for example, or businesses wanting to secure a loan to purchase some new machinery). The credit squeeze occurs because the government borrows money from the same sources (in the same markets) that businesses and families use; all potential borrowers are in competition with one another.

The result of such competition is that some businesses and families wanting to borrow money cannot get loans, and those who do obtain loans end up paying higher interest rates on them. This is just exactly what is *not* needed if there is a problem of insufficient aggregate demand, since, as we have seen, higher interest rates discourage both investment and consumer buying, thus reducing aggregate demand. This type of negative outcome will at least partially offset the positive effects of the increased spending.

Thus, spending and borrowing by the government may have the effect of reducing the spending and borrowing of families and businesses. This phenomenon is called "crowding out": the government's borrowing "crowds out" families and businesses from the markets where they all compete to borrow money, and this has the effect of reducing the spending of families and businesses. The positive effect of deficit spending on aggregate demand must therefore be measured taking into consideration its "crowding out" effects.

The extent of crowding out effects is hotly debated among economists. Clearly, the effects depend on how seriously government borrowing reduces the availability of credit to consumers and businesses and, in turn, on how much a reduction of credit availability affects consumer spending and investment spending. Unfortunately, there is not much agreement on these issues.

If, rather than borrowing, the government simply prints more money to finance its spending, it does not compete in credit markets against families and businesses. Still, crowding out may occur because the government will spend the money it has printed, and these dollars will then compete with the spending of families and businesses for the available supplies of goods and services. As we explain in the next chapter, when governments, families, and businesses spend

beyond their means, either through borrowing or, in the case of the government, through printing additional money, inflation is likely to be the result. This is one reason why the government does not generally finance its spending by printing money.

What Determines the Interest Rate?

The interest rate (i) is simply the price of renting money. If the interest rate is 10 percent per year, the borrower who borrows $100 must pay back $110 at the end of the year. The $10 difference is the annual rent on $100.

There is no single interest rate. Interest rates vary depending on the length of the loan, its purpose, and the likelihood that the borrower will pay it back, among other things. But because interest rates tend to go up and down together, we will discuss "the interest rate" as if there were an average of all interest rates.

Very simply put, the interest rate is determined by supply and demand, as are other prices. The relevant supply in this case is the supply of money available for lending, and this is called the *supply of loanable funds*. The demand in question is the *demand for loanable funds*. The interest rate is thus determined by how much money individuals, firms, and the government want to borrow and how much individuals, firms (especially banks), and the government want to lend.

The supply of loanable funds depends on various factors that affect how much money banks, households, firms, and governments, both at home and abroad, want to lend. This depends on government policies, such as those regulating the lending practices of banks; on behavior of households—how much they want to save at various interest rates, for example; and how much people or governments abroad wish to lend in the U.S. market, such as by buying U.S. bonds issued by government or by corporations. A change in any of these determinants will shift the loanable funds supply curve.

The demand for loanable funds depends on how much governments, households, and businesses want to borrow. Governments borrow to finance deficits or refinance existing debt, to build local infrastructure like schools or water systems; and so forth. Households borrow to buy goods on credit (autos, homes, and so on); and businesses borrow to make investments or finance their operations. A change in any of these determinants will cause the loanable funds demand curve to shift left or right.

Wherever the two curves happen to be on any given day, the interest rate will move toward the market-clearing level—the level at which the two curves intersect and the supply of loanable funds (the money available for borrowing) equals the demand for loanable funds (the amount of money various economic actors want to borrow).

In short, the interest rate is the price of borrowed money, and this price will be determined in the same way that prices are determined in most other markets. A change in any of the factors determining the location of either the supply or the demand curve—for example, people, businesses, or government changing their decisions about how much to lend or borrow—will cause the market-clearing interest rate to rise or fall.

Figure 17.5 shows how the demand for loanable funds and the supply of loanable funds interact to determine the interest rate. The demand-for-loanable-funds curve indicates how much money all the potential borrowers want to borrow at each interest rate. For example, if the interest rate is at the level "a" in the figure, borrowers will

want to borrow a total of "Da" dollars. The demand for loanable funds stems from many different sources. These include consumers wanting to finance a car purchase or to get a mortgage to buy a house; businesses wanting to borrow money to make an investment (such as building a new plant); and governments (local, state, and federal) wanting to (deficit) spend more than they are collecting in taxes. Businesses must have enough money on hand to do business, and consumers must have enough money to keep up the flow between income and purchases; and sometimes they need to borrow for these purposes. Businesses might also want to borrow money to buy another company or to buy land for speculative purposes (individuals might do this as well). And the list goes on. If the interest rate falls, the demand for loanable funds will increase, as is indicated by the fact that the demand curve slopes downward to the right.

The supply-of-loanable-funds curve is also indicated on the graph. It should be read the same way as the demand-for-loanable-funds curve. If the interest rate is at "a" percent, individuals, businesses (including banks), and the government will want to lend a total of "Sa" dollars. If the interest rate goes up, they will want to lend more, so the supply curve slopes upward to the right.

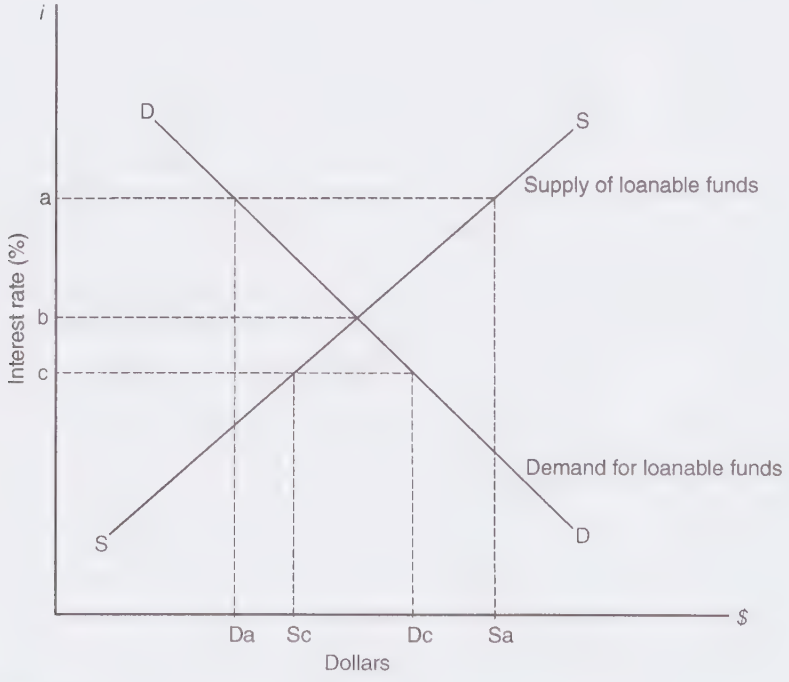


FIGURE 17.5 The determination of the interest rate.

This figure shows how the rate of interest (*i*) is determined by the interaction of supply and demand in the market for *loanable funds*. Like the supply curves for most goods and services, the loanable funds supply curve is upward sloping to the right, indicating that a rise in the interest rate will call forth a greater supply of loanable funds (people, businesses, and government will be interested in lending more money), while a fall in the interest rate will bring about a contraction in the supply of loanable funds. The loanable funds demand curve is downward sloping to the right, indicating that when the interest rate falls the demand for loanable funds will increase (people, businesses, and governments will seek to borrow more money), and conversely, when the interest rate rises the demand for loanable funds will contract.

The borrowers and lenders need not be citizens, businesses, or governments within the same country; they may be from anywhere in the world. For example, the supply of loanable funds in the United States (as well as the demand for them) is not determined solely by what happens there. If the rate of interest is low in the United States, many foreign borrowers may choose to borrow money there rather than in their own countries. This would add to the demand for loanable funds in the United States. On the other hand, when the interest rate is high in the United States, foreign banks and other lenders from outside the country will want to lend money to borrowers in the United States, raising the supply of loanable funds there.

A very important determinant of the supply of loanable funds is the monetary policy conducted by a country's central bank, for example, the Bank of England in the United Kingdom and the Federal Reserve System ("the Fed") in the United States. (The monetary policy of the countries that have joined together in the European Union [EU] is conducted by the European Central Bank.) The U.S. central bank (the Fed), like those of other countries, is an entirely independent and autonomous agency. Its own board of governors has the power to decide, entirely on its own, what the monetary policy of the United States will be. And the Fed's board of governors is not an elected body. The only (indirectly) democratic aspect of this body is the fact that its chair is appointed every four years by the U.S. president (in the year preceding each presidential election year). All the other members of the board of governors are either representatives of regional branches of the Fed (themselves not popularly elected) or individuals (usually bankers) selected by the current members of the board.

A nation's central bank regulates its banks and has the power to affect the total amount of money (supply of money) available for lending and spending in an economy. In particular, the central bank can make it easier or more difficult for banks and other lending institutions to make loans, and this directly affects the supply of loanable funds. Thus, the position of the supply-of-loanable-funds curve on any particular day is strongly influenced by monetary policy.

As explained in Chapter 18, the directors of a nation's central bank usually put a possible threat of inflation at the top of their list of concerns. They may come to believe that the elected leaders of the nation—for example, members of the Congress of the United States—are engaging in so much deficit spending (B) that the sum of the components of aggregate demand ($C + I + B + X$) exceed the economy's current capacity to supply goods and services. In this case the central bankers may try to head off inflation by curtailing the growth of the nation's supply of money. This would prevent the supply-of-loanable-funds curve from moving to the right, or it might even move it to the left. In any case, with the demand-for-loanable-funds curve moving to the right, as it will in a growing economy, the result will be a rise in the interest rate. And as explained earlier, a rise in the interest rate will tend to restrain the borrowing and spending decisions of consumers and businesses, thereby weakening aggregate demand and slowing or reversing the growth of employment. In this manner the Fed may influence the market for loanable funds in a way that intensifies its (already) counteracting effect on expansionary fiscal policy.

The supply of, and demand for, loanable funds determines the interest rate in the following way. Going back to Figure 17.5, imagine that the interest rate is at "c" and the demand for loanable funds is "Dc," while the supply is "Sc." Because "Dc" is greater than "Sc," some borrowers will not get to borrow the amount they want at that interest rate. Noticing this unsatisfied demand, banks and other lending

institutions will reason that they could charge a higher rate. As banks raise their rates, the supply of loanable funds will increase, and the demand will decrease. In a way similar to the process of equilibration described in Chapter 8, the interest rate will tend to rise to "b" percent, the level at which the amount demanded and the amount supplied are equal. At this point the market for loanable funds will clear: there will be neither excess supply nor excess demand.

Many factors other than the interest rate influence the supply of and demand for loanable funds. The expectations of suppliers or demanders with regard to rates of employment, growth, inflation, and various other variables, including political stability and war or peace in the future, will affect their willingness to lend or their eagerness to borrow money. When suppliers' expectations about such matters change, the result will be a shift in the supply-of-loanable-funds curve; when demanders' expectations change, the result will be a shift in the demand-for-loanable-funds curve.

Borrowing and the Exchange Rate

Government borrowing can adversely affect spending even if it does not suppress the demand for consumer goods and investment goods. It can reduce a nation's net exports because it can push up the value of the nation's currency (in the case of the United States, the value of the dollar), thereby discouraging exports and encouraging imports. U.S. net exports, for example, will fall when the value of the dollar rises, because this makes U.S. exports more expensive to buyers in the rest of the world while making imports from the rest of the world cheaper.

How does borrowing by a government in a particular country raise the value of that country's currency? We have seen that large-scale government borrowing, for example by the United States, will tend to raise the U.S. interest rate. The higher interest rate will cause people and businesses in the rest of the world to want to invest their money in the United States in order to get a higher rate of return on their wealth. But in order to invest in U.S. securities, for example, they will have to buy dollars in global currency markets. In doing this they will increase the demand for dollars, and this will shift the demand curve for dollars to the right. Unless there happens to be an increase in the supply of dollars at the same time (people wanting to sell dollars and buy other currencies), the increased demand for dollars will bid up the value of the dollar relative to other currencies. It has often happened that when the interest rate has risen in the United States (perhaps as a result of an increase in government borrowing), foreign investors seek to buy U.S. Treasury bonds since their interest rate is now higher than the return on other securities. This means that they increase their demand for dollars, offering to pay for them with their British pounds, Japanese yen, European euros, or whatever—and this bids up the value of the dollar. This happened during and after 1979, when the interest rate was sharply raised. It happened again in 1994, with the federal funds rate rising from 7.2 percent to 9.2 percent over the course of the year. Sometimes such episodes cause so much flow of money to the United States that a currency crisis (a sharp drop in the value of the local currency) ensues in the countries from which money has flowed out. This happened to Mexico at the end of 1994, causing a large loss in value of the Mexican peso, leading to deep economic crisis.

This brief look at the relationship between borrowing and the exchange rate brings us to another important conclusion. Because government borrowing in a country raises domestic interest rates, it also tends to raise the value of that

country's currency. This, in turn, will tend to increase imports, make exporting more difficult, depress net exports, and reduce aggregate demand (of which net exports is one component). The result is that the effects of the borrowing required to finance deficit spending will tend to counteract the employment-generating effects of it.

The Conflict Between Monetary and Fiscal Policy

The conflict between the two kinds of macroeconomic policy can be summarized as follows. When a government engages in deficit spending (fiscal policy) it must borrow in the market for loanable funds enough money to finance the deficit (assuming that it does not just print the money required). This borrowing competes with the borrowing of both consumers and businesses in the same market, raising the interest rate and "crowding out" at least some of the nongovernmental borrowers. The resulting increase in the interest rate will not only diminish aggregate demand by depressing spending by individuals and businesses on consumer and investment goods, it will also tend to raise the country's foreign exchange rate, encourage imports, and discourage exports—and the resulting decrease in net exports will also tend to reduce aggregate demand. Both of these negative effects on aggregate demand will tend to counteract the employment-generating effect of the original amount of deficit spending.

Moreover, these counteracting effects will be amplified if the central bank (the Federal Reserve System in the United States) decides that the amount of deficit spending is threatening to cause inflation. In this case (and this often happens), the central bank will respond with a monetary policy that limits the supply of loanable funds by restricting the growth of the money supply. The lesson is clear: government borrowing can have a positive influence on the level of aggregate demand and employment, but the effectiveness of such a fiscal policy can be partially counteracted by its other effects. The positive effect of an expansionary fiscal policy on the rate of employment may be nullified if the Fed pursues a monetary policy that has the effect of raising the interest rate even further than it might otherwise be raised by the government's borrowing in the market for loanable funds.

Institutions for Achieving Full Employment

The potential benefits to be gained from the use of macroeconomic policy—especially fiscal policy—were the focus of the previous chapter. In this chapter we have analyzed the difficulties such policies run into in practice. We now weigh both the benefits and the difficulties. There are many controversial issues in this area, but some points are clear.

First, in the absence of government policies to promote high employment, unemployment in a capitalist economy may at times reach unacceptable levels. And the negative human consequences of unemployment will be magnified by the uncertainties associated with the business cycle.

Second, government policies can affect the level of aggregate demand and thus reduce the amount of unemployment. Government policies can also smooth out at least some of the gyrations of the business cycle.

The third point on which there is general agreement is that fiscal and monetary policy *alone* cannot bring about lasting increases in employment (reductions in

unemployment). Without institutional changes full-employment policies continue to face the obstacles outlined in this chapter:

- the high-employment profit squeeze
- the inflationary pressures that occur at high employment levels
- the negative effect of high employment on net exports
- the tensions between monetary and fiscal policy

All of the obstacles reviewed here will be present in a capitalist economy, but some countries have been able to achieve lower unemployment rates than the United States has in recent years. Figure 17.1 shows that both Sweden and Japan had significantly lower rates of unemployment than did the United States, the UK and France, on average, during the period from 1971 to 2012. Comparisons with other countries remind us that unemployment is not, like the weather, something we simply have to accept and endure. Some countries have higher rates of unemployment over considerable periods of time because they have institutions and policies that produce that result. What changes are needed to achieve higher levels of employment?

Institutional Obstacles to Full Employment

To answer the question about institutional change we must understand exactly which institutional structures in U.S.-style capitalist economies are responsible for the persistence of high unemployment. There are two main ones.

The first is the *private and uncoordinated nature of economic decision-making.* Each person's decision about how much to buy and each employer's decision about how much to produce and how many people to employ is made independently. There is no reason to expect that all of these independently made decisions will result in a level of aggregate demand that would require exactly the number of hours of employment equal to the number of hours workers would like to work.

However, the uncoordinated nature of economic decision-making does not by itself mean that we must always have unemployment. It just means that the amount of labor demanded is usually not going to be equal to the amount of labor supplied. Excess supply (unemployment) is one possibility; excess demand is another, even though we have no term to describe it.

The distinct nature of the labor market is significant here. Other markets operate so that excess supply or excess demand tends to be eliminated by competition among the demanders and suppliers (recall the analysis of the beer market in Chapter 8). The ways in which the labor market is different from other markets help to explain why the uncoordinated decisions of economic actors are not likely to result in a clearing of this market. (See the analysis of the labor market in Chapter 12, in particular the discussion in that chapter of why businesses usually choose to offer wages that are above the market-clearing wage.)

The other important institutional structure that works against the achievement of full employment in a capitalist economy has to do with the fact that the capital goods used in production are privately owned (see Chapter 5). Since the ownership of capital goods is highly unequal (see Figure 5.2), the worker in a

capitalist economy generally works for someone else rather than for herself or himself. Also, the goods or services produced are owned by the owners of the capital goods (the owners of firms) rather than by their employees. When workers do not own what they produce, they have little reason to do more than what is necessary for them to keep their jobs. Only the threat of unemployment keeps them on their toes.

If workers owned the tools of their trade and were thus able to work for themselves, each worker would enjoy the full benefits, in the form of higher income, of his or her own hard work. By the same token, each worker would pay the price of less income for slacking off at work. We can summarize these ideas by saying that unemployment exists because of the *class nature of the production process*, that is, unemployment is necessary to motivate and control workers who do not own the products of their labor.

In sum, there are two main reasons for the persistence of unemployment in a capitalist economy. First, the private and uncoordinated nature of economic decision making results in *a mismatch between supply and demand in the labor market.* Second, the class nature of the production process requires unemployment as a way of regulating *the conflict between capital and labor* on terms generally favorable to business.

The mismatch between supply and demand in the labor market seems at first to be just an accident (resulting from the lack of coordination). And there will be some situations in which increases in the rate of employment would be desirable from the standpoint of all involved. Employers would like to see increases in the rate of employment because such increases would mean operating at higher rates of capacity utilization (u) and hence would mean higher rates of profit (all other things being equal). From workers' and consumers' standpoints, increases in the rate of employment would mean more jobs and more consumer goods to enjoy.

As we have seen, however, increases in the rate of employment are not beneficial to all parties in all situations. This is because a decrease in the rate of employment—an increase in unemployment—is sometimes desired by business owners to tip the balance of power between labor and capital in favor of capital. In this case unemployment is not just a mistake: it serves the purpose of restraining wage increases and pressing employees to put forth the maximum amount of work effort for each dollar of wage paid. It thus benefits business at the expense of both workers and consumers.

The Handshake: Ways to Reach Full Employment

The two reasons for unemployment suggest quite different remedies. The remedy for the first is to coordinate economic decisions to eliminate the mismatch. Thus, when a mismatch occurs the government could intervene in some way to eliminate it. The remedy proposed by John Maynard Keynes and adopted by many governments since his time is to let economic decision-making remain private but have the government adjust the overall level of aggregate demand with fiscal policy (deficit spending), thus correcting for mismatches that become too large.

The remedy for the second reason for unemployment—the class nature of production—is more difficult. Business owners are happy to let the government have a role in the economy as long as it means that macroeconomic policy will ensure a steady level of demand for their products. However, businesspeople tend to oppose a sustained reduction of unemployment, particularly if it involves a very low rate of

unemployment. And, of course, they oppose changing the class nature of production because it is what gives them the right to hire and fire employees and own the products that they produce.

In the United States, macroeconomic policy is quite effective in dealing with unemployment as a mismatch between supply and demand in the labor market. But it is not effective—nor is it intended to be effective—in eliminating unemployment, for if successful, this would tip the balance of power in favor of labor. Macroeconomic policy has succeeded in smoothing out the business cycle, but it has not succeeded in eliminating unemployment.

To summarize, unemployment is an economic problem, but its solution is in part political. When nothing is done about it, as the laissez-faire followers of Adam Smith would prefer, unemployment is a permanent and painful waste in a capitalist economy.

Where unemployment has been largely eliminated—whether in Nazi Germany or in modern-day Sweden or in the United States during World War II—it has not been by means of Adam Smith's invisible hand. Rather, full employment has been brought about with a fist (as in Nazi Germany) or with a visible handshake (as in Sweden). The fist refers to the power of business to get workers to work on acceptable terms using the forcible power of the state rather than the threat of unemployment. The handshake refers to deals struck between business and labor to share the increased output associated with higher employment in ways that (a) are fair and (b) also allow for a high enough rate of profit to stimulate investment.

If the invisible hand approach does not solve the problem and the fist is unfair and undemocratic, the handshake is the only viable way to achieve full employment. But what would be the main elements of a handshake approach?

First, in order for high employment to be achieved, the high-employment profit squeeze and its associated inflationary consequences would have to be eliminated. But these two threats could be eliminated only if unemployment were no longer used to regulate the class conflict between workers and their employers.

One way to regulate class conflict is for employers and employees to find a way to determine both pay and work effort that does not rely on the worker's fear of job loss. In Sweden, for example, wages and work rules have been negotiated between the leaders of the major unions and the representatives of employers. During times of high employment the unions had the power to gain much higher wages, but they generally did not use it. Their reasoning was that if they did, they might price Swedish goods out of the world market, resulting in eventual income losses for themselves. Swedish unions were willing to exercise restraint in good times because they were powerful enough to insist that in bad times employers not use their authority to lay off workers and reduce wages. As a result, Sweden achieved very low levels of unemployment.

Another approach to regulating class conflict would be to change the class nature of the production process. This would mean going beyond the handshake approach, but it could be done keeping economic decision-making decentralized and preserving market competition as a way to allocate resources. The difference would be that production would be organized in worker-owned and democratically run firms of the type described in the last section of Chapter 13. Since workers would be the owners of the products they and their coworkers produced, they would have no reason to work less hard or demand more pay when unemployment was reduced. If they shirked or demanded too much, they would be using their bargaining

power against themselves; if they squeezed profits, they would be reducing their own income. If most workers in an economy worked for themselves in their own firms, the high-employment profit squeeze would be greatly attenuated.

Even in a democratic economy, however, macroeconomic policies would have to be used wisely. If deficit spending expanded aggregate demand to the point at which all workers were employed and all productive capacity was in use, inflation would most likely be the result. In this situation the demand for many products would be greater than the available supply of them, and worker-owned firms, like their capitalist counterparts, might well raise the prices of their products. This would most likely lead to a more generalized inflation. So workplace democracy is not a cure-all.

The second important element in the handshake approach to full employment is that an equitable solution must be found for the conflict of interests between employers and workers in the labor market. If workers are going to forgo using the bargaining power that high employment gives them, employers must also give up the bargaining power that high unemployment gives them. What labor agrees to do in the handshake must have its counterpart in a corresponding commitment on the part of business—otherwise it will be no deal. Employers must commit themselves to providing employment security both at the level of the firm (promising to refrain from laying off their employees) and as a macroeconomic strategy (supporting government policies to achieve high employment).

The third element in the handshake approach would be the expansion of built-in stabilizers, for example, unemployment insurance. This would not only strengthen income-support programs; such programs also have the effect of automatically dampening business cycle fluctuations because they pump more money into the hands of consumers when income from employment is shrinking, and they work the other way as well. More deliberate countercyclical (macroeconomic) policies, legislated changes in taxes or spending programs, for example, could also be used to maintain a stable level of aggregate demand. In the context of the whole bundle of "handshake" compromises, the effects of such policies would be beneficial to almost everyone, so both business and labor should be able to agree on them.

The fourth element of the handshake approach results from the fact that a handshake across the capitalist-labor bargaining table will not be enough: handshakes across national boundaries will also be necessary. Coordination among nations pursuing high employment policies must be secured through international agreements among the major countries.

To be effective, international macroeconomic policy agreements would have to discourage strategies whereby one country tries to export its unemployment problem to some other country by engaging in export promotion or import restriction to secure higher employment in its own economy. These agreements would have to enable countries to coordinate their policies of expanding aggregate demand so that each nation's expansionary policies would contribute to the success of similar policies in other nations.

The fifth and final element of the handshake involves policies that facilitate economic adjustment without imposing the costs of adjustment on particular workers or unemployed people. Some industries or sectors will be expanding while others are contracting. In responding to changing economic conditions, the economy must be flexible, but the costs of its flexibility must be distributed fairly.

To promote flexibility, a basic principle of a full-employment economy must be that everyone seeking work has a right to find employment but not to stay permanently

in the job he or she currently holds. The economy must adapt to new technologies, to changes in consumer tastes, and to the new forms of international and other competition outlined in Chapter 15. This means that many workers will have to move from job to job and from industry to industry. But the government can ensure that any worker who loses a job will have income support during the transition from one job to another, will have access to adequate training for a new job, and will receive support, as necessary, for moving from one area of the country to another. The cost of such programs—in Sweden they are called "active labor market policies"—should be financed out of general tax revenues. Virtually all of a nation's individuals and businesses benefit from the continuous modernization of the economy, while the adverse effects of its modernization are felt only by particular workers. Thus, it is only fair that the beneficiaries of a flexible economy be required to support those who must pay the price of its flexibility in job loss, income loss, possible loss of self-respect, and severance of ties of friendship in the workplace and the neighborhood. (Of course, care will have to be taken to make sure that such programs of support are limited and monitored in order to prevent their abuse.)

Handshakes do not mean the end of conflict, of course, either between nations or between classes. But they can be a way of regulating conflicts in a less wasteful and brutal manner than either the fist or the invisible hand.

Conclusion

With the exception of the fourth element of the handshake, international cooperation in the making of macroeconomic policies, the five institutional and political prerequisites for full employment are close to being achieved in some countries. Norway, Austria, and Sweden are examples. As a result, these countries have had very low unemployment rates. Unfortunately, the same cannot be said about many other countries.

The United States achieved relatively high levels of employment (low unemployment) during the decade beginning in 1994 not by the handshake but by a variant of the invisible hand. Figures 17.2 and 17.4 showed that during the business cycle expansion of the 1990s the high-employment wage push and the high-employment profit squeeze did not make their appearance until well into the expansion, when the unemployment rate had fallen below 5 percent. Thus, the economic expansion could proceed without interruption from these usual impediments to high employment.

Three aspects of U.S. economic institutions have made it more difficult for workers to push up wages when there is a low rate of unemployment. The first is the decline in union membership (see Chapter 7). This has weakened labor's ability to use the bargaining power that usually comes with low unemployment.

The second is the continued segmentation of labor markets and growing inequality of earnings among workers. Primary labor market workers—those in relatively well-paying jobs in manufacturing, for example—who are laid off or fired cannot realistically expect to find jobs with similar pay and benefits. These workers will most likely experience a period of unemployment, followed by reemployment in a position paying considerably less, possibly in the secondary labor market.

The third is the overall decline in the social safety net—such as the programs that give people income and other forms of support when they have lost a job. Welfare programs are now far less generous and support fewer families than in the past,

following the dismantling of the Aid to Families with Dependent Children program in the mid-1990s. Disability payments were scheduled to be cut back in 2016. Unemployment benefits are on average less generous, and in some states last less long, than in the past.

An important result of labor market segmentation, combined with the weakening of the social safety net, is that the cost of losing a primary labor market job is now very high—*even when the unemployment rate is low.* Even at the peak of the business cycle, the threat of firing continues to be a powerful weapon in the hands of an employer in the core business sector. The loss of a machinist's job will probably be a financial disaster—even if the machinist knows that he or she can find work as a cashier at the supermarket.

Few governments have made long-lasting commitments to the goal of full employment. As we have seen, unemployment is not simply a mistake that can be corrected by government action: it is also a weapon in the conflict between workers and employers. Business owners do not want low unemployment to be used as a weapon against them, and they occasionally urge the government to adopt policies that will increase unemployment.

Chapter 18 discusses the causes and effects of the Great Recession of 2007–2009, as well as of crises generally. In Chapter 19 we further explore the issue of conflict between capital owners and workers over the role of the government in the economy.

Suggested Readings

Roger Alcaly, *The New Economy* (New York: Farrar, Straus & Giroux, 2003).

Duncan Foley, *The Unholy Trinity: Labor, Capital and Land in the New Economy* (London: Routledge, 2003).

Doug Henwood, *After the New Economy* (New York: New Press, 2003).

Paul Krugman, *The Great Unraveling: Losing Our Way in the New Century* (New York: Norton, 2003).

Robert Pollin, *Contours of Descent: U.S. Economic Fractures and the Landscape of Global Austerity* (New York: Verso, 2003).

Joseph E. Stiglitz, *The Roaring Nineties: A New History of the World's Most Prosperous Decade* (New York: W. W. Norton, 2003).

18 Financial and Economic Crisis

A generation was scarred by the Great Depression of the 1930s. The stock market crashed in 1929. The economy collapsed, with average income falling steeply every year for four years, and remaining low until World War II. By 1933 the banking system had stopped working. One out of every four members of the labor force was unemployed at the height of the calamity. People left their homes and went on the road; some lived in tents; many had little to eat and relied on the generosity of others.

This led to enormous changes: New laws regulated the banking system, created Social Security and unemployment insurance, and established the Securities and Exchange Commission to regulate corporations. In the mid-1930s the Wagner Act protected labor organizing across the country. Yet only with the advent of World War II and new demand from Europe for war-related supplies did the U.S. economy truly emerge from depression.

But was the Great Depression a one-time event? Hardly. Similar economic collapses happened in the 1800s. The National Bureau of Economic Research (NBER), a nongovernmental organization which among other things dates business cycles) says that the economy was in recession over two-thirds of time from 1865 to 1879, including the entire five years 1874–1878 following the Panic of 1873. Then there was the so-called "Great Depression" of 1894–1898 following the Panic of 1893, which brought a stock market collapse, bank failures, an unemployment rate estimated at 20 percent, and widespread poverty.

After the 1930s, though, it was more than four decades before a downturn of similar severity occurred. The 1982–1983 recession was quite deep, though the economy was never on the brink of a major, wrenching collapse. But the Great Recession of 2007–2009 was the closest thing to the Great Depression since World War II. There was a terrifying moment in mid-September 2008 when many truly believed the financial sector could blow apart (possibly within hours or days) and take the whole economy down with it. Government intervention made the collapse less total than it could have been, but the consequences were still devastating: 9 million people became newly unemployed, and about the same number of persons fell below the poverty line. Several huge financial institutions became insolvent (had negative net

worth) and were either bought out or closed down. Home building and home sales fell to their lowest level in many decades.

Such disasters lay waste to the work done over years or decades to build enterprises, create employment, cultivate skills, and generate steady demand. We therefore urgently need to understand why they happen and how to prevent them.

This chapter has two purposes. One is to trace the unfolding of the Great Recession, with its causes and consequences, to give a detailed example of how a major crisis can occur in a capitalist economy. The other is to explain why capitalist economies are in general prone to crisis and recession, particularly if they have large, lightly regulated financial sectors. The blow-by-blow account of the crisis will explain why the economic actors made the decisions they did, based on the economic incentives they faced. A later section will take account of the previous history of crises and explain why in capitalist economies crises happen so often. The chapter concludes with reasons why better government regulation of financial sector activities would help to reduce the chances of future such major crises—as well as reasons why the financial sector has fiercely resisted re-regulation, with some success.

Several main ideas are central to this chapter. The first is that capitalism is prone to instability and crisis, which tends more often to originate in the financial sector, the less that sector is regulated. The second is that although regulation and sensible fiscal and monetary policy can reduce the risk of crisis, there are also persistent forces seeking to undo such regulation. When these forces eventually prevail, they leave the economy open to major upheaval. The third is that increasing concentration of wealth and power, especially in the financial sector, as has happened over the last several decades, makes matters worse. These ideas are expressed in the following eight points:

1. The house price bubble represented a market failure. Decision-makers made the decisions one would expect, given the incentive structures facing them; but these incentives did not lead them to make decisions that benefited society. Instead, the incentive structure caused all to play their part in creating the bubble, whose bursting caused massive damage.

2. Among other things, these incentive structures turned deception into a profitable strategy for at least three groups: mortgage lenders, issuers of securities called *derivatives*, and credit rating agencies that rated the riskiness of these securities. As has long been recognized, in a capitalist economy deception is often profitable if not regulated. Government regulation is therefore crucial to prevent it.

3. Countries where financial firms' activities are not adequately regulated are vulnerable to reckless behavior by such firms, leading to economic crises. Prolonged depressions in the nineteenth century and the 1930s finally led to regulation of the financial sector in the United States, which played a great role in reducing the length and severity of downturns for roughly four decades during and after World War II. Financial deregulation beginning in the 1980s opened a Pandora's box of reckless financial dealings, increasing the tendency toward severe crises.

4. Crises can also originate in the nonfinancial sector, for example due to overinvestment in a major industry such as home construction, creating more productive capacity (or output) than is needed to meet existing demand. Such overinvestment easily happens in a capitalist economy, since investment

spending is uncoordinated among firms. A related contributing factor to stagnation or crisis can be increasing income inequality, slowing growth in consumer demand, and causing a recession due to underconsumption—another kind of mismatch between supply and demand.

5. Economist Hyman Minsky described why over the long term, deep crises would tend to recur: Deep crisis provokes new regulations, which soon lessen the depth and length of crises. But this in turn leads to complacency, deregulation, increasingly risky firm behavior, and increasing firm indebtedness. Eventually a business failure triggers default on debt, with a domino effect that precipitates another crisis. This is Minsky's *financial fragility hypothesis*.

6. As mergers and acquisitions create a very few giant firms, especially in the financial sector, these firms acquire growing political power, using it to accelerate deregulation. Growth in firm size and power also leads to *moral hazard*: "too big to fail" firms act more recklessly, embracing high risks for high returns on the assumption that if their strategies fail, they will be bailed out by the government.

7. The damage done by a crisis depends partly on the strength of the social safety net: unemployment insurance, welfare benefits, old age pensions, good personal bankruptcy laws, availability of health care at reasonable cost, and so on. The social safety net in the United States has been greatly weakened in recent decades.

8. Although strong government regulation could minimize the chance of a new crisis rooted in the financial sector, powerful financial firms have fought such re-regulation in the form of the Dodd-Frank Act and the rules made to implement it, and pressed for its total repeal. Some economists have urged breaking up "too big to fail" banks by capping the size of their assets relative to GDP. Others have proposed establishing public banks (such as in the post office) for some purposes, or nationalizing some existing banks.

The Great Recession and the Subprime Crisis

The Great Recession is officially dated from December 2007 to June 2009, according to the NBER.[1] But it was not until September 15, 2008—the day the giant financial firm Lehman Brothers went bankrupt—that most people felt the shock that financial markets were heading over a cliff. Many financial institutions were close to insolvency and stock prices dropped sharply, calling the public's attention to the gravity of the situation.

Figure 18.1 shows the sudden drop in GDP. The crisis ended up raising the unemployment rate to 10 percent by the standard U-3 official measure, and to 17 percent by the more inclusive U-6 measure that some economists prefer (see Chapter 16). In every month from January 2009 through February 2012, between 13 and 16 million people were unemployed. The average growth rate of GDP remained lower than in previous business cycles for many years even after the recession officially ended. (See Figure 16.1.)

[1]The Bureau of Economic Analysis reports GDP data only quarterly, but the NBER uses several monthly data series not widely available to establish the month in which a recession started and ended.

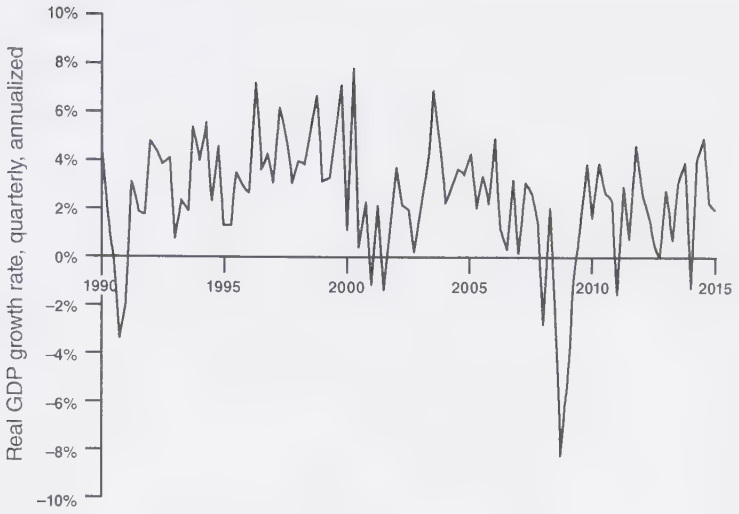


FIGURE 18.1 Real quarterly GDP, growth rate.

A recession is defined as two quarters of declining GDP. However, the National Bureau of Economic Research, which dates recessions, does not always follow this definition, but considers a variety of indicators in making its decisions.

Source: U.S. Bureau of Economic Analysis, National Income and Product Accounts Table 1.1.1; available at http://www.bea.gov/.

While Lehman Brothers went bankrupt, other large financial firms and banks (Bear Stearns, Merrill Lynch, Washington Mutual, Wachovia, and more) became insolvent and were acquired by still other giant firms. Many other major financial firms and banks came perilously close to bankruptcy and were bailed out by the Treasury Department and the Federal Reserve. Several kinds of securities markets froze and then were rescued by the Fed. In all, the combined expenditures of the U.S. Treasury Department and the Fed to rescue troubled banks as well as other financial and nonfinancial institutions reached $14 *trillion*. Yet the public suffered: From 2007 to 2010, nine million people were added to the ranks of the poor. And because most ordinary families' biggest investment is in their homes, many families lost their savings: between 2007 and 2011, one-fourth of families in the United States lost at least 75 percent of their wealth, and more than half of all families lost at least 25 percent of their wealth. For some, they could no longer retire as planned; for others, college had to be postponed; still others lost ownership of their homes.[2]

The crisis developed in three stages. First, home prices began rising rapidly, as many households were persuaded to take out mortgage loans that in the medium run they likely could not afford. Next, home prices leveled off and then began falling, which caused securities markets to unravel and brought the financial institutions holding these securities to the brink of insolvency and beyond. Third, there was the bailout and recovery stage just described. Collateral damage hit ordinary people throughout these phases, as foreclosure, eviction, unemployment, poverty, and other

[2]Fabian Pfeffer, Sheldon Danziger, and Robert Schoeni, "Wealth Disparities before and after the Great Recession," University of Michigan, April 1, 2013, available at https://www.russellsage.org/sites/all/files/PfefferDanzigerSchoeni_InterimReport_2013.pdf.

social ills took their toll in an economy with a weakened social safety net. A recent change in the law also made it harder for people to file for bankruptcy.

In some ways the crisis spanned a decade or more. The rot had begun to set in at least as early as 2005, when home prices—after rising rapidly since 2001—began to level off; but it was not until 2007–2008 that the problems reached the level of persistent front-page news. Yet early on, by 2002–2003 and in some places even earlier, many state regulators had already identified a growing problem with unfair, possibly illegal, lending practices, and tried to regulate, investigate, and prosecute any such crimes. But states' enforcement efforts were largely prevented by the federal government. In light of this, it could be argued that, as early as 2003, the stage was already set for a dangerous house price bubble and eventual catastrophe.

The pain lasted long after the recession, by official measures, had ended. The number unemployed for fifteen weeks or more as a percent of the civilian noninstitutional labor force rose from 1.5 percent in March 2007 to a peak of 5.9 percent in March 2010, and was still elevated at 3.5 percent in March 2014, higher than at any point during 1994–2008. In 2010, 46 million people were in poverty, 9 million more than in 2007; by 2014 the number in poverty was still 46 million. In short, even though the stock market had recovered by 2013, the devastating consequences of the crisis had by no means ended. And by 2014–2015 there were still historically low levels of new home construction, persistent long-term unemployment, and an interest rate very close to zero, making it hard for the Federal Reserve to stimulate investment spending.

Understanding the Great Recession

Because the recent crisis was triggered by the bursting of a house price bubble, to truly understand the crisis requires understanding how and why people buy homes, and the different ways they can borrow the money to buy a home. We begin there, and then explore what led up to the crisis and how it unfolded. We examine the incentive structure facing each key group of economic actors, and why, therefore, they did what they did. Later in the chapter we explore more deeply why these events occurred and draw conclusions about why a capitalist economy, especially a lightly regulated capitalist economy with a large financial sector, is prone to crises.

Why to Buy a Home

Many people who are not already homeowners probably would like to buy a home if they could afford it. As a homeowner, I can have great freedom to decorate my home as I please, to use it more or less as I please, and to remodel it as I please (consistent with the building code, and after getting a construction permit)—in short, the freedom not to have to answer to a landlord. Another advantage is that U.S. tax law allows homebuyers to deduct from their taxable income the interest they pay on a home mortgage loan. This can reduce the income tax bite considerably, particularly near the beginning of a thirty-year mortgage when most of the monthly payment consists of interest. Still another advantage is that the monthly mortgage payment is typically fixed in money terms so that gradually, over the period of such a thirty-year mortgage, the mortgage payment can shrink greatly as a percent of the buyer's income.

Historically, buying a home has been seen as a sound investment, because it is a form in which to hold wealth likely to yield a reasonable annual return. This is partly

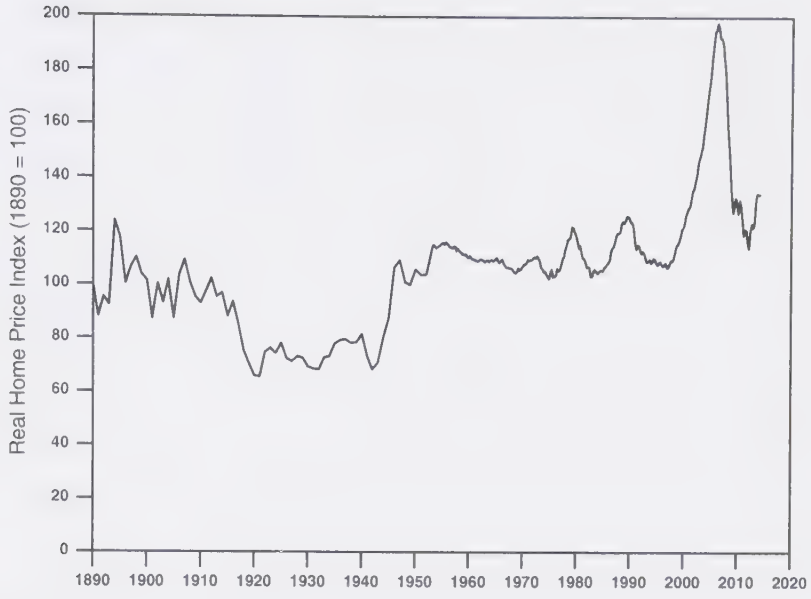


FIGURE 18.2 Real Home Price Index, national, Case-Shiller, 1890–2014.

A nominal home price index, which in recent decades is one compiled by the Federal Housing Finance Agency, is deflated by the CPI-U, the urban Consumer Price Index. The base year is chosen as 1890.

Source: The data are for part of Figure 2.1 in Robert Shiller, *Irrational Exuberance*, 2nd edition (Broadway Books, 2005). Available updated by Shiller at http://www.econ.yale.edu/~shiller/data.htm.

because home prices have risen fairly steadily over decades about as fast as the overall price level. This means that the real price of homes—their value in terms of the alternative basket of goods that could have been bought with the same money—remained at least constant. Figure 18.2 shows a measure called the Real Home Price Index, which is a measure of the nominal price of homes deflated by the Consumer Price Index (CPI), with the index set to 100 in the year 1890. The index rises if the price of homes is on average rising faster than the price of the goods in the basket used to compute the CPI. The fact that the index was at about the same level in 1997 as in 1947 means that in those 50 years, on average home prices rose about as fast as the prices of everything else. In some periods and some places, of course, they rose faster, and in these times and places, buying a home was an attractive option relative to other ways to hold wealth.

Home Prices Start to Rise

In the late 1990s, home prices started to rise in earnest. Although the upward trend stalled for a few months just before the 2001 recession, it quickly resumed. An extra push came in early 2001 when the Fed lowered the federal funds rate from over 6 percent to under 2 percent to help speed recovery (see Figure 18.3). The federal funds rate is a key interest rate, namely, the rate at which banks lend money to each other. This brought down interest rates on conventional mortgages from over 8 percent in 2000 to below 6 percent by January 2003. And this made houses cheaper to buy and, along with aggressive marketing of new, apparently more attractive types of loans, helped to accelerate sales of new homes, as shown in Figure 18.4.

Figure 18.4 shows that annual sales of new single-family homes peaked in 2005 at an astonishing 15.3 million new homes sold. But then came a dizzying fall to the

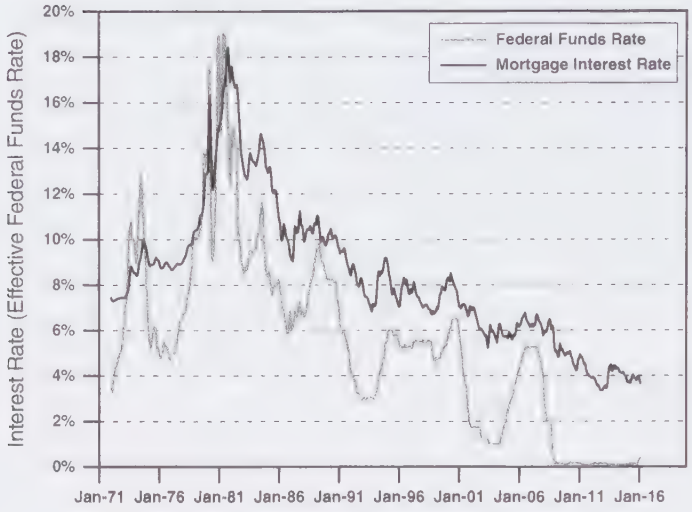


FIGURE 18.3 Interest rates, 1971–2016.

The bottom line is the federal funds rate, the rate at which banks lend to one another; the top line is the interest rate on conventional thirty-year mortgages.

Source: Federal Reserve Economic Data. FEDFUNDS: Effective federal funds rate, monthly (first day of the month), not seasonally adjusted; MORTG: mortgage interest rate, monthly (first day of the month). Available at https://research.stlouisfed.org/fred2/series/FEDFUNDS and https://research.stlouisfed.org/fred2/series/MORTG

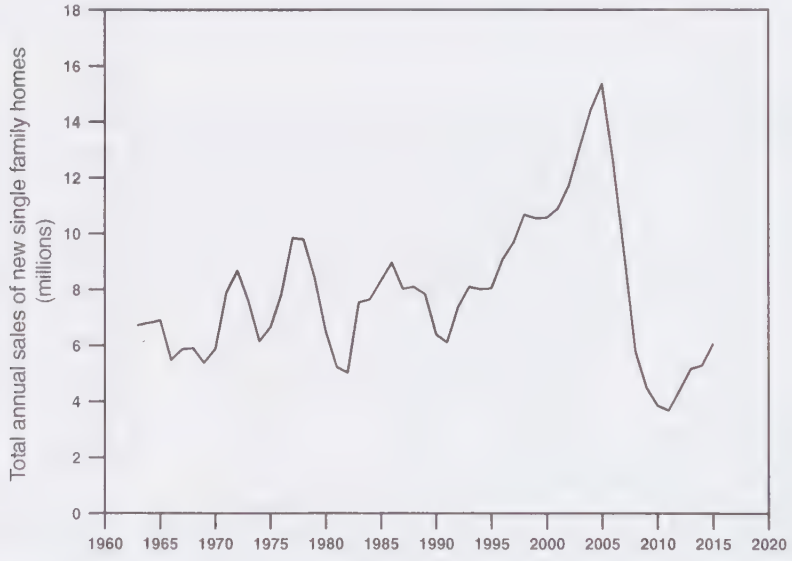


FIGURE 18.4 Annual sales of new single-family homes.

Home sales are pro-cyclical: they grow in an economic boom, and they often decline shortly before a recession and during the recession. The 2007–2009 Great Recession was no different, except that the slump in home buying was worse than at any time since at least 1963, and by early 2016 had already lasted for at least six years after the recession officially ended in 2009. Sales of new single-family homes did start rising in 2012, but by 2015 had barely even reached their previous (1991) low. Of course, existing homes were also being sold, but it is sales of new homes that stimulate the economy, by stimulating employment in the construction sector and economic activity in sectors that supply construction industry inputs.

Source: Federal Reserve Economic Data, annual data aggregated from HSN1F: New One Family Houses Sold: United States, Thousands, Monthly, Seasonally Adjusted Annual Rate. Available at https://fred.stlouisfed.org/series/HSN1F.

lowest level in at least forty-seven years. Despite the formal end of the recession in mid-2009, and a steady rise in GDP thereafter, new home sales long remained at historically low levels: 6 million or fewer new homes sold every year from 2009 through 2015, and in two of these years (2010–2011), less than 4 million sold. This resulted in the largest drop in construction employment recorded in seventy-five years, a loss of 2.2 million construction jobs from the 2006 peak of 7.7 million to 5.5 million in 2010 (see Figure 18.5). Even in 2015, employment in construction was still more than 1 million below its peak, and the lagging construction sector was a drag on GDP growth.

Rising new home purchases led in the late 1990s to an increase in the home ownership rate: the percent of all U.S. households who were buying a home, or had already bought and paid for it. The home ownership rate rose from 64 to 69 percent during the 1995–2005 decade, and then reversed its steps, falling back down to less than 64 percent by 2015, as many home buyers defaulted on their mortgages, were foreclosed on, and lost their homes, while new purchases fell to low levels.

What was unprecedented was the enormous size of the house price bubble. Figure 18.2 shows that the real home price index rose from 108 in 1997 at a breakneck pace to a peak of 198 in the first quarter of 2006. But then it briefly hovered, and in early 2007 the bubble burst and the index dropped rapidly back down to 127. The size of the bubble was an anomaly: no bubble of such size had happened in the entire 125-year history shown. But why was there a housing price bubble in this period at all? Why was it so large? What made the bubble burst? And how and why did it affect the rest of the economy so harshly? To understand all this, it helps to understand how to buy a home.

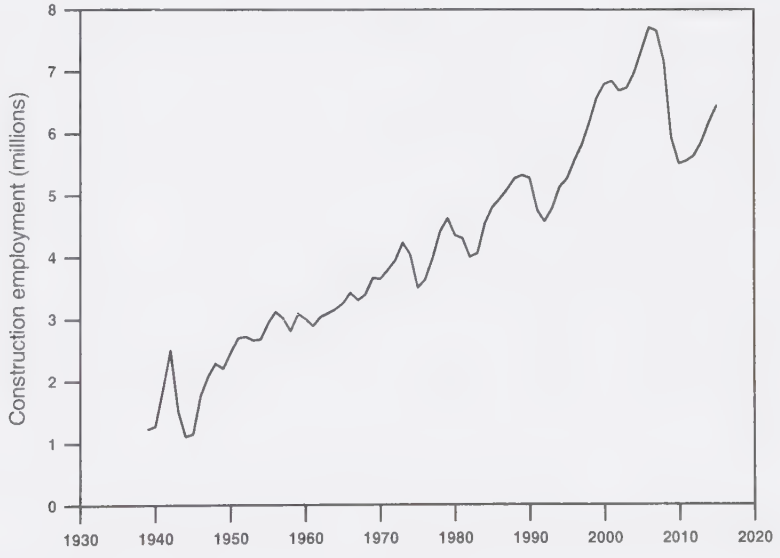


FIGURE 18.5 Total employment in the construction sector.

There was a surge in sales of new homes in the late 1990s and in the early 2000s. This was partly due to easing the criteria for qualifying for home loans, and partly to new types of mortgages, such as Adjustable Rate Mortgages (ARMs) that offered low monthly payments for a few years before the payments jumped up to higher levels. All these increased construction of new homes and therefore employment in the construction industry. Data shown are for July 1 in each year.

Source: FRED, series USCONS: "All Employees: Construction, Thousands of Persons, Monthly, Seasonally Adjusted," available at https://fred.stlouisfed.org/series/USCONS.

How to Buy a Home

In the past, a standard rule of thumb has been that a family should expect to have saved enough money to make a 20 percent down payment on a home they wish to buy. They can then seek to borrow the remaining 80 percent from a bank or other mortgage lender. A 20 percent down payment protects the lender against a fall in the house price of up to 20 percent—and this encourages the lender to approve the loan in the first place.

But what if the buyer were to put down only $5,000 of his own money on a home priced at $100,000? The buyer would borrow the remaining $95,000—including a first mortgage of $80,000, plus an additional borrowing (a *second mortgage*) of $15,000 to be able to pay the other three-fourths of the $20,000 down payment. If it should happen that the price of the house falls to below $95,000, the borrower will then own none of the value of the home (will have no *equity* in the home). If that happens, one possibility is that the borrower may decide to stop making loan payments. Although this happens somewhat rarely, it did happen on a fairly large scale in 2007 and later.

Lenders have often preferred to try to protect themselves against this possibility by requiring borrowers to buy mortgage insurance if they do not pay the whole down payment out of their own funds. The mortgage insurance is required in part to make it possible for the lender to sell the loan to the large government-sponsored enterprises (GSEs) and home mortgage brokers known as Fannie Mae (Federal National Mortgage Association) and Freddie Mac (Federal Home Loan Mortgage Corporation). For the lender, an advantage of reselling a mortgage is that the lender can then use the money to make further home mortgage loans.

When the buyer borrows some of the down payment, this *second mortgage* results in a larger monthly payment than if the buyer had put down a full 20 percent down payment in cash or check. Of course, in principle the buyer needs to choose a home whose price is low enough that he or she can afford that larger monthly payment. One might think that the lender would also want the borrower not to buy a house whose monthly payments the borrower could not really afford; and until the early 2000s, this was on the whole what did happen. But over a very few years, this changed. According to state regulators, many lenders deliberately lent to borrowers who they knew quite well would have trouble making the monthly payments over the full life of the loan.

Securitization of Mortgages

Why did this happen? It was overwhelmingly because over this period a growing share of all mortgages was being *securitized*: parts of the mortgage payments (the stream of payments by borrowers) were bundled with payments from other mortgages into *mortgage-backed securities* (MBSs) that were then sold on securities markets to pension funds and other organizations and individuals—any investor who would buy them. By selling the mortgages to MBS issuers, the mortgage lenders insulated themselves from harm if the borrowers failed to make the monthly mortgage payments on time. Once the lender sold the mortgage off to someone else, often within a month or so, a late payment became entirely the problem of the new owner of the mortgage or the MBS.

CREDIT RATING AGENCIES. Unfortunately, potential investors in MBSs could not easily find out how risky they really were. The issuers of these securities did pay to have them rated by well-known credit rating agencies. But they were often given credit ratings by these agencies without a look at the data on the underlying

mortgages that would reveal their risk level. If ratings agencies assigned them AAA ratings without taking account of (or even knowing) the riskiness of the underlying mortgages, then there was no way for prospective buyers of MBSs to know this. It was therefore easier to sell the MBSs than it would have been to sell the individual mortgage loans bundled in them. Prospective buyers, having little information about them, might have been wiser not to buy them at all. But the ratings were assigned by prominent rating agencies: Moody's, Standard & Poor's, and Fitch, and promoted by other respected financial firms. For years, nobody thought there was a problem.

How did mortgage lending change over time, and why did home buyers sometimes accept such risky mortgages? To understand this, it is useful to understand a few more technical terms and procedures in mortgage lending.

Prime and Subprime Mortgages

A standard mortgage loan for a borrower who has a good job, a good income, enough assets, and a good credit history is called a *prime* mortgage. The interest rate is the best rate currently available (in 2016 about 3 to 4 percent, but during 2001–2006 varying from 5 to 8 percent, as shown in Figure 18.3). In many prime mortgages, the borrower agrees to pay equal monthly payments for the (often thirty-year) life of the mortgage. A minority of prime mortgages were Adjustable Rate Mortgages (ARMs) that started out at a lower interest rate and could rise to a higher interest rate after a few years, depending on what happened to market interest rates generally. The advantages and disadvantages of ARMs will be described below.

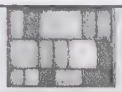
If a homebuyer's creditworthiness is too low for a prime mortgage, she may be eligible for a *subprime* loan, and these come in various types. In one type of subprime mortgage, the borrower has to pay a significantly higher interest rate, and therefore higher monthly payments, but the mortgage is otherwise ordinary: a fixed-interest-rate mortgage with equal monthly payments for (say) thirty years. But several types of *subprime* mortgages have monthly payments that may change, or that do change, during the life of the mortgage. Many of these are various types of ARMs. Some lenders offered a low "teaser" interest rate in the first several years, which after a specified number of years could, or would, increase suddenly to a higher interest rate, so that the monthly payments would suddenly go up (or "reset").

Some mortgages were interest-only for some initial period, postponing the principal payments and giving the borrower the hope that before principal payments were scheduled to begin, the value of the home would rise and the buyer could refinance the loan to lower the monthly payments. But what exactly is *refinancing*?

Refinancing

Why did so many homebuyers agree to loans whose monthly payments would suddenly increase after several years, such as ARMs and interest-only mortgages? On the face of it, if a home buyer can really only afford the lower monthly mortgage payment for the first two to three years, and can ill afford it after it resets to a much higher monthly payment, why would the buyer agree to the mortgage at all?

The answer lies in the way *refinancing* works in a world of rising home prices. Refinancing means simply paying back an existing loan in full, by taking out a new loan for a similar amount. A homebuyer sometimes finds it useful to do this, mainly for two reasons. One is that the new loan may offer better terms than the existing



Refinancing a Home

Refinancing a home is a way, with luck, to exchange bad mortgage terms for better terms, or perhaps to get some money out of a house to fund something such as a college education. Suppose Dalia, with a so-so credit score, income, and assets, buys a modest home for $100,000. The left hand side of the table below shows how she finances the purchase, how much equity she has at first, and how much she owes. Suppose the house price rises to $120,000 after two years. The right hand side shows how she refinances the loan after two years, and how her equity and the amount she owes will change.

At the time of purchase she pays $2,000 of the down payment out of her own funds, and this is initially her *equity* in the home (the part of the home's value that she actually owns). Suppose that within two years the house price rises by $20,000, which is added to that equity. She makes the monthly payments (mostly interest, little principal), so that perhaps she pays off only $3,000 of principal during the two years. (Interest is not shown in the table.) By 2004 she has $25,000 in equity ($2,000 that the down payment paid for + $3,000 of paid-off principal + $20,000 house price increase), and owes $95,000 on the mortgage. She takes out a new loan for $95,000 on better terms, pays off the remaining $95,000 in loans, and starts making monthly payments on the new thirty-year loan. But now, because she owns $25,000 of equity in the home, which is 20 percent of its value, this equity improves her creditworthiness, which gets her a lower interest rate loan. Thus she benefits from having refinanced.

But what if, instead, home prices *fall* after she takes out the original mortgage? If the price falls by $5,000 (to $95,000) after she takes out the original mortgage, she could lose both the $2,000 down payment and the $3,000 she paid in two years on the principal. Then her equity will be zero—negative if the price falls further. She cannot refinance, because the lender has no buffer against a further fall in the price of the home.

Purchase (2002)		**Refinance (2004)**
$100,000	House price	$120,000
$20,000	Down payment	Equity = 20% = $25,000 = $20,000 price increase + $2,000 down payment (2002) + $3,000 principal paid off (2002–2004)
$80,000	First mortgage	$95,000
+		
$18,000	Second mortgage	Not needed
+		
$2,000	Cash down payment	Not needed
$100,000		$120,000

loan—a lower interest rate, or a constant monthly payment instead of one that jumps upward after several years. The box, "Refinancing a Home" gives an example of successful refinancing.

When home prices have been rising for several years, buyers tend to assume the best-case scenario: that prices will keep rising, and that therefore they will later be able to refinance on more favorable terms any mortgage with undesirable terms they agree to now. With such expectations, optimistic homebuyers, or even ordinary homebuyers, might agree to ARMs, even ones with fairly awful terms, such as large increases in monthly payments after two to three years. Mortgage lenders rarely cautioned them that home prices were unlikely to keep rising indefinitely; their salespeople were paid to make as many loans as they could, not to discourage people from taking out mortgage loans they could not afford. And even though in many states there were laws on the books preventing lenders from making loans that borrowers could not afford, still, in practice, states were prevented by the federal government from enforcing these laws, as we will see.

But why did *lenders* agree to such risky mortgages? One might expect mortgage lenders to be steely-eyed businesspeople, making loans based strictly on the homebuyer's ability to pay, even when the homebuyer herself is willing to agree to a burdensome loan. Part of the answer has already been given: mortgage lenders planned to sell off the mortgages quickly to someone else to be packaged into MBSs and resold. Let us clarify the historical evolution of the use of mortgage-backed securities and then identify two further causes that reinforced these trends. This will clarify what benefits the various players expected to get, and what they did to ensure that they could get those benefits.

Deregulation

From 1934 until 1999, a Depression-era law known as the Glass-Steagall Act separated banks into two types. Commercial banks took deposits from the public and issued loans to people and businesses. Investment banks sold the service of arranging the sale of securities for corporations, and invested in securities. Glass-Steagall said that commercial banks could not do what investment banks did, and investment banks could not do what commercial banks did. The two were distinct.

There were good reasons for the 1934 law. An investigation by the Pecora Commission of the U.S. Senate had uncovered some embarrassing practices. Some banks had advised their customers to buy foreign securities, which the evidence suggests that the banks knew were worthless, or likely soon would be. In some cases a bank that did this had already bought the securities and wanted to unload them. In other words, banks that took public deposits appeared at times to be giving their customers bad investment advice to enhance their own profits. Revelation of such unsavory practices helped to inspire Congress to separate commercial banking from investment banking.

This separation was gradually eroded in the 1980s and 1990s. In the 1980s, Fannie Mae and Freddie Mac (the GSEs) were authorized to begin issuing mortgage-backed securities. It created these out of the mortgages it owned, by packaging together streams of payments from a group of mortgages for sale to investors. Selling MBSs was seen as a way to get more funds for financing home loans, by attracting funds from sources outside of banks and mortgage lenders. Lenders who sold mortgages to the GSEs, could then use the proceeds of the sale to lend to more homebuyers, allowing more people to buy homes. In fact, a few years earlier, private entities had also been authorized by law to issue similar mortgage-backed securities (MBSs). Changes

in laws later in the 1980s laid the basis for further development in MBS markets. As early as 1987, 47 percent of mortgages were already being packaged into securities.

Another major change took place in 1999. In 1998, giant Citicorp arranged a merger with Travelers Group, another giant, forming a new firm to be called Citigroup. But under the Glass-Steagall Act the merger was not legal because Citicorp was a commercial bank and Travelers owned an investment bank. Under existing law the merged firm might have been able just to sell off the investment banking business within a couple of years and remain legal. Instead, under heavy lobbying Congress repealed Glass-Steagall in 1999. This made the merger completely legal and gave free rein from then on to banks both to resell mortgages and to repackage them into MBSs, as well as to deal in all kinds of securities. Lenders could now resell mortgages, even in many cases within weeks after they originated the mortgages. So mostly they did so.

A 2010 study done at the Federal Reserve Bank of Chicago found that, out of a large subset of all subprime mortgages originated between 2004 and 2007, 60 to 70 percent were sold off in less than a month. This left the mortgage originators with little incentive to check borrowers' creditworthiness. That is why the repeal of Glass-Steagall is widely regarded as a key cause of the housing price bubble and ensuing disaster.

Another reason why mortgage lending grew so rapidly was that at first there seemed to be no problem. Initially there were still many borrowers who qualified for mortgages on the best terms available. Even among borrowers who had been given subprime mortgages, *delinquency* and *default* rates were low in 2001. (*Delinquency* generally refers to late payments; *default* is more serious, and can mean that payments are at least sixty or ninety days late, or that foreclosure proceedings leading to eviction have already been initiated.) Hence buyers of MBSs and lenders who made the mortgage loans that were resold and went into MBSs did not see a developing problem.

Moreover, sales of MBSs were brisk. Investors burned by the bursting of the high-tech stock price bubble in 1999–2001 were happy to invest in other kinds of assets, such as MBSs. And at the time, with rising home prices and rising home sales, MBSs looked like good investments.

In addition, when issuers of MBSs contracted with credit rating agencies to rate the riskiness of these securities, they made it clear that they expected high ratings or would take their business to other rating agencies. Indeed, as the Securities and Exchange Commission now explains, when agencies are paid by the issuers of securities, "this creates a potential conflict of interest" and "the credit rating agency may be influenced to determine more favorable (i.e., higher) ratings than warranted to retain the . . . issuers as clients."[3]

This identifies an enormous problem that developed (about which MBS buyers were not necessarily warned): agencies often gave these MBSs a AAA rating (riskless), when they actually carried significant risk, shown by their ultimate collapse in value. In fact, when a rater employed in such an agency asked to see the actual mortgages underlying these securities in order to check their riskiness, the rater was told that the information was not available, but also was not needed for the rater to be able to rate the securities. And MBSs rated AAA were easy to sell, meaning that issuers were continually seeking to buy more new mortgages to slice and dice into MBSs. So lenders aggressively sought out potential borrowers, inventing new types of mortgages and reducing or fabricating documentation of loans.

---

[3]SEC, Office of Investor Education and Advocacy, Investor Bulletin, "The ABCs of Credit Ratings," p. 3, available at http://www.sec.gov/investor/alerts/ib_creditratings.pdf

Liars' Loans

Once the most creditworthy borrowers had already taken out mortgages, the quality of available borrowers deteriorated, and lending standards deteriorated with them. Rather than reduce the number of mortgages made, and miss out on the lucrative resale of mortgages to issuers of MBSs, lenders simply made more poor quality loans to less creditworthy borrowers. From 2002 through 2005, there is some evidence that the share of subprime mortgages among all mortgage originations grew. Worse, lenders sometimes filled out incomplete loan applications for prospective borrowers, or failed to document the borrower's qualifications, or wrote down lies, exaggerating the borrower's income or assets. One reason why mortgage brokers failed to document or verify the incomes or assets stated on the loan documents may well have been to keep from being successfully prosecuted for violating laws against predatory lending. Incomplete documentation was evidently thought less likely to bring prosecution than complete false documentation.

So the term *liar's loans* became common in the industry, as did *ninja loans* ("no income, no job, no assets"). From the lenders' standpoint, as long as the mortgages would be quickly resold, this was not a problem; the risk would be passed on to the buyers of MBSs. The issuers of MBSs evidently did not care much whether the borrowers were good risks or not, because repackaging the mortgages as MBSs would conceal their riskiness. And the credit rating agencies were going to give most of them high ratings in any case. Sows' ears became silk purses. Lipstick was painted on pigs. Lenders actively persuaded people to take out loans who by all the odds were unlikely to be able to repay them.

Some state laws against predatory lending existed before this, but in 2002–2003, a number of states passed new laws designed to prevent predatory lending. There is no one clear definition of *predatory lending*—but one thing it means is giving someone a loan that will clearly not benefit that person, because the borrower cannot reasonably be expected to repay the loan in light of the borrower's income, net assets, and credit score. The box "Predatory Lending" in this chapter describes the efforts states made to remedy the problem, and what happened next. In brief, a federal agency took aggressive action to halt investigation and prosecution of predatory lending at the state level, going so far as to file suit against state officials to stop such investigation and prosecution. Predatory lending was mostly allowed to continue, despite creative efforts by state regulators to find ways around the limits imposed by the federal government.

As mentioned above, studies of the evolution of mortgage lending in this period show that most mortgages were being resold to securitizers, as shown in Figure 18.6. Normally, the only mortgages that can be resold for securitization to GSEs Fannie Mae and Freddie Mac are loans that meet a certain quality standard. Lower-quality, riskier mortgages can normally only be sold to private securitizers. In the 2010 study, some of whose findings are shown in the figure, out of prime mortgages originating in 2004–2006, between 70 and 83 percent were sold to GSEs within the year. Of subprime mortgages originated in the same years, about 90 percent were sold to private securitizers within the year, and almost none to GSEs. In all cases, after a year the lender held onto less than 10 percent of mortgages. The data also show that, among subprime mortgages, 60 to 70 percent were sold off to private securitizers almost immediately—within the first month after origination.

In the same study, a much larger share of subprime mortgages than of prime mortgages were the riskier ARMs. In 2004, 73 percent of subprime mortgages were ARMs, while only 26 percent of prime mortgages were. These shares declined

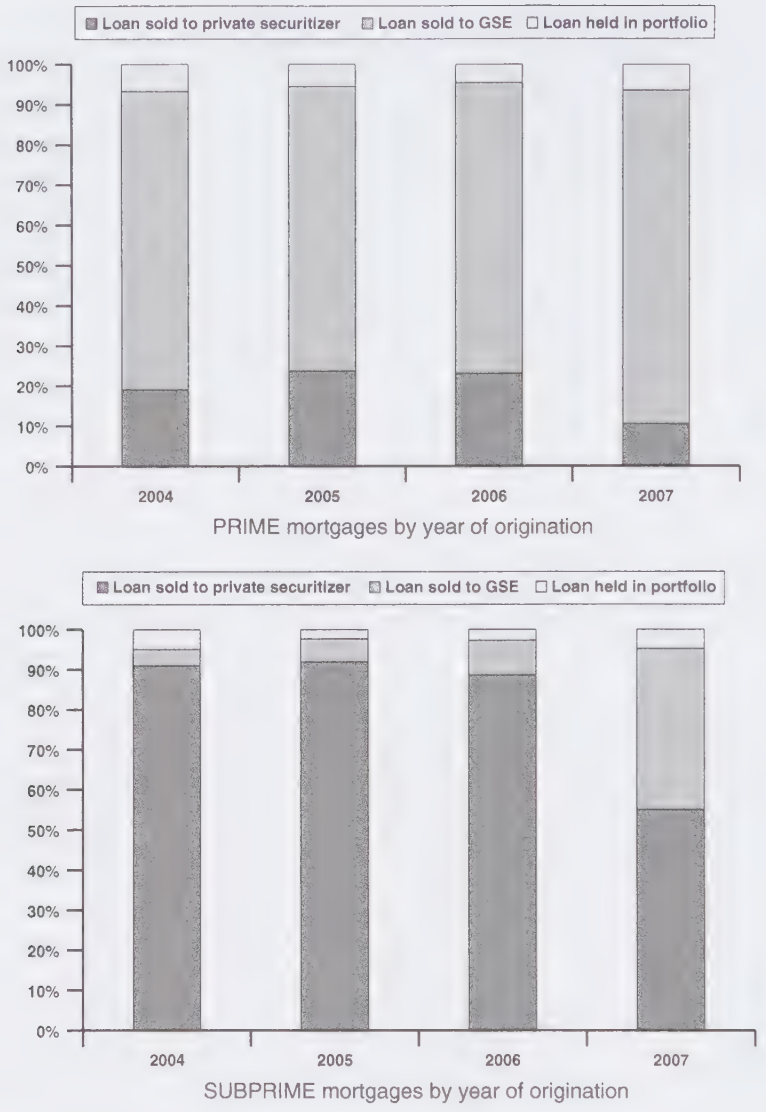


FIGURE 18.6 Prime and subprime mortgages sold to other financial institutions by twelve months after origination.

A 2010 study found that mortgage originators sold off nearly all the mortgages they made during 2004–2007 within the first year after the mortgages were originated—well over 90 percent (the dark gray and light gray portions of the columns shown). However, they sold prime mortgages mainly to government-sponsored enterprises (GSEs) Fannie Mae and Freddie Mac (light gray), while they sold subprime mortgages mainly to private securitizers (dark gray), probably because the GSEs were not willing to buy most subprime mortgages because of their riskiness or lack of documentation. From 2006 to 2007 the number of subprime loans originated fell by more than half, probably in large part because the private securitizers greatly reduced their purchases of subprime mortgages.

Source: Gene Amromin and Anna L. Paulson, "Default Rates on Prime and Subprime Mortgages: Differences and Similarities," *Profitwise News and Views*, September 2010, Consumer and Community Affairs Division, Federal Reserve Bank of Chicago. Available at https://www.chicagofed.org/publications/profitwise-news-and-views/2010/pnv-september2010.

somewhat by 2007. Default rates were rising over time for both prime and subprime mortgages (whether ARM or not), but were much higher for subprime mortgages. For subprime mortgages originating in 2004, just 23 percent had defaulted twenty-one months after origination, while for subprime mortgages originating in 2006, fully 44 percent had defaulted twenty-one months after origination (see Figure 18.7). Of

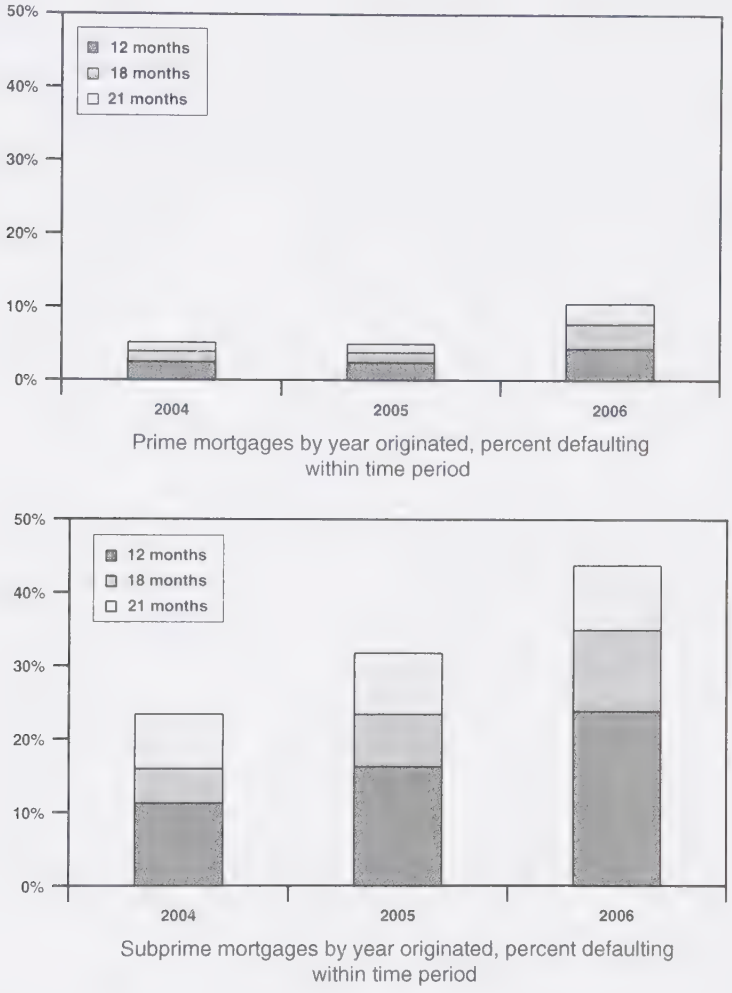


FIGURE 18.7 Increasing default rates on both prime and subprime mortgages, 2004–2006.

For both prime and subprime mortgages, each stacked column shows, for the year of origination, what percent of mortgages were in default twelve months after origination (darkest gray segment), what additional percent of mortgages originating in the same year were in default eighteen months after origination (lighter gray segment), and what additional percent of mortgages originating in the same year were in default twenty-one months after origination (lightest segment). The database included only loans serviced by the large mortgage servicers.

Source: Gene Amromin and Anna L. Paulson, "Default Rates on Prime and Subprime Mortgages: Differences and Similarities," *Profitwise News and Views*, September 2010, Consumer and Community Affairs Division, Federal Reserve Bank of Chicago. Available at https://www.chicagofed.org/publications/profitwise-news-and-views/2010/pnv-september2010.

course, this means the latter group defaulted after house prices had actually begun to fall. For prime mortgages the default percentages were much smaller, but also increased from 2004 to 2006 originations. This gathering storm warned would-be buyers of MBSs that they were becoming bad investments.

Collateral Damage

The number of new single-family homes sold per month peaked in mid-2005 and began dropping sharply by the end of the year. Real home prices peaked in the first quarter of 2006 and also began falling steeply. Suddenly refinancing was out of the question. The rate of mortgage delinquencies rose sharply from 2007 to 2010, as shown in Figure 18.8. For the one hundred largest banks, mortgages with delinquent payments rose to about 12 percent of all mortgages they held, while for smaller banks the increase was only to a little over 4 percent. One main reason was that unemployment ballooned along with the general economic crisis.[4]

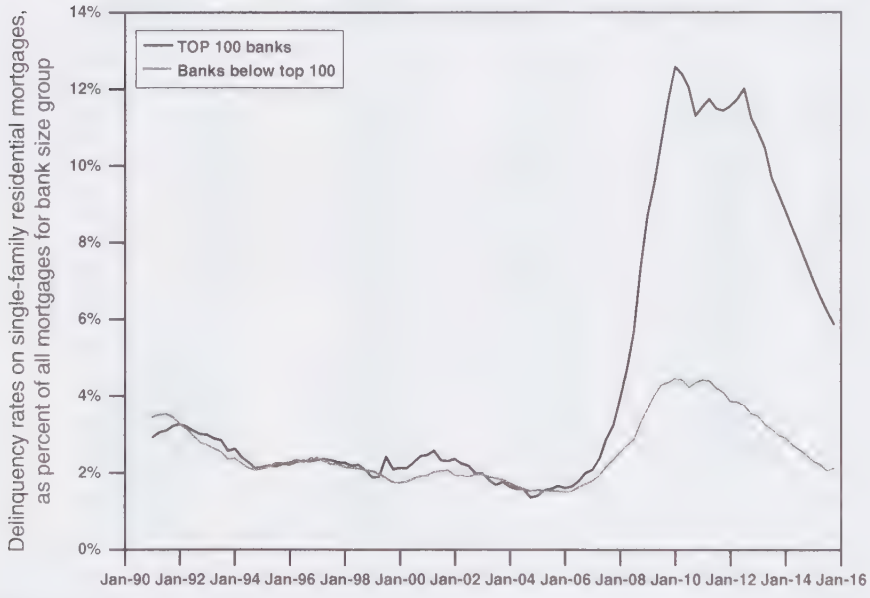


FIGURE 18.8 Mortgage delinquency rates on single-family residences, by bank size.

This figure shows delinquency rates on single-family residential mortgages, as a percent of all outstanding mortgages for the bank size group. The heavy line shows delinquency rates for the top 100 banks, and the thin line for banks below the top 100. After home prices stopped rising in 2005, mortgage delinquency rates grew for all mortgages. But mortgages held by the biggest 100 banks had much larger increases in delinquency rates than mortgages held by smaller banks.

Source: Federal Reserve Economic Data (FRED), "Delinquency Rate On Single-Family Residential Mortgages, Booked In Domestic Offices, Top 100 Banks Ranked By Assets, Percent, Quarterly, Seasonally Adjusted," available at https://fred.stlouisfed.org/series/DRSFRMT100S; and "Delinquency Rate on Single-Family Residential Mortgages, Booked in Domestic Offices, Banks Not Among the 100 Largest in Size (By Assets) (DRSFRMOBS)," available at https://fred.stlouisfed.org/series/DRSFRMOBS.

[4]Alan S. Blinder, *After the Music Stopped* (New York: Penguin Books, 2014).

By 2007 the market for privately issued MBSs had dried up. Fannie Mae and Freddie Mac bought up some of the remaining subprime mortgages that in earlier years would likely have been sold to private securitizers. Naturally, the value of existing MBSs began dropping because of the shrinking likelihood both that mortgage borrowers would be able to continue making regular payments, and that borrowers could refinance to get more affordable monthly mortgage payments. Thus entities that held MBSs as assets watched the value of these assets shrivel. This included banks, other financial institutions, pension funds, firms, and various kinds of government bodies, both in the United States and abroad.

The nominal house price index, which was 190 in the second quarter of 2006, dropped in less than three years to 129. This meant that anyone who had, for example, bought a house for $190,000 in May–June 2006 and still owed more than $129,000 on it was underwater in early 2009, with less than zero equity in their home. By the end of 2009 over 25 percent of all mortgages in the United States were underwater. (Even by 2015, it was reported that 13 percent of all mortgages were still "seriously" underwater, meaning that the amount still owed on the loan was 25 percent or more above the estimated market value of the home.[5])

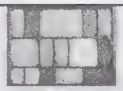
Derivatives of Other Kinds

A *derivative* is a contract—a type of security—derived from some underlying entity or entities. MBSs are one kind of derivative, derived from the mortgage payments packaged into them. But there are also other kinds of derivatives that were important in the lead-up to the crisis. As mortgages were securitized (packaged) into MBSs, so were the MBSs themselves sometimes securitized further into even more complex and obscure derivatives called Collateralized Debt Obligations (CDOs). Often mortgages were packaged with other kinds of debt (credit card debt, student loans, and so on). Their value was difficult for buyers to assess, so buyers were stuck with trusting the ratings given them by credit rating agencies.

Derivatives are often exceptionally risky precisely because their value rests on the underlying streams of income packaged into them—and because typically little or no information is available about those income streams. The lack of information about the worth of derivatives is a reason why their market value can suddenly change when new information comes to light, and that makes them inherently risky. Brooksley Born, who in the late 1990s was chair of the Commodity Futures Trading Commission, warned that derivatives ought to be regulated, and asked Congress to put the CFTC in charge of regulating them. Instead, Congress responded in 1999 by explicitly prohibiting regulation of derivatives. Born resigned from her position.

Holders of either MBSs or CDOs often insured themselves against a drop in their value by buying Credit Default Swaps (CDSs) (a form of insurance sold by the giant insurance company AIG). But when large numbers of borrowers stopped paying their mortgages after the bursting of the house price bubble, AIG suddenly owed enormous payments as a result of these claims—but did not have the money to meet them. In the end, AIG was deemed "too big to fail," and so $182 billion was put together to

[5]Data from RealtyTrac, as reported in Diana Olick, "13 Percent of Homeowners Are 'Seriously' Underwater on Mortgages," available at http://www.nbcnews.com/business/real-estate/13-percent-homeowners-are-seriously-underwater-mortgages-n401081



Predatory Lending

There is a name for practices like aggressively promoting a loan agreement to a potential borrower who very likely will not be able to make the payments: *predatory lending*. About half the states in the United States have some form of law against predatory lending, some of which laws already existed in the early 2000s. Several states, including Georgia and North Carolina, passed new laws in that period. So what happened? Did states simply neglect to enforce these laws? Not exactly.

State officials who were alarmed by the rise of predatory lending went to Washington, DC, in April 2003 to inform John D. Hawke, the head of the Office of the Comptroller of the Currency, and ask him to act. Despite their pleas, he denied there was a problem, and then informed all fifty states that under an 1863 statute, only the federal government had the right to investigate possible predatory lending by nationally chartered financial institutions. When some states tried to investigate state-chartered mortgage lenders, some of the lenders arranged to be acquired by nationally-chartered lenders, and the OCC forced the states to leave the lenders alone. Yet the OCC did nothing to investigate predatory lending by nationally chartered institutions.

New York governor Eliot Spitzer wrote an op-ed in February 2008 that was meant to call attention to the federal government's hypocrisy. Spitzer had been widely known as a kind of "Mr. Clean," prosecuting corporate wrongdoers and advancing from district attorney to special prosecutor to attorney general to the office of governor.

In 2003, when Spitzer was New York's attorney general, he and other state attorneys general noticed that mortgage lenders were frequently "misrepresenting the terms of loans, making loans without regard to consumers' ability to repay, making loans with deceptive 'teaser' rates that later ballooned astronomically, packing loans with undisclosed charges and fees, or even paying illegal kickbacks." Some state governments reached settlements with local lenders, who agreed to stop such practices; and as noted above, some passed new laws. But, Spitzer wrote:

> In 2003, during the height of the predatory lending crisis, the OCC invoked a clause from the 1863 National Bank Act to issue formal opinions preempting all state predatory lending laws, thereby rendering them inoperative. The OCC also promulgated new rules that prevented states from enforcing any of their own consumer protection laws against national banks. The federal government's actions were so egregious and so unprecedented that all fifty state attorneys general, and all fifty state banking superintendents, actively fought the new rules.

And, he said, the OCC actually filed a federal lawsuit against Spitzer's office to halt its investigation of discrimination in mortgage lending.

Spitzer's op-ed did not reach his intended audience in the way he had hoped, however. On February 13, news media broke the scandal that Spitzer had used high-priced prostitutes on several occasions far outside his home state. The story featured lurid details that made Spitzer a laughingstock, forcing him to resign from the governorship. So when the op-ed was published on February 14, hardly anyone took note.

Sources: Eliot Spitzer, "Predatory Lenders' Partner in Crime," Op-Ed, *Washington Post*, February 14, 2008; Robert Berner and Brian Grow, "They Warned Us: The Watchdogs Who Saw the Subprime Disaster Coming—and How They Were Thwarted by the Banks and Washington," *Business Week*, October 20, 2008, 36–42.

bail it out from both the Federal Reserve and the Temporary Asset Recovery Program (TARP) passed by Congress. The federal government used some of it to buy stock in AIG, and by 2010 the Treasury Department actually held 92 percent of the voting stock.[6] In effect, it was nationalized, although temporarily; however, by 2012 it had sold off its shares, reportedly at a profit, and AIG was once again privately owned.

What happened with AIG is a classic example of *moral hazard*: a situation in which a person or firm that is insured acts with greater recklessness than if uninsured. For instance, someone who rents a car must have or buy insurance for the rental. But once insured, he has less reason to be careful with the car, creating a *moral hazard* problem for the insurance company.

So it was with executives at many large financial firms, those with confidence that they were "too big to fail." These executives acted as if they were implicitly insured by the government. That is, they made risky investments, acting as if they believed that in the best-case scenario they would reap great profits, and if their risky strategies led to disastrous failure, their firms would be bailed out and they would lose little or nothing. Mostly, it turned out that they were right.

Bailouts and Buyouts

In all, the total money spent on bailouts was about $14 trillion, about half of it from the U.S. Treasury and about half from the Federal Reserve. Some of the bailouts were loans and were subsequently paid off.

A series of other collapses of major financial institutions followed the bursting of the house price bubble and the linked MBS price bubble. Large financial firms (banks, brokers, mortgage lenders, and so forth) like Merrill Lynch, Washington Mutual, Wachovia, and Countryside were bought out by other large banks, a process described in detail in Alan Blinder's *After the Music Stopped* (2014). But on September 15, 2008, for the first time a collapse in a major financial institution, Lehman Brothers, could not be resolved by persuading some other financial institution to buy it, even at a rock-bottom price. Lehman Brothers therefore went bankrupt, which unnerved other investors and roiled markets. Stock prices began to drop rapidly, and financial sector troubles also began to weaken manufacturing, employment, and the economy as a whole.

One factor in deepening the recession and increasing unemployment was that as the economy declined, revenue to state and local governments also declined. Many were bound by law to maintain balanced budgets, so they could no longer continue spending at previous levels. They were forced to cut back on contracting with local businesses and in some cases to reduce their staff.

Unemployment and Poverty

The standard U-3 unemployment rate, which was still low at 4.4 percent in May 2007, edged up very slowly at first, but in May 2008 started rising fast, from 5 percent that month to 10 percent in October 2009 (see Figure 18.9). Meanwhile, the U-6 unemployment rate, which counts people as unemployed if they have part-time

[6]Steven Davidoff Solomon, "Who Owns A.I.G.? A Continuing Story," October 7, 2010, http://dealbook.nytimes.com/2010/10/07/who-owns-a-i-g-a-continuing-story/?_r=0. See also Alan S. Blinder, *After the Music Stopped*, 136–7.

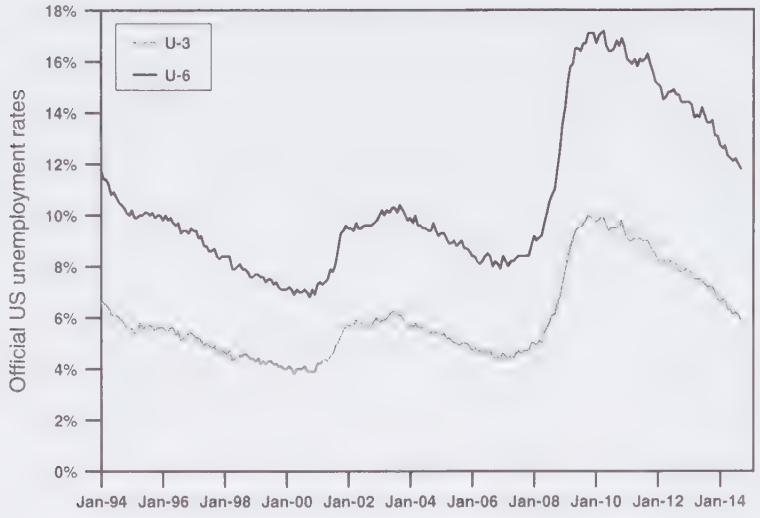


FIGURE 18.9 Two official U.S. unemployment rates.

The Bureau of Labor Statistics computes six different unemployment rates. The one usually cited is the U-3 (the lower thin line), which counts persons as unemployed if they are able and willing to work, not working, but have sought work in the last four weeks. The higher U-6 unemployment rate also includes among the unemployed those who with part-time work who want full-time, and those "marginally attached to the labor force," including those who have sought work in the last year but not the last four weeks—a group that the U-3 measures counts as "not in the labor force" (and so not unemployed).

Source: Vernon Brundage, "Trends in Unemployment and Other Labor Market Difficulties." Department of Labor, Bureau of Labor Statistics, *Employment & Unemployment* 3, no. 25 (Nov. 2014). Available at http://www.bls.gov/opub/btn/volume-3/trends-in-unemployment-and-other-labor-market-difficulties.htm.

work but want to work full-time, and also counts discouraged workers, reached over 17 percent (see Chapter 16).

From 2007 to 2010, nine million people fell into poverty in the United States. The 46 million who were below the poverty line in 2010 brought the poverty rate to 15 percent, up from the 11 percent rate in the year 2000. While on two past occasions (1982–1983 and 1993) during or shortly after recessions, the poverty rate had also risen from 11 percent to 15 percent, in both those cases the poverty rate had declined quickly afterward, as Figure 18.10 shows. After the Great Recession, however, the poverty rate still remained at 14.8 percent in 2015.

Lessons from the History of Economic Crises

Capitalist economies are clearly prone to crises. Since the mid-nineteenth century the U.S. economy has never gone more than ten years between recessions. What causes such economic downturns, or crises? How well can they be tamed by regulation and by suitable choice of policies? We turn now to these important questions.

Over the last century and a half, only during 1940–1980 were economic crises consistently shorter and shallower than the average for the whole period. Government regulation of the financial sector was a major reason for this change. In those

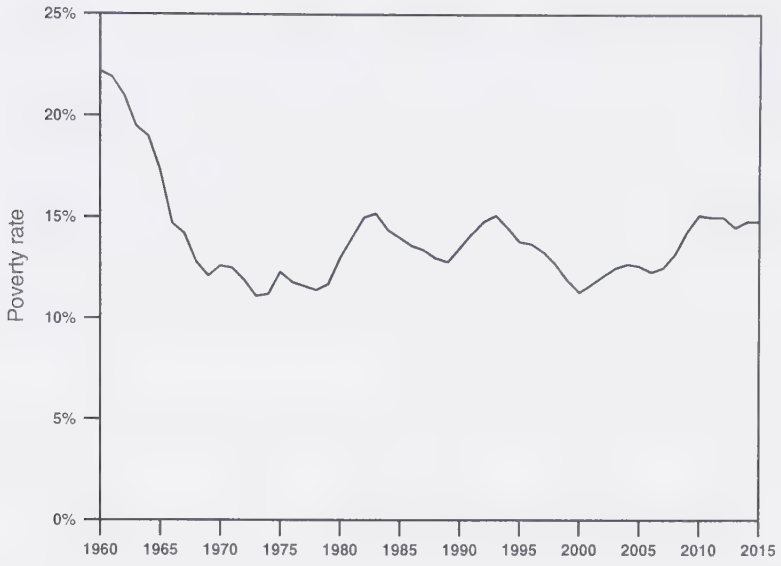


FIGURE 18.10 The poverty rate in the United States. Persons in poverty as a percent of all persons.

The poverty rate rose leading up to the deep 1982–1983 recession, and was rising during the 1990–1991 recession as well, though it peaked in 1993 after the recession had ended. What is striking is that following each of these two peaks in the poverty rate, there was an immediate decline as the economy recovered. But when the poverty rate rose during the 2007–2009 Great Recession, and hit 15 percent in 2010, it remained nearly at that level for the following four years.

Source: U.S. Census Bureau, *Income and Poverty in the United States, 2014*, P60–252, p. 44, Appendix Table B-1: Poverty Status of People by Family Relationship, Race, and Hispanic Origin, 1959 to 2014. Available at www.census.gov/content/dam/Census/library/publications/2015/demo/p60-252.pdf.

postwar decades, regulations that had been imposed on the financial sector during the 1930s were still strong and fairly strictly enforced. Once these were progressively relaxed in the 1980s and 1990s, turbulence in the financial sector grew. The result was the bankruptcy of many savings and loan institutions by the late 1980s, followed by a bailout of several hundred billion dollars. This was followed in 2000–2001 by the bursting of the dot-com stock price bubble and another recession. And as we have seen, deregulation of mortgage lending and refusal to regulate derivatives were important causes of the subprime mortgage crisis and the Great Recession.

This tendency toward periodic crises is intensified by another tendency in capitalist economies, the tendency for wealth and power to become concentrated in very few hands, especially in the financial sector. Indeed, without anti-trust regulation, this concentration would have happened far faster, as the rise of trusts (monopolies) in the late 1800s, discussed in Chapter 6, makes clear. The resulting political influence of large financial firms fosters a strong tendency toward both government deregulation and what economist Hyman Minsky termed *financial fragility*. Financial fragility arises from a situation in which, in order to make larger profits, firms (including financial firms) tend to take on an increasing amount of risk (in the form of either debt or risky assets) relative to their net worth. When many firms overextend themselves in the same way, and owe each other considerable sums, the whole

system becomes financially fragile, raising the danger that one bankruptcy will make the whole system collapse.

In order to explain why crises happen, we first explore crises that originate in the nonfinancial sector, and so can occur even when the financial sector is highly regulated, such as in the post–World War II decades. Next we discuss why crises originating mainly in the financial sector can be larger and more destructive, largely because of the ways in which markets for assets differ from markets for goods. We then go on to consider the how the emergence of giant firms with enormous political power exacerbates these problems and make them harder to reverse. An example is the powerful resistance financial firms have mounted to the Dodd-Frank Act of 2010 intended to re-regulate financial sector activities.

Nonfinancial Causes of Crisis: Overinvestment, Underconsumption

There are several ways in which a crisis can originate in the nonfinancial sector. For one, we have already seen in Chapter 17 that a crisis may be caused when high employment and rising wages squeeze profits.

Crises can also originate in the nonfinancial sector from a temporary, but sometimes severe, mismatch between productive capacity and demand in a major sector, which affects the rest of the economy. Such a mismatch often takes the form of an episode of *overinvestment.*

A capitalist economy by its very nature does not centrally coordinate the amount different firms spend on investment. So when demand for, say, autos is growing, automakers may all compete to enlarge their market share (their sales as a percent of all sales in the industry). They invest in expanded production facilities and try to increase their sales more than do their rivals. But then total investment in the industry is likely to create more total productive capacity than is needed to meet demand. When some find they have overinvested, they are likely to halt further investment for a considerable period while they wait for demand to catch up with productive capacity. But in a large, key sector of the economy like new home construction or auto production, such a halt in investment can have substantial ripple effects that can lead to a general downturn. In fact, a downturn in the home construction sector has quite often preceded recessions in the United States.

Underconsumption as a Cause of Stagnation or Crisis

While overinvestment can create a persistent mismatch between supply and demand, slow growth of demand can sometimes also create such a mismatch. An increase in income inequality can be the cause. Suppose a poor household loses a dollar of income and a rich household gains a dollar. The poor household would have spent most of that dollar, while the rich household will typically spend much less of it. This means that a shift of income from poor to rich will slow, perhaps even halt, the growth of consumption spending.

That slower growth in consumer demand can also discourage firms from investing in new capacity, because they doubt that consumer demand will grow enough so that they can sell much additional output. Both stagnating consumer demand and its effect in discouraging investment spending tend to produce general economic stagnation, which could prolong a crisis, or even help to cause one. In this case, stagnation or crisis can be attributed to *underconsumption* that flows from rising income inequality.

In recent decades, the effect of stagnating wage incomes on the level of consumption has been masked or delayed as households have taken on increasing levels of debt. This postpones a crisis of underconsumption, but can also make that crisis worse when it hits.

Asset Markets Differ from Goods Markets

What else makes capitalist economies crisis-prone? One factor is proliferation of financial asset markets, and especially derivative markets.

Asset markets, particularly financial asset markets, differ in important ways from goods markets, and the two behave quite differently. In goods markets, the value of a good is relatively easy for a would-be buyer to determine, by inspecting it or looking for reviews online or in print. It is often reasonable to assume that the producer mass-produces the good to exact specifications, so that other people's experience with the same good should be a good guide to its quality.

In contrast, visually inspecting a financial asset only gives a little information about it. Its value depends heavily on the state of the economy in which it exists, a state that can change fairly quickly—as happened in 2005–2008. Since assets provide a stream of future income, their value depends on the state of the economy in the future—something about which we can try to make intelligent guesses, but must expect that those guesses will often be wrong.

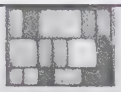
There are, of course, goods such as houses whose usefulness and durability can be checked by inspection—but the *asset* dimension of the house—how much its price is likely to rise in future—cannot be easily checked, and often is not known to anyone, because we cannot see very far into the future.

Because the quality or value of assets is not nearly as transparent as the quality or value of goods, it is easier to deliberately deceive potential buyers about the value of assets. This is what happened, for example, with MBSs and CDOs that were overrated by credit rating agencies. And where deception is profitable and inadequately monitored or regulated, it becomes likely to happen, as we have seen in this chapter.

Lack of solid information also adds to the risk that an upward price trend will suddenly reverse itself. When the information available to would-be buyers is limited and hazy, one additional piece of negative information can easily turn market optimism to market pessimism. As those who play the stock market say, greed can suddenly turn to fear, turning an upward price trend suddenly into a downward price trend. Asset prices have an intrinsic tendency to be volatile.

For instance, a house price bubble can grow based in large part on the assumption that house prices will continue to rise. But when house prices stop rising, then the mortgage lending that kept house prices rising also largely ceases. Refinancing then becomes nearly impossible and rates of delinquency and default rise. As house prices start to fall, they further reduce the likelihood that homebuyers will be able to continue regular monthly payments on their mortgages in future, because refinancing is not an option. This reduces the value of mortgage-backed securities backed by the stream of payments from those mortgages.

The asset markets in which a crisis is rooted need not be housing, nor securities based on housing. Sometimes it is the bursting of a stock market price bubble that sets off a crisis, as in 1929, or in 1999–2001 as the dot-com high-tech stock market bubble burst. Financial assets are often the soil in which the seeds of crisis grow. The box "Bubble Trouble" lists a very few of the major asset price bubbles that have occurred over several centuries in the United States and the world.



Bubble Trouble

The first major recorded asset price bubbles occurred even before economies became predominantly capitalist: the almost simultaneous seventeenth-century Mississippi Bubble in France, recounted in the film *The Ascent of Money* and historian Niall Ferguson's book of the same name, and the South Sea Bubble in Britain. The most common asset price bubbles are stock market bubbles and real estate price bubbles (residential or commercial real estate, or both). Following is a selected short list of such bubbles:

When	What	Where
1719–1720	Mississippi Bubble	France
1720–1721	South Sea Bubble	Britain
(up to) 1837	Panic of 1837 (real estate)	U.S.
1840s	Railway Bubble	Britain
(up to) 1873	Panic of 1873 (railway securities)	U.S.
1921–1926	real estate bubble	Florida
1925–1929	stock market bubble	New York
1986–1992	real estate and stock price bubble	Japan
1996–1997	real estate bubble	Thailand
late 1990s–2001	stock market dot-com bubble	U.S.
2001–2009	house price and derivative bubble	U.S.
2000–15	real estate price bubbles	many European countries
2015–16	stock market bubble	China

Sources: Robert Z. Aliber and Charles P. Kindleberger, *Manias, Panics, and Crashes*, 7th edition (John C. Wiley & Sons, 2015); Niall Ferguson, *The Ascent of Money* (New York: Penguin, 2009). *Financial Times*, various issues.

Why Are Asset Markets Prone to Price Bubbles?

To get a clear understanding of why asset price bubbles happen, it is worth exploring in more depth how assets differ from goods, and how asset markets differ from goods markets. An important way in which goods differ from assets is that we buy them for different reasons. We buy goods in order to use them: food to eat, pens to write with, footballs to play with, furniture to sit on, cars to drive, books to read. In contrast, we buy assets not because of their usefulness but because we believe that they will provide a stream of income or be a safe way to hold or increase wealth.

People may buy a share of stock for the dividends they hope it will pay, or buy a bond for the annual fixed coupon payments it brings. They may buy an apartment building (if they can afford it) for the rental income from it, if it seems likely to be enough to more than cover the costs of taxes and maintenance.

But a second reason to buy an asset is the expectation that its price will rise and it can then be sold to reap a *capital gain*. Chapter 10 gave the example of merchants

buying cheap and selling dear, but also mentioned that investors quite often speculate in (buy and sell) financial assets (stocks, securities, other). Their hope is that the price of the asset they buy will rise faster than the price of everything else. If its price rises at the same rate as the CPI, it is a relatively safe way to store wealth. If it rises faster, it can then be sold for a capital gain, or used as collateral for a loan that could be used to buy more assets.

The prices of assets fluctuate more than the prices of most manufactured goods. Why is this so? There are two main causes. As mentioned above, one has to do with information. It is often easier to find out the true qualities of a good than of an asset. A prospective buyer of a good can sometimes inspect it where it is sold, or look at reviews online or in print. Manufactured goods are typically produced in a way intended to ensure that each unit of the product is the same as any other. Quality is tested, and defective units are reworked or rejected. To be sure, this is partly due to regulations like product safety laws. But producers may also worry about their brand name's reputation; for instance, a car with widely publicized defects can lose sales with astonishing speed.

As also mentioned briefly above, the value of a financial asset cannot usually be assessed by looking at it, but rather depends heavily on how the circumstances under which it is traded might change over time. If house prices are rising, a mortgage-backed security might look like a valuable asset; but if house prices will then fall, it might not. Thus the value of an asset depends on conditions in the whole economy in a way that the value of a good does not—and how these conditions will evolve is hard to predict.

Lacking good information about the value of assets, investors tend to follow current trends. The fact that an asset's price is rising is a signal to buy; the fact that its price is falling is a signal to sell. But this creates sharp fluctuations. Having little information, market participants may tend to assume that others have such information. Or they simply think that if they fail to buy when others are buying, they will miss an opportunity for capital gains; and when they fail to sell when others are selling, they are likely to be left holding assets of very little value. Keynes described stock market participants as displaying "herd behavior." Such behavior rarely happens in markets for manufactured goods, whose prices are far more stable; and if it does happen in markets for things we view as goods, it is often because the good has to some extent taken on the qualities of an asset, even if only temporarily.

Indeed, goods can in some cases also be assets, either temporarily or somewhat permanently. If a good is bought for use, it can nevertheless also increase in value over time, sometimes predictably so. In that case it is also an asset. An example is a fine painting. I may buy it because I love its artistry, and want to display it in my living room to show to my guests. In that sense it is a good. But if the artist is achieving growing recognition, the value of the painting may also rise over time, and it may then be regarded as an asset. Most homebuyers use housing both as a good and an asset, but there are also some players in housing markets who mainly treat housing as an asset.

To explain further the house price bubble shown in Figure 18.2, we turn to an explanation of feedback loops.

Feedback Loops

We have seen that the incentive structures facing those involved in the house price bubble initially all worked in the direction of driving house prices up. The dynamics of asset price bubbles are an example of a *feedback loop.*

The classic example of a feedback loop comes from a sound system in which a microphone is mistakenly placed in front of the speaker (amplifier) to which it is connected. When it is turned on, we hear an ear-splitting shriek because the amplified sound re-enters the microphone and gets amplified again, and this is repeated many times in a split second. The cause-effect relationship circles back on itself, forming a closed loop:

> mike converts sound input to electrical signal → amplifier amplifies the sound → mike converts that to electrical signal . . .

In general, when X causes Y, which causes X, we have a feedback loop. The chain of causation can also have more than two elements (X → Y → Z → X, for example), but must loop back to the first cause to constitute a feedback loop. Another name for a feedback loop is a *vicious circle* or a *virtuous circle*—although the very same feedback loop can be viewed as a virtuous circle from the standpoint of some in the economy, but a vicious circle from the standpoint of others. A key purpose of identifying a vicious circle is to establish that it takes outside intervention to interrupt it and prevent its unfortunate results.

On the upside of the house price bubble, the feedback loop—at least temporarily—was this:

> home prices keep rising → borrowers happily borrow and investors happily buy MBSs → mortgage lenders happily lend → more houses are bought → home prices keep rising

Prospective borrowers were willing to accept mortgages on unfavorable terms because home prices were rising rapidly, and they believed they could resell or refinance their homes within two to three years and get more favorable terms. Mortgage lenders (banks or others) were eager to lend to such borrowers, as long as home prices were rapidly rising so that it would be easy to resell the mortgages to MBS issuers for a good price. More lending caused more home purchases, which caused home prices to keep rising, and rising home prices in turn encouraged the sale of MBSs and therefore more mortgage lending.

But home prices eventually reached a level at which fewer prospective homebuyers could afford them. Qualified borrowers became more and more scarce; lending standards deteriorated; mortgages were offered on riskier terms; mortgage brokers wrote more "liar's loans"; and the virtuous circle (or feedback loop) supporting rising home prices turned fairly quickly into a vicious circle (another feedback loop) of falling home prices. When home prices were falling, some buyers either defaulted and their homes were foreclosed on, or they tried to sell their homes before they lost all the equity in them. Sales of foreclosed homes and sales by homebuyers both helped to accelerate the decline in home prices. If asset price bubbles end up in feedback loops, or vicious circles, of this kind, then what is the outside force that can interrupt them? The answer is government regulation, which we discuss shortly.

Reflexivity

In financial markets there is a particular feedback loop that also gives momentum to asset price bubbles. Billionaire philanthropist George Soros calls it *reflexivity*, and it

has to do with the interaction between the real world and human beings' perception of that world. (See Soros, "Reflexivity. . ." in the Suggested Readings.)

On the upside of a house price bubble, many people buy homes, and the price of each home rises after it is bought. Once the price has risen, mortgage lenders treat the home as if its intrinsic value had risen—rather than as if a speculative bubble had temporarily raised its market price, but was likely to burst and lower that price in future. Lenders now regard the house price as representing collateral valued at the new price, and they grant larger mortgage loans than before on that basis, making the purchase possible.

But the longer the price bubble continues, the higher the price of each home rises above the value it would have in a non-bubble market—and the more likely the home's price will fall suddenly back to that level (or lower) in future.

This mutual causation between reality and perception, by which price increases raise the perceived value of collateral and therefore make lenders willing to grant larger mortgage loans, is what Soros labels *reflexivity*. And this also contributed to the formation of the house price bubble. The same tends to happen in any asset price bubble. The main point is that causation goes both from reality to perception, and from perception to changes in reality. When something changes in the economy, it also changes the perceptions of the human beings who may take action toward that economy; and their changed perceptions thereby react upon the economy itself.

Misinformation

Previous chapters have already said much about what capitalists do to make a profit. One strategy that has not been much discussed, though, is that they can exaggerate the value of what they are selling. Advertising often involves minor exaggerations. Major exaggerations, amounting to outright deception, are defined as fraud and are clearly illegal.

Sometimes these regulations are enforced consistently enough to make it unprofitable to deceive potential customers. But history is full of examples of capitalist enterprises selling dangerous or unhealthful products as if they were perfectly fine, and profiting by doing so. Unfortunately, competition can drive out honesty. When one firm cuts corners by producing harmful products without telling the public, firms that spend the cost and effort to produce and sell safe products in the same market will tend to have higher costs and therefore lower profits.

Government regulation is therefore necessary. That is why federal, state, and local agencies were created to inspect the safety of food, medicines, toys, autos, buildings, and so on. Even so, there are occasional outbreaks of contamination in meat from major firms, and pharmaceutical products are approved that, according to their television ads, have serious side effects. And over time, *regulatory capture* tends to take place, whereby the industries being regulated by an agency place their friends in the agency, or influence its activities, so that enforcement is blunted. To serve the public interest, the tendency for capitalists to profit from deception must be countered by effective regulation, but firms often resist such regulation. Indeed, the conflict over whether foods should be labeled to indicate that they contain genetically modified organisms (GMOs) is a good example of a case in which capitalist enterprises find it more profitable not to inform consumers about what is being sold.

Moreover, every country has its swindlers and con artists for whom deception can be highly profitable. The sharp fall in the stock market in fall 2008 was what forced well-known investment fund manager Bernie Madoff to admit that he had defrauded his clients for years. His investment fund, which had given his clients high and steady annual returns over many years, was a sham: it never invested his clients' funds at all, but simply operated what is called a Ponzi scheme: Returns to existing clients were being paid entirely out of money invested by new clients. As long as business was growing, he could keep this a secret. But as the stock market and other markets collapsed in fall 2008, Madoff could no longer find enough new clients to continue the scheme, and so confessed, turned himself in, and was sentenced to jail time. There is a kind of loose parallel with the house price bubble: As long as new money was coming in and driving house prices higher, everything worked well; but when new money dried up, the whole thing collapsed.

In unregulated markets in a capitalist system, when stretching the truth is profitable and not penalized, the truth will often be stretched. In the house price bubble, as we have seen, deception was one element both in mortgage lending and in sales of mortgage-backed securities. Consumers were often led to believe that the mortgage loans they sought were less risky than they actually were, and at the same time buyers of MBSs were told that these securities were less risky than they actually were. Moreover, by 2007 the large majority of real estate appraisers reported that they had been pressured by lenders to appraise homes at higher values than they thought were appropriate.

Outright fraud was also a factor in the subprime crisis, just as it was in the savings and loan debacle of the 1980s. As a result of the S&L crisis, the staff of the Office of Thrift Supervision (which supervised the S&Ls) was responsible for analyzing evidence that led to more than one thousand felony convictions related to fraud.[7] After the year 2000, prosecutions for mortgage fraud by large firms were rare, although years after the Great Recession officially ended, a number of big banks did settle with the government for billions of dollars to avoid prosecution.

Deregulation and Financial Fragility

Another reason why capitalism is crisis-prone was explained by economist Hyman Minsky in writings that spanned several decades and built on the work of John Maynard Keynes. His best-known works were published in the 1980s. Minsky observed that a major crisis typically cleared much debt out of the economy and typically led the government to impose new regulations designed to prevent similar crises from happening. For instance, in the Great Depression banking was regulated, the Federal Deposit Insurance Corporation was created to insure consumer bank deposits, the Securities and Exchange Commission was formed to regulate corporate behavior and reporting, and so forth. After a crisis ended, however, both these trends tended to reverse themselves gradually over time: Firms tended to accumulate debt and the regulatory regime tended to be weakened.

Immediately after a crisis, firms became more cautious about taking on debt and engaging in risky transactions. But over time firms became more willing to borrow

[7]William K. Black, "How to Rob a Bank," TedX talk, video, Kansas City, Missouri, https://www.youtube.com/watch?v=-JBYPcgtnGE

and incur risk in order not to miss profitable opportunities. Decades after a crisis, caution faded, and regulation seemed less necessary. As new profitable opportunities arose, pressure from business convinced government to deregulate. And when new kinds of financial assets were invented, as happened with repeatedly with different types of derivatives, the government became convinced that they did not need to be regulated at all.

Increasing deregulation allowed firms to take on more debt, or more risky assets, or to adopt risky business strategies, and yet hold lower capital reserves than in the past to cover the risk of losses. What reinforced this process was a decline in enforcement of existing laws, and a change in regulations implementing those laws, which in some cases actually began to reverse the intent of existing laws. (See the box later in this chapter, "No Skin in the Game: Loophole Eats Rule.")

Over time, the combination of deregulation and increasingly risky financial structures within and among firms led to what Minsky called *financial fragility*: a situation in which each firm is stretched nearly to the limit of the debt it can handle in good times, and in addition firms are linked together in a network of payment obligations so that if one fails, a domino effect can cause many others to fail. A single major bankruptcy then can endanger the whole financial sector. A condition of financial fragility tends to reappear at intervals of several decades, and to cause crises. (Such a network of obligations is also possible in principle among nonfinancial firms, but the financial sector is more vulnerable to such problems, and has been the epicenter of the market earthquakes triggering crises.)

Large Firms Exacerbate the Problem

These are the reasons why capitalist economies are generally vulnerable to periodic crises. But there is also a reason why that crisis tendency is worse in the current period, and likely to get still worse unless adequate regulatory steps are taken. In a capitalist economy there is a long-term tendency for wealth and power to become highly concentrated in a few very large firms. This is particularly true in the financial sector, where there are two enormous advantages to bigness.

The first advantage is that a financial sector firm that is very large is almost certain to be seen as "too big to fail," and therefore, if it faces ruin, is almost certain to be bailed out by the government or the Federal Reserve. As is widely acknowledged, this creates what, as mentioned earlier, economists call *moral hazard*: a situation in which the firm's managers and other personnel know that if they take great risks, they will either reap high returns, or, if they face large losses, they will be bailed out. Faced with the possibility of large gains from success, and minimal losses from failure, executives of these firms will lean toward choosing high-risk, high-return strategies. This insulation from downside risks also allows big banks to borrow at more favorable interest rates than their smaller competitors.

The second advantage is that concentration of economic power in large financial firms motivates and makes it possible for these firms to seek greater political influence. Such influence both allows them to achieve deregulation and to become prime candidates for future bailouts. Chapter 11 discussed the advantages of *economies of scale* based on large fixed costs and much smaller constant marginal costs as one explanation for why firms grow large. But there can also be scale economies in gaining political power. Setting up a lobbying campaign or a political donation apparatus may have substantial startup costs, but then approaching one more senator or

promoting one more piece of legislation has a small marginal cost. And the payoffs can be enormous.

Total expenditures by the financial sector (that is, Finance, Insurance, and Real Estate, or FIRE) on lobbying Congress grew from about $200 million a year in 1998 to nearly $500 million in each year from 2010 through 2015. Campaign donations also grew at a rapid rate. Figure 18.11 shows the rapid rise in financial sector campaign donations, with donations in two-year presidential election cycles shown separately from donations in two-year nonpresidential election cycles. Donations rose steeply over twenty or more years, focusing on key committees. In the 2014 election cycle, 37 percent of campaign donations to members of the House Financial Services Committee ($31 million out of $84 million) came from the FIRE sector (from both PACs and individuals). For the Senate Banking, Housing, and Urban Affairs Committee, the FIRE share was 25 percent, or $48 million out of a total of $193 million.[8] Indeed, seeing the political advantages of bigness, firms may embrace chances to grow through mergers and acquisitions, so that they can have more influence.

There has also been an active revolving door between the private and public sector, with financial sector executives moving into high positions in the executive branch, and government officials moving into lucrative positions in the private financial sector. The office of the secretary of the treasury has been held (1995–1999) by former Goldman Sachs executive Robert Rubin, who joined Citigroup after he left office, as well as another former Goldman Sachs executive Henry (Hank) Paulson

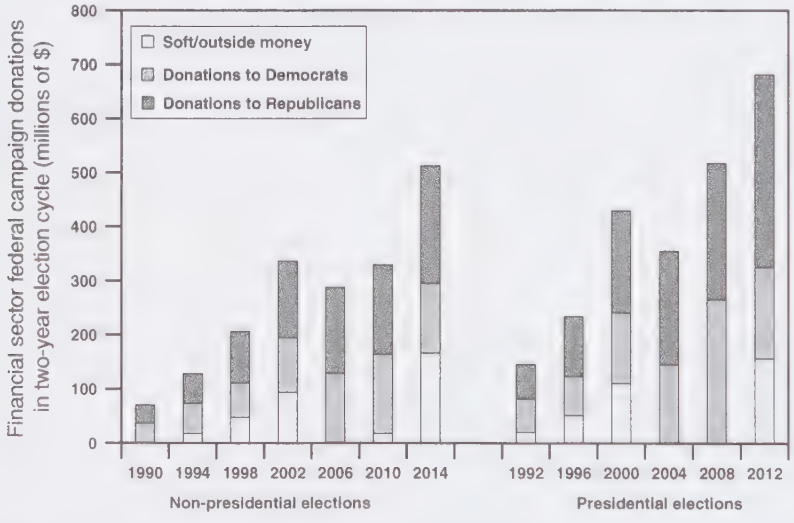


FIGURE 18.11 Financial sector campaign donations to federal candidates and soft/outside money organizations, by two-year election cycle.

Source: Center for Responsive Politics, Finance/Insurance/Real Estate: Long-Term Contribution Trends, available at http://www.opensecrets.org/industries/indus.php?Ind=F; https://www.opensecrets.org/industries/totals.php?ind=F; House Financial Services Committee, 114th Congress, 2014 election cycle, available at https://www.opensecrets.org/cmteprofiles/overview.php?cmteid=H05&cmte=HFIN&congno=114&chamber=H; and Senate Banking, Housing, and Urban Affairs Committee, 114th Congress, 2014 election cycle, available at https://www.opensecrets.org/cmteprofiles/overview.php?cmteid=S06&cmte=SBAN&congno=114&chamber=S.

[8]All data are from the Center for Responsive Politics at https://www.opensecrets.org.

(2006–2009); John Snow (2003–2006) resigned as treasury secretary to become chair of Cerberus Capital Management.

That the revolving door between the financial sector and government has brought benefits to financial firms is evident in part from the repeal of the Glass-Steagall Act under Treasury Secretary Robert Rubin, and his subsequent leadership positions in Citigroup. Further evidence of benefits to financial sector firms from gaining influence in government emerges from comparing the government's response to the savings and loan crisis of the late 1980s with its rather different response to the subprime crisis. In the S&L crisis, the Office of Thrift Supervision, working with the FBI, investigated S&L officials and charged over one thousand with crimes, prosecuted them, and in over 90 percent of cases obtained convictions. After the recent crisis, however, no executives of Wall Street firms were prosecuted. Instead, the firms themselves agreed to multi-billion-dollar settlements. The total amount of such mortgage-crisis-related settlements actually paid is unclear. Some accounts put it at over $100 billion. Meanwhile, average Wall Street bonuses continued at a high level, as shown in Figure 18.12.

Weakening the Social Safety Net

In the last two decades, while deregulation of the financial sector has been proceeding apace, the social safety net on which people could fall back in case of personal or economy-wide disaster has shrunk. In the mid-1990s, the welfare system was weakened, with the end of Aid to Families with Dependent Children (AFDC) and its replacement with Temporary Aid to Needy Families (TANF). Families that under

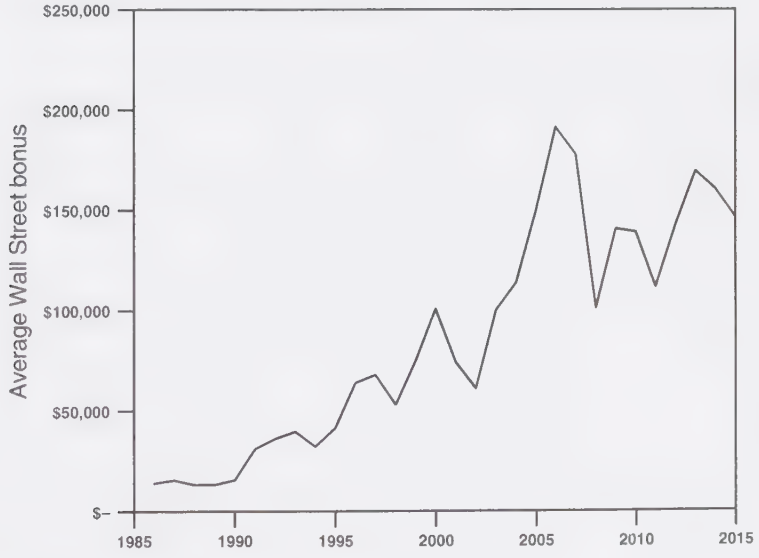


FIGURE 18.12 Average Wall Street bonus, by year.

All employees are included in averaging the total pool of bonuses, but many top executives actually receive bonuses in the millions or tens of millions of dollars, and sometimes more.

Source: New York State Comptroller, "New York City Securities Industry Bonus Pool." Available at http://www.osc.state.ny.us/press/releases/mar16/nyc_security_bonus_pool.pdf.

AFDC could receive welfare payments when they needed them were restricted under TANF to a maximum of five years of benefits over their entire lifetimes. Some states imposed shorter maximum periods.

In 2005, Congress reduced the options available to households for going into bankruptcy and getting rid of their debts. The new law limited the circumstances under which a debtor could file for personal bankruptcy under Chapter 7 and get rid of all his or her debts, and pushed debtors to file instead under Chapter 13, which required that the debtor repay some of the debts before earning the right to cancel the rest. After the law went into effect, bankruptcies fell sharply to less than one per thousand of population. But then almost at once the Great Recession began, so that personal bankruptcies rose again to about two per thousand per year, remained there for 2009–2011, and then slowly began to recede back to about 1.5 per thousand by 2012. Thus many who could have filed for bankruptcy in earlier years, clearing their debts and starting afresh, faced new obstacles in doing so by the time the crisis hit. Many families saw their personal plans disrupted, as they lost their college or retirement savings, lost health insurance, surrendered their cars when they couldn't pay car loans, lost jobs, and became anxious as their personal financial situations moved closer to the edge of disaster.

The Great Recession was over by official reckoning in 2009, but its effect has been harsh and long-lasting for millions. From 2007 to 2010, nine million people joined the ranks of the unemployed, including many construction workers. It left several million people unemployed for years and four million people working part time who wanted or needed to work full time. It robbed many of their savings when they had to abandon homes they were in the process of buying. It pushed nine million more people into poverty, and they remained there for years. It lowered the median household income, and caused a sharp fall in median household net worth, affecting those in the middle income range, as well as many below and some above.

Regulation in a Capitalist Economy

What should be done? If for the foreseeable future most nations are going to have capitalist economic systems, clearly one necessary strategy is strict regulation, especially of financial sector activities.

For several decades after World War II the U.S. financial sector was remarkably tame. In most states, interest rates were capped at fairly low levels by state laws. Interstate banking was strictly regulated. Commercial and investment banking were strictly separated. There were few exotic or hard-to-comprehend, hard-to-value financial assets. Local banks made local mortgages and hung onto them. Banking was fortunately, as described by economist Gerald Epstein, "boring." The largest banks were well known and fairly powerful, but nowhere near the size or influence of the largest banks today. The transition "from boring banking to roaring banking" (Epstein's phrase) brought few, if any, benefits for the general population, and many hardships.

Regulating the Financial Sector

The financial sector contains ticking time bombs of unregulated activities that will, with near certainty, lead to further crises unless the sector is strictly reregulated. What kinds of regulations are needed?

Breaking up the big banks and other large financial institutions in order to end the "too big to fail" phenomenon has been proposed by many economists and policymakers, and is feasible. If size is part of the problem, size can be reduced—if big banks' political opposition can be overcome. In 2015 the total assets of the four largest banks were about half of U.S. GDP by U.S. accounting standards (about the same as during the crisis). These banks are "too big to fail," according to Simon Johnson, MIT economics professor and former chief economist of the International Monetary Fund. He and co-author James Kwak (*13 Bankers*, 2011) propose simply putting a size limit on the assets of any one bank and creating a procedure for them to meet that limit. Standard Oil was, after all, broken up by the courts over one hundred years ago in the public interest, using anti-trust law; this is not a new concept. Johnson and Kwak propose capping commercial bank assets at 4 percent of GDP, and investment bank assets at 2 percent of GDP (since their activities are riskier).

In addition, where incentives reward risky behavior when it reaps profits, and do not penalize it when it causes enormous harm, the obvious solution is to change the incentives facing the actors. There are many ways to do this, all or most of which could be used. Risky behavior by the highest decision-makers of major financial institutions should be made costly enough, both for them individually and for their firms, that they are deterred from taking those risks. The way it was during the boom, decision-makers received high salaries and large bonuses; during the collapse, they paid neither a monetary penalty nor faced prosecution for crimes such as fraud. Although the institutions themselves (not the individuals who profited) later paid tens of billions of dollars in settlements to the government related to their activities leading up to the crisis, it is doubtful whether such settlements have much deterrent value.

Measures should be taken to limit creation of very risky securities, as well as to discourage sale of securities whose buyers have too little information to assess their value. Newly invented securities could be prohibited from sale unless expressly permitted after a rigorous examination and approval process. Banks could be required to perform much more rigorous and realistic *due diligence* on derivatives they issue, including stress testing under conditions of a widespread collapse. (*Due diligence* refers to meticulous examination of the pros and cons of a potential investment or transaction.) Such a process might make it unprofitable to create the riskiest securities (perhaps not a bad thing). Over-the-counter sales of derivatives could be prohibited, that is, all derivatives could be required to be traded on exchanges only, which could ensure that only the most transparent securities could be traded. Again, if the added costs made some risky securities unprofitable, and caused their demise, that might not be a bad thing.

To ensure that large losses on risky securities would not lead to insolvency, banks could be compelled to hold capital reserves equal to a substantial share—some recommend about 8 percent, some more—of the value of all their assets (whether they are held on or off balance sheet). Such a requirement might take away the allure of holding high-risk, high-return securities; and this, too, might not be a bad thing. However, when Lehman Brothers failed, it had substantial capital reserves; but they were not enough to prevent failure.

Executives of financial institutions could be required to repay a large share of their boom-time bonuses or excessive salary payments when the same decisions that made temporary high profits for the firm then led to large losses in later years. They could also be required to contribute large amounts in case their firms

requested and received government bailouts. This might help to deter reckless decisions.

Investment banks should be actively regulated, instead of just being asked for voluntary compliance with standards (as the SEC was doing leading up to the crisis). And some experts have proposed establishing public banks, by letting the post office assume certain banking functions, or by nationalizing one or more banks. (As we saw, AIG actually *was* in effect nationalized when it was bailed out, but it was re-privatized as soon as this seemed feasible.)

The Dodd-Frank Act

In 2010 Congress passed the Dodd-Frank Act, a new law regulating the financial sector. Its provisions do in principle reduce the likelihood or the severity of future crisis, but there are two important limitations. One is that some of the provisions are not as strong as might be needed to deter risky behavior. The other is that the financial sector has been able to slow down by several years the process by which the government arrives at final regulations for implementing the law—by sending thousands of lengthy public comments to agencies in response to proposed regulations, comments that had to be read before the process could go forward. Such strategies delayed issuance of some regulations by as much as five years. Even by early 2016, many final regulations had not been issued, and financial firms were pressing for repeal of the entire Act. Perhaps more important in the long run, however, is not so much the slowdown in regulations, but the watering-down of them.

In principle, Dodd-Frank improved the situation. It authorized the Federal Deposit Insurance Corporation, the agency that reorganizes insolvent banks, to liquidate other types of insolvent financial institutions. In principle, it prohibited taxpayer bailouts of supposedly "too big to fail" institutions. It put the Federal Reserve in charge of regulating and supervising so-called "systemically important financial institutions" (SIFIs), a polite term for "too big to fail." It also created the Consumer Financial Protection Bureau, which by 2016 had already had responded to over a million complaints and provided $12 billion in relief to consumers who had been wrongly treated by financial institutions.

As passed by Congress, Dodd-Frank required issuers of risky mortgage-backed securities to hold at least 5 percent of such securities themselves, in principle forcing them to be more cautious and in effect to have "skin in the game." However, as regulations were issued to implement the law, the government widened an existing loophole by defining a large category of securities as exempt. See the box, "No Skin in the Game: Loophole Eats Rule."

Inequality, Concentration, and Crisis

To avoid or at least minimize crises, a democratic government needs to recognize who are the small number who are able to profit from crisis, and who are the vast majority who suffer losses from crisis. In a democracy, government should represent the large majority. But when income and wealth become concentrated in a very small group, they can more easily ensure that they benefit (or do not lose) from an asset price bubble that bursts and causes crisis. The increasing size of the financial sector and its increasing influence and presence in government have shifted the center of gravity, speeding up the cycle that Minsky identified as gradually recreating financial fragility after a crisis. This constitutes a feedback loop of wealth and power:



No Skin in the Game: Loophole Eats Rule

Those determined to effectively regulate the financial sector succeeded in putting into the Dodd-Frank Act of 2010 a rule that the securitizer—the bank or financial institution that issued mortgage-backed securities—had to keep 5 percent of those securities. This was intended to motivate the institution to avoid creating securities that were too risky, because it would have "skin in the game."

In 2014, however, a loophole written into Dodd-Frank ended up (as co-sponsor Rep. Barney Frank put it) "eating the rule." The loophole was that some mortgages—those that were considered exceptionally safe because they required substantial down payments—would be exempt from the 5 percent "skin in the game" requirement. The securitizer could repackage them into MBSs without being required to retain any of them. This supposedly "supersafe" category of mortgages was called Qualified Residential Mortgages, or QRMs.

But after an aggressive lobbying campaign, the final regulation issued in 2014 put most mortgages into the exempt category. It allowed even mortgages with *no* down payment at all—clearly quite risky mortgages—to be called QRMs, meaning that despite their risk, the institution that securitized them would not be compelled to hang onto a single one of them. Unfortunately, this opened the door to the same kinds of recklessness that led to the subprime crisis in the first place. Minsky was right: after a crisis forces for deregulation grow ever stronger; but he surely never imagined that the process would happen quite that fast.

Source: Floyd Norris, "Banks Again Avoid Having Any Skin in the Game," *New York Times*, October 23, 2014, available at http://www.nytimes.com/2014/10/24/business/banks-again-avoid-having-any-skin-in-the-game.html?_r=0

Firms with enormous assets and profits are both able and motivated to secure political influence, and that influence in turn protects and enhances their profits. It is no doubt seen as a virtuous circle by the firms, but from the standpoint of the public it looks more like a vicious circle.

Some would argue that the best solution to the problem of crisis, and to other problems, is a substantial redistribution of wealth and income that would greatly level the political playing field, along with reforms to government and elections that would allow the majority to be better represented. Chapter 19 discusses what government does in a capitalist economy, how it is financed, and how this has changed over time.

Suggested Readings

Anat Admati and Martin Hellwig. *The Bankers' New Clothes: What's Wrong with Banking and What to Do about It* (Princeton: Princeton University Press, 2013).

William K. Black, *The Best Way to Rob a Bank Is to Own One: How Corporate Executives and Politicians Looted the S&L Industry*, with postscript on current crisis (Austin, TX: University of Texas Press, 2013).

Alan S. Blinder, *After the Music Stopped: The Financial Crisis, the Response, and the Work Ahead* (New York: Penguin Books, 2014).

John Cassidy, *How Markets Fail: The Logic of Economic Calamities* (Picador, 2010).

James Crotty and Gerald Epstein, "Crisis and Regulation: Avoiding another Meltdown," *Challenge* 52, no. 1 (January/February 2009): 5–26.

Gerald Epstein, "From Boring Banking to Roaring Banking" (July/August 2015), http://www.dollarsandsense.org/archives/2015/0715epstein.html

Charles H. Ferguson, et al., *Inside Job*, video (Culver City, CA: Sony Pictures Home Entertainment, 2011).

Mark Jarsulic, *Anatomy of a Financial Crisis: A Real Estate Bubble, Runaway Credit Markets, and Regulatory Failure* (New York: Palgrave Macmillan, 2010).

Simon Johnson and James Kwak, *13 Bankers: The Wall Street Takeover and the Next Financial Meltdown* (New York: Vintage, 2011).

Robert Z. Aliber and Charles P. Kindleberger, *Manias, Panics, and Crashes*, 7th edition (New York: Palgrave Macmillan, 2015).

David M. Kotz, *The Rise and Fall of Neoliberal Capitalism* (Cambridge, MA: Harvard University Press, 2015).

Chris Martenson, "Bubbles," https://www.youtube.com/watch?v=0F7SCbrU5sQ (n.d.).

Hyman Minsky, *Can "It" Happen Again?* (New York: Routledge, 1982).

Hyman Minsky, *Stabilizing an Unstable Economy* (New Haven, CT: Yale University Press, 2008).Matthew Sherman, "A Short History of Financial Deregulation in the United States," (Center for Economic and Policy Research, July 2009) http://cepr.net/documents/publications/dereg-timeline-2009-07.pdf

George Soros, "Fallibility, Reflexivity, and the Human Uncertainty Principle," *Journal of Economic Methodology* vol. 20, no. 4, special issue, "Reflexivity and economics; George Soros's Theory of Reflexivity and the Methodology of Economic Science," http://dx.doi.org/10.1080/150178X.2013.859415.

George Soros, *The Crash of 2008 and What It Means: The New Paradigm for Financial Markets* (New York: PublicAffairs, 2009).

Jennifer Taub, *Other People's Houses* (New Haven, CT: Yale University Press, 2015).

Martin H. Wolfson and Gerald A. Epstein, eds., *The Handbook of the Political Economy of Financial Crises* (New York: Oxford University Press, 2014).

Government and the Economy in Transnational Capitalism

19

In 2012, editor and writer André Schiffrin, living in New York, learned that he had stage four pancreatic cancer. He began getting chemotherapy at New York's Memorial Sloan Kettering, a world-renowned cancer center. But he wanted to make his usual summer trip to Paris, where he and his family owned an apartment. So he and his wife went, and he continued chemo treatments there in the French health system. His daughter, Anya Schiffrin, wrote about the experience for Reuters news agency.

Her mother, worried that their ignorance of French might make it hard to arrange for care, soon found the French system of universal single-payer care was far more patient-friendly than the U.S. system. Her father had just the "plain-vanilla" care available to any French person; yet he was able to find just the right specialist and get care easily.

In New York, Anya and her mother had spent a full day each week in a waiting room while he waited for a blood test, then for its results, then for his chemotherapy treatment to be done. The doctor was always behind schedule. But in France the nurse came to their apartment two days before chemo to take her father's blood. No waiting room, no waiting. The hospital visit for chemo was just ninety minutes; when he got there, they were ready for him. When he later needed to see specialists, they put him in a room and had the specialists come to him, and they coordinated his care with one another, something rare in U.S. hospitals. When the chemo drug stopped working, the doctor switched to another drug that in New York Schiffrin's insurance plan had refused to approve. And when he became unable to walk, his doctor prescribed a wheelchair and sent it to his apartment at no charge. They offered to send someone to his house every day, to massage his neck to ease the pain. The forty-five-minute appointment with the doctor to plan her father's care cost just 18 euro (about $25 in 2013). Home visits several times a week began when he was too weak to come in, and continued until he passed away in December 2013.[1]

The French health care system resembles a Medicare-for-all system, a mixture of public and private, a mixture of single-payer and private care. (In a pure single-payer

[1]Anya Schiffrin, "The French Way of Cancer Treatment," Reuters, February 12, 2014, available at http://blogs.reuters.com/anya-schiffrin/2014/02/12/the-french-way-of-cancer-treatment/.

system, the government collects taxes and spends them to provide health care to everyone.) No one can be denied health care due to a pre-existing condition. The French system excels in both costs and results. In 2014, in France total health spending was 11.5 percent of GDP; in the United States, it was 17.0 percent, the highest in the world. This was $4,508 per person in international dollars in France, but more than twice as much ($9,403 per person) in the United States. The difference is nearly $5,000 per person per year. The French system avoids most avoidable hospitalizations by delivering care early, when it is needed—and is tops in the world in doing so. In the United States, the rate of avoidable hospitalizations per 100,000 of population is 2.5 times that in France, and the United States ranks nineteenth on this measure. (A high ranking, like that of France, means that few people were hospitalized who could have avoided hospitalization by getting earlier and more timely treatment).[2]

Why do Americans spend so much more on health care and get so much less? The question is well worth answering, since solving this problem and making the U.S. system work as well as the French could be worth around $5,000 per year to every man, woman, and child in the United States. But health care—to which we return shortly—is just one piece of the larger picture of the relationship of the government to the economy, the main topic of this chapter. In fact, there are really three players involved: the government, private businesses, and the people.

It is useful to compare the relationship of corporations to government with the relationship of citizens to government. There are many parallels, but also important differences. Both citizens and corporations pay taxes to government. Citizens sometimes receive transfer payments from government, such as Social Security benefits, unemployment benefits, or welfare; corporations sometimes receive subsidies. The government regulates both citizens' and corporations' activities. Both citizens and corporations seek to influence decisions taken by legislatures, government officials, or courts. Employees staffing government agencies perform services and receive wages; corporations with government contracts also perform services and receive payment. Government has all of these relationships with both human beings and firms.

However, the democratic right to vote belongs exclusively to human beings—to citizens, not to firms. Democracy is a process by which the will of the majority of citizens is expressed in government decisions. The point of democracy is that citizens elect a government, which then should do what is best for the whole people. Corporations have no direct right to democratic representation, although legally they have the right to express their views. Yet in some ways, and especially in particular sectors, over the years the government-business connection has increasingly come to dominate the triangle.

In Chapter 6 we saw that early nineteenth century corporations were originally chartered for a temporary period and expected (in principle) to act in ways that benefited, or at least did not harm, society. But over time corporations gained more autonomy. Government was in principle charged with regulating them, but corporations found it profitable to press the government to favor their interests, and backed that up with campaign donations and other rewards.

---

[2]Anne Underwood and Sarah Arnquist, "Health Care Abroad: France," *New York Times*, September 11, 2009, available at http://prescriptions.blogs.nytimes.com/2009/09/11/health-care-abroad-france/?_r=1#h%5BIqtTas,2,3%5D&mtrref=undefined&gwh=50D69DED54C1C0B540A6FF963BA24C57&gwt=pay#h[IqtTas,2,3]; and World Health Organization data at http://www.who.int/countries/usa/en/ and http://www.who.int/countries/fra/en/.

In consequence, government has often served the interests of corporations, rather than placing the public interest paramount. The term *regulatory capture* has become commonplace: The businesses that a government agency was created to regulate succeed instead in gaining effective control of that agency, so that it no longer effectively regulates them. This chapter looks at this and other examples of corporate influence over government. This main idea is expressed in the following nine points:

1. Firms sometimes can increase their profit more by spending a dollar on influencing government rules or contracting decisions than by spending a dollar on producing products and services of better quality, or of the same quality at lower cost. So increasingly they spend on influencing government. Supreme Court decisions like the 2010 *Citizens United* ruling have allowed large donations by corporations and the wealthy to influence political campaigns, accelerating this trend.

2. Over time, corporate influence has grown in three major areas: tax policy and enforcement, the regulatory regime, and government contracting.

3. Feedback loops exist at least in taxation and regulation: Firms obtain tax breaks or more favorable rules, raising their profits and the war chest they use for campaign donations, hiring lobbyists, and operating a *revolving door* (hiring the government personnel who formerly regulated them or administered their contracts). All these help firms to get still more favorable rule changes, and strengthen corporate influence on government.

4. Partly as a result of such feedback loops, the burden of taxes has continued to shift from corporations and the wealthy to working households. Meanwhile, the benefits of government spending have shifted to some extent from working households to corporations and the wealthy, partly through the channel of government contracting with private businesses. This may help explain why so many people think government is "too big": They get less than they should in services or benefits for the taxes they pay.

5. When the government contracts out to providers of goods and services, it has only weak incentives to monitor quality closely. This gives contractors the incentive to skimp on quality to save on cost and increase profit, as well as to seek more lucrative types of contracts, like "cost-plus" contracts that reimburse all costs and pay a fixed profit on those costs, or contracts that guarantee a certain level of payment to a firm whether or not it actually produces the level of services paid for.

6. As firms grow larger, bidding for government contracts becomes less competitive, with fewer bidders and more *sole-source contracts* (a contract with a single firm and no bidding process). *Bid-rigging* (illegal collusion among bidders to pre-determine the outcome of a bidding process for a contract) also is a persistent problem. Weakening of competition discourages keeping quality up and costs down.

7. Recent research shows that the policy preferences of economic elites are reflected in policies adopted by government, but that the preferences of the average citizen are not.

8. The growth of supranational institutions and rules has strengthened the influence of TNCs over governments and reduced the scope for democratic decision-making, in part through the introduction of *investor-state dispute settlement* rules in various bilateral and multilateral agreements, and in part through the rules of the World Trade Organization.

9. In principle, a truly democratic society will elect representatives who represent the interest of the majority, which means choosing rules that prevent growing inequality. In practice, increasing inequality can make the political process so heavily influenced by corporations and the wealthy that the interests of the 99 percent have a declining influence on rule-making. This can happen, for example, because representatives of the interests of the 1 percent change the rules governing the rule-making process itself.

More Spending, Less Health

We asked earlier in this chapter why the French pay about $5,000 less per person for health care than do Americans, and yet appear to get better care. But it is not just France. The United States pays more for health care, gets fewer resources for its money, and ends up with poorer results in many respects than any other country at a similar level of average income. Figure 19.1 shows the trend in U.S. health spending, public and

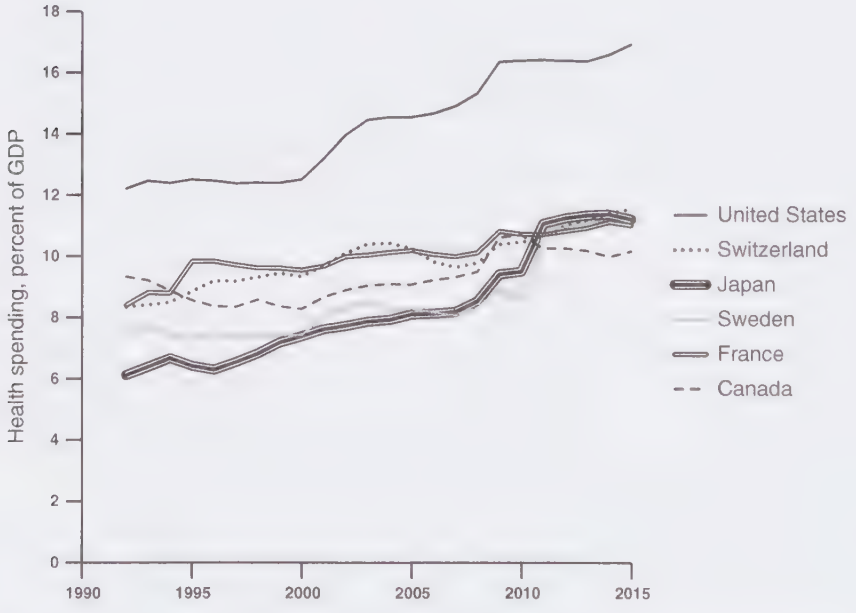


FIGURE 19.1 Health spending as percent of GDP, six countries.

The U.S. spends a far greater share of GDP on health care (including both public and private spending) than any other OECD country. In 1970 it was also tops, but only by a little, spending just over 6 percent of GDP on health. Its lead has grown, decade by decade. In 2015 the U.S. spent 16.9 percent of GDP on health, while the next four highest spending countries in 2015, as shown, all spent between 11.1 and 11.5 percent of GDP on health. In that year Canada, which is also shown, spent just 10.2 percent of GDP on health care.

Source: OECD Health Statistics 2016, Frequently Requested Data, available at http://www.oecd.org/els/health-systems/health-statistics.htm.

private combined, as a percent of GDP, compared to the countries that were the next highest spenders on health care in 2015. The gap between the United States and other countries was already substantial by 1992, but started growing rapidly after 2000.

What explains this large health spending gap? The balance of power between the public and the private sector in the United States is key to the answer. One factor is the sizeable gap between the United States and other countries in spending on pharmaceuticals per person. Figure 19.2 compares the United States with the next five biggest per capita spenders in this category among Organization for Economic Co-operation and Development (OECD) countries. The OECD is a predominantly European group, but includes former British colonies like the United States, Canada, and Australia, and a few upper-middle income developing countries like Korea and Mexico. Thus it is an excellent set of countries for comparison with the United States. As with health care overall, the gap in spending between the United States and the next-highest country has grown over time. One explanation may be that the United States in 1997 became one of the very few OECD countries that allows "direct-to-consumer" (DTC) advertising of drugs on electronic (online, cable) and broadcast (radio, television) media. Research confirms that this significantly stimulated drug sales. (If it had not, drug companies would likely not be advertising so much.)

Despite high U.S. health spending per person, the United States actually has fewer health resources of various kinds than many OECD countries. In 2012 the United States had just 2.3 physicians per 1,000 of population, fewer than most other

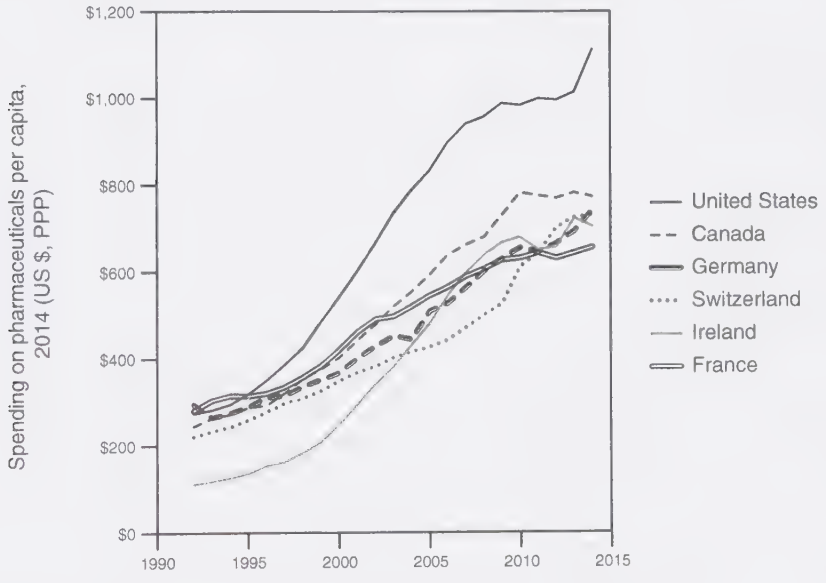


FIGURE 19.2 Expenditure on pharmaceuticals per capita, in U.S. dollars at PPP.

The six countries in which spending on pharmaceuticals per capita is highest are shown. In 1992, U.S. spending per person was about the same as most of the countries shown, except Ireland; but since 1996 it has been the highest spender. One possible contributing factor is that the U.S. is one of the few countries that has permitted direct-to-consumer advertising of pharmaceutical products on television, beginning in 1997. In 2014 the U.S. spent $1,112 per capita; Canada spent about $772, and France, the lowest one shown, spent $656.

Source: OECD Health Statistics 2016. Current expenditure on pharmaceuticals (prescribed and over-the-counter medicines) and other medical non-durables, per capita, US$ purchasing power parities (current prices, current PPPs), available at http://stats.oecd.org/Index.aspx?DataSetCode=SHA

OECD countries, and only 0.31 general practitioners per 1,000 of population, compared to 1.24 in Canada and 1.55 in France. (No wonder it is so hard to get an appointment!) The United States also had just 2.7 hospital beds per 1000 persons in 2013, well below the OECD average of 4.7. Health outcomes are also worse in the United States. Among OECD countries, sixteen have longer life expectancy, and all but three have lower infant mortality rates.

U.S. shortages of hospital beds and physicians tend to push up prices for these in the United States, which raises insurance premiums. A system that relies heavily on individual insurance, as the U.S. system does, is also necessarily more expensive, simply because it requires a labor-intensive billing process estimated to use up 15 percent of the U.S. health care dollar[3]—or about half a trillion dollars per year. If health care were simply paid for out of tax revenues and then provided free to everyone (universal single-payer health care), billing would be unnecessary. Many other countries have some version of universal health care. Yet in the 1990s, when a universal health care system (not single-payer, though) was proposed in Congress, the health care and insurance industries mobilized massively and defeated it in 1994. President Obama's Affordable Care Act (ACA) introduced some welcome changes: It included children on their parents' health insurance to age twenty-six as long as they lived at home, and it prohibited insurers from denying coverage due to preexisting conditions. But it remained more expensive than necessary, due in part to billing costs and drug costs. Senator Bernie Sanders advocated universal health care ("Medicare for all") in his presidential campaign, a demand taken up by grassroots opponents in the epic 2017 battle over health care legislation.

Government in the United States: Too Big?

Many U.S. conservatives assert that government is "too big." This often accompanies a claim that the private sector should play a greater role in the economy. Conservative Grover Norquist is famous for saying that he would like to make government small enough so he could drown it in the bathtub.

Drowning the government in the bathtub would, of course, raise the question of how most roads, bridges, ports, and airports would be built and maintained; how schools would be built and their teachers paid; whether half of seniors, lacking an old age pension, would fall back into poverty; and whether those who fall seriously ill would face destitution or premature death. Governments at all levels perform vital functions, including protecting the public from contamination of food, drugs, air, water, and soil. The private sector would only do these things if it were profitable—so would they just not be done? How well would the economy work to promote the general welfare?

Government Spending: Comparing the United States with Other Countries

It is useful to look at how big government spending actually is in the United States as a share of GDP, compared to other countries at similar income levels. The measure used in Figure 19.3 includes government spending at both federal and lower levels.

[3]Aliya Jiwani, David Himmelstein, Steffie Woolhandler, and James G. Kahn, "Billing and Insurance-Related Administrative Costs in United States Health Care: Synthesis of Micro-Costing Evidence," *BMC Health Services Research* 14 (2014): 556, available at http://www.biomedcentral.com/1472-6963/14/556

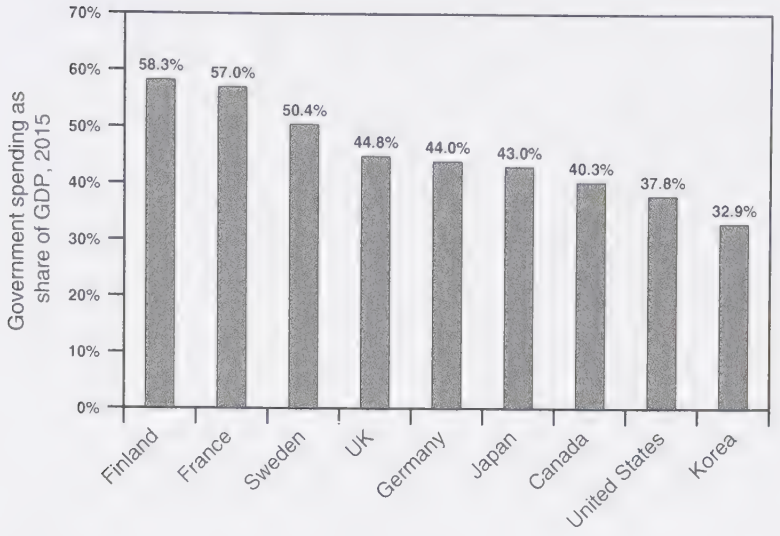


FIGURE 19.3 Government outlays as a share of GDP, 2015.

Government spending in the U.S., including federal, state, and local spending, was a lower share of GDP than in twenty-six of thirty-two OECD countries that reported this statistic for 2015. Only in Latvia, Australia, Ireland, Switzerland, and Korea was government spending a lower percent of GDP in 2015 than in the U.S.

Source: Organization for Economic Co-operation and Development, Economic Outlook 2015, Table 29: General Government Outlays as a Percent of GDP (variable YPGTQ), available at http://stats.oecd.org/Index.aspx?DataSetCode=EO#.

The federal portion of U.S. government spending is measured by *discretionary spending*, which was $1.1 trillion in 2015. Discretionary spending is funded out of general tax revenue, mainly personal and corporate income taxes, and is authorized each year in a budget bill passed by Congress. (The *nondiscretionary* part of the federal budget is Social Security, Medicare, and such long-term programs whose provisions are planned years in advance, although in principle Congress has the power to change them at any time.)

When the full list of thirty-two OECD countries that reported data in 2015 is ranked from highest to lowest share of total government spending in GDP (federal, state, and local), it turns out that the United States is twenty-seventh out of the thirty-two. Its government is not "big" compared to countries at a similar level of income. U.S. federal discretionary spending has actually been a declining share of all government spending. During World War II it was over 90 percent of all government spending (including state and local), but this fell to below 40 percent by 1996, and has mostly remained around there since.

Federal employment has also been remarkably stable. The number of civilian executive branch federal employees was actually *lower* in 2010 (2.8 million) than in 1970 (3.0 million), despite a near doubling of the U.S. population. This is due to some combination of computerization of certain functions, contracting out, and transferring some functions to state governments. Figure 19.4 shows that as a percent of all U.S. employment, federal civilian employment fell from 3.8 percent in 1970 to 2.0 percent in 2010. Over the same period, the number of uniformed military employees

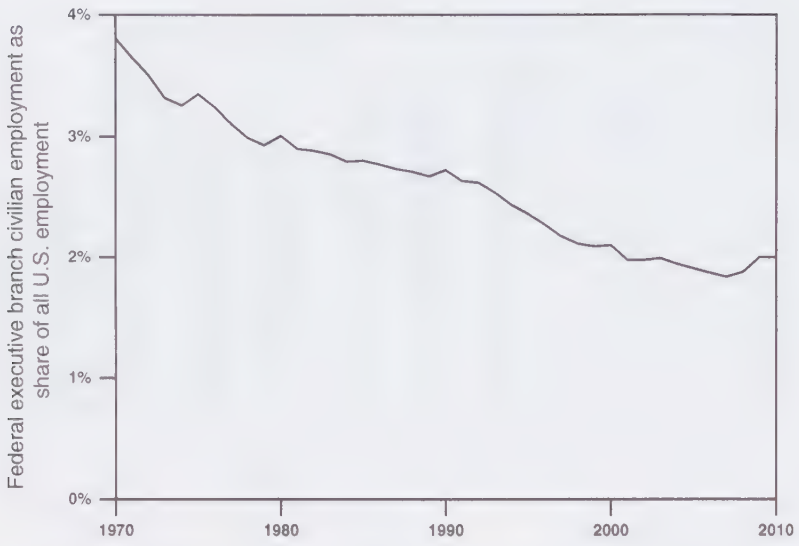


FIGURE 19.4 Federal government civilian executive branch employment as percent of total U.S. employment, 1970–2010.

The total number of civilian federal employees peaked in 1970 and 1989 at just over 3 million, but as a percent of all employment declined steadily from 3.8 percent to just 2.0 percent in 2010. One important reason for this trend is the contracting out of federal government services; another is that federal funds have been granted to state governments to operate many programs.

Source: U.S. Statistical Abstract 2012 (131st edition), Table 496: Federal Civilian Employment and Annual Payroll by Branch: 1970 to 2010. Available at https://www.census.gov/library/publications/2011/compendia/statab/131ed/federal-govt-finances-employment.html

also fell, particularly over the 1990s, and stayed fairly low due to contracting out of many military functions in the Iraq and Afghanistan wars.

U.S. government spending patterns differ from those in other countries. Notably, U.S. military spending as a share of GDP exceeds that in any other OECD country except Israel. Figure 19.5 also shows that, beginning in World War II, U.S. military spending has never fallen below 60 percent of the discretionary federal budget.

In some countries taxes are high and spending on social services is high, but because the spending is on programs that bring obvious and fairly generous benefits to families, there is less complaining about the level of taxation than there is in the United States. The large U.S. military budget of over $0.7 trillion[4] in 2016 left only about $0.4 trillion in discretionary spending for federal government services.

Old-Age Pensions

One important benefit is old-age pensions. OECD data show that the size of benefits varies quite a bit from one country to another; the United States offers lower pensions than do most countries with comparable income levels. Figure 19.6 shows, for a male worker making the median earnings, what percent of earnings is replaced by

[4]This figure is from the Bureau of Economic Analysis; other sources may count the military budget a little differently, but the trend is still similar.

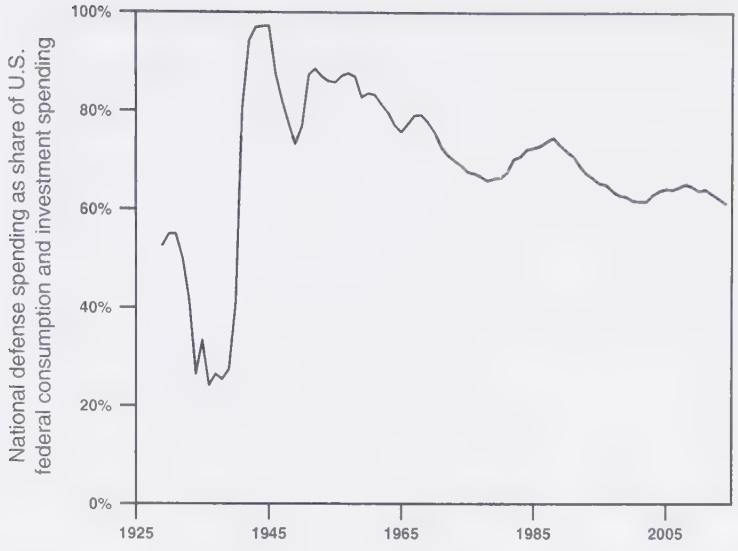


FIGURE 19.5 U.S. National defense spending as a percent of federal consumption and investment spending, 1929–2014.

The share of U.S. defense spending in all federal discretionary spending was 97 percent in 1943–1944, and since World War II this share has been consistently greater than 60 percent. In 2014, federal consumption and investment spending was $1,220 billion and national defense spending was $748 billion, according to the U.S. Bureau of Economic Analysis. The ratio of the second to the first was 61 percent. These categories differ somewhat from the categories presented in the official federal budget.

Source: U.S. Bureau of Economic Analysis, National Income and Product Accounts, Table 1.1.5: Gross Domestic Product (national defense spending divided by all federal consumption and investment spending), available at http://www.bea.gov.

pension payments. (A person whose annual earnings are at the median earns less than the pay of the highest 49.9 percent of workers, and more than the pay of the lowest 49.9 percent, roughly speaking.) For the United States in 2013, just 41 percent of earnings were replaced, while for the *average* OECD country 57.9 percent of earnings were replaced, and for a few countries, including the Netherlands, Israel, and Denmark, that number was over 80 percent. In other words, a retiree who had been making $50,000 a year in wages in the United States would on average receive just 41 percent of that ($20,500) per year in social security benefits. In the Netherlands the pension would be $45,700.

People in the United States pay somewhat less of their income as taxes than elsewhere, but the tangible benefits they get for that money are less than what residents of many European countries get. This may help to explain why so many people think government is "too big" in the United States: They perceive that they get less than they should in services or benefits for the taxes they pay. This could be because (1) working households contribute an increasing share of all income taxes (individual and corporate), while corporations' share has fallen; (2) the rules governing the economy have tended to shift increasingly in favor of what is profitable for business rather than what promotes the general welfare; and (3) when government contracts out to businesses for services, it often pays more than it would if the government itself produced the services.

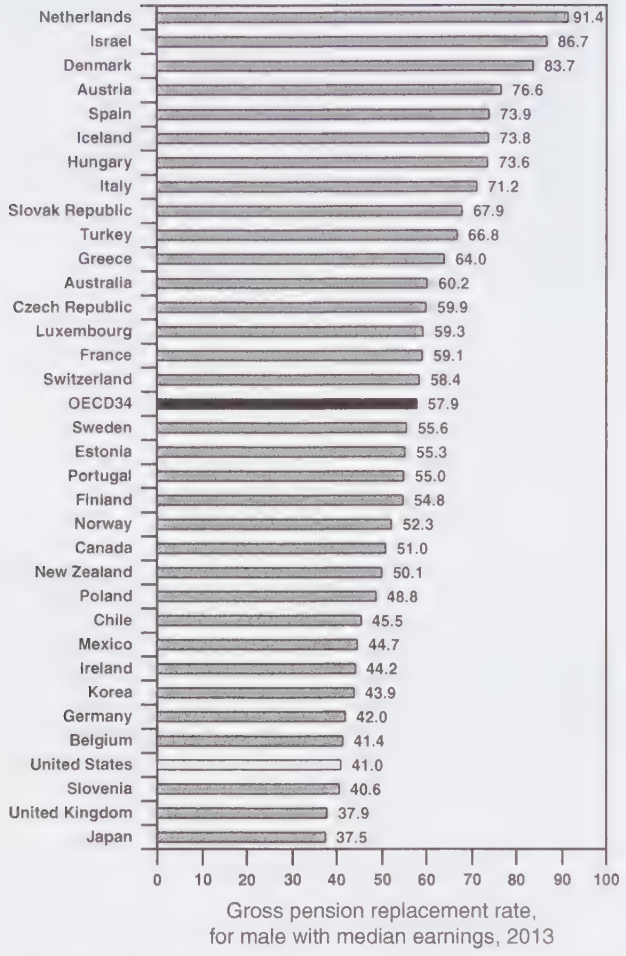


FIGURE 19.6 Percent of earnings replaced by pension, OECD countries, 2013.

In the average OECD country, the old-age pension replaced 57.9 percent of the earnings of a male with median earnings in 2013. In the United States, it replaced 41.0 percent. Only three countries (Slovenia, the UK, and Japan) replaced a smaller portion of earnings.

Source: Pensions at a Glance 2013: OECD and G20 Indicators, OECD Publishing, Paris, Chapter 4, Table 4.1. Gross Pension Replacement Rates by Earnings, available at http://dx.doi.org/10.1787/pension_glance-2013-en. Data table alone is available at http://dx.doi.org/10.1787/888932907224

Influencing the Rules

If businesses succeed in influencing the rules under which an industry operates, they can often increase their profitability. Recall from Chapter 6 that Congress passed the law creating Medicare, Part D (the drug insurance portion of Medicare) in 2003 with a provision forbidding the federal government from bargaining with drug companies for discount prices. Under previous laws, several federal agencies had already established a discounted drug price list (the Federal Supply Schedule), from which the Veterans Administration, the Department of Defense, the Public Health Service, the Bureau of Prisons, and others could purchase drugs. This could have been extended to Medicare, but was not. The prices the VA was paying, for example, were just 58 percent of the

prices paid under Medicare Part D. One researcher estimated that Medicare could save about $20 billion annually if allowed to negotiate discounted drug prices.[5]

To get these results, the drug industry used not only lobbying and campaign donations, but also the revolving door by which government officials are hired by industry, and industry executives are appointed to government positions. After the 2003 vote, Congressman Billy Tauzin, credited with getting the bill passed, was hired as the chief lobbyist for the pharmaceutical industry in 2004, a job said to carry a salary of about $2 million.[6] Such tactics have long been successfully used by several industries, including the arms industry. While there are laws designed to prevent such revolving doors, they typically do no more than create a short waiting period between leaving a government job and carrying out the full range of lobbying activities for private business.

Likewise, some sectors have been quite successful in getting their executives appointed to government positions—particularly the financial sector, as described in Chapter 18. The traffic goes both ways, and sometimes the same person moves from government to industry and back to government. Dick Cheney was a congressman, then rose to secretary of defense during 1989–1993, then became CEO of Halliburton Corporation, then vice president under President George W. Bush. During his 2001–2009 term, Halliburton's subsidiary KBR became the largest government contractor ($40 billion) for the Iraq and Afghanistan wars.[7] The revolving door between the private and public sectors is one way that the interests of corporations have come increasingly to influence government decision-making. (Halliburton sold off KBR in 2008, after government auditors had noted numerous problems with KBR's performance.)

Lobbying, even apart from the revolving door, has also been an effective strategy, which helps to explain why the total amount reportedly spent on registered lobbyists more than doubled from $1.5 billion in 1998 to $3.5 billion in 2009, thereafter falling back to $3.2 billion in 2015.

As for spending on federal election campaigns, the cost of presidential and congressional races combined more than doubled from $3.1 billion in 2000 to $6.3 billion in 2012, and continued rising thereafter. Contributions from business to these campaigns rose from $1.2 billion in 2000 to $2.7 billion in 2012.[8] (None of these figures is adjusted for inflation.) The Supreme Court's *Citizens United* decision helped make it possible for this figure to grow more rapidly.

Some business sectors spend far more than others on lobbying, campaign donations, and setting into motion the revolving door. In addition to the pharmaceutical industry, the financial sector is also a big spender. It has reaped substantial profits from changes in rules such as the repeal of the Glass-Steagall Act, the prohibition on regulation of derivatives, and the bailout from the subprime mortgage crisis. The military goods and services industry is another big spender on lobbying, as shown in Figure 19.7.

---

[5]Walid F. Gellad et al., "What if the Federal Government Negotiated Pharmaceutical Prices for Seniors? An Estimate of National Savings," *Journal of General Internal Medicine* 23, no. 9 (June 2008): 1435–40, available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2517993/pdf/11606_2008_Article_689.pdf or DOI: 10.1007/s11606-008-0689-7.

[6]"The Drug Lobby Scores Again," *New York Times* editorial, December 17, 2004.

[7]Commission on Wartime Contracting in Iraq and Afghanistan, *Final Report to Congress*, August 2011, available at https://cybercemetery.unt.edu/archive/cwc/20110929213815/http://www.wartimecontracting.gov/.

[8]Center for Responsive Politics, available at https://www.opensecrets.org/lobby, https://www.opensecrets.org/overview/cost.php, and https://www.opensecrets.org /overview/blio.php.

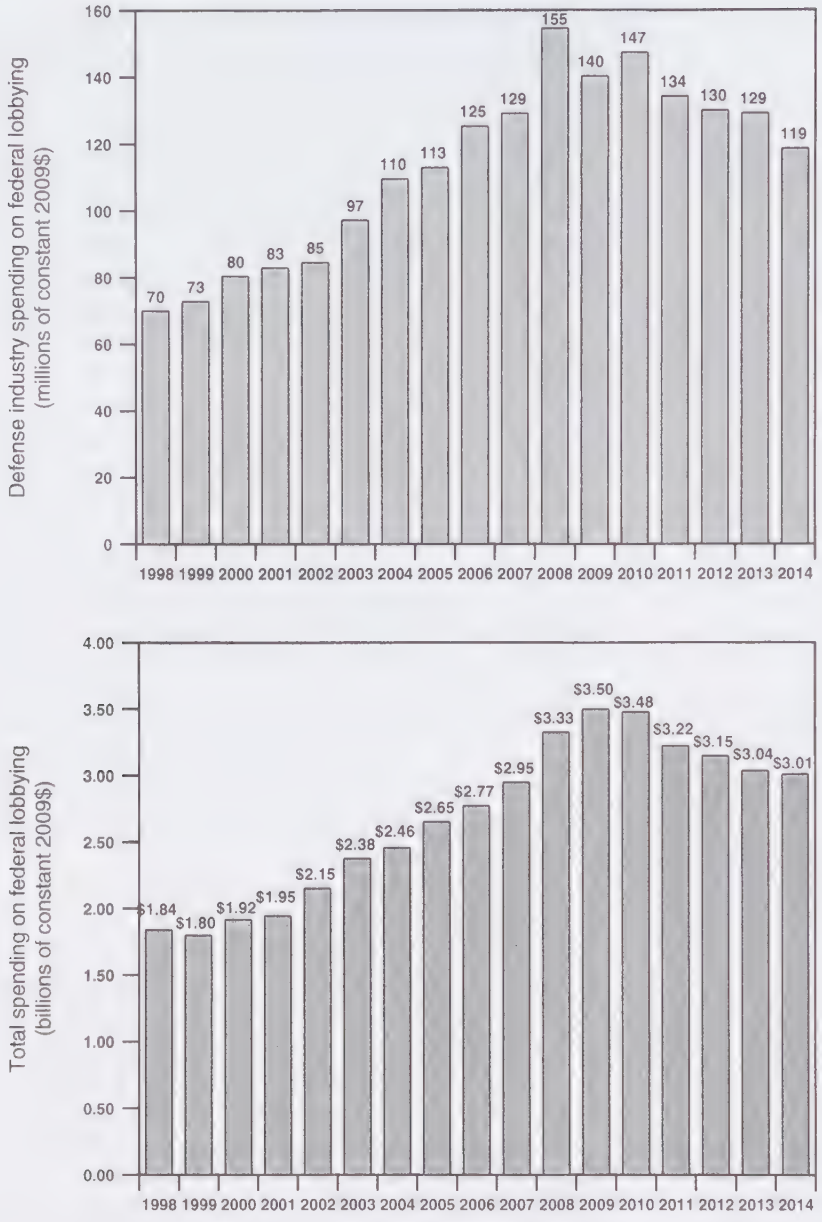


FIGURE 19.7 Annual spending on federal lobbying, defense industry (top) and total (bottom), 1998–2014.

The Center for Responsive Politics compiles these data, based on the information that lobbyists have been required to submit to the U.S. Senate.

Source: Center for Responsive Politics, Lobbying Database, available at https://www.opensecrets.org/lobby/.

Taxation

Taxation is another area of government rule-making in which corporate interests have gained influence since World War II, as shown by the declining share of federal corporate income tax revenue in all federal income tax revenue. This has been driven in part by the rise in corporate use of tax havens to avoid taxes, in part by other

devices for reducing tax liability, and in part by a decline in IRS efforts to audit the largest businesses and those most likely to owe additional taxes.

During World War II, corporate income tax revenues were briefly more than 60 percent of all income tax revenues (corporate plus individual). Just after the war they declined to about a 40 percent share, and then fell steadily over time, until during the three decades from 1983 through 2015 they averaged just 18 percent. (Taken as a percent of *all* federal tax revenue, including Social Security payroll taxes and the like, they are usually in the single digits.)

One reason for this decline is the fall in corporate income tax rates in recent decades, although personal income tax rates on the highest incomes also fell to some extent at the same time. Federal discretionary spending of just over $1 trillion is funded by general tax revenues, which come mainly from three kinds of taxes: personal income taxes, capital gains taxes (which are reported in IRS data as part of personal income taxes), and corporate income taxes.

TAX RATES AND TAX PAYMENTS TO THE FEDERAL GOVERNMENT Income tax rates in the United States started out low more than a century ago, rose during wartime and the Great Depression, and in recent decades were reduced for corporations, high-income taxpayers, and capital gains, though generally not quite to their lowest historical levels in 1913 or the 1920s.

A low personal income tax was imposed briefly during the nineteenth century, and then was declared unconstitutional by the Supreme Court in 1895. In 1913 a constitutional amendment legalized income taxes, and in that year the personal income tax was set at 7 percent for the highest earners. With the need to raise funds for World War I, the highest marginal tax rate (for the highest earners, on the last dollars of income) quickly rose to 77 percent in 1918, and remained nearly that high through 1921. By the late 1920s it was back down to about 25 percent. It rose again during the Great Depression, and again during World War II: The highest tax rate imposed on the highest income bracket reported by high-income taxpayers reached 94 percent in 1944–1945 and was still over 90 percent during 1951–1964. Thereafter it was lowered in several stages from 70 percent to 50 percent, and after 1987 it dropped below 40 percent.

The corporation income tax rate and the capital gains tax rate were also higher during and after World War II and are now also lower than they have been since World War II. During 1913–1921 there was no separate capital gains tax; it was the same as the top personal income tax rate, and peaked at 77 percent in 1918. In 1921 it was made a separate tax and reduced to 12 percent; thereafter it was typically under, and usually well under, half the top personal income tax rate. From 2003 on the capital gains tax rate was 15 percent for most taxpayers, lower than at any time since the 1930s.

The corporation income tax was initially the lowest of the three, and then in 1922–1933 was close to the capital gains tax rate at about 12 percent. It was raised during the Great Recession, but only gradually. The top rate was still 40 percent or lower during most of the 1940s, less than half the top personal income tax rate. From the 1950s through most of the 1970s the top corporate income tax rate was close to 50 percent, and then was reduced to about 35 percent, where it remained from the late 1980s to 2016.

In other words, all three tax rates were initially low, higher in wartime, and relatively high from around the 1930s or 1940s through about the 1970s. The 1980s began a period of substantial decline in all three tax rates, with some variation. All

three taxes are now about as low as they have been over the last century, except that in the late 1920s all three were briefly lower than their recent levels.

TAX RULES AND ENFORCEMENT. The statutory tax rates alone do not tell the whole story. Tax rules and enforcement are also key. During the last half-century, tax accountants and attorneys have invented numerous ways to avoid or reduce income taxes, both personal and corporate. A variety of *tax shelters*—certain types of schemes for reducing taxes—use several kinds of legal strategies to lower income earners' tax liabilities. In addition, Congress has in recent decades reduced IRS funding and personnel, despite the fact that doing so increases the government deficit because it results in losses in tax revenue that exceed than the savings in IRS budget dollars.

As for corporate income taxes, numerous arrangements for tax avoidance exist and are widely used by corporations. (Similar tax avoidance strategies exist for those individuals wealthy enough to afford the services of a wealth manager or specialized tax accountant or attorney. The word *tax evasion* is used for illegal ways to pay lower taxes; *tax avoidance* is the word for methods that are legal.) Many corporations take advantage of the existence of *tax havens*. For example, a pharmaceutical company can set up an affiliate in a tax haven and assign ownership of its drug patents to that affiliate; the affiliate then charges the U.S. firm a substantial annual royalty fee for use of the patents, and the U.S. firm deducts the fee from its taxable income, reducing the amount of taxes it owes. The fee that flows to the tax haven affiliate is taxed there at a low, often zero, level. By shifting its taxable income to a tax haven in this way, the TNC reduces its total tax payments and so increases its total global after-tax profit. The trend shown in Figure 7.3 for U.S. corporations' international profit to rise as a share of all their profit partly reflects movement of production and partly deliberate profit-shifting abroad as part of a strategy of tax avoidance.

Another tax avoidance strategy is for a U.S. transnational corporation to put its funds into a tax haven affiliate, which then lends back the funds to the U.S. firm, creating annual interest payments that it can deduct from its U.S. taxable income. This is known as *earnings stripping*, and is another way to reduce U.S. tax liability. Payment of U.S. taxes on income earned abroad and kept abroad are also deferred indefinitely, reducing taxes paid, if not taxes ultimately owed. Later in this chapter we see how U.S.-based TNCs have tried to leverage their unrepatriated profits to try to get Congress to pass tax holidays, with some success. Figure 19.8 shows that due to very effective use of such methods and others, the decline in the ratio of corporate to individual federal income tax revenues has been dramatic.

WEAKENING ENFORCEMENT In addition, IRS enforcement efforts have been weakened over several decades. Some of this is a result of laws passed by Congress that placed limits on IRS activities, for example in the 1990s. Some of it is an increasing unwillingness to uncover and pursue tax fraud, because of pressure from many directions: some members of Congress, some wealthy taxpayers, and some organized pressure groups.

In recent decades the IRS has changed its enforcement of the tax laws, auditing high-income taxpayers less and middle- or lower-income taxpayers more.[9] In addition, IRS funding has been substantially reduced, and partly as a result, audits of large corporations have been relaxed. This is true although, at the current level of IRS spending, every dollar spent by the agency produces much more than a dollar

[9]Charles Lewis, Bill Allison, and the Center for Public Integrity, *The Cheating of America* (New York: William Morrow, 2001).

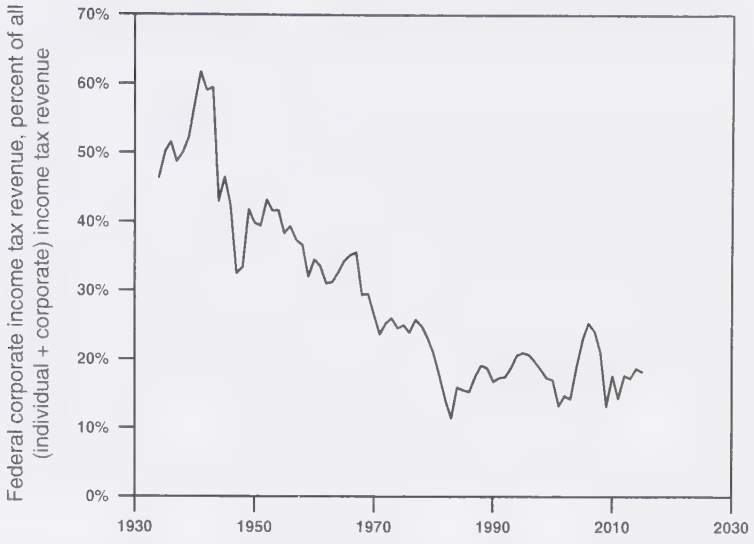


FIGURE 19.8 Federal corporate income tax revenue as a share of all federal income tax revenue (individual and corporate), 1934–2015.

Source: U.S. Office of Management and Budget, Historical Tables, Table 2.2: Percentage Composition of Receipts by Source: 1934–2021, available at https://www.whitehouse.gov/omb/budget/Historicals.

of additional tax revenue, so that funding cuts actually worsen the budget deficit. In September 2015, due to budget cuts, the IRS had 27 percent fewer revenue agents than it had in 2010. (It is revenue agents who audit complicated business returns.) IRS Commissioner John Koskinen told the non-partisan Tax Policy Center in 2016 that the personnel cuts brought a loss in tax revenue of $7–8 billion for 2010–2015, far more than the salaries of the approximately four thousand lost agents.[10] If Congress were to set IRS funding based on maximizing additional tax revenue per dollar of IRS spending, it would increase the IRS budget up to the point that a dollar spent on enforcement resulted in a dollar of additional tax revenue.

The IRS also reduced audits of big corporations (with at least $250 million in assets) and giant corporations (at least $20 billion in assets). In fiscal year (FY) 2015 revenue agents spent 34 percent fewer hours auditing big corporations than in FY 2010, and the recommended additional tax to be levied on them dropped from $23.7 billion to $8.5 billion. As for giant corporations, audit time spent by revenue agents dropped by 47 percent in the same period, and recommended additional taxes fell by 71 percent.

The IRS budget and personnel cuts that Congress mandated were so extraordinary that in November 2015, seven former IRS Commissioners whose terms had spanned from 1961 to 2002 wrote a joint letter expressing concern at the extent of the cuts, and called for restoring the IRS budget.[11] Yet cutbacks in IRS audits of both big corporations and the wealthy had already begun by the 1990s, partly due to congressional and public pressure, and partly due to a new law that made it harder for the IRS to investigate and prosecute those taxpayers (corporate or individual) most likely to yield much additional tax revenue because they have considerable income.[12]

---

[10] "IRS Auditing of Big Corporations Plummets: Potential Annual Revenue Loss $15 Billion," TRACIRS, Syracuse University, March 15, 2016, available at http://trac.syr.edu/tracirs/latest/416/.
[11] "IRS Auditing of Big Corporations."
[12] David Cay Johnston, *Perfectly Legal* (New York: Penguin Group, 2003).

Rules that Affect Business Costs and Profits

Businesses not only seek to influence government to reduce the taxes they pay; they also want the rules that affect their costs and conditions of doing business to be as favorable to their profitability as possible. Some business executives oppose increasing the minimum wage, prefer not to be compelled to install pollution-control equipment, and would like not to face class-action lawsuits, discrimination complaints, health and safety inspections, the requirement of environmental impact statements, and so on. Yet all these are the result of rules that protect the public against real dangers. There is unavoidable constant contention over what rules will govern the relationship between business and government, and how strongly those rules will be enforced. In a later section we discuss further the kinds of rules that affect particular sectors. Apart from seeking to influence rules, another way in which some businesses hope to influence government is that they seek to become government contractors, a topic to which we turn next.

Contracting Out Government Services

A government provides many services directly, when its agencies hire employees to offer them. But governments at all levels also contract out to private businesses for some goods and some services. A town government may contract with a business to remove snow from the streets in winter. State or federal agencies may contract with a firm to answer phone calls from the public asking basic first-inquiry questions about their services.

On the other hand, contracting out can be on a far more sophisticated level and require technically skilled personnel: U.S. intelligence agencies regularly contract out intelligence activities, and over 400,000 intelligence contractors have top secret security clearances (as did whistleblower Edward Snowden). The Department of Defense (DOD) expects that contracts for development, production, and maintenance of the new F-35 Joint Strike Fighter (with Lockheed Martin and various other firms) will be around $1 trillion over the lifetime of the plane. Senator John McCain, in a November 2016 letter, referred to "the troubled performance, continued delays, and persistent cost overruns of this program," and the U.S. Government Accountability Office in 2016 judged that it still faced "affordability and oversight challenges."[13] The GAO report urged that in view of this, Congress should choose contractual arrangements for future aspects of the program in which more oversight, rather than less, would be required. But all such legislation is subject to lobbying, and lobbying for DOD contracts typically means pointing out to members of Congress which jobs in their districts will be affected if contracts are not approved. Usually such lobbying is remarkably effective.

In the 2014 fiscal year (Oct. 2013–Sept. 2014) the Department of Defense entered into contracts for $285 billion, an amount that was well over a third of the sum total of DOD spending that year, and was 8 percent of all federal spending. This was also nearly two-thirds of the total $445 billion in contracts entered into by all federal

[13]Ana Radelat, "Pentagon-Lockheed Fight over F-35 Price Doesn't Extend to Pratt," CTMirror, November 4, 2016, available at http://ctmirror.org/2016/11/04/pentagon-lockheed-f-35-contract-fight-doesnt-extend-to-pratt/; U.S. Government Accountability Office, "F-35 Joint Strike Fighter: Continued Oversight Needed as Program Plans to Begin to Develop New Capabilities," GAO-16-390, April 2016, available at http://www.gao.gov.

government agencies and recorded in the government's system for keeping track of these contracts.

Contracting for military services in the context of fighting wars is particularly difficult. There is often pressure to give sole-source or no-bid contracts to speed up the process of acquiring the services or goods. But without competitive bidding, prices for services can become highly inflated. And where costs are difficult to define in advance, contracts are often structured as "cost-plus" contracts: The contractor reports costs and is reimbursed for them, plus a fixed amount of profit over and above cost. Such arrangements reportedly led in the case of the Iraq war to contractors inflating costs by training employees at lavish resorts, with high food and recreational expenses; leasing vehicles at a price that cost far more than buying them; charging $100 for laundering one person's bag of laundry; and junking costly vehicles and equipment for minor defects.[14]

From 2002 to 2011, the DOD paid contractors about $200 billion for all kinds of services in U.S. military operations in Iraq and Afghanistan. In 2011, the final report of the official Commission on Wartime Contracting in Iraq and Afghanistan found that there was "massive waste, fraud, and abuse" by some contractors, conservatively estimated at $31–60 billion over the ten-year period, or 15 to 30 percent of the entire amount spent.[15]

There are sometimes good reasons for government to contract out goods or services. However, there are also pitfalls. There is no guarantee of saving money through contracting: intelligence contractors reportedly cost about twice as much as government employees, and military contract employees, at least those who are U.S. citizens, have typically been paid far more than uniformed soldiers.

If problems in contracting are more noticeable in defense procurement, they nevertheless affect all kinds of government contracting, because the incentives created by selling to the government are in many cases fundamentally problematic.

Government Contracting: Profits and the Public Interest

Decades ago, on the Bay Area Rapid Transit system in the San Francisco Bay area, there was an ongoing problem with vandalism on the trains that carried passengers all over the region. Seats were slashed from time to time, and a private contractor was paid to repair them. But there was an odd consistency to the method and timing of the slashing. Eventually the security force became suspicious, increased its surveillance, and—you guessed it—confirmed their suspicion that the seat-slashers were sent by the very same contractor who was being paid to do the repairs. This true story points to the perverse incentives built into some kinds of government contracting.

In most cases the role of the profit motive in inflating government spending is not quite this obvious. There would be no problem if government contractors only maximized profits by operating more efficiently, using technology to good effect, and motivating and training workers well. But contractors sometimes find ways to maximize profits that turn out to be inconsistent with the public interest.

One key issue is competition. In a truly competitive market, with many small firms, one way to maximize profit in the short term is to produce the same product as other firms, at the lowest possible cost. But a firm's profit will never be large until it distinguishes itself from other producers somehow. It can do so by making a higher

---

[14] Robert Greenwald, *Iraq for Sale: The War Profiteers*, video, 2006.

[15] Commission on Wartime Contracting, *Transforming Wartime Contracting: Controlling Costs, Reducing Risks*, Final Report to Congress, August 2011, available at www.wartimecontracting.gov; p. 5.

quality or more desirable product or service, or finding a unique way to cut cost without cutting quality. Alternatively, it can do so by entering a less competitive niche market: selling the product to the government, for example, under an initial contract, and then lobbying the government officials likely to decide whether the contract will be renewed. It can seek to expand its business by encouraging legislation or court decisions that open up greater or more lucrative opportunities for whatever it is selling. That is just what the pharmaceutical industry did with the Medicare Part D legislation forbidding the federal government from negotiating with drug companies for discount prices.

Decline in Competitive Bidding for Government Contracts

At least in European countries, no-bid and one-bid government contracts have become more frequent in recent years. As both the size of firms and the size of contracts offered for bid has risen, there has been an increase in the dollar value of contracts that are awarded after the government receives just one bid. The absence of competition makes it hard to ensure the quality of goods or services and hard to control cost.

In data based on government contracts over a certain threshold amount, the share of such contracts that only received a single bid doubled in Italy from 15 percent of the total in 2006 to 30 percent in 2015, and similarly at least doubled in Spain, Belgium, France, and the UK in the same period, ending up in 2015 between 10 percent and 23 percent in these four countries.[16] Data show that larger tenders do get fewer bids, on average, and the bidders on larger tenders are usually larger firms, which are better equipped to handle them.

Similar concerns exist in the United States. Many of the problematic contracts identified by the Commission on Wartime Contracting in 2011 were no-bid contracts, never put up for competitive bidding in the first place. Yet even when the urgency of the goods or services needed does not initially allow time for competitive bidding, the government can (and often should) structure the contract so that it is opened soon afterward to competitive bidding to ensure future quality performance at reasonable cost. In 2011, the Department of Defense gave $115 billion in sole-source contracts, the large majority of all federal no-bid contracts.

Similar issues arise in government contracting with private businesses to operate prisons: maximizing profit can conflict with serving the public interest.

Government Contracting for Private Prison Services

Government decisions to contract out prison management to private firms have generated considerable controversy in recent decades. In the last several decades the United States has been unique among the countries of the world in the high percentage of its population that is incarcerated. It is also unusual in the percentage of them who are held in privately run prisons or detention centers.

Forty years ago the United States did not stand out in this respect; the startlingly rapid rise in incarceration and in contracting out of prison services took place mostly from the 1980s through about 2008. From 1925 to 1972, the population of state and

[16]"Government Contracting Is Growing Less Competitive, and Often More Corrupt," *The Economist*, November 19, 2016.

federal prisoners was fairly steady at about 1 for every thousand of adult population, a rate roughly similar to that in other countries. But thereafter, both the number of state and federal prisoners and the total number of persons incarcerated (including in local jails and in detention centers) grew steadily and rapidly. This happened despite the fact that a long-term downward trend in the rates of both property crimes and violent crimes began in the early to mid-1990s, and by 2015 reached levels (measured at the rate per 100,000 of population) so low that they had not been observed since the 1960s.

The main reason for the increase in incarceration was not any rise in crime, but rather an increase in the number of crimes that resulted in prison terms, along with a large increase in the length of those terms. This came from a dramatic change in the rules that governed sentencing decisions, as well as a change in other rules, programs, and priorities that determined both sentences and police priorities. These changes were introduced as part of the "War on Drugs" beginning in the 1980s.

Laws and Norms On Sentencing

Court decisions and new laws both drove these changes. In the early and mid-1980s, the Supreme Court, in two cases of first-time drug offenses involving selling cocaine, upheld very long sentences (forty years in one case, life in the other). The 1984 federal Sentencing Reform Act was passed, with a provision often referred to as "truth in sentencing" mandating that federal prisoners serve at least 85 percent of their sentence before becoming eligible for release. Many states passed similar "truth in sentencing" laws applying to convictions under state law. Parole for federal prisoners was also ended in 1987, lengthening incarceration and increasing the number of inmates at any one time.[17]

In 1986 the Anti-Drug Abuse Act passed by Congress established mandatory minimum sentences of five or ten years for relatively minor federal drug offenses such as selling small amounts of cocaine, and for possession of crack cocaine, although in most high-income countries the sentence for such crimes is less than a year. Similarly, "Three strikes and you're out" laws caused some to receive harsh mandatory sentences of several decades to life after conviction for a third crime that by itself would likely have led to a sentence of at most a few years.[18]

The three changes together increased the total number of prisoners and greatly increased the length of time each prisoner was locked up, so that states began to run out of prisons before they were able to build more. This allowed private prison operators both to increase their volume of business and to maintain high and relatively predictable capacity utilization.[19] Figure 19.9 shows the jump in average federal prison time from 1988 to 2012, comparing the total time served by those released in each of the two years. Over all categories of offenses shown, average prison time more than doubled, from 17.9 to 37.5 months. Over those years, the number incarcerated in federal prisons increased at a rate that closely matched the rate of increase in time served.

In 2014, the total number of persons incarcerated was 2.3 million, including 1,351,000 in state prisons, 211,000 in federal prisons, 646,000 in local jails, and about 90,000 in a wide variety of other types of institutions, including 33,000 in immigration detention centers. The majority of those in local jails had not

---

[17]Michelle Alexander, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*, revised edition (New York: The New Press, 2012).

[18]American Civil Liberties Union, *Banking on Bondage: Private Prisons and Mass Incarceration*, November 2, 2011, available at https://www.aclu.org/banking-bondage-private-prisons-and-mass-incarceration.

[19]Alexander, *The New Jim Crow*. This is the main source for the information in the first three paragraphs in this section.

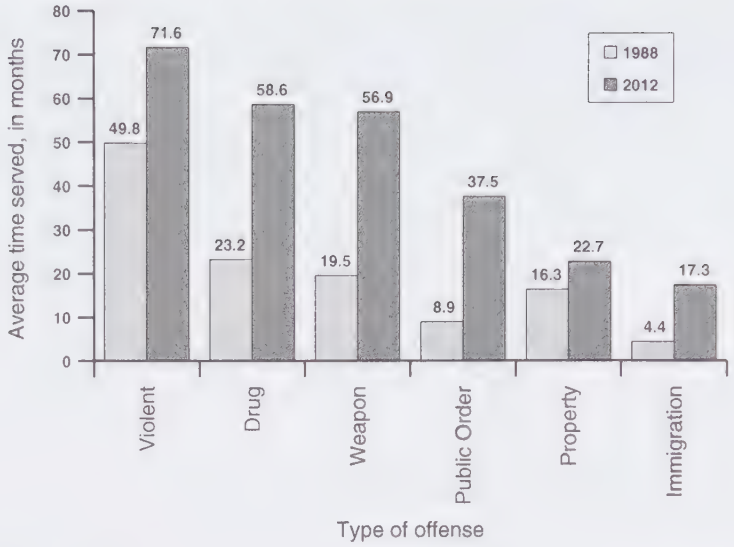


FIGURE 19.9 Increase in average federal prison time served, 1988–2012.

The Federal Bureau of Prisons reports that the length of time actually spent in federal prison for those released in 2012 was on average twice what it was for prisoners released in 1988. For all categories of offenses combined, this number rose from an average of 17.9 months in 1988 to 37.5 months in 2012. In four of the six categories (drug, weapon, public order, and immigration) the length of incarceration was from 2.5 to 4.2 times as great for those released in 2012 as in 1988; for those convicted of violent crimes and property crimes, it less than doubled between the two years. Included in "Public order" offenses are violations of tax law, racketeering, extortion, bribery, perjury, and more.

Source: "Prison Time Surges for Federal Inmates: Average period of confinement doubles, costing taxpayers $2.7 billion a year." Issue Brief, November 15, 2015. Public Safety Performance Project, Pew Trusts, available at http://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2015/11/prison-time-surges-for-federal-inmates. Based on Bureau of Justice data sources.

yet been convicted, and some were being held just for small fines that they were unable to pay.[20]

This total number incarcerated was nearly ten out of every thousand of the adult U.S. population, the highest incarceration rate in the world. If instead we count only the number of *prisoners*, referring to those held in state or federal prisons designed for those persons already sentenced to terms of one year or more, then that number was 1,561,000 in 2014, or about seven per thousand of adult population.

The number of prisoners in for-profit prisons grew rapidly from 69,000 in 1999 to a high of 137,200 in 2012; in this peak year, nearly 9 percent of all prisoners were in private facilities. This includes immigration detention centers: half of all immigration detainees were in centers run by private companies that contracted with the federal government.

An important effect of the drug war measures has been that many African-Americans, as well as Latinos, have been categorized as felons (even if their crimes were nonviolent), and so in many states have been prevented from voting. State laws vary about whether felons can vote even after they have served their time. (In fact,

[20]Peter Wagner and Bernadette Rabuy, "Mass Incarceration: The Whole Pie 2016," Prison Policy Initiative, available at http://www.prisonpolicy.org/reports/pie2016.html.

some states also prevent *anyone* in prison from voting, even if convicted of a misdemeanor.) In 2015, one in thirteen African-American adults were not legally able to vote because of these provisions. And because the large majority of those incarcerated are men, the percentage of African-American men who were denied the vote was even higher. Those who are not allowed to vote are ordinarily not allowed to serve on juries, either.[21]

Rethinking Private Prisons and Mass Incarceration

In the last decade or so, both state and federal governments have become concerned about the rising cost of housing so many prisoners. Increasingly they became aware of research findings that releasing many nonviolent offenders earlier would likely do nothing to harm public safety, and could even increase it by easing reintegration of released prisoners into the workforce and therefore into society. In some states this has been done quite successfully, as the Public Safety Performance Project of the Pew Center for the States has pointed out.[22]

By 2007, several states had made successful efforts to begin reducing their prison populations. Some states modified the legal requirement that prisoners serve 85 percent of their sentences, for example permitting exceptions that allowed officials to use their judgment to reduce sentences in individual cases. State prisoner populations leveled off at about 1.4 million during 2007–2010 and then began to decline. Federal prison populations peaked in 2012–2013 at about 220,000 and also slowly began falling, but in 2013, about 30,000 federal prisoners were still housed in private, for-profit prisons, a number that had been rising since 2003.[23]

In August 2016 the Department of Justice (DOJ) declared that the federal prison system should reduce its contracts with private prison corporations. After measures such as reducing sentences for low-level nonviolent drug offenders had brought down the total number of prisoners from 220,000 in 2013 to 195,000 in 2016, the DOJ had decided to minimize use of for-profit prisons to house federal prisoners. Privately run prisons, it declared, neither offered the same level of correctional services, nor the same level of safety and security, as government-run prisons. They lacked programs of education, job training, and rehabilitation that helped inmates reintegrate into society after release—programs that were well developed in government prisons—and so failed to maintain public safety as well as government prisons do. They also had higher rates of assault than government-run prisons.[24]

The DOJ's policy change came without warning, and so sparked a sudden drop in the stock prices of major correctional services providers Correction Corporation of American (CCA) and the GEO Group. But the action it took was fairly mild. No contracts would be prematurely terminated, but contracts were scaled back, and some were not to be renewed. By May 1, 2017 the number of privately held federal prisoners was to be halved compared to 2013. Staff were urged to consider further

---

[21]Katie Rose Quandt, "1 in 13 African-American Adults Prohibited from Voting in the United States," March 24, 2015, available at *billmoyers.com/2015/03/24/felon-disenfranchisement/.*

[22]See http://www.pewtrusts.org/en/projects/public-safety-performance-project/research-and-analysis for numerous such reports.

[23]Sally Q. Yates, Deputy Attorney General, "Phasing Out Our Use of Private Prisons," Department of Justice, August 18, 2016, available at https://www.justice.gov/opa/blog/phasing-out-our-use-private-prisons.

[24]U.S. Department of Justice, Office of the Inspector General, "Review of the Federal Bureau of Prisons' Monitoring of Contract Prisons," August 2016, available at https://oig.justice.gov/reports/2016/e1606.pdf.

reductions. The mandate only directly affected a small segment of the private prison "market"—federal prisoners—but could have had indirect effects on other segments.

The issues the DOJ raised are pervasive in all types of private prisons. Looking carefully at the way private prison operators function tells us what kinds of problems arise and why. Many of the principles that apply to contracting out of prison management apply more generally to contracting out of all sorts of services, and even for contracting out of some goods production, such as by the military goods and services industry.

Selling to Government, Maximizing Profit

When conventional economics textbooks describe how suppliers and demanders behave, they are mostly talking about businesses selling to households. They say little about businesses selling to governments. Yet government contracting is big business—$445 billion at the federal level in 2014 alone in the United States, plus contracting out by state and local governments.

When businesses sell to governments, they have the same profit motive, driven by the same determinants of total profit R, as described in Chapter 10. But when firms seek contracts with government, they may adopt other strategies to influence these determinants. In trying to increase total output (Z), advertising matters less; and lobbying, campaign donations, and favorable think tank reports matter much more. A business can increase the quantity of output sold by getting more contracts, or by getting more favorable contracts. An example later in this chapter describes how some private prison operators have won favorable contracts that reduce their risk by guaranteeing their revenues even when they house few prisoners.

A for-profit business that sells correctional services is like any other for-profit business. It maximizes profit by minimizing costs, raising its price when this will increase profit, and trying to increase its capacity utilization—the number of prisoners it houses at a point in time—by trying to keep every prison bed filled in an existing physical facility that it runs. These are the same ways to maximize profit that we saw in Chapters 10–12, using the determinants of the profit rate.

One difference is that government contractors are selling their services not to an individual household or business that will carefully inspect what they are buying, but to a government that is unlikely to scrutinize its "product" as closely as would other kinds of buyers. This tends to be true with many kinds of government services, but with correctional services, it is partly because the government is not the recipient of the services; the prisoners are. This creates both an information problem and a motivation problem.

The information problem for the government is that prison services are delivered to prisoners away from the public eye. Those who know most about what happens there are the prisoners themselves (and the staff, who may hesitate to reveal deficiencies for fear of losing their jobs). Prisoners have good information about whether the contractor is delivering services of the appropriate quantity and quality. But prisoners have little direct voice or power over whether the private prison operator's contract is renewed, or even terminated prematurely, or whether the prison operator is publicly reprimanded for providing inadequate services.

Prisoners do usually have some indirect voice, however, such as the right to submit grievances, a right that they sometimes exercise. Grievance records are available to government investigators. If questioned directly by a government monitoring team, prisoners may be able to reveal problems; but such investigations may tend to

happen only intermittently. In principle (if prisoners are able to get an attorney) they also have the right to sue. In some cases such lawsuits have had a significant impact.

The power that inmates do sometimes exercise is the power to rebel ("riot")—but often at great cost to themselves, as some are usually injured, sometimes seriously, or even killed. Insurrection captures the attention of the public and of government officials, often provoking an investigation into prison conditions. If prisoners rebel because of bad or insufficient food, or inadequate medical care, the prison operator may be called to task; possibly its stock price will fall. If prisoners with acute illness are found not to have been given access to emergency medical services, or prisoners with chronic conditions have not been given adequate, consistent care, this can in principle help to cause nonrenewal of contracts.

But apart from dramatic events, or revelations by enterprising reporters or advocacy groups, government monitoring often falls short of what is needed to alert authorities to mismanagement, or to cause them to take action to correct deficiencies. There are exceptions, to be sure; for instance, an unannounced official inspection of Ohio's Lake Erie Correctional Institution in January 2013, one year after the Correctional Corporation of America bought it from the state of Ohio to run it as a private prison, garnered some attention and led to the making of a short documentary film about the problems there.[25] In follow-up inspections the situation was found to have somewhat improved.

Thus the information problem for the government is that the private prison's operations are mainly known only to the prison staff and the prisoners. The government may do assessments, whether regular or only occasional. But when the whole purpose of contracting out was to save money, more monitoring (at government expense) will reduce these savings, and may make contracting out a more expensive option than performing the services in-house.

For-profit prisons are typically paid a fixed fee per inmate per day, and their profit depends on several factors. One is the extent to which they can keep all the beds filled with inmates. A second is the level of the fixed fee for immigration detainees, private companies typically receive a higher daily fee than for other kinds of incarceration, so profits have generally been higher for these facilities. A third factor is what services the correctional facility provides; if the for-profit prison provides fewer rehabilitation services, educational programs, and the like, its profits will be higher. A fourth factor is how much it pays its workers and how much it pays for the food it provides. For-profit prisons are sometimes understaffed, and there is often very high turnover so that most staff get paid low entry-level wages.

A fifth factor in determining a contractor's profit is how much money it can make from the prisoners themselves. One way is to arrange for them to work either in the prison or on a work-release program. Pay rates are far below the minimum wage, and in some states work can be compulsory but entirely unpaid. Often the prison company deducts a sizeable portion of this pay as reimbursement for the cost of housing and feeding them. It is quite common for them to end up being paid less (often much less) than a dollar an hour. In fall 2016, a loosely coordinated series of prisoner strikes in a number of states raised the issue of low wages, among others. Another way for operators of private prisons to increase profit is to arrange for a

---

[25]Craig Knowles, *Prisons for Profit*, video, ACLU of Ohio, 2013, available at http://www.acluohio.org/prisons-for-profit. See also Correctional Institution Inspection Committee, Report on the Inspection and Evaluation of the Lake Erie Correctional Institution, January 22–23, 2013, available at http://www.acluohio.org/wp-content/uploads/2013/02/2013_0123LakeErieCorrectionalInstitutionFullReport.pdf.

phone company to provide a pay phone at very high cost, and then require the company to pay the prison management firm a cut of the phone revenues.

Summarizing these options in the terms of the analysis in Chapter 10 of the determinants of the profit rate, to maximize profit firms may raise P_z or Z or both, or lower P_m or M or both. (As discussed in Chapters 10 and 11, it is not always advantageous to raise P_z, but there exist situations in which it can raise profits.) The other possibility is to gain revenue streams from other sources, like the phone revenues or a share of the prisoners' pay for work.

Medical Services in Private Prisons

Among cost factors, a quite important and potentially large one is the quality and quantity of medical care that prisoners receive. Medical care costs include both materials costs P_mM (equipment, supplies) and labor costs wH (regular medical staff at the prison such as a nurse or physician, plus the labor cost of emergency services).

Many private prison operators begin with a cost advantage over government-run prisons, in that they often secure in their contract a provision that they will not house prisoners with exceptional medical needs.[26] This allows them to avoid costs that government prisons then will have to assume. It also means that any claim that private prisons cost less per prisoner than government-run prisons must take account of the greater responsibilities that government prisons bear.

Numerous reports of poor management or outright abuse in medical services in for-profit prisons have appeared, beginning at least as early as the 1990s. Prisoners are supposed to receive appropriate medical care provided at the expense of their caretakers. Because prisoners tend to have more chronic illness, more serious illness, more substance abuse problems, and other conditions, they need to have these conditions identified and an appropriate plan of monitoring and treatment provided. But in a number of reported cases, for-profit prisons have delayed or denied emergency medical care to prisoners who clearly needed them, causing them in some cases to die of treatable illnesses. They also failed sometimes to give proper or consistent medical care to prisoners they knew had chronic conditions, some of whom died as well.

Seeking to maximize profit does tend to cause such cost-cutting strategies. It has been common for private prisons to save money by hiring medical personnel with too low or narrow a level of knowledge and experience to handle life-threatening emergencies, or to handle mental health issues. For example, the largest for-profit prison operator, Corrections Corporation of America (CCA), ran a work farm in Tennessee. A licensed practical nurse with a year of training, the only medical staff on hand when a pregnant prisoner reported being in terrible pain, refused to send her to the emergency room. She was only permitted by CCA officials to go to the hospital twelve hours later—too late, since she died from the undiagnosed complications of her pregnancy.[27] (CCA was renamed CoreCivic in 2016.)

The 2016 decision by the Department of Justice to reduce contracts with private prisons was supported by a report released by the Inspector General of the Department.

---

[26]Katie Rose Quandt, "Why There's an Even Larger Racial Disparity in Private Prisons Than in Public Ones," *Mother Jones*, Feb. 17, 2014, available at http://www.motherjones.com/mojo/2014/01/even-larger-racial-disparity-private-prisons-public-prisons.

[27]E. Bates, "Prisons for Profit," *The Nation* (January 5, 1998): 13–14, cited in Elizabeth Alexander, "Private Prisons and Health Care: The HMO from Hell," in *Capitalist Punishment*, ed. Alan Coyle, Allison Campbell, and Rodney Neufeld (Atlanta, GA: Clarity Press, Inc., and London: Zed Press), 67–74.

The report reviewed the DOJ's monitoring procedures and the data it used to assess private prison contractors' performance. It found that these private businesses cut corners by providing too little medical care, hiring staff with a low level of training and experience, providing few rehabilitation programs to help prisoners reintegrate into society, and in general delivering lower quality services than did government prisons.[28]

Changing the Rules

As we have seen, changing government rules (laws, regulations, policies, and programs) is one strategy for firms to maximize profits. Private correctional services companies benefited from three major changes of rules beginning in the 1980s: mandatory minimum sentencing laws, "truth in sentencing" laws, and "three strikes and you're out" laws, which together made it easier for private prison operators both to keep beds filled, and to predict how many beds would be filled.[29] High capacity utilization is one way to keep profits high.

RULES WRITTEN INTO CONTRACTS In recent years another new type of contract provision has gained popularity and helped solve the capacity utilization problem: the "occupancy guarantee" or "lockup quota." A 2013 study by the advocacy group In the Public Interest found that over half the private prison contracts they reviewed had a clause guaranteeing that a given percent of beds in the particular facility would be filled at all times; the usual guaranteed minimum was 90 percent or higher, with some contracts specifying 100 percent.[30] This means that government-run prisons will send enough prisoners to the private prison to fill the quota—or if the government fails to fill those beds, the private prison company will be paid exactly as if the beds had been filled.

In the case of one immigrant detention center, the South Texas Family Residential Center, a four-year contract guaranteed payment as if 100 percent of beds were filled. In effect, CCA was receiving a fixed monthly fee for operating the facility. So in early 2016, even though the 2,400 beds in the facility were only half occupied, the Corrections Corporation of America (CCA) was getting paid for 2,400 inmates. No wonder, then, that in 2015 CCA made 14 percent of all its U.S. profits from that one single facility out of the seventy-four facilities it operated.[31] Perhaps the only thing that can be said for such an arrangement is that it reduces the incentive for CCA to lobby for laws that detain more immigrants, or detain them for longer periods; with such a contract, CCA evidently makes more money, the fewer detainees it houses. (However, it still has the incentive to build additional detention centers, if it can get contracts for them.) But government-run detention centers would accomplish the same thing: they too lack the incentive to encourage more and longer detentions; contracting out has no advantage in this regard.

---

[28]U.S. Department of Justice, Office of the Inspector General, "Review of the Federal Bureau of Prisons' Monitoring of Contract Prisons," August 2016, available at https://oig.justice.gov/reports/2016/e1606.pdf.
[29]American Civil Liberties Union, *Banking on Bondage: Private Prisons and Mass Incarceration*, November 2, 2011, available at https://www.aclu.org/banking-bondage-private-prisons-and-mass-incarceration.
[30]In the Public Interest, "Criminal: How Lockup Quotas and 'Low-Crime Taxes' Guarantee Profits for Private Prison Corporations," September 2013, available at https://www.inthepublicinterest.org/wp-content/uploads/Criminal-Lockup-Quota-Report.pdf.
[31]Chico Harlan, "Inside the Administration's $1 Billion Deal to Detain Central American Asylum Seekers," *Washington Post*, August 14, 2016, available at https://www.washingtonpost.com/business/economy/inside-the-administrations-1-billion-deal-to-detain-central-american-asylum-seekers/2016/08/14/e47f1960-5819-11e6-9aee-8075993d73a2_story.html?utm_term=.7c0fe92aa743.

In short, contracting out private prisons has been found to produce inferior correctional services at excessive prices, with no clear benefit, and quite possibly significant harm, to the public.

Taxing Transnational Corporations: Who Is In Control?

It is often hard to know which is more in control: government or corporations. Just as government contractors gain control through lobbying, campaign donations, and the revolving door, and private prison operators sometimes appear to be gaining control over the government officials who are supposed to be monitoring them, so too we find similar trends in the international arena.

In 2004, Congress passed the American Jobs Creation Act, with provisions intended to promote investment in productive activities in the United States to create more jobs. One provision declared a tax holiday to encourage transnational corporations to repatriate the profit earnings that they had been holding abroad to continue deferring paying taxes on these funds. Unfortunately, the tax holiday actually ended up *reducing* government revenues over the next decade and failing to produce investment or job creation.

Recall from earlier in this chapter that corporations with operations outside the United States must declare each year how much taxes they owe on their foreign earnings, but can defer paying taxes on these earnings until they return that money to the United States as dividends. As long as the money is held abroad, it can be reinvested abroad and payment of the taxes owed can be put off indefinitely. Citizens for Tax Justice estimated in 2016 that at least $2.4 trillion of such earnings were being held abroad, and that the deferred unpaid taxes amounted to around $700 billion.[32] For the year in which corporations took advantage of the tax holiday (2005, mostly), the law reduced every dollar of tax owed on earnings repatriated in that year to just 15 cents. Part of legislators' expectations were fulfilled, and hundreds of billions of dollars were indeed repatriated. But neither the investment nor the jobs materialized. According to detailed studies published in 2011 by the Senate Permanent Subcommittee on Investigations and the congressional Joint Committee on Taxation,[33] large firms did not use the money to invest or create jobs, but instead laid off tens of thousands of workers, and spent money buying back their own stock.

As we saw in Chapter 7, when a corporation buys back some of its own stock, its major stockholders, such as the CEO and other executives, earn substantial capital gains. Figure 19.10 shows the dollar amount of U.S. stock buybacks in each year from 2002 to 2007 by the eighteen firms that repatriated the most profits from abroad under the tax holiday. In 2005 there was a rise in stock buybacks which continued through 2007. Procter & Gamble, IBM, Hewlett-Packard, and Pfizer together spent $40 billion on stock buybacks in 2006 alone. Figure 19.10 also shows the rise from

[32]Citizens for Tax Justice, "Fortune 500 Companies Hold a Record $2.4 Trillion Offshore," March 3, 2016, available at http://www.*ctj.org/pdf/pre0316.pdf.*

[33]"Repatriating Offshore Funds: Tax Windfall for Select Multinationals," United States Senate Permanent Subcommittee on Investigations, Committee on Homeland Security and Governmental Affairs, Majority Staff Report, December 14, 2011, available at http://www.gpo.gov/fdsys/pkg/CPRT-11SPRT70710/pdf/CPRT-112SPRT70710.pdf; and Letter from Thomas A. Barthold, Chief of Staff, U.S. Congress Joint Committee on Taxation, to U.S. Representative Lloyd Doggett (4/15/2011), available at http://doggett.house.gov/images/pdf/jct_repatriation_score.pdf.

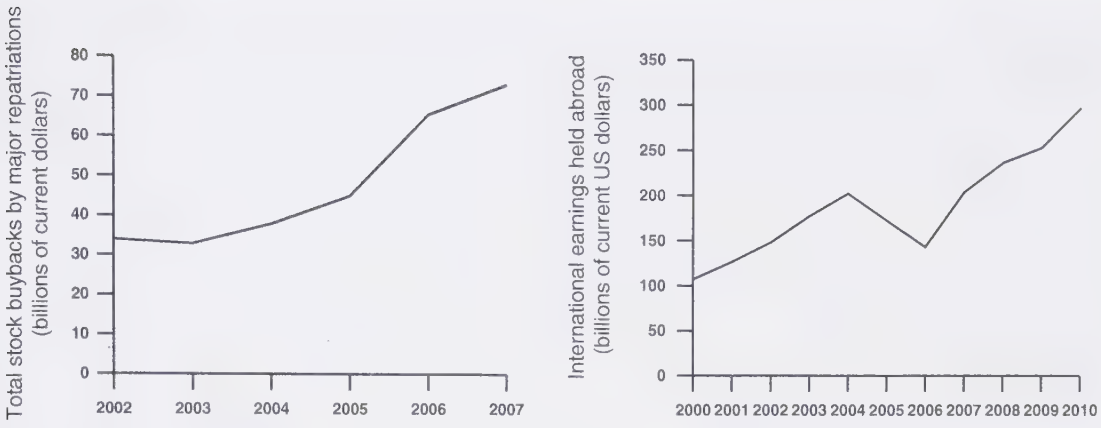


FIGURE 19.10 Transnational corporations' response to the tax holiday in 2005.

The federal American Jobs Creation Act of 2004 allowed U.S. transnational firms that were holding unrepatriated profit earnings abroad to bring those funds back to the U.S. and pay 15 percent of the taxes they owed. Congress thought the money would be used for investment and job creation. Instead, many firms laid off workers and, as the left hand panel shows, spent large amounts to buy back their own stock. Afterward, U.S. TNCs increased the dollar amount of their global profit earnings that they held abroad, as the right hand panel shows.

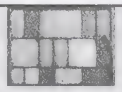
Sources: U.S. Senate Permanent Subcommittee on Investigations, Committee on Homeland Security and Governmental Affairs, Majority Staff Report, "Repatriating Offshore Funds: Tax Windfall for Select Multinationals," United States Permanent Subcommittee on Investigations, Committee on Homeland Security and Governmental Affairs, Majority Staff Report, October 11, 2011, available at http://www.gpo.gov/fdsys/pkg/CPRT-112SPRT70710/pdf/CPRT-112SPRT70710.pdf

before to after the tax holiday in profits earned abroad each year that this group of firms kept abroad.

A congressional Joint Committee on Taxation study found that this later rise in unrepatriated profits abroad, and the related tax revenue deferred, more than offset the 2005 gains in revenue—producing a net revenue loss from the tax holiday. TNCs may well have figured that parking more profits abroad would increase their leverage with U.S. lawmakers and enable them to get additional tax holidays later on. But "Fool me once, shame on you; fool me twice, shame on me." When a new tax holiday was proposed in 2014, the fact that the 2004 tax holiday had not fulfilled its promise of increased investment, jobs, or even tax revenue, was sufficient to discourage Congress from approving a new tax holiday.

Democracy and Inequality

For democracy to work well, government must respond to the views of most of society's citizens. But when the power and influence of owners of capital (the wealthy) grow large, and the power and influence of average citizens shrinks, democracy works less well. The box "Democracy Disconnect" describes research that finds that the average person's opinion has no discernible effect on policy-making, while elite opinion closely resembles the policies that are actually chosen.



Democracy Disconnect

Does the opinion of Joe or Josephine Public influence legislation?

Hardly at all, says a study by Princeton University political scientists Martin Gilens and Benjamin Page published in 2014. But a much higher-income group's opinion does count.

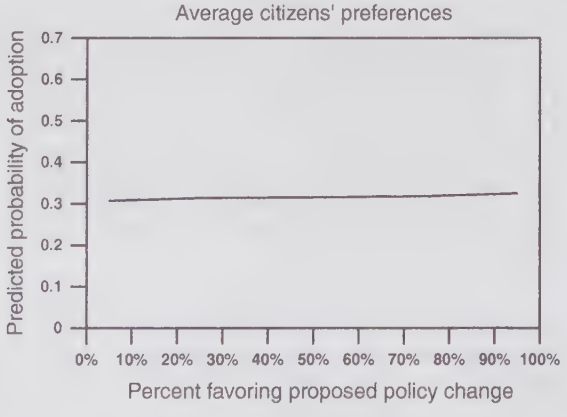
The pair of researchers used Gilens' database on 1,779 different public issues for which opinion polls existed between 1981 and 2002 in which people were asked both their views and their income levels, and in which interviewees were asked to take a clear-cut yes or no position on the specific policy change being asked about. Gilens and Page and their research assistants then ascertained, for each of the 1779 issues, whether the policy change actually happened or not during the several years following the opinion poll.

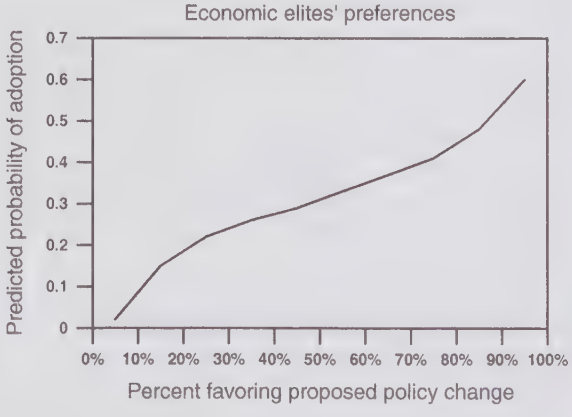
The question they asked of the data focused on two groups: the "average person," with median income (in the middle, when the interviewees were ranked by income) and "the economic elite," those at the 90th percentile (incomes higher than 90 percent of all those interviewed, and below 10 percent of them).

Their statistical analysis asked two key questions. The first was, "How does the percent of *average citizens* that approve a policy affect the likelihood that it will become law?" Their statistical analysis showed that the answer was, "Hardly at all." The second question was, "How did the percent of *economic elites* that approved a new policy affect the likelihood that it would become law?" They found that the answer to the second question was, "The greater the percent of economic elites that approved a policy, the more likely it was to be adopted."

The left-hand graph below shows their finding that if 5 percent of average persons preferred a policy change, that change had about a 30 percent chance of being adopted—but if a large majority of average citizens preferred a policy change, it had almost exactly that same 30 percent chance of being adopted. A policy embraced by a high percentage of average-income persons was no more likely to be adopted than one favored by few such persons.

The right-hand graph shows that for economic elites (the 90th income percentile) the situation was very different. Proposals that only 5 percent of elites favored were infrequently adopted, while proposals that 95 percent of economic elites favored had about a 60 percent change of being adopted. So either economic elites' views directly influenced policy makers, or policy makers tended to see the world in the same way as economic elites.





Source: Martin Gilens and Benjamin Page, "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens," *Perspectives on Politics* 12, no. 3 (2014): 564–81.

Feedback Loops Between Inequality and Power

Inequality has risen so much in the United States that we are back to its level in the 1920s. Such high levels of inequality are bound to weaken democracy. The box "Democracy Disconnect" shows the result, but what exactly are the dynamics? How did we get to this point? Two big factors are who pays the taxes and who gets the benefits of government spending.

The tax burden has shifted from corporations and wealthy households toward middle class and poor households. This is partly a consequence of the enormous reduction in progressivity of the income tax system. During World War II and for several decades afterward, the incomes of everyone advanced together. But as the rich increasingly moved their wealth and income to tax havens, they paid less tax. The tax on increase in wealth through realized capital gains was lowered to 15 percent, and then hedge fund managers' incomes, by a twist of definition, came to be viewed legally as capital gains. Thus these managers, some with annual incomes of a billion dollars or more, pay a lower percent of their income to the IRS than do ordinary middle class people. Or, as wealthy Leona Helmsley famously told a woman who worked for her in the 1980s, "We don't pay taxes; only the little people pay taxes." Schemes for reducing taxes proliferated, and the IRS reduced its audits of corporations and the wealthy, as we have seen.

The rise in inequality has both made it harder for the majority to assert their views, and easier for corporations and the wealthy to assert theirs. The average household has fewer resources than in the past: real wages have stagnated since the 1970s, and both median income and median net worth fell sharply as a result of the Great Recession of 2008–2009. Not only do many people have less money; many people also have less free time than in the past, on average, because given their lower wages, many of them have to try to work extra hours to make a living. The rich and the owners of capital now have more money to spend on putting forward their viewpoint and donating to candidates who will do the same.

Supranational Rule Changes

Corporations have also gained leverage over national governments through international agreements among countries. So far, when we have talked about government, we have been referring to federal, state, and local government. But although there is no world government as such, there have long been international agreements that restrict what national governments can do, and international organizations that implement such agreements. Formation of the World Trade Organization in 1995 in particular allowed corporations to limit foreign governments' regulation of their investment activities, making them freer to seek maximum profits with less regard for the public interest than in the past. International agreements have also sometimes offered a way for foreign corporations to exert influence on the U.S. government in ways that sidestepped longstanding democratic processes and benefited U.S. corporations.

We already know from Chapter 15 that the intellectual property rights provisions in the World Trade Organization agreement have increased the profits of pharmaceutical companies and other firms, both inside and outside the United States. TNCs have also benefited from investment provisions in the WTO restricting the requirements that national governments could impose on foreign investors. The

Brazilian government successfully created an auto industry in the 1950s by prohibiting imports of foreign cars and so persuading auto makers to locate plants in Brazil; today, such strategies are outlawed by the WTO.

TNC Power over Foreign Government Decisions

One important kind of rule giving corporations leverage over governments has been the *investor-state dispute settlement* (ISDS) procedures written into many trade and investment agreements starting in the late 1980s. Such rules are best explained by an example.

The United States and Mexico are both signatories to the North American Free Trade Agreement (NAFTA). Chapter 11 of NAFTA (its ISDS provision) allows a foreign investor to file a complaint and seek damages against a government of a country in which it has an investment, if that host government takes an action that is "tantamount to expropriation" of the investor's property.

Shortly after NAFTA went into effect in 1994, the U.S. company Metalclad, which had bought a toxic waste dump in a town in the San Luis Potosí district of Mexico and planned to operate it. Metalclad began construction on the site although the town government refused to grant it a permit. The town's demand was that before getting a construction permit, Metalclad had first to clean up the leaking toxic waste containers left in the dump by its previous owners, since the leaking chemicals were blamed for a sharp increase in cancer, skin rashes, and other illnesses in the town.[34]

Instead of cleaning up the mess, Metalclad sued Mexico under Chapter 11, demanding compensation for the future profits it lost when the permit was denied. Chapter 11 mandated that such a case be judged by an arbitration panel chosen from a list of designated "international experts." The procedure was entirely outside national court systems; proceedings were not public, and the public had no right to see the hearing record, nor to file an *amicus* ("Friend of the Court") brief. In the end Mexico was compelled to pay $16 million to Metalclad.

In another case, the government of Canada paid a $13 million settlement to U.S. chemical manufacturer Ethyl Corporation, whose fuel additive Canada had decided to ban due to its environmental toxicity. The ban would have reduced the firm's future profits, and Canada settled, and removed the ban, in order to avoid going through the hearing process, although convinced that the ban was needed to protect public health. While settlements were small in these two cases, they have been larger in some other ISDS cases under various agreements.

The ISDS concept originated in the old legal principle of *eminent domain*: that a government has the right to expropriate property for public purposes as long as it pays the property owner fair compensation. To build a public road a government can seize private land that lies in the path of the road, provides it pays fair compensation. Courts may adjust the payment upward if the landowner proves that it is below the property's value.

Nationalization, or expropriation, falls into the same category: It is legal, as long as it is for a public purpose and fair compensation is paid. Mexico expropriated American oil companies' operations in 1938; and after World War II the UK nationalized coal companies and France nationalized the Renault auto company. Compensation was paid. Yet in some such cases firms have expressed outrage, and in at least

[34] Bill Moyers, *Trading Democracy*, video, 2002.

one case (Guatemala) a government that expropriated the property of a U.S. business (United Fruit Company) was shortly afterward overthrown by a U.S.-backed coup. Firms have sought increasing protection against expropriation.

But ISDS clauses now found in many trade and investment agreements go further. First, "tantamount to expropriation" has been interpreted to mean that not only physical or financial assets, but also an intangible—a firm's future profits—is property, protected from reduction in value caused by a government decision. The historical origin of this idea evidently lies in Canada's announcement in 1980 of a National Energy Programme that, among other things, informed oil firms drilling in a certain territory in eastern Canada that if they discovered oil they would have to turn 25 percent of it over to PetroCanada, a "crown corporation" (analogous to a state-owned enterprise). This announcement enraged the mostly American oil companies that were already drilling in that territory, because it reduced the value of their enterprises. They fought fiercely—and successfully—to reverse the policy, and the United States then insisted on putting an ISDS clause into the Canada–U.S. Free Trade Agreement of 1988. That clause was copied into the North American Free Trade Agreement (NAFTA) with both Canada and Mexico, and in essence into agreements with other countries.[35]

The second innovation was that only *foreign* investors—investors from one country party to a particular trade/investment agreement, investing in another country also party to that agreement—had the right to complain under the ISDS clause.

The third was that the venue in which the case was to be heard was completely separate from the host country's judicial system, and operated without considering possible conflicts of interests of hearing panel members, without reference to any domestic legal principles, without requiring or permitting input from any source other than the complainant and the host country government, and without review by the country's highest court.

ISDS provisions in trade and investment agreements have thus given transnational corporations a new tool allowing them in some cases to nullify or prevent rule changes that cut into their profits. Governments that face complaints sometimes simply concede without a fight, and the public hears little or nothing about it. Corporations have used ISDS to try beat back many government policies that benefit the public, including environmental protections, local and state laws requiring "green" business practices, government-subsidized package delivery, laws requiring labeling cigarettes with the harm caused by smoking, and more. Democracy thus suffers.

Conflict over Rules

Rules decided by the legislative, judicial, and executive branches of government shape what capitalist enterprises can and cannot do, as well as what workers and citizens can and cannot do. As we have seen, this means that capitalists have in principle a strong reason to seek political influence over the rules that are adopted. Most business enterprises, especially the largest ones, find it useful to spend effort and funds trying to influence the rules in ways that increase their profitability. Their methods have included lobbying, campaign donations, and the revolving door between employment in government and in business, as well as supporting conservative think tanks to produce reports in support of rules that increase corporate

[35]Stephen Clarkson, *Canada and the Reagan Challenge: Crisis and Adjustment, 1981–85*, new updated edition (Toronto: James Lorimer & Company, 1985).

after-tax profits, intensive media campaigns to sell their ideas, and other means. In this chapter we have seen examples in pharmaceuticals, arms production and contracting, taxation, and the prison industry, as well as examples both at the national and the supranational level. Wealthy individuals similarly seek rules that minimize the taxes they owe, protect their investments, and enhance their opportunities to expand their wealth. So it is that wealth confers power, including power to shape the rules governing the economy; and likewise, power confers wealth. This tends to create growing inequality in both wealth and power.

Yet while such tendencies can be strong, they are not inevitable or irreversible. At times a broad consensus arises against them, or those within the structures of power act conscientiously to curb abuses or excesses. Certain types of defense contracting were reduced after the devastating report of the 2011 Commission on Wartime Contracting mentioned earlier in this chapter. Some state governments also reversed the trend toward increasing incarceration in the last decade, and were able to achieve reductions in their prison budgets. The federal prison system also began reducing the number incarcerated, mainly by releasing nonviolent offenders. On the other hand, it is possible for these reverses themselves to be rolled back when an election changes those in control of the government.

Naturally, it is not only owners of capital who have motives to influence the rule-making process. The 99 percent similarly have strong reasons to mobilize and exert political pressure in favor of rules (as well as appointments to public office) that will sustain and enhance health and education, and provide income support during old age, disability, or hard times; clean air and water; labor rights, including overtime pay, paid sick leave, paid vacations, paid leave for childbirth, occupational health and safety; and other ways to bring better lives for all.

In 2017 there was extraordinarily widespread protest against changes of rules in a direction that the majority judged would not be to their benefit, such as repeal of the Affordable Care Act, as well as reductions in Medicaid, Medicare, and Social Security.

The various problems with contemporary capitalism just discussed have led many to ask: Can the rules governing the economic system be changed so that we obtain more good outcomes and fewer bad ones? For instance, could we introduce changes that would allow us to continue to enjoy the high productivity and diversity of goods that capitalism creates, but not suffer its harsh inequality and environmental destruction? Quite simply, what would be a better set of rules than those under which the economy operates now?

Proposals for Alternative Rules

In recent years at least three distinguished economists have written books calling for an overhaul of the rules governing the U.S. economy, and for adopting new rules that would reverse the trend toward growing inequality of wealth and power and share the benefits of economic growth more widely.

The book *Rewriting the Rules of the American Economy*, by Nobel Laureate and Columbia University professor Joseph E. Stiglitz (and other authors associated with the Roosevelt Institute), raises such issues and proposes alternative rules. The authors write:

> Today's inequality is not the inevitable result of the evolution of capitalism. Instead, the rules that govern the economy got us here. And we can change those rules . . . (Stiglitz et al. 2015, x)

They add that inequality is a choice: many other countries that, like the United States, have capitalist economies and some form of democracy have more social benefits and less inequality than does the United States. The trend toward greater inequality in the United States should therefore be reversible through the democratic process.

These authors propose, for instance, to re-regulate the financial sector, both to reduce the chance of another disastrous financial crisis like that of 2008 and to level the playing field for smaller financial firms. The authors also propose raising the marginal income tax rate on the highest incomes. They call for taxing corporations on their global income, which might eliminate tax avoidance through tax havens and cause corporations to pay their fair share. And they advocate raising the minimum wage, setting full employment as the goal of macroeconomic policy, and establishing universal health care.

Robert B. Reich, a professor at the University of California, Berkeley, and former secretary of labor in the Clinton administration, is the author of *Saving Capitalism: For the Many, Not the Few* (2015) on the same topic: the centrality of the rules governing the economy, and the need to adopt rules that share more widely the benefits of a growing economy. Reich advocates for rules like those proposed by Stiglitz, with similar goals. One goal is to regulate pharmaceutical patents to reduce the prices of needed drugs, by shortening the duration of patents and copyrights; banning pharmaceutical companies from paying generic drug producers to delay marketing cheaper generic varieties of drugs (a ban that is common in other countries). He also calls for stronger and more comprehensive antitrust laws and enforcement. He proposes legislation stopping corporations from requiring employees, franchisees, or contractors to sign agreements saying that any disputes will be settled only through arbitration. He would set the minimum wage at half the median wage and then adjust it periodically to preserve its purchasing power.

Reich also offers proposals to "reinvent the corporation," to restrain CEO pay and encourage pay raises proportionate to productivity growth. He suggests copying "codetermination" practices in Germany, mandating that up to half of an enterprise's decision-making boards consist of employee representatives. In fact, Reich calls for a kind of "stakeholder society": corporate governance that takes account of the interests of all the stakeholders (workers, community, and the environment)—all those who are affected by a corporation's activities—not just shareholders seeking high and steady dividends or rising share prices. One strategy would be to encourage the formation of "benefit corporations" (a category that already exists) whose articles of incorporation specify that they will consider all such relevant stakeholders. Another idea is to change the rules to grant corporate personhood—with constitutional rights and limited liability—solely to organizations that share power with their employees.

Reich also proposes guaranteeing a universal basic income (UBI) to every person, just enough for a decent living. This idea, he notes, was proposed by the conservative economist F. A. Hayek in 1979. Some advocate an UBI simply to prevent technological change and growing robotization from reducing low-skilled workers to destitution. A basic income paid to everyone would have a very low administrative cost, and not require intrusive investigation into recipients' lives.

The third such book, by Dean Baker, codirector of the Center for Economic and Policy Research, a Washington, DC, think tank, is *Rigged: How Globalization and the Rules of the Modern Economy Were Structured to Make the Rich Richer* (2016), and it offers similar critiques and solutions. Many proposals resemble Reich's. Baker focuses on five main ones: making full employment a key target for macroeconomic policy, eliminating financial sector waste, ending patent and copyright monopolies,

reforming corporate governance and restraining CEO pay, and allowing more competition in the market for physician services such as by allowing Medicare funds to be spent on medical procedures outside the United States. Baker estimates that these policies could yield $2.0–$3.7 trillion a year in new benefits—half of it from targeting full employment—or about $6,000–$12,000 per person per year.

Capitalism and Inequality

Is there a built-in tendency toward increasing inequality in capitalist economies? Much information presented in this and previous chapters suggests that the answer is yes. That is, there is no mechanism in capitalism itself that ensures mostly equal outcomes, and, as we have seen, many mechanisms that tend to produce, reinforce, and exacerbate unequal outcomes. Hence specific rules and programs—such as progressive taxation, minimum wage laws, estate taxes, and public provision of education and health care—are required to counter capitalism's tendency towards inequality. Capitalist economies other than the United States have achieved greater equality mainly because they have adopted more laws, norms, and institutions that counteract tendencies toward polarization of wealth and power.

Government intervention can potentially give every child a chance to grow up healthy and well educated, increasing the opportunity for all children to make a good living as adults. U.S. Supreme Court Justice Sonia Sotomayor is a good example: Her mother was widowed when she was a child, and relied both on hard work and public assistance to raise her. But the United States, especially in the last two decades, has done far less to give all children good health and education than have other countries at similar levels of income or lower. Figure 19.11 indicates, in each country shown, what percent of all children were in poverty in 2009. In this diagram the definition of poverty is a relative one: A child lives in poverty if she lives in a household whose income is below 50 percent of the household median income, after adjustment for household size. In the United States, 23.1 percent of children were in poverty in 2011, while in Canada the share was only 13.1 percent, and even poorer countries like Greece had a lower share of children in poverty than did the United States.

In many capitalist economies, government policies and programs have long aided poor families in ways such as providing universal health care at little or no cost, paid for mostly by the government out of tax revenues. This is a highly redistributive program that, along with universal public education, serves to promote the general welfare. Both programs together help to ensure that most children have a relatively good chance to grow up healthy and well educated. But the benefit is both to the children and to society, because a good start in life helps children to avoid unemployment and poverty later in their lives. Good social policies that favor children allow many to escape poverty.

In the United States, giving a minimal level of support to children to ensure good nutrition and a decent living was the rationale for the Aid to Families with Dependent Children (AFDC) program that was legislated in 1935 along with Social Security. But AFDC was terminated in 1996 and replaced by the much less generous Temporary Aid to Needy Families (TANF) program. As we saw in Chapter 7, in 1995 fully 62 percent of children living in poverty in the United States were receiving public assistance, while by 2014 that share had fallen to only 17 percent.

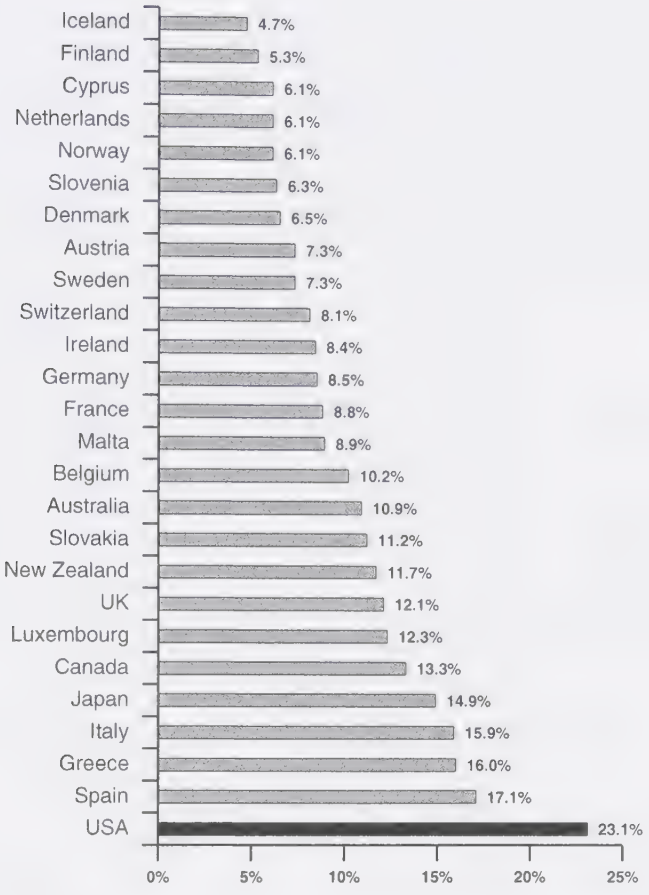


FIGURE 19.11 Child poverty rate, comparison of selected OECD countries, 2009.

The U.S. child poverty rate in 2009 far exceeded that in countries of similar, and even lower, income levels. In the data, a child is defined as in poverty if the income of the household in which the child lives is below 50 percent of the median household income, after adjusting for household size. Countries with productivity less than half that of the United States were excluded from the data.

Source: Economic Policy Institute, *The State of Working America*, 2012, from UNICEF Innocenti Research Centre Report Card 10 (Adamson 2012), Figure 1b, "Child Poverty Rate."

In short, market economies tend to generate poverty and inequality unless the rules governing the economy prevent that; government can also direct income and services to poor and working people and so restore greater equality. Helping all children to meet their full potential and become productive members of society is one way to reach this goal.

Democracy and Inequality

It is easy to see that wealth confers power, including the power to change the rules, and power in turn confers wealth. But should it not be possible to stop and reverse this dynamic through democratic processes? On the face of it, the majority of the population, who for decades have shared little in the benefits of growth, should have

every reason to vote for egalitarian policies that redistribute wealth, income, and social services toward that same majority. Why has this not happened in recent decades? What stands in the way of a vibrant democracy that keeps the power of the wealthy in check?

The answer has many parts. Concentration of ownership in the media is a factor. Another is gerrymandering of congressional districts that allows a party that gets a minority of the votes in a state to gain a majority of seats in the state legislature, and in the House of Representatives.[36] Sophisticated strategies for marketing candidates and proposed rule changes have also made it more difficult for voters to know how to vote to protect their own interests. The U.S. political culture in recent decades has also degenerated, with candidates rarely presenting a political platform with clear stances on issues, and often failing to discuss issues at all. Instead, campaigns create images and appeal to emotions. Politicians make promises that they abandon once in office. When voters are kept in the dark, they cannot rationally choose the candidate that will benefit them. Voter suppression through arbitrary purges of the rolls and burdensome voter ID requirements, as well as insufficient voting locations that create long lines are also problems; and many voters simply stay home. And the Voting Rights Act, which long helped to ensure the right of everyone to vote, was eviscerated in 2013 when the Supreme Court decided to invalidate a key provision of it. Yet election winners behave as if they had a mandate, approving legislation that they rarely or never advocated during their campaign.

Do other countries have better arrangements? Is there some other set of rules surrounding democratic elections that would yield better outcomes? Some would say that a parliamentary system, which allows both smaller parties and coalitions, has advantages over a two-party system that marginalizes third parties; others point to sometimes unfortunate outcomes in such systems. Many countries at income levels similar to that of the United States offer public assistance for campaign spending. In some, free television air time is also provided (or else TV advertising is prohibited to all candidates). This might be an improvement.[37] Recent polls have found support for campaign finance reform across the political spectrum.

Conclusion

An economic system is shaped by the rules that govern its economic transactions. But these rules are in turn shaped both by the rules that govern its *political* system, and by the structures of both economic and political power that decide who most influences the rule-making process. That is why we need a political economy analysis—one that takes account not only of competition, but also of command and of change—to fully understand the way the economy works and the way it evolves.

There has long been sentiment for some kind of alternative system. Some advocate for socialism, and by that they can mean a wide variety of economic and social

---

[36]Rob Richie, "Republicans Got Only 52 Percent of the Vote in House Races," *The Nation*, November 7, 2014.

[37]Paul Waldman, "How Our Campaign Finance System Compares to Other Countries," *The American Prospect*, April 4, 2014, available at http://prospect.org/article/how-our-campaign-finance-system-compares-other-countries.

arrangements, including a society in which cooperative production units organized in an egalitarian way are the main form of production. Some advocate for communism ("from each according to their ability, to each according to their needs"), anarchism (minimal government), or utopian communities. Any one of these would mean a far greater change in the rules than the changes discussed above. While the possibilities seem great, a thorough analysis of the reasons why most socialist or communist regimes have moved back in the direction of capitalist forms of organization is needed—but we will not attempt it here. Such an analysis would also need to take account of the principles in this book.

Suggested Readings

Michelle Alexander, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*, revised edition (New York: The New Press, 2012).

Eileen Appelbaum and Rosemary Batt, *Private Equity at Work: When Wall Street Manages Main Street* (New York: Russell Sage Foundation, 2014).

Dean Baker, *Rigged: How Globalization and the Rules of the Modern Economy Were Structured to Make the Rich Richer* (Washington, DC: Center for Economic and Policy Research, 2016); available at http://deanbaker.net/books/rigged.htm.

Andrew Coyle, Allison Campbell, and Rodney Neufeld, eds., *Capitalist Punishment: Prison Privatization and Human Rights* (Atlanta, GA: Clarity Press, Inc.; London: Zed Books, 2003).

Joel Dyer, *The Perpetual Prisoner Machine* (Boulder, CO: Westview Press, 2000).

Thomas Ferguson, *Golden Rule: The Investment Theory of Party Competition and the Logic of Money-Driven Political Systems* (Chicago: University of Chicago Press, 1995).

David Cay Johnston, *Free Lunch* (New York: Portfolio, Penguin Group, 2007).

David Cay Johnston, *Perfectly Legal* (New York: Penguin Books, 2003).

Charles Lewis, Bill Allison, and the Center for Public Integrity, *The Cheating of America* (New York: William Morrow, 2001).

Robert S. MacIntyre, Matt Gardner, Kayla Kitson, Alexandria Robins, and Michelle Surka, *Offshore Shell Games 2016: The Use of Offshore Tax Havens by Fortune 500 Companies* (Citizens for Tax Justice, October 2016, ctj.org).

Nancy MacLean, *Democracy in Chains* (New York, Viking, 2017).

Jane Mayer, *Dark Money: The Hidden History of the Billionaires Behind the Rise of the Radical Right* (Anchor, 2016).

Robert B. Reich, *Saving Capitalism: For the Many, Not the Few* (New York: Alfred A. Knopf, 2015).

T. R. Reid, *The Healing of America: A Global Quest for Better, Cheaper, and Fairer Health Care* (New York: The Penguin Press, 2009).

Kenneth Sheve and David Stasavage, *Taxing the Rich: A History of Fiscal Fairness in the United States and Europe* (Princeton, NJ: Princeton University Press, 2016).

Peter Singer, *Corporate Warriors: The Rise of the Privatized Military Industry* (Ithaca, NY: Cornell University Press, 2007).

Joseph E. Stiglitz, *The Price of Inequality: How Today's Divided Society Endangers Our Future* (New York: W. W. Norton, 2013).

Joseph E. Stiglitz, with Nell Abernathy, Adam Hersh, Susan Holmberg, and Mike Konczal, *Rewriting the Rules of the American Economy: An Agenda for Shared Growth and Prosperity* (New York: W. W. Norton, 2015).

Richard Wilkinson and Kate Pickett, *The Spirit Level: Why Greater Equality Makes Societies Stronger* (New York: Bloomsbury Press, 2011, reprint edition).

Mark Zepezauer, *Take the Rich Off Welfare* (Cambridge, MA: South End Press, 2004).

List of Variables

(1) Prices

P_z = Price of a unit of output

P_m = Price of materials and of the wear and tear on machines, structures, and equipment during the production process

w = "Wage rate" = payment to workers (wages + benefits) per hour of labor

P_{im} = Price of an imported good in the currency of the importing country

P_{imf} = Price of an imported good in the currency of the country where it was produced

(2) Totals

(a) Flows: Physical or money totals for a firm or for the economy as a whole, for a given period of time, such as a year.

Z = Physical quantity of *total output* produced during a year, in the sense that price times quantity ($P_z Z$) is the money value of total output; Z can be used either for a firm or for the whole economy

Z^* = *Potential output*, that is, the physical quantity of *total output that could be produced* during a year *using the current productive capacity of a firm*, in the sense that price times quantity ($P_z Z^*$) is the money value of total potential output.

M = *Physical quantity* of *materials used* during a year, plus the value of *depreciation of machinery*, structures, and equipment

H = Total *hours of labor employed in a firm* during a year

Y = Money value of *net output* produced in a year, either by a firm or in the whole economy; $Y = P_z Z - P_m M$, where $P_z Z$ is the value of total output and $P_m M$ is the value of materials used up and wear and tear on machines, structures, and equipment during a year. Y is called *net output* or *value added*. When Y is used for the economy as a whole, it can also be called *net national product*.

W = Total *wages*, *salaries*, and benefits paid in a year; for a firm, $W = wH$; for the whole economy, $W = wN$

R = Total *profit* (money value) made by a firm during a year = $Y - W$

F = Fixed cost, or overhead = sum of all costs that do not change when quantity of output Z changes

N = Total *hours of labor employed in the economy* during a year

LS = *Labor supply* = total hours of labor workers would like to work during a year

AD = *Aggregate demand*

AS = *Aggregate supply*

C = Money value of *consumer goods and services* bought in a year in the whole economy

I = Money value of business spending on *investment goods and services* during a year

G = Money value of total *government spending on goods and services* during a year

B = Money value of *government borrowing* from the public to finance government spending during a year

X = Money value of *net exports* = total value of exports minus total value of imports during a year

(b) Stocks: A **stock** is a quantity at a given moment in time (such as January 1, 2017), and is measured simply in physical or monetary units.

K = Money value of *capital stock*

(3) Ratios

Ratios are totals divided by other totals; they can always be interpreted as "the amount of one thing per unit of something else" (for example, the ratio X/Y means "the amount of X per unit of Y").

z = Money value of output per hour of labor employed = $P_z Z/H = ef$

m = Depreciation of machines, structures, and equipment, plus quantity of materials used up, per labor hour = $P_m M/H$

y = Money value of net output per hour of labor employed = Y/H

e = Work effort, or work done, in an hour of labor; intensity of work

f = Efficiency of labor = value of output produced per unit of work done

r = Rate of profit = money amount of profit divided by value of capital goods owned

= $R/K = (Y - W)/K$

atr = After-tax profit rate = (profit minus taxes paid)/K

i = Rate of interest (percent per year), which is the cost of borrowing money, say, for the purpose of investment; i = (money value of interest paid)/(principal of the loan)

u = Capacity utilization rate = actual output as a percent of potential output during a time period, or Z/Z^*

ulc = Unit labor cost = wage cost per unit of output = $wH/Z = W/Z = w/z = w/(ef)$

umc = Unit materials cost = $P_m M/Z$

uc = Unit cost = average cost of producing a unit of output = $(P_m M + wH)/Z = umc + ulc$

up = Unit profit = $R/Z = P_z - uc$

c = Fraction of wage and salary income spent on consumer goods and services = C/W (assumes profit recipients pay for their consumption only out of wage/salary incomes)

j = Profit effect on investment = how much total investment (I) increases (in units of a currency) when total profit (R) increases by one unit (measured in the same currency)

v = Value of the dollar = foreign exchange rate = the number of units of a foreign currency that can be obtained in the global currency market in exchange for $1

(4) Other

ww = Weekly wage

ud = Unemployment duration (in weeks)

ui = Unemployment insurance paid per week

cjl = Cost of job loss

Sources of Economic Information

Many of the facts in this book are taken from the following sources, including U.S. federal government agencies and the Federal Reserve System; international organizations such as the Organization for Economic Co-Operation and Development, the United Nations and its various agencies, and the World Bank; and some academic and non-governmental organizations. Some sources are available in libraries. A typical website includes not only data, but often extensive discussions of the data, as well as suggestions about other useful data. To find links that may no longer work, one possibility is to learn to use the Wayback Machine, a vast web-accessible archive that includes web pages that were posted at some time in the past. More information is at https://archive.org/about/faqs.php.

U.S. federal agencies and the Federal Reserve

U.S. President's Council of Economic Advisors

Economic Report of the President (Washington, DC: U.S. Government Printing Office), annual; current and past reports available at https://www.gpo.gov/fdsys/browse/collection.action?collectionCode=ER.

U.S. Bureau of the Census (**census.gov**): data on population, income distribution and poverty by race, gender, age, and other demographic characteristics; income by source; businesses of all sizes

Statistical Abstract of the United States, ceased publication in 2012, available at https://www.census.gov/history/www/reference/publications/statistical_abstracts.html (and in some libraries).

Historical Statistics of the United States: Colonial Times to 1970 (Washington, DC: U.S. Government Printing Office, 1975).

U.S. Bureau of Economic Analysis of the Department of Commerce (**bea.gov**)

U.S. Bureau of Labor Statistics of the Department of Labor (**bls.gov**): data on wages and employment by race and gender and many other labor topics

U.S. Congressional Budget Office (**cbo.gov**): information and analyses of the economic impact of legislation before and after it is approved and implemented)

U.S. Congressional Research Service: excellent reports on policy issues; google "Congressional Research Service Reports" or ask your librarian, since CRS does not directly supply reports to the public

U.S. Government Accountability Office (**gao.gov**): analyses of the effectiveness of government programs; reports are online

Federal Reserve System (**https://www.federalreserve.gov/**): data on money and banking
Federal Reserve Economic Data (FRED), St. Louis (**https://fred.stlouis.org**): Data and graphs for a wide range of economic data on the U.S. economy; some foreign data; see Archival FRED (**https://alfred.stlouisfed.org**) for historical data series

International organizations

United Nations (**un.org**): see specific UN agency, such as:

United Nations Development Program (**undp.org**), *Human Development Report* (New York: Oxford University Press, annual, available at http://hdr.undp.org/en/global-reports).
Food and Agricultural Organization (**fao.org**): data and reports on agriculture and food, by country and crop
International Labor Organization (**ilo.org**): data and reports on employment, working conditions, labor rights, and other labor issues

The World Bank (**worldbank.org**),

World Development Report (New York: Oxford University Press, annual, available at http://www.worldbank.org/en/publication/wdr/wdr-archive).
World Development Indicators DataBank: thousands of data series, by country, at http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators

International Monetary Fund (**imf.org**): Country-specific data from Afghanistan to Zimbabwe at **http://www.imf.org/external/country/index.htm**
Organization for Economic Cooperation and Development (**oecd.org**)

Statistics on member countries: **http://stats.oecd.org/**

Non-governmental think tanks and research institutes

Economic Policy Institute (**epi.org, http://www.epi.org/data/**): Data and policy analysis

Lawrence Mishel, Josh Bivens, Elise Gould, and Heidi Shierholz, *The State of Working America*, Economic Policy Institute (Ithaca, New York: ILR Press, an imprint of Cornell University Press, latest print edition 2012, available online at **http://www.stateofworkingamerica.org/**).

Center on Budget and Policy Priorities (**cbpp.org**): reports, data on current budget issues
Center for Economic and Policy Research (**cepr.net**): reports, data on economic issues
Center for Responsive Politics (**opensecrets.org**): data on money in politics
Center for Tax Justice (**ctj.org**): analyses of tax issues, corporate use of tax havens
Institute for Policy Studies (**ips-dc.org**): manages website **inequality.org**, with data and articles on inequality
Institute on Taxation and Economic Policy (**itep.org**): analysis of tax issues
Public Citizen, Trade Data Center (**https://www.citizen.org/our-work/globalization-and-trade/trade-data-center**): trade-related job losses, by state; other
Tax Justice Network (**taxjustice.net**): reports on tax issues, including use of tax havens
Tax Policy Center (**taxpolicycenter.org/**): reports on current tax policy issues
Urban Institute (**urban.org**): reports on a wide range of current policy issues
WorldWatch Institute (**worldwatch.org**): online information on environmental issues, and print reports like the annual series *The State of the World* (Washington DC: Island Press)

Academic information and data sources

Brennan Center for Justice (New York University) (**https://www.brennancenter.org/**): data on crime, criminal justice, and incarceration
National Bureau of Economic Research (non-government) (**nber.org**; **nber.org/data**)

The NBER has industry data and macroeconomic data, like beginning and ending dates for each recession, as well as working papers and other resources.

Penn World Tables, now at the University of Groningen's Groningen Growth and Development Centre (**http://www.rug.nl/ggdc/productivity/pwt/**): economic data on inputs, output, income, and productivity for 182 countries, beginning in 1950; GGDC ("Historical Development" tab) also has the historical economic database developed by Angus Maddison and others

University of Texas Inequality Project (**utip.lbj.utexas.edu/**): data on world inequality

Resources for Economists on the Internet (**rfe.org/**)

Designed for students, teachers, and practitioners of economics; links to academic economics programs, organizations, conferences, listserves, and other resources

Data visualization websites, some interactive

International Monetary Fund DataMapper (**imf.org/external/datamapper**)

The Observatory of Economic Complexity, MIT (**atlas.media.mit.edu/en**)

Visualizing Economics (**visualizingeconomics.com**)

World Bank Data Viz (**worldbank.tumblr.com**)

Newspapers and periodicals

The following provide extensive, informative reporting on economics and the economy; a few articles may be free, and your college library may have a subscription: *Financial Times* (**ft.com**), *The New York Times* (**nytimes.com**), *The Washington Post* (**wapo.com**), *The Los Angeles Times* (**latimes.com**), *The Wall Street Journal* (**wsj.com**), *The Economist* (**economist.com**), *Dollars and Sense* (**dollarsandsense.org**), *The American Prospect* (**prospect.org**), *Challenge* (**tandfonline.com/toc/mcha20/current**), *Bloomberg Business Week* (**bloomberg.com/businessweek**), *Fortune* (**fortune.com**), and *Forbes* (**forbes.com**).

Other

Center for Popular Economics, Amherst, MA (**www.populareconomics.org**): offers week-long summer courses and weekend courses at other times, to teach analysis of current economic and political issues directed toward advocating change.

economicsnetwork.ac.uk/links/data_free.htm: links to data on the British economy

Glossary

ACCUMULATION is profit-driven investment, including the process of mobilizing, transforming, and exploiting the inputs required in capitalist production and in selling the output.

An **AFFILIATE** is a corporate entity in which the parent corporation holds enough shares to give it control, which usually means it holds more than 50 percent of the affiliate's voting stock (some shares of stock carry with them the right to vote on company policies; others do not).

The **AFTER-TAX PROFIT RATE** is the profit rate made by firms after they have paid all required taxes.

AGGREGATE or **POPULATION-LEVEL OUTCOMES** refer to the economic totals, averages, and relationships that are generally studied by economists.

AGGREGATE DEMAND is the total demand for goods and services during some period, say, a year.

AGGREGATE SUPPLY is the total supply of goods and services produced during some period, say, a year.

AVERAGE COST is the total cost of producing a certain number of units of a good or service divided by the total number of units produced.

BARRIERS TO ENTRY are obstacles that make it more difficult or costly for new firms to enter a market; examples include technical secrets, initial investments that are very large, and exclusive marketing arrangements.

BELIEFS are one's understandings of the actions necessary to bring about particular outcomes.

A **BOND** is an "IOU" that contractually commits the issuer of the bond (a government or a corporation) to making regular "coupon payments" to the owner of the bond, of a specified amount, at specified time intervals, as well as to make a final payment of the face value of the bond on the maturity date of the bond.

A **BREAKTHROUGH** occurs when a firm discovers or develops a new method of doing business, such as a new way of organizing work, a new product, or a new market.

BUILT-IN STABILIZERS are automatic changes in government spending and taxation that dampen the business cycle without conscious policy decisions directed toward this end. Examples are unemployment insurance and income taxation, especially progressive income taxation.

BUREAUCRATIC CONTROL is a system of control that uses job ladders, seniority rewards, and other organizational incentives in order to extract work from workers.

The **BUSINESS CLIMATE** refers to political, ideological, psychological, and other influences on the expected profit rate. It is sometimes termed the investment climate.

The **BUSINESS CYCLE** is a periodic expansion and contraction of output and employment usually taking place over a period of three to ten years.

BUSINESS CYCLE CONTRACTIONS (or **RECESSIONS**) are periods of increased unemployment and reduction in output and income.

BUSINESS CYCLE EXPANSIONS are characterized by increases in employment and income.

CAPACITY UTILIZATION is the ratio of the quantity of output a firm is actually producing to the quantity

of output it could produce if its resources were fully in use.

CAPITAL GOODS are goods used in production—machines, buildings, and the like—that are durable and will be used up only over the course of years.

A **CAPITAL GOODS-SAVING TECHNICAL CHANGE** is a new technology that reduces the capital goods and materials necessary to produce a given amount of total output.

CAPITALISM is an economic system in which employers, using privately owned capital goods, hire wage labor to produce commodities for the purpose of making a profit. The **CAPITALIST CLASS,** or **CAPITALISTS,** are those who own capital goods used in production and exercise control over the labor of others; they receive their income in the form of profits or other payments (such as interest and rent) for the use of their capital goods.

The **CAPITALIST EPOCH** began in some parts of Europe around AD 1500, when capitalist organization of labor processes first appeared. It continues to the present in most of the world.

CAPITALIST PROFITS are profits that result from a production process.

A **CARTEL** is a combination of business firms or states operating in concert to regulate the production, pricing, and marketing of goods by its members.

CHANGE, or the time dimension in economics, refers to the historical evolution of people and economic systems.

A **CLASS** is a group of people who share a common position in the economy with respect to the production and control of the surplus product.

A **CLASS RELATIONSHIP** exists between the producers of the total product, including the surplus product, and those who command the use of the surplus product.

COLLECTIVE BARGAINING occurs when, in negotiating wages and other employment conditions, all workers in a firm or occupation are represented collectively by a union; employers may also be collectively represented by an employers' association.

COMMAND, or the vertical dimension in economics, refers to aspects of economic relationships in which power plays the predominant role.

COMMAND RELATIONS are relationships between superiors and subordinates in which the superior exercises substantial power over the subordinate.

For a merchant, **COMMERCIAL PROFITS** result from selling something for more than it cost to purchase ("buying cheap and selling dear").

A **COMMODITY** is any good or service that is produced with the intention of selling it in order to make a profit.

COMPETITION, or the horizontal dimension in economics, refers to aspects of economic relationships in which voluntary exchange and choice among a large number of possible buyers and sellers play the predominant role.

COMPETITION FOR PROFITS is the struggle for survival and expansion among firms seeking new ways of doing business, new markets, new products, and other possibilities for profitable investment.

COMPETITIVE MARKETS are those with many actual or potential demanders and suppliers.

A **COMPLETE CONTRACT** is one that fully specifies—in ways that the courts will enforce—everything that each party to the contract is to do as a result of the contract.

CONSTRAINTS are the limits on the actions that an individual or a society can take.

The **CONSUMER PRICE INDEX (CPI)** is a measure of the average prices a typical family pays for the goods and services it buys, relative to the prices of the same or similar items in a reference year, called the base year.

A **CONTRACT** is an agreement, either written (explicit) or unwritten (implicit), that commits two or more parties to taking certain actions, such as making payments and delivering goods or services.

A **CONTRACT SPECIFYING WORK TO BE DONE** is an agreement between an employer and a worker that specifies payment for actual work activities instead of for work time.

CONVENTIONAL ECONOMICS. *See* **NEOCLASSICAL ECONOMICS.**

COORDINATION BY COMMAND takes place when interactions are governed by orders specifying precise behavior.

COORDINATION BY RULES takes place when interactions are governed by general principles of behavior.

COORDINATION FAILURE occurs when markets or other types of coordination by rules fail to coordinate an economy in such a way as to produce outcomes that are desirable.

The **COST OF JOB LOSS** is the loss of income a worker experiences as a result of quitting or being laid off from a job.

COUNTERCYCLICAL POLICIES are policies aimed at dampening the business cycle.

DEFICIT SPENDING occurs when the government finances its purchases by borrowing from the public.

A **DEMAND CURVE** indicates, for each possible price, how much of the good or service demanders are willing and able to buy.

DEMOCRACY is a process with three characteristics: the exercise of power is accountable to those affected by it, civil rights and personal liberties are guaranteed, and citizens have relatively equal access to political resources and influence.

A **DEMOCRATIC FIRM** is one that is owned by its employees and run by people who are elected by the employees.

DEPRECIATION is the cost (due to wear and tear) of restoring the capital goods that suffered wear and tear in producing last year's output.

DESKILLING means changing a production process in such a way as to make it possible to employ workers with fewer skills.

The **DETERMINANTS OF THE PROFIT RATE** are the things upon which the profit rate depends; they determine how high the profit rate will be.

DISCRIMINATION means treating someone differently simply because that person belongs to a certain group.

A share of stock represents a share of ownership of a corporation and entitles its holder to receive a **DIVIDEND**—a payment per share decided by the board of directors—whenever a corporation decides to pay out a portion of its total profits as dividends to its stockholders. The amount of the dividend varies. See also **STOCK.**

DIVISION OF LABOR (or **ECONOMIC SPECIALIZATION**) exists when people are not economically self-sufficient but instead produce things used by others and use things produced by others.

ECONOMIC CONCENTRATION is the extent to which the economic activity of an industry or the whole economy is conducted by the largest firms.

ECONOMIC MAN *(Homo economicus)* refers to the assumption that human beings are calculating, amoral, and self-interested.

ECONOMIC SPECIALIZATION. *See* **DIVISION OF LABOR.**

An **ECONOMIC SYSTEM** is a set of relationships among people that organizes the labor processes all societies need to sustain life.

ECONOMICS is the study of how people interact with one another, with nature, and with the other things they require in order to produce their livelihoods.

ECONOMIES OF SCALE exist when an increase in the number of units of output produced—an increase in the *scale* of production—brings about a fall in the average cost, that is, a fall in the cost per unit of output.

An **ECONOMY** is a collection of labor processes.

The **EFFICIENCY OF LABOR** refers to how much output can be produced as a result of a certain level of work effort.

The term **EFFICIENT** is applied to a labor process if the effort, time, intelligence, creativity, raw materials, natural environment, information, and machinery used in it are applied in a way that enhances people's well-being by equipping them with the things and the free time needed to lead a flourishing life.

The **EMPLOYMENT EFFECT** is the change in the number of hours of labor employed that results from the direct and indirect effects of a change in business investment or government deficit spending.

The **EMPLOYMENT MULTIPLIER** is the hours of new employment directly and indirectly created by an additional dollar of investment or other spending.

EQUALIZATION OF PROFIT RATES refers to the process by which competitive pressures on firms in different industries, different geographical regions, or different markets push their profit rates toward a common, or average, level.

EQUILIBRIUM refers to a situation—a price and quantity exchanged—in which there are no forces internal to the situation pushing it to change.

The **EQUILIBRIUM LEVEL OF EMPLOYMENT** is the amount of employment that corresponds to a particular macroeconomic equilibrium (when there is no excess demand or excess supply in product markets, all taken together).

EXCESS DEMAND exists when at a particular price more of some good or service is demanded than is supplied.

EXCESS SUPPLY exists at a particular price when at that price more of some good or service is supplied than is demanded.

The **EXPECTED PROFIT RATE ON INVESTMENT** is a firm's estimate of the future profit rate that it thinks will be earned on its investment.

EXTERNALITIES occur when some of the effects of a market exchange are not reflected in the price and are thus "external" to the participants in the exchange.

The **EXTRACTION OF WORK FROM WORKERS** is the process of transforming the labor time that an employer has purchased into work done.

FAIRNESS means that people in an economic system suffer the burdens and enjoy the benefits of that economic system equitably.

The **FALLBACK WAGE** is the wage at which an employee has no preference for keeping his or her current job as opposed to being fired or quitting; it varies with the employee's income prospects in the absence of the current job.

FEUDALISM was the dominant economic system in Europe in the Middle Ages; lords obtained the surplus product through rents and other customary obligations owed by the serfs.

FIXED COSTS are costs that a business pays regularly and that do not vary when the total number of units of output produced varies.

Government **FISCAL POLICY** uses taxes and spending to regulate the level of total output and employment.

A **FOREIGN EXCHANGE RATE** is how much of one currency another will buy. The exchange rate of the dollar to the euro is the number of euro a dollar will buy, and is also known as "the **VALUE OF THE DOLLAR** in euro."

FULL EMPLOYMENT is a situation in which almost everyone seeking work readily finds it.

A **HIERARCHY** is an organization of power in which superiors have command over subordinates.

The **HIGH-EMPLOYMENT PROFIT SQUEEZE** occurs when the high demand for labor creates both labor and materials cost increases, which in turn reduce profits.

HORIZONTAL ECONOMIC INTERDEPENDENCE is based on specialization and is not necessarily based on unequal advantage or command.

INCOME can be earned from property or labor and is considered a "flow."

An **INCOMPLETE CONTRACT** is an agreement between two or more parties that leaves out certain aspects of an exchange and requires or imposes significant enforcement costs upon one party or another.

An **INCOMPLETE LABOR CONTRACT** is a contract (explicit or implicit) between an employer and a worker that is incomplete in the sense that it specifies the wage rate but does not specify the exact tasks to be performed or the amount of effort to be provided by the worker.

The **INDEPENDENT PRIMARY LABOR MARKET** includes those jobs with highly elaborate bureaucratic or professional career patterns; it contains mainly the jobs of craft, technical, professional, and lower-level supervisory workers.

INDEPENDENT PRODUCTION OF COMMODITIES is an economic system in which the producers own the capital goods needed in production and use (primarily) their own labor.

An **INEFFICIENT TECHNICAL CHANGE** is one that can bring about an increase in the output of a production process only by using proportionally more of at least one input.

INFLATION is a general increase in prices, often measured by the Consumer Price Index (CPI).

The **INFRASTRUCTURE** of an economy consists of its roads, railways, airports, harbor facilities, bridges and tunnels, water and sewage systems, public utilities and electricity grids, and communications networks such as telephone lines and the internet.

The **INTENSITY OF LABOR** is how much work effort producers must expend per hour of work, or more simply, how hard they work.

The contractual payment made by the government to owners of U.S. Treasury bonds, bills, or notes is called **INTEREST**. The contractual payment made by a corporation to owners of its bonds or to banks that have provided loans to the corporation is also called **INTEREST**.

The **INTEREST RATE** is the cost of borrowing money; for a firm that borrows money, such as from a bank, it is the interest payment divided by the **PRINCIPAL**, which is the part of the loan amount that has not yet been repaid.

INVESTMENT means spending money to add to, or improve, a firm's productive equipment, software, facilities, or workforce skills in order to increase productive capacity and productivity.

JOB LADDERS link together a series of related jobs, in which a worker over the years climbs from one job to another and gains access to jobs higher on the ladder only by first succeeding in the lower jobs.

LABOR is any activity performed by people that contributes to production or reproduction.

The **LABOR EXTRACTION CURVE** describes for each wage rate the intensity of work that the worker chooses to perform.

A **LABOR MARKET** is a market in which workers sell their labor time (not work itself) in return for a wage; employers are the demanders, and workers are the suppliers, of labor time.

A **LABOR PROCESS** is any activity performed with the intention of producing something, which may be either goods or services.

A **LABOR-SAVING TECHNICAL CHANGE** is a new technology that increases the total output produced with a given amount of labor.

LABOR TIME measures the number of hours worked; it does not measure how much work gets done, since there are many different levels of work effort (intensities of work) possible.

LAISSEZ-FAIRE is an approach to economic policy that advocates a very limited role for the government, confining its activities to national defense and the enforcement of laws and contracts.

A **LAYOFF** refers to a firm's temporary or permanent dismissal of workers for any of several reasons.

A **LOCKOUT** occurs when an employer locks the workers out of the workplace and closes down production in order to force workers to accept the employer's terms for wages, work pace, or other working conditions.

A **LONG SWING** occurs over a period of thirty to fifty years. The first part generally coincides with the "consolidation" phase of a particular social structure of accumulation (SSA) and is characterized by relatively high rates of investment and economic growth and relatively little unemployment. The remainder of a long swing is associated with the "decay" phase of an SSA; it is a period of stagnation of economic growth with relatively little investment and high unemployment.

A **MACROECONOMIC EQUILIBRIUM** exists when aggregate demand equals aggregate supply (no excess aggregate demand or excess aggregate supply exists) and there are no forces tending to change the situation other than accidental ones or ones coming from external sources.

MACROECONOMICS is about how the decisions of individuals, families, firms, and governments produce outcomes—such as economic progress or stagnation, inflation or unemployment—for society as a whole.

MARGINAL COST is the increase in the total cost incurred by a firm when it increases its total output of a commodity by one unit.

A **MARKET** refers to all the buying and selling activities of those persons willing to trade a good or service as long as the price is within an acceptable range; a market consists of suppliers potentially wanting to sell and demanders potentially wanting to buy.

The **MARKET CLEARING PRICE** is the price at which buyers want to purchase exactly the quantity that sellers want to sell.

A **MARKET EXCHANGE** is a transfer to another party of the title to a good or of the right to receive a service, in return for some form of payment on mutually acceptable terms.

A **MARKET FAILURE** is said to take place when the spontaneous interactions of buyers and sellers on markets each pursuing their own objectives results in outcomes that are different from the decision that society as a whole would make if it had all the information it needed to make the choice.

A **MARKET SHARE** is one firm's sales as a percent of the total sales in an industry.

MATERIALS are goods that are used up during the process of production.

MICROECONOMICS deals with decisions that individuals, families, and firms make (and why they make them).

MIDDLE CLASSES in capitalist society possess one but not both of the attributes of capitalists; therefore, they stand between capitalists and workers.

MONETARY POLICY refers to government influences on the rate of interest intended to regulate the level of investment, output, employment, and other macroeconomic outcomes.

MONOPOLY POWER is the ability of one or a few firms in an industry to exercise substantial control over the market price and other aspects of competition, usually by excluding other firms.

MULTINATIONAL CORPORATION (MNC). See **TRANSNATIONAL CORPORATION (TNC).**

The **NECESSARY PRODUCT** is the portion of the total product that is needed to maintain the inputs in the labor process—workers, materials, and tools—at their current level or in their current condition for the next round of the production process.

NEOCLASSICAL or **CONVENTIONAL ECONOMICS** is an economic theory emphasizing the horizontal dimension of markets and voluntary exchange.

NET EXPORTS are equal to the monetary value of total exports of goods and services minus the monetary value of total imports of goods and services.

NET OUTPUT. See **NET PRODUCT.**

NET PRODUCT (or **VALUE ADDED**, or **NET OUTPUT**) is the total product minus materials and capital goods used up in the course of producing the total product.

NET WORTH, or **NET ASSETS,** is the sum of all of a person's assets minus all outstanding debts.

The **NEW MIDDLE CLASS** consists of those who do not own the capital goods used in their own labor processes but who do regularly control the labor of others; it includes managers and supervisors.

NONPRICE COMPETITION involves sales efforts, style changes, or other marketing strategies to increase the appeal of a product without lowering its price.

The **OLD MIDDLE CLASS** consists of those who do own the capital goods used in their own labor processes but who do not regularly control the labor of others; they are self-employed or are small business employers.

OLIGOPOLY, or **SHARED MONOPOLY,** is a market situation in which several firms together, but no one firm by itself, can exercise substantial monopoly power.

An **OPPORTUNITY COST** is the value of the best opportunity given up (forgone) in order that whatever activity was chosen could be undertaken.

The **OUTPUT MULTIPLIER** is the additional output generated directly and indirectly by a dollar of additional investment or other spending.

PARALLEL PLANTS are plants owned by the same employer and producing the same product but located in different geographical regions in order to weaken the workers' ability to bargain collectively.

A **PIECE RATE** is a form of wage payment in which a worker is paid for each unit of output produced instead of for work time.

A **POLITICAL BUSINESS CYCLE** occurs when recession and/or expansion are in part intentionally created by governmental economic policy, as officials attempt to generate a business cycle for their own or others' interests.

POLITICAL ECONOMY is a term we use for a theory that analyzes capitalism in terms of the three dimensions of competition, command, and change.

A has **POWER** over B if by imposing costs on B (or threatening to do so) A can cause B to act in a way that is to A's advantage.

PREFERENCES are the relative values one places on various outcomes that one's actions might bring about.

PRICE COMPETITION is a form of, or strategy for, competition in which firms attempt to attract customers primarily by offering lower prices.

PRINCIPAL. See **INTEREST RATE.**

PRIVATE COSTS are the costs borne by the user of a good or service (a person or a company), while the total costs borne by all members of a society are termed **SOCIAL COSTS.** At best, prices only measure private costs.

PRIVATE PROPERTY is a social institution (or rule) that gives individuals or firms the right to use, lend, or sell things such as land, buildings, and artistic or intellectual creations of any kind; it means that one can have or use such things only if one has made, rented, purchased, or been given them.

PRODUCTION is a labor process whose output is a good or a service.

PROFIT is the form of the surplus product in a capitalist economic system; it is what is left over, out of sales revenues, after wages, the costs of materials used up, and wear and tear on machines have been paid.

The **PROFIT EFFECT ON INVESTMENT** is the amount by which investment will increase for each $1 increase in total profits.

PROFITABILITY refers to how much profit is derived from a labor process.

A **PROFIT-LED EMPLOYMENT** situation exists when rising wages reduce employment.

PROPERTY INCOME is income that is received in the form of profit, rent, interest, or dividends as a result of owning an asset such as a business, a piece of land, an existing structure, a bond, or a share of corporate stock.

PROPERTY RIGHTS establish the owner's right to control the property, to decide who uses it for what purpose, and to benefit from its use or sale.

A firm's **RATE OF PROFIT** is the total amount of profit divided by the value of the capital invested in the firm, which is approximately the value of the capital goods owned by a firm.

The **REAL WAGE,** or **REAL INCOME,** refers to wages or income corrected to take account of the effects of inflation.

RENT refers to payments that firms are required to make to the owners of land, office space, buildings, or other facilities in return for the right to use or occupy them.

REPRODUCTION is a labor process whose output is people; it includes not only biological reproduction but also such activities as child rearing, training, feeding, and caregiving.

RETAINED EARNINGS are the part of total profit that a corporation does not pay out as dividends, but keeps for future investment or other purposes.

A **RUNAWAY SHOP** is a workplace that an employer has moved from an area where workers are strong to an area where workers are weak in order to escape having to meet workers' demands.

The **SALES EFFORT** consists of all those activities by a firm that relate to the selling of the firm's product.

SAVING is income minus consumption spending.

The **SECONDARY LABOR MARKET** includes jobs in workplaces that lack the formal organization (such as collective bargaining agreements, bureaucratic control, or professional or craft patterns) of primary markets; it contains jobs such as those of nonunionized service and retail workers, clerks in the private sector, seasonal workers, and nonunionized employees of small businesses.

SELF-INTEREST refers to a disposition to consider only how one's actions will affect oneself, not how they may affect others.

The **SEPARATION OF CONCEPTION FROM EXECUTION** is one method for deskilling work in which the workers who plan production are different from those who carry it out.

SHARED MONOPOLY. *See* **OLIGOPOLY.**

SIMPLE CONTROL is a system of control that focuses on the supervisors' personal exercise of workplace rewards and sanctions to maintain the work pace.

SLAVERY was the dominant economic system in the U.S. South before the Civil War; slaveholders obtained the surplus product by owning all of the inputs (including slaves) and the output of slave production.

SOCIAL COSTS are the costs of producing a good or service that are borne by society as a whole. *See also* **PRIVATE COSTS.**

The **SOCIAL ORGANIZATION OF THE WORKPLACE** refers to the way in which jobs are defined, work tasks are assigned, supervisory power is delegated, and other social aspects of the workplace are organized.

A **SOCIAL STRUCTURE OF ACCUMULATION (SSA)** is the institutional setting within which accumulation occurs; it structures relations among capitalists, between capitalists and workers, among workers, and between government and the economy. Two phases of an SSA can be distinguished: first, its "consolidation" and, second, its "decay."

SPEEDUP is an effort by an employer to increase the pace of work.

SPILLOVERS. *See* **EXTERNALITIES.**

The **STAGES OF CAPITALISM** in the United States are distinct phases in the development of U.S. capitalism, with each being defined by a particular social structure of accumulation.

STOCK. See **DIVIDEND.**

The **SUBORDINATE PRIMARY LABOR MARKET** includes most government employment, as well as private sector jobs in workplaces organized according to collective bargaining agreements; it includes mainly the jobs of the traditional unionized industrial working class.

A **SUPPLY CURVE** indicates, for each possible price, how much of the good or service suppliers wish to sell.

The **SURPLUS PRODUCT** is what remains out of the total product after the necessary product has been deducted.

A **SYSTEM OF CONTROL** is an employer's strategy or method for governing the workplace to facilitate the extraction of work from the workers.

A **TAX HAVEN**, or offshore financial center, is a jurisdiction that offers strict banking secrecy laws and low or zero tax rates on foreign-owned business activities.

TECHNICAL CONTROL is a system of control that incorporates a work pace designed into the machinery of production.

TECHNICAL EFFICIENCY: A production process is said to be **TECHNICALLY EFFICIENT** if the output of it cannot be increased without using more of at least one of the inputs.

TECHNICAL PROGRESS is a change in the relationship between inputs and outputs that permits the same output to be produced with less of one or more of the inputs.

TECHNOLOGY is the relationship between inputs and outputs in a labor process.

The **TOTAL COST OF INVESTMENT** consists of two parts, the cost of the capital goods purchased and the opportunity cost of the money used to purchase them.

The **TOTAL PRODUCT** is the total amount of goods and services produced in an economy during a given period of time.

A **TRANSNATIONAL CORPORATION** (TNC), or multinational corporation (MNC), is a firm earning profits on its global operations by locating its corporate entities, called **AFFILIATES**, in multiple countries, in whatever combination of sites yields the highest profit for the global corporate structure as a whole, based on the wages, materials costs, markets, government policies, taxes and subsidies, and local markets for the relevant outputs that exist in various locations.

UNEMPLOYMENT COMPENSATION consists of regular payments made to some unemployed workers from a government insurance fund to which employers contribute.

A **UNION** is an organization of workers established with the intention of providing a unified and stronger voice on behalf of the members' interests.

UNIT LABOR COST is the labor portion of the average cost of producing each unit of output; it is the total wage bill (hourly wage times total hours worked, plus benefits), divided by the number of units of output produced.

UNIT MATERIALS COST is the portion of the average cost of producing each unit of output that is spent on materials used in producing the output, plus wear

and tear on the machines and tools used. It is therefore the total cost of materials and wear and tear on machines, divided by the number of units of output produced.

VALUE ADDED. See **NET PRODUCT** or **NET OUTPUT.**

The **VALUE OF NET OUTPUT** is the monetary value of the total output minus materials and machine costs.

VALUE OF THE DOLLAR. See **FOREIGN EXCHANGE RATE.**

VERTICAL ECONOMIC INTERDEPENDENCE exists when one person controls the labor and the products of another; it is based on a relationship of unequal advantage and command.

WAGE LABOR is work performed under the direction of an employer in return for a wage or salary.

A **WAGE-LED EMPLOYMENT SITUATION** exists when rising wages increase employment.

The (hourly) **WAGE RATE** is used here to refer to total wages and benefits paid to a worker, divided by the total number of hours worked.

WEALTH (or **NET WORTH,** or **NET ASSETS**) of a person or firm refers to the total value of the assets owned by the person or firm, minus the total value of the outstanding debts, or **LIABILITIES** owed by the person or firm.

WORK EFFORT refers to a worker's level of exertion (intensity of work) on the job during a certain period of time, such as an hour.

The **WORKING CLASS,** or **WORKERS,** are those who perform wage labor; they neither own the capital goods used in their labor processes nor command control over the labor of others.

Index

A
Accounting practices, 219–220
Accumulation
  defined, 134
  process of, 135–136
  as source of change, 134–137
Accumulation process, 135–136
Adjustable Rate Mortgages (ARMs), 463, 467–469
Affiliates, 118–119
  defined, 119
Africa, 198–199
  output, 9
African Americans, 341–346. *See also* Discrimination; Racial discrimination; Slavery
  discrimination against, 317–319
  earnings, 344
  income inequality, 327
  infant mortality, 113–114
AFSCME, 315
After-tax profit rate, 215
  defined, 215
Aggregate demand (*AD*), 393–394, 394, 435–437, 441–442
  coordination, 442
  curve, slope of, 393, 406, 418
  defined, 393
  and employment/unemployment, 383–421
  exports and imports, 435–439
  investment, and monetary policy, 410–416
  with investment constant, 395–397
  wages and unemployment, 417–420
Aggregate economic concentration, 263–264
Aggregate or population level outcomes, defined, 31
Aggregate supply (*AS*), 392–393
  defined, 392
  equal to aggregate demand, 395–399
Agricultural subsidies, 372–373
Agriculture, U.S. productivity, 4, 5
Agrochemicals mergers, 267
Aid to Families with Dependent Children (AFDC), 129, 165
AIG, 471, 473
Alexander the Great, 23
Altruistic preferences, 36
American Federation of Labor (AFL), 148
American Jobs Creation Act, 516
American South, 93–94
Antitrust, 266
  enforcement, 267–268
AOL-Time Warner merger, 269
Apple Computer, 233, 251, 262
Arapesh of New Guinea, 43, 102
Archer Daniels Midland (ADM), 252
Asia, West compared, 7–8
Asian tigers, 8
Assembly lines, 304–305, 305–306, 311
Asset markets
  bubble trouble, 477–478
  goods markets versus, 477
  price bubbles, 477–482
Asset price bubbles, 477–478
  feedback loops, 479–480
Assets, types of wealth, 334–335
Austria, 452
Auto industry, 243–247
  Brazil, 374
  decline in U.S. production, 264
  economies of scale, 361–362
  global vehicle production, 1961–2011, 362
  Mexico, 374
  self-driving car, 250
  subsidies, 263
  transnational, 375
  wage rates abroad, 251
  WTO restrictions on nations, 520
Average cost, 196–197
  defined, 175

B
Bailouts and buyouts, 473
Baker, Dean, 523–524
Bakery example, 238–239
Bangladesh, 352, 355
  garment workers, 280, 380
Bankruptcy, personal, 486
Banks and banking
  and the Federal Reserve System, 415–416
  government regulation of, 415–416, 465

interest rates and, 415
liars' loans, 467–470
long term trends, 486
predatory lending, 467, 472
supply of loanable funds, 443–446
Barriers to entry, defined, 252
Barter, defined, 170
Battuta, Ibn, 169, 352
Bay Area Rapid Transit, 507
Behaviors, 37–38
Beliefs, 31, 36
defined, 32
Belonging, sense of, 329
Bid-rigging, 493
Big Mac Index, 371
Billionaires, 104–105
Blacks. *See* African Americans
Boeing Corporation, 263
Bonds, 255
defined, 211
financing with, 218
Book publishers, 269
Borrowing, 442–447
and the exchange rate, 446–447
government, 425, 442–447
Botswana, 373
Brazil, 243–247, 263, 358–359, 362, 365, 374, 440
Breakstone/Firestone, 310–311
Breakthroughs, 237, 260
defined, 237
exploiting, 262–263
as temporary monopoly, 250–251
Britain, 185, 199–200, 439. *See also* United Kingdom (UK)
economic take-off, 6–7
real wages, seven centuries, 6, 8
Bubbles, 477–482
Budget deficit, 394
Budget surplus, 394
Built-in stabilizers
defined, 410
and unemployment benefits, 409
Bull market boats, 77
Bureaucratic control, 153, 305–307
defined, 305
Burke, Edmund, 41
Business climate
defined, 414
and investment, 413–415
Business cycle contraction, defined, 408
Business cycle expansion, defined, 408
Business cycle peak, 408
Business cycles, 143, 384–385, 428
built-in stabilizers, 407–410
dampening, 425
defined, 386, 409
and employment multiplier, 408
expansion, 386, 419
recession phase, 386
Buyouts, 473

C

Call centers, 23–24
Campaign financing, 124–125
Canada
energy program, 521
health care, 494–496
ISDS cases, 520
Canal construction, funding, 121
Capacity utilization, 248–250
defined, 249
and investment, 257–258
Capital flight, 366
Capital gains, 478–479
stock buybacks, 516–517
tax rate, 503
Capital goods, 76–78, 79, 108
capacity utilization, 249–250
concentration of ownership of, 106, 108
defined, 72
employer control and, 286
in grain growing, 79
in independent production of commodities, 138
ownership of, 106
profit and, 214
and the profit rate, 211, 220–222
saving technical change, 81–83, 83
stockholding, 336
Capital-goods saving technical change, 81–83, 83
Capitalism, 96, 354, 407–408
and change, 53–54
definition of, 2, 96, 110
effect on government, 114
firm as command economy, 275–276
inequality and, 524–525
lessons from history, 474–478
needs unemployment, 424
in neoclassical economies, 49–50
profits and surplus product, 108–110
"race" to accumulate, 136–137
regulating, 486–489
role of crisis and recession, 455
social class and, 91
stages of, 144–147
United States, history of, 137–141
Capitalist class, 138–139
defined, 107, 326
surplus product and, 110
Capitalist consumption, 78
Capitalist epoch, defined, 2
Capitalist profits, defined, 209
Capitalists, 109–110, 272, 273
competition among, 263
conflicts with workers, 276–281
defined, 107
investment versus consumption, 411
policy toward unemployment, 431
power of, 114
over workers, 114
and profit, 207–213
recession seen as good, 434
Carbon dioxide emissions, 17–18, 20–21, 205

Carbon tax, 194
Carson, Rachel, 17, 21, 22
Cartels, 253, 262
  price-fixing and other forms, 266–268
Cash crops, 100
Catholic Church, 321
CEOs, 219, 332
  astronomical pay, 166
  corporate stock buybacks, 166–167, 516–517
Change, 50, 53–54
  accumulation and, 135–136
  capitalism and, 53–54
  neoclassical economics on, 57
  surplus product and, 86–87
Charter, corporate, 115, 218
Cheney, Dick, 501
Child-rearing practices, 42
Children
  government support for, 524–525
  inequality of opportunity for, 327
  infant mortality, 496
  mortality rates compared, 356–357, 360
  poverty rates compared internationally, 525
  reproduction activities, 83
China, 355–357, 362, 373
  early technology, 26
  economic growth, 9, 11
  economic institutions, 6
  productivity and, 372
  purchasing power of yuan, 371
  technology transfer, 364
Christie's, 253
Cities, growth of, 12–15
Citigroup, 466, 484
Citizens for Tax Justice, 378
Citizens United v. Federal Election Commission, 125, 493, 501
Civil liberties, 62
Civil War, 2
Class. *See also* Class relationships
  defined, 91
  and economic systems, 93–96
  structure in the United States, 139–141
Class conflict, 87
Class relationships, 90, 272, 273
  defined, 91
  inequalities, 326–327
  precapitalist societies, 96
Class society, 90
Climate change, 17–22
Clothing sweatshops, 24, 380
Coal industry, 299–300
Coal miners, 299–300
Coca-Cola Company, 35
Cod fishing company example, 223–225, 279
Collective bargaining, 314
  defined, 280
Collusion, 252
  price-fixing and other forms, 262
Command, 52–53, 54, 68
  commercial profit-making, 210
  decision-making and, 63
  defined, 52
  employer-worker relationship, 272
  firm as system of, 273
  limitations of conventional economics, 56
  replacing competition, 118
  vertical independence, 70
Command relations, defined, 275
Commercial profits, 208–209
Commodity, 96–102
  goods and services not included, 330
Commons
  social regulation of the, 204
  tragedy of the, 203–205
Communism, 527
  economic growth, 11
  end of, 165
  end of European, 1–2
Competition, 50, 51–52, 180, 234–235, 235–237. *See also* Cooperation
  animal species, 45
  defined, 51
  dynamic, 255
  dynamics of, 258–259
  employer-worker relationship, 272
  and equalization of the profit rate, 259–261
  forms of, 237
  forms of (overview), 237
  global/international, 234, 425, 441–442
  government contracts, 507–508
  high employment policies, 435
  labor market and, 285
  limits price increases, 426
  nonprice, 254
  price, 177, 237
  for profits, 133, 136–137, 234, 235–237
  and scale of output, 261
  virtuous and vicious circles in, 361
Competition for profits, 136–137
  defined, 136
Competitive capitalism, 144–145
Competitive markets, 170, 171
  defined, 171
Complete contract, 55
Complete contracting assumption, 55, 56
Concentration of ownership of media, 269, 526
Conflict
  international, 87
  between producers and owners of the surplus, 84–86
Congo, 355, 373
Constant prices, 392
Constitutions, 41
Constraints, 31, 32–33
  defined, 32
Consumer goods, 17, 72, 73
Consumer Price Index (CPI), real wages, 426–427

Consumption (*C*)
effect of unemployment on, 408
fraction of wages spent on (*c*), 418–419
income and saving, 394–395, 396
material life, 6–11
spending, 394
Consumption of the producers, 76
conflict over, 84–86
Containerized shipping, 24
Contract
defined, 55
system of labor payment, 284–285
Contract specifying work done, defined, 284
Control
of labor, and its product, 69
of the surplus product, 87, 90
systems of, 300, 302–307, 304
Control rights, 110
Conventional economics, 49–50. *See also* Neoclassical economics
defined, 55
Cooperation
benefits of, 200–202
as human trait, 31
motives for, 45–47
use of labor, 40–41
Coordination, 181–187
by command, 184–187
problem, 408
by rules, 184–187
Coordination failure, 200–205
defined, 200
overinvestment, 476
Copyrights, 250
Corn production, 40
Corporate capitalism, 145–146
Corporate income tax, as share of all income tax revenue, 505
Corporate personhood, 116–117, 216–217
Corporate stock buybacks, 166–167, 516–517
Corporate tax inversion, 163
Corporate tax rates and payments, 503–504
Corporations. *See also* Firms; *specific corporations*
alternative rules, 523
government influence, 492
leaders of largest, 26
organization of, 115–119, 217
original charters, 492
overview of regulatory capture, 484–485
profit rate, 215–216
relationship with government, 492
transnational, 118–119
Cost of job loss (*cjl*), 286–287
in a segmented labor market, 453
avoiding, 288–295
defined, 286
Cost of materials input, 220–221
Costs
private, 194
social, 194
Countercyclical policies, defined, 408
Cox's capper, 320
Craft union, 147–148, 148
Credit Default Swaps (CDSs), 471
Credit Mobilier, 122–123
Crowding out, 442
Cuba, 355
Cultural differences, 31
human nature, 36–38
Culture, 38
Current prices, 392

D

Darwin, Charles, 46–47
Day care experiment (Israel), 29–30
Debt after economic crises, 482–483
Decentralization in invisible hand, 190
Decreasing average costs, 175
Default rates, 466, 469
Deficit spending (*B*), 442, 446–447
defined, 400
effects on employment, 400, 406
Delinquency and default rates, 466, 469, 470
Delinquency rates, 466, 470
Delta Airlines, 149, 314
Demand
derived demand, 417
excess, 389, 391
unemployment and, 386
Demand curve, 171–173
defined, 171
location of, factors determining, 180
movement along, 178–179, 180
shift of, 178
Democracy, 62–64
defined, 62
inequality and, 517–518, 525–526
and political economy, 50
role of government, 22–23
and three-dimensional economics, 62–64
twentieth-century, 128–130
Democratic firm, 300–301
defined, 322
Democratic government, 488
collective provision for general welfare, 119
Denmark, 369
Depreciation, 72, 75, 411
defined, 75
Depression. *See* Great Depression; Great Recession; Recession
Deregulation
debt and, 483
financial fragility and, 482–486
political power and, 475
transnational capitalism, 163–165
under Reagan, 129–130, 163–164

Derivatives, 164
due diligence, 487
MBSs as, 455
Derived demand, 417
Deskilling, 306, 312–313
Determinants of the profit rate, 208, 236
defined, 220
intermediate inputs per unit of output, 226–227
total profit and, 220–223
unit labor cost as a, 227–229
Direct investment, 378–379
Discrimination, 153, 317–319
conventional economic view of, 318–319
defined, 317
differential pay, 342–346
effect of capitalism on, 317–319
effects of, 341–349
hiring, 342
income levels, 326, 327
job, 153
racial, 153, 317–319, 341–349
sexual, 317–319
Dissaving, 394–395
Dividend, defined, 109, 212
Division of labor, 69–70
defined, 69
Dodd-Frank Act, 456, 488
loophole eating the rule, 489
Drive system, 304
Drugs. *See* Pharmaceuticals
Due diligence, 487
Dynamic cost advantage, 258–259
Dynamic equilibrium, 234–235

E

Earned income, 104
Earnings, 344
gaps, 344–345
Earnings stripping, 504
eBook industry, 196–197
Economic concentration, 234, 235, 261–268
defined, 261
industry trends, 265
Economic crises, 454–455
lessons from history, 474–478
overinvestment, 476
overview, 475
regulation after, 482
underconsumption, 476–477
Economic development
foreign investment and, 373–380
other factors, 369–373
requisites for, 366–369
Economic growth
defined, 330
and poverty in the global economy, 352–381
Economic institutions
for achieving full employment, 447–452
China and Britain, 6–7
importance of, 381
patterns of interaction, 42
preferences, 42–45
Economic interdependence, 69
"Economic man," 55
cooperation and, 46
defined, 30
neoclassical view, 55
reconsidered, 34–36
Economic models, 78
Economic specialization. *See* Division of labor
Economic success, elements of, 338–341
Economic systems
and capitalism, 50–51
class relationships and, 90
and classes, 93–96
defined, 50
distinctions among, 94–96
and efficiency, 59–60
and fairness, 60–62
three-dimensional economics and analysis of, 49–65
in the United States, 137–141
Economics
aggregate or population level outcomes, 31
defined, 54
interpretation in, 58
Economies of scale, 56–57
defined, 175
falls in average cost, 175
labor cost per unit of output, 226
large firms, 263
as market failure, 196
market failure and, 197–198
other advantages of large size, 247–248
political power, 483–484
price competition, 243–247
transnational corporations, 361–362
Economy
defined, 70
labor processes and, 70
Ecosystem. *See also* Environment
threats to, 17–22
Ecuador, 374–375
Edgeworth, F. Y., 34
Education
earnings gaps, 346
female, progress, 360
income and wealth, 336, 337
international comparisons, 356–357
progress, 360
socializing role, 42
workforce, 368
Efficiency, 301, 302
defined, 59
and three-dimensional economics, 50, 59–60
Efficiency of labor (f), 228, 236
Ely, England, 89–90, 91, 137
Eminent domain, 520
Emissions rights, 205
Employee compensation, 221–222
Employees. *See also* Workers
differences among, 327
Employers, 272, 300–301. *See also* Capitalists
conflict between workers and, 276–281
differences among, 326–327

power over workers, 115
systems of control summarized, 307
Employment, 430, 439–441
and aggregate demand, 383–421, 395–399
construction sector, 1930–2020, 461
contract, 282
discrimination, 317–319
during 1990s boom, 433–434
equilibrium level of, 399, 401–403, 406
defined, 390
federal, 497–498
full, 389, 406, 421, 447–452
globalization, impact, 156–165
historical trends, 132–133
job loss, 154–155
multiplier, 404–407
shifts in production, 159–161
Employment effect
defined, 404
on net exports, 436–437
Employment multiplier, 404–407
defined, 405
size of, 408–409
Employment situation
profit-led, 419
wage-led, 418, 419–420
Energy
-using sectors, 236
geothermal, 204
Enron Corporation, 220
Environment, 17–22, 60, 63
inputs from nature, 74
negative externalities, 194, 198
Equal opportunity, 61–62
Equality, 60–61
Equality of opportunity, 326, 338–341
Equalization of profit rates, 235, 259–261
Equilibrium, 234–235, 237–238, 298, 389, 399
defined, 176
macroeconomic, 389, 398
output, 399
Equilibrium level of employment, 390, 399, 401–403, 406, 432
Equity, 464
Ethics, 46
Euro, 438
Europe, 352
economic take-off, 6–7
precapitalist epoch, 3
European Union (EU), 262, 266, 268
Excess demand, 389, 391
defined, 176
Excess supply, 389, 390–391, 398–399
defined, 176
in the labor market, 417
Exchange rate, 437–439
and borrowing, 446–447
Expected capacity utilization, 258
Expected profit rate, 255–257, 411–412
and business climate, 414
defined, 257
and investment, 257, 411
Expected profit rate on investment, defined, 257
Exports
demand for, 436
imports and aggregate demand, 435–439
worldwide growth, 25
Expropriation, 520–521
Externalities
defined, 56
market failure and, 193–195
negative, 193–194
Extraction of work from workers, 282–283
defined, 283

F

Fair Labor Standards Act of 1938, 125
Fairness, 330
defined, 60
and three-dimensional economics, 50
Fallback wage, 288–289, 289–290, 428
defined, 288
Family
background and economic success, 338–341
transformation of, 16–17
Fannie Mae and Freddie Mac (GSEs), 465, 467, 471
Fast food industry, 304
Faulkner, William, 271
Federal Deposit Insurance Corporation, 488
Federal employment, 497–498
Federal Reserve System, 415–416, 447
Dodd-Frank Act, 488
Feedback loops, 479–480
inequality and power, 519
reflexivity, 480–481
taxation and regulation, 493
Feudalism, 50, 89, 91, 94
defined, 94
production and consumption, 136–137
Finance, Insurance, and Real Estate (or FIRE) sector, 484
Financial crises, 454–455
Financial fragility
defined, 483
deregulation and, 482–486
Financial fragility hypothesis, 456, 475–476
Financial sector
campaign contributions, 484, 501
deregulation, 164–165
Dodd-Frank Act, 488
inequality, concentration, and crisis, 488–489
lobbying by large firms, 484, 501–502
regulating, 486–487
Financialization, 163–165
Financing development, 366, 367–368
Finland, 149

Firms, 417–418
advantage of large, 261–268
capacity utilization, 248–250
capitalist, as a command economy, 275–276
capitalist compared to democratic, 323
competition, 236–237
in competition, 250, 254
debt and investment, 411–412
democratic, 322–323
entry and exit of, 260–261
influencing the rules, 500–506
and investment, 254–258
monopoly power or shared monopoly, 252–254
profit rates of, 259–261
research by, 262
types of organization, 115–117, 216–217
First copy costs, 196–197
Fiscal policy, 406, 410, 425, 442–447
defined, 400
and unemployment, 400–407
Fishing, 203–204
Fixed costs, 221
defined, 243
overhead, 243–244
Flourishing life, 58–59
Flow, 334
Food insecurity, 100–101
Ford Camaçari auto assembly plant, 243–247, 263
Ford Motor Company, 375
Foreign aid, 367
Foreign exchange rate, 437–439, 446–447
defined, 438
Foreign investment
development and, 373–380
direct investment, 376
U.S. direct investment abroad, 2014, 379
Foreign markets, access to, 369
Fossil fuels, 19, 20
Four-firm concentration ratio, 261
France
medical systems compared, 491–492, 494–496
unemployment rate, 424
Franklin, Benjamin, 16
Free trade agreements, 361. *See also* North American Free Trade Agreement (NAFTA); World Trade Organization (WTO)
Freedom
economic, 328–329
elements of, 328–329
Full employment, 421
defined, 389
institutions for achieving, 447–452
political obstacles to, 434
unit cost, output and profit rate, 432

G

Garment workers, 280, 380
Gates, Bill, 15
Gender, income and wealth, 336–337
General Agreement on Tariffs and Trade (GATT), 162
General Dynamics, 263
General Electric (GE), 253
General strike, 148
Genetic influences, 38, 340, 341
Germany, 1–2, 373
fascism and unemployment, 450
health care, 494–495
tariffs, 19th century, 439
unemployment rate, 424
Gerrymandering, 526
Gilded Age, 122
Glass-Steagall Act, 465, 466
repeal, 130, 466, 485
Global warming, 17–18, 194, 202–203
Globalization, 23–26. *See also* Transnational capitalism
deregulation, 163–165
disadvantages, 25–26
fragmenting production, 161
rules for the global economy, 162–166
unions, 303
U.S. capitalism today, 155–165
wage levels compared, 353
winners and losers, 160
Gompers, Samuel, 314
Goods markets versus assets markets, 477
Government(s)
borrowing, 425, 442–447
and capitalist development, 22–23
consent of governed, 22
contracts to large firms, 263
countercyclical policies of, 408
demand (G) for goods and services, 394
democratic, 114, 488
discretionary spending, 497
distribution of income, wealth, and welfare (history), 120–125
effect on business climate, 413
expenditure, 113–114
fiscal policy, and unemployment, 400–407
functions, 130–131
influencing the rules, 500–506
military contracting, 506–507, 522
military spending, 498, 499, 501–502
monetary and fiscal policy, 442–447
need for well-functioning, 368
new roles for, 22–23
nondiscretionary spending, 497
other interventions to promote development, 366
outlays, as share of GDP, 497
policies, 247, 420
policies affecting capital owners or wealthy, 123–124
policies affecting workers, 123

proposals for alternative rules, 522–524
relationship with citizens, 492
role in technology transfer, 363
as rule enforcer, 267–268, 504–505
rules of organization, 114–119
rules that affect business costs and profits, 506
in the social structure of accumulation, 146
spending, 400, 406, 410
spending, compared internationally, 496–498
subsidies to large firms, 263
too big?, 496–500
trade policy, 439–440
transnational corporations, 374–375
as vehicle for promoting general welfare, 120
Government contracts, 493
bidding processes, 508
contracting out services, 506–507
decline in competitive bidding, 508
maximizing profit, 512–514
medical services in private prisons, 514–515
private prison services, 508–516
profits and the public interest, 507–508
revolving door, 501
Government intervention
benefits to business and public, 122–123
child welfare and, 524–525
Grain, used as example, 162
Grain growing, 79, 213–214
trends, 158
Great Depression, 454
corporate capitalism and, 146
costs of, 454–455
as personal disaster, 129
profits, 207
Great Recession, 130, 383. *See also* Housing; Mortgages
bankruptcy, 486
causes, 164, 165
collateral damage, 470–474
liars' loans, 467–470
mortgage-backed securities (MBSs), 465–466
poverty, 473–475
prime and subprime mortgages, 463
significance, 454
subprime crisis, 456–458
understanding, 458–474
unemployment, 473–474
Greenhouse gases (GHG), 18, 19
Gross domestic product (GDP), 391
Gross National Income (GNI), 391
Gross national product (GNP), 391

H

Hadza people, 38
Haiti, 355
Halliburton, 263, 303, 501
Hamilton, Alexander, 366
Handshakes, 450–452
Harambee system, 39
Harlan County, Kentucky, 300
Hayek, Friedrich A., 187
Health, economic aspects of, 328
Health care, 328
compared internationally, 491–492, 494–496
endemic disease and, 198–199, 359, 361
private prisons, 514–515
single-payer, 491–492, 496
Health insurance mergers, 267
Hierarchy
defined, 302
markets and, 321–322
workplace, 302
in the workplace, 302
High-employment profit squeeze, 425, 426–434, 431–434, 441, 448, 450
defined, 431
High-employment wage push, 428–429
Hispanics. *See also* Discrimination; Racial discrimination
workers, 153
History, 54
democracy in twentieth century, 128–130
neoclassical economics on, 57
HIV/AIDS, 365
Hobbes, Thomas, 182
Hobsbawm, Eric, 15
Homeownership, 336
Homo economicus. *See* Economic man
Honda, 251
Horizontal dimension, 50, 51–52
coordinating economic activity, 183
Horizontal economic interdependence, 69
Hospital equipment example, 249–250
House price bubble, 455, 477
Household production, 15, 98, 132–133, 330
early 19th-century, 183
not measured in total output, 391
Households
described, 16–17
income and wealth, 336–337
incomes in U.S., 333–334
as independent producers, 132
as productive units, 16–17, 44, 132, 132–133, 330
saving and net worth, 394–395
wealth in U.S., 336
wealth ownership, 334–335
Housing, 334
annual sales of new single-family homes, 460
delinquency and default rates, 466, 469, 470

Housing (*continued*)
employment in the construction sector, 1930–2020, 461
equity, 464
feedback loops, 479–480
Great Recession, 458–474
how to buy a home, 462
mortgage interest rates, 1971–2016, 460
nominal house price index, 471
progress, 360
real home price index, 1890–2014, 459
refinancing mortgages, 463–465
rising prices before Great Recession, 459–461
why to buy a home, 458–459
Human capital, 335
Human nature, 31
cooperation, 46–47
cultural differences, 36–38
Hume, David, 30
Hunter-gatherer societies, 37–38, 44
Hyundai, 361–362

I
Ibn Battuta, 40, 169
Idle capacity, 383
Immigration
for-profit correctional facilities, 513
globalization, and, 157–158
opportunity, 338
trends, 157–158
Imported inputs, real price of, 84, 85
Imports, 425
demand for, 436
exports and aggregate demand, 435–439
Incarceration rates, 508–512
Incentive structure, 455
changing, 487
Income
compared internationally, 355–357
consumption and saving, 394, 396
cross-country comparison of, 355–356
defined, 104
distribution. *See* Income distribution
effect on net exports, 436
as flow, 334
international comparisons, 360
living standards and, 330
measuring well-being, 330–331
parents' and children's related, 338
productivity and, 359–366
sources analyzed, 107
types of, 211–213
universal basic income (UBI), 523
voter turnout related, 127–128
wealth compared, 334
Income distribution
among countries, 354, 357, 358
effect on aggregate demand, 386
equality of, 372
inequality, 327, 331–334
measuring inequality, 330–331
in quintiles, 331, 333–334, 334, 337
top one percent, trends, 331
Income inequality, 327
quintile shares, 333–334
United States, 331–332
Income rights, 110
Income tax
rates under Reagan, 30
trends, 503
Income transfers, 36
Incomplete contracts, 55, 284
Incomplete labor contract, defined, 284
Independent primary labor market, 153–154
Independent producers, 64, 283
Independent production of commodities, 138–139, 139, 141
defined, 138
India, 184, 349, 355–356
call centers, 23–24
comparative inequality, 358
environment, 21
imperialism in, 7–8
Indigenous populations, 13
Indonesia, 37, 418
Industrial protection, 366
Industrial unions, 148
Industrial Workers of the World (IWW), 148, 310, 311
Inefficient technical change, 321
defined, 319
Inequality, 325–327, 325–351
alternative rules, 522–523
capitalism and, 524–525
democracy and, 517–518, 525–526
efficiency impeded by, 372
feedback loops between power and, 519
global, 11–12
growing, 331–334
impact of crises on, 488–489
of income, 327
measuring, 330–331
of opportunity, 338–341
racial discrimination, 341–346
trends, 331–334
uneven development, 354
within and among nations, 354, 357, 358
Infant mortality, African Americans, 113–114
Inflation, 448
Information, 181, 187
market failures, 192
and motivation, 186
prices as decentralized system of, 190
Infrastructure, 369
Innovation, 262. *See also* Technical change
competition, 133
Inputs, 59, 70, 71

Intellectual property rights (IPRs), 162, 364–365
WTO role, 519
Intelligence and economic success, 340
Intensity of labor (*e*), 85
defined, 83
Intensity of work (*e*), 286, 289, 291. *See also* Intensity of labor; Work effort
Interactions, nature of, 119
Interest, 212
defined, 109
Interest rate (*i*), 255, 414
and investment, 411–413
mortgage interest rates, 1971–2016, 460
what determines, 443–446
Intermediate goods, 223
Intermediate inputs, 221, 221–222
per unit of output as profit rate determinants, 226–227
Internal Revenue Service (IRS), 504–505
International economy, 354. *See also* Globalization; Transnational corporations
International trade, and macroeconomic policy, 439–442
Investment (*I*), 254–258
aggregate demand, and monetary policy, 410–416
and the business cycle, 408–409
capital goods, 228
capitalists' power to attract, 119
to compete, 235
and competition, 133–134, 258–259
costs of, 254–257
defined, 78, 411
direct, 378–379
effect of employment, interest rate and business climate, 414
effect of lower profit rate on, 426, 434
effect of wage increase on, 417–418
and equalization of profit rates, 261
expected profit rate on, defined, 257
expected return from, 255
falling productive, 166–167
interest rate and business climate, 414–415
part of aggregate demand, 394
profit effect on (*f*), defined, 413
spending, effect on employment, 407
of surplus product, 133–134
in tax havens, 354
total cost of, defined, 256
what determines, 416
Investment banks, 488
Investor-state dispute settlement (ISDS), 494
TNC power, 520–521
Invisible hand, 182–183, 187–191, 450
problems with, 191, 200
Involuntary unemployment, 297
Iowa Beef Packers (IBP), 305–306
IQ, influences on, 340
Iran, 369–370
Ireland, 494–495
Islam, 23, 26

J

Japan, 355–357, 362, 363
natural resources, 369
output, 7–8
trade policy, 440
unemployment rate, 424, 448
Job ladders, defined, 307
Job segregation, 346–347
Johnson, Lyndon, 113
Just price, 35

K

Kahneman, Daniel, 35
Kenya, 38, 39, 355–357
Keynes, John Maynard, 400, 421, 479
Keynesian policies, 146, 400–401
King, Martin Luther, Jr., 113
Kohn, Melvin L., 42–43
Korea. *See* South Korea
Kyoto Protocol (1997), 19–20

L

Labor
defined, 73
household, 60
intensity of, 83
time, 83, 85
wage and, 283–284
Labor accord, 146, 151–152
rise and fall of, 151–152
Labor discipline, 282–285
Labor effort (*e*), measuring, 60
Labor extraction, unemployment and unit labor cost, 426
Labor extraction curve, 290, 293
profit maximization and, 291, 293
Labor-management relations, 283
Labor market, 273, 281–282, 285, 297, 398, 448–449
defined, 285
discrimination in U.S., 2001–2002, 343
excess supply in, 417
Labor processes, 68
capitalist, 300
and capitalist profits, 209
and class, 93
defined, 70
pancakes example, 73, 97–98
social organization of, 300
and technology, 307–309
Labor-saving technical change, 81–83
defined, 81
Labor supply (*LS*), 398
Labor time, defined, 282
Labor unions. *See* Unions
Laissez-faire, 328, 450
defined, 183
Lamalera whale hunters, 37, 38–40, 41

Land grants, 121
Large firms, periodic crises and, 483–485
Latin America, output, 9
Latinos. *See* Hispanics
Layoff, 398
  defined, 154
Lehman Brothers, 473
Leisure time, 330
Liabilities, income inequality and, 334
Liars' loans, 467–470
  risky loans, 480
Life expectancy, 355–356, 360
Lighting technology, 10–11
Limited liability, 115–116, 217–218
  defined, 115
Living standards, 327, 330. *See also* Well-being
  compared internationally, 354, 355–357, 358–359
  during 1990s boom, 428–430
  elements of, 327
  evolution of, 6
  globalization and, 353
  inequality and, 355
  measurement, and inequality, 330–331
  measuring, 329–331
  qualitative improvements, 10–11
Loanable funds, supply of and demand for, 443–446
Lobbying by large firms, 248, 483–484, 489, 501
  conflict over rules, 521–522
Lockheed Martin, 263, 267
Lockout, 301
Long swings (or long waves), 143, 384, 384–385
  defined, 143
Ludlow Massacre, 300
Luxury goods, 73–74, 77, 87

M

Macroeconomic equilibrium, 398
  business climate, 413–415
  with deficit spending, 402
  defined, 389
Macroeconomic policy, and international trade, 439–442
  coordination among countries, 441–442
  dilemmas, 422–453
Macroeconomics, 389
  defined, 64
Madoff, Bernie, 482
Management, 219, 283
  rights, 309
Managers, 141
Manufacturing, transnational comparisons, 376
Marginal cost, 196, 197, 244
  defined, 174
Mark-up, 241–242
Market, 51, 170–171
  assumptions about, 55–57
  defined, 170
  exchange, 57, 275
  and hierarchy, 300, 321–322
  message and motivation, 189
Market clearing, 188
  lack of, in labor market, 298
Market clearing price, defined, 176
Market exchange, defined, 170
Market failures, 182, 192–200, 200–205
  coordination failure and, 200
  defined, 191
  house price bubble as, 455
  when they occur, 198
Market integration, 41
Market power, 196
Market share, defined, 254
Market socialism, 101
Marx, Karl, 41, 135
  vertical dimension, 70
Material life, 6–11
Materials, defined, 72
Materials cost, 236, 430, 433
  push, 426, 430
  unit (*umc*), 430, 433
The Matrix Revolutions debut, 23
Meat packing industry, 305–306
Media industry concentration, 269, 526
Medicare drug program, 124, 500–501, 508
Mega-yachts, 77
Mergers and acquisitions, 264, 267
  moral hazard, 456
Message, excess demand as market, 190
Metalclad, 520
Mexico, 355–356, 362, 367, 374, 520
Microeconomics, defined, 64
Microsoft, 233
Middle class
  middle classes defined, 140
  new and old, 140–141
Middle income, 333
Migration, 13
  globalization, and, 157–158
Military contracting, 506–507, 522
Military interventions, costs of, 226–227
Mill, John Stuart, 40
Millennium Development Goals, 359
Minimum wage
  reasons for raising, 420
  trends, 149
Minorities. *See* African Americans; Discrimination; Hispanics; Racial discrimination
Minsky, Hyman, 456, 475, 482
Molly Maguires, 300
Monetary policy, 415–417, 425, 442–447
  defined, 410
  investment, and aggregate demand, 410–416
Monopoly
  as market failure, 196
  trusts and risk, 117–118
Monopoly power, 237, 252–254, 260, 261–268
  defined, 237
Monsanto, 162, 364–365

Moral hazard, 456
AIG, 473
too-big-to-fail firms, 483
Mortgage-backed securities (MBSs), 465–466, 471, 477
Dodd-Frank Act, 488
Mortgages
Adjustable Rate Mortgages (ARMs), 463
delinquency and default rates, 466, 469, 470
home buying, 462
interest rates, 1971–2016, 460
liars' loans, 467–470
mortgage-backed securities (MBSs), 465–466, 471
prime and subprime, 463
prime and subprime by year of origination, 468–469
refinancing, 463–465
second, 462
securitization, 462–463
underwater, 471
Mosaic, 327
Motivation, 185, 186, 188–189
and information, 186, 188–189
price as market motivation, 189–190
provided by market, 185, 189–190
self-interest, 30
Multiplier, employment, 404–407, 408–409
Multiplier, output, 406–407, 407

N

National Bureau of Economic Research (NBER), 454
National Labor Relations Board (NLRB), 148–149
National Nurses United, 315
Native Americans, 137
Natural environment. *See also* Environment
depreciation of, 392
destruction of, and capitalist development, 17–22
and efficiency, 60, 63
Natural resources, 369
as development factor, 369
Necessary product, 75–77
defined, 76
Negative externality, 193–194
Neoclassical economics, 55–57, 234, 318–319
contrast with political economy, 65
defined, 55
economic man, 34
preferences, 33–34
self-interest, 34
three assumptions summarized, 57
view of competition, 234
view of discrimination, 318–319
view of power, 54
Net assets, 104, 336
Net domestic product (NDP), 392
Net exports (X), 435–436
defined, 394
and employment, 436
high employment effect, 448
promoting, 439–441
Net investment, 110, 411
Net National Income (NNI), 392
Net output, use of concept, 392
Net output per hour of labor, unemployment and, 388–389
Net product, 75–76, 76, 79
defined, 76
Net worth, 104–105, 336
defined, 336
New Deal, 129
New middle class, 326
defined, 141
New World, 3
raw materials of colonies, 7
New York Stock Exchange, 171
Newton, Isaac, 57
Ninja loans, 467
Nixon, Richard, 211
Nonfinancial sector, crises in, 455–456, 476–477
Nonmarket work, 98–100
Nonprice competition, defined, 254
North American Free Trade Agreement (NAFTA), 130, 162–163, 520, 521
Norway, 452

O

Obedience and discipline, 42
Occupational Safety and Health Act of 1974, 125
Occupations
gendered, 346
preferences related, 37
shifts, 13
and women's and men's wages, 346–348, 347
Office of the Comptroller of the Currency (OCC), 472
Office of Thrift Supervision, 485
Oil and petroleum, 117–118, 246, 430
impact on development, 369–370
Oil industry, prices, in 1970s, 236
Okun, Arthur, 372
Old-age pensions, 498–500
Old middle class, 140–141, 326
Oligopoly, defined, 237
Olympia Veneer Plywood Cooperative, 102–103
On Human Work, papal encyclical, 275
Opportunity, equality of, 338–341
Opportunity cost, 274
defined, 256
Organization of Petroleum Exporting Countries (OPEC), 236, 253–254
Orma herders, 38, 39
Other-regarding behaviors, 30
Output, 59, 70, 71. *See also* Net output; Total output
produced per unit of work effort, 309
Output multiplier, 406–407, 407

Outsourcing, 23–24, 353
Overhead, 244
Overinvestment, 455–456
  as cause of crises, 476
Ozone layer depletion, 22

P

Panic of 1873, 454
Panic of 1893, 454
Papal encyclical, 275
Papau-New Guinea, 37
Paraguay, 37–38, 39
Parallel plants, defined, 301
Pareto optimality, 60
Paris Climate Agreement (2015), 20
Partnerships, 115, 217
Patents, 364–365
Paulson, Henry (Hank), 484–485
Peak, of business cycle, 408
Pensions, 498–500
Per capita GDP
  real, 392
  world over time, 9
Per capita output, 392
Permafrost, 20
Personality traits, 44–45
Peru, 37, 40
Pharmaceuticals, 198–199
  alternative rules, 523
  compulsory licenses, 365
  expenditures per capital compared internationally, 495
  intellectual property rights, 365
  Medicare Part D, 124, 500–501, 508
  mergers, 267
Piece rate system, 284–285
  defined, 284
Planning, economic, 187
Plant closing, 63
Political business cycle. *See* Business cycle, political
Political economy, 49. *See also* Three-dimensional economics
  contrasted with neoclassical economics, 65
  defined, 49
  surplus product, 68
Political expenditures, 412
Political issues
  control of surplus product, 87
  lobbying by large firms, 248, 483–484, 489, 501–502
  ownership of capital goods, 110
  rules and, 526
  women politicians, 349
Pollution reduction, 22
Ponzi scheme, 482
Poor countries, 355–357
  disadvantages with economies of scale, 361
  technology gap, 363–364
Poor Laws, 186
Poor people, 186
  and economic success, 338–341
Population
  effects of growth, 26
  growth of cities, 12–15
  world growth, 3
Population explosion, 12–13
Positive externalities, 195
Poverty
  data for families, 337
  and economic growth in the global economy, 352–381
  global, 12
  government support for children, 524–525
  Great Recession, 473–475
  in U.S., 352
Power
  in a democracy, 62
  A has power over B, defined, 52
  control of surplus, 87
  of employers over workers, 115, 285, 425, 426, 449–450
  feedback loops with inequality and, 519
  global economy, 25–26
  importance of recognizing, 54
  ownership of capital goods, 110
  political, of large firms, 263, 475, 483–484, 489
Power gap, 354
Precapitalist epoch, 2–3, 407
  competition, 135
  Europe, 3
Predatory lending, 467, 472
Preference for reciprocity, 35–36
Preferences, 31, 35–36, 38
  defined, 32
  economic institutions, 42–45
Price
  of capital goods (P_c), 236
  competition, 237
  as decentralized system of information and motivation, 190
  goods and services and well-being, 330
  of imports in foreign currency, (*Pimf*), 438
  invisible hand and, 193–194
  just price, 35
  market clearing, 176, 180
  variables listed, 529
Price competition, 238–250
  defined, 237
Price leadership, 237
Price of imports (*Pim*), 84, 85
Primary labor market, 153–154, 452
Prime and subprime mortgages, 463
Principal, 255
Prisoner's Dilemma, 200–202
  global warming and, 202–203
Prisons. *See* Private prison services
Private costs, defined, 194
Private prison services, 508–516
  changing the rules, 515–516
  medical services, 514–515
  rethinking, 511–512
  sentencing laws and norms, 509–511
Private property
  among the Arapesh, 102
  defined, 102
Private property solution, 204

Privately owned capital goods, 96, 102–107
Production, 70–71
  defined, 71
  process, class nature of, 449
  recent shifts, 159–161
Production-reproduction cloverleaf, 72, 183
Productivity
  of democratic firms, 322
  development and the surplus product, 370
  early modern Europe, 7, 9
  and income, 281
  income and, 359–366
  inequality and, 354, 372
  rates differing, 371–373
  vicious circles and, 359, 361
Profit (*r*), 109–110, 207–232, 430–434, 433
  after-tax profit rate, 215
  capitalist, 209–211, 211
  changing strategies, 142–143
  commercial, 208–209
  competition and, 133
  defined, 110
  effect on investment, 413, 419–420
  exchange and command in profit-making process, 210
  financial sector, 164
  finding selling price at which profit is highest, 240
  government contracts, 512–514
  how spent, 109–110
  international shifting, 155–156
  from investment, 256
  labor extraction curve, 291, 293
  output and wages, 291–297
  and the production process, 209–211
  profit-maximizing price, 241
  reason for existence, 207–213
  social structure of accumulation, 142
  squeeze, high-employment, 426
  surplus as, 91
  and surplus product, 109
  uses of, 412
Profit effect on investment (*j*), 419–420
  defined, 413
Profit-led employment situation, defined, 419
Profit per unit of output, 224–225
Profit rate (*R*), 213–215, 220–232, 277–278, 426, 431–434
  analyzing conflicts, 229–232
  conflicts summarized, 225–232
  defined, 213
  determinants of, 220–223
  and employment level, 450
  equalization of, 235, 259–261
  equation, 222, 239
  expected, and investment, 412
  on investment, 254–257
  labor determinants of, 227–229
  at Philip Morris, 214–215
  and technical change, 307–309
  in the United States, 1948–2014, 215–216
Profit shifting, 377
Profitability
  defined, 319
  efficiency and, 59
  versus efficiency, 319–321
Property income, 104, 211
  calculating U.S. total, 212
  defined, 104
Property rights, 102, 103–104
  Arapesh, the, 43
  defined, 93
Proprietorships, 217
Public borrowing, 368
Public good, 30
Public health, development and, 369
Purchasing Power Parity (PPP), 371

Q

Qualified Residential Mortgages, or QRMs, 489
Quality of life, measuring, 329–330

R

Race, income and wealth, 336
Racial discrimination, 153, 317–319, 341–349
  biology or history, 341
  earnings gaps, 342–346
  inequality, 341–346
  labor market (U.S.) 2001–2002, 343
  legislation outlawing, 343
Railroad financing, 121–123, 122
Rana Plaza garment factory fire, 380
Rate of profit. *See* Profit rate
Ratios, variables listed, 530–531
Reagan, Ronald, deregulation, 129–130, 163–164
Real, or inflation adjusted prices, 392
Real Home Price Index, 459
Real income, 426
Real per capita GDP, 392
Real price of an imported input (or real price of imports), 84, 85, 437
Real wage
  defined, 426
  money wage and, 426–428
  purchasing power, 427
Recession, 386
  business cycle contractions as, 408
  wholesome?, 434
Redistricting congressional seats, 125
Refinancing, 463–465
Reflexivity, 480–481
Regulated capitalism, 145–147, 146–147
Regulation
  in capitalist economy, 486–489
  financial crises and, 455
  overview of regulatory capture, 484–485
  preventing misinformation, 481–482

Regulatory capture, 481
overview of activities, 493–494
Reich, Robert B., 523
Remittances, 367–368
Renewable resources, 74
Rent, defined, 211
Reproduction, 39
defined, 71
Respect, 329
Retained earnings, 213
Revolving door, 484–485
drug industry, 501
feedback loops, 493
government contracting, 501
Rice, 40
Rich and poor, 353, 358–359
inequality of opportunity, 338–341
United States, 331–332
Rich and poor countries, 11–12, 373
Roaring Twenties, 332
Robinson, Joan, 272
Roderick, David, 302
Roman Empire, 23
Rubin, Robert, 208, 484–485
Rules and institutions
national and international, 354
overview, 381
WTO restrictions on nations, 364
Rules of the game, 96, 111, 114, 115, 136, 142
alternative, 522–524, 526
conflicts over, 522
contention between labor and capital, 124–125
corporate interests and, 502–505
enforcement of, 504
firms influencing, 500–506
global corporations, 162–165, 354, 364
Runaway plants/shop, 301, 303, 353
Rural-to-urban migration, 13, 15
Russia, 363

S
S&L crisis, 485
Safety net. *See* Social safety net
Sales effort, defined, 254
Samsung, 251
Samuelson, Paul, 272
San Francisco, California, 204
Sarbanes-Oxley Act, 220
Saving, 394, 419
defined, 394
income, and consumption, 394, 396
Savings and loans deregulation, 129–130
Schiffrin, André, 491
Schools, economic analysis, 98
Science, 26
Second mortgages, 462
Secondary labor market, 155, 452
Securities and Exchange Commission, 454
Segmented labor markets, 152–155, 327
and cost of job loss, 453
independent primary labor market, 153–154
primary labor market, 153–154
secondary market, 155
subordinate primary market, 154–155
Segregation. *See* Discrimination
Self-direction, 42
Self-interest, 34–36
defined, 30
knavery related, 42
neoclassical economics, 34
in organizations, 45
Sen, Amartya, on economic man, 34
Seniority, 306
Separation of conception from execution, defined, 313
Serfs, 94
Servants, family economy and, 132
Sexual discrimination, 153, 317–319, 346–351
Sharecropping, 40
Shared monopoly, 252–254
defined, 237
sales effort, 254
Shepherd, Mark, Jr., 211
Sherman Anti-Trust Act, 118
Silent trade, 169
Simple control, 303, 304
Single-payer health care, 491–492, 496
Sit-down strike, 149
Skilled workers, 16
Slavery, 2, 64, 93–94
and class, 91, 93–94
commodity production, 101
control of surplus through, 90
defined, 93
in the early U.S. economy, 137–138
family economy and, 132
southeast Asian seafood industry, 192–193
and U.S. class structure, 139
Sleeping sickness, 198–199
Small-scale societies, 37–38, 42
Smith, Adam, 67, 70, 208, 450
coordination, 182
invisible hand, 182, 187–191
moral emotions, 46
self-interest, 30, 188
Social costs, defined, 194
Social norms, 40
Social organization of the workplace, defined, 302
Social relationships, 50
Social safety net, 22
during crises, 456
weakened, 165, 452–453, 485–486
Social structure of accumulation (SSA), 134, 141–144, 384, 425
changing nature of, 142–143
consolidation and decay, 143–144
defined, 142
key relations in the, 145
need for viable, 167
Socialism, 526–527
state, 50

Socialization, 42, 44
Sole-source contracts, 493
Soros, George, 480–481
Sotheby's, 253
South Korea, 247, 263, 361–362, 373, 440
  government strategies, 366
  technology transfer, 363
Soviet Union (former USSR), 181
  collapse, 165
  economic growth, 11
  surplus, 68
Special-interest groups, 205
Specialization, economic. *See* Labor processes
Speedup, 236, 310, 311, 312
  defined, 310
Spillover effects. *See* Externalities
Spitzer, Eliot, 472
Sri Lanka, 24
Stability and peace, 366, 367
Stages of U.S. capitalism, 144–147
Stagnation, 476–477
Standard of living. *See* Living standards; Quality of life; Well-being
Standard Oil of New Jersey (SONJ), 118
Standard Oil Trust, 117, 253
Static equilibrium, 235
Steel making, 302
Stiglitz, Joseph E., 522–523
Stock market, 422
  herd behavior, 479
Stock ownership, 105–107
  corporate stock buybacks, 166–167, 516–517
Stockholding, 336
Stocks, 334, 336
  appreciation, 219
  defined, 213
  early issues, 115
  financing with, 218–219
  list of variables, 530
Strike, sit-down, 149
Subordinate primary labor market, 154–155
Subprime crisis, 456–458, 463
  liars' loans, 467–470
  misinformation and fraud, 482
  prime and subprime mortgages by year of origination, 468
Substitutes, 173
Supervision, effects of more, 295–296, 304, 305, 307
Supply, excess, 389, 390–391
Supply curve, 173–175
  defined, 173
  location of factors determining, 180
  movement along, 179–180
  shift of, 179–180
Surplus
  defined, 69
  size of, 69
Surplus product, 67–69, 74–78, 90, 109
  change and, 86–87
  and classes, 90–91
  competition and investment, 133–134
  conflict over size of, 84–86
  defined, 67, 75
  political power and, 87
  and producer's economic well-being, 84–85
  productivity and development, 370
  and profits, 108–110
  technical change and, 84–85
  ways to increase, 79–84, 83
Sustainable Development Goals, 359
Sweden, 372
  buying power, 355
  leisure time, 330
  living standards, 355–357
  opportunity compared, 340
  unemployment rate, 424, 448, 450, 452
  union restraint, 450
Switzerland, 494–495
Systems of control, 300, 302, 302–307
  defined, 304

T

Tariffs, 439
Tax audits, 504–505
Tax avoidance, 504
Tax breaks
  feedback loops and, 493
  profit-led employment situations, 420
Tax evasion, 504
Tax havens
  corruption related, 368
  defined, 377
  investment in, 354
  pharmaceutical TNC example, 504
  transnational corporations, 377–379
Tax holidays, 163, 164, 167, 368, 516–517
  government revenues and, 368, 516–517
Tax shelters, 504
Taxes, 409–410
  benefits in return for, 499
  corporate tax rates and payments, 503–504
  government distributional effect, 123
  government rules and corporate interests, 502–505
  rules and enforcement, 504
  shifts, 519
  transnational corporations (TNCs), 163
  trends since Reagan era, 130
  weakening enforcement, 504–505
Taylor, Frederick Winslow, 302
Technical change, 70, 308
  capital-goods-saving, 81–83, 83
  capitalist development and, 3–6
  and conflict, 309–314
  defined, 70
  inefficient, 319
  labor-saving, 81–83
  profitability versus efficiency, 319–321
  surplus product and, 84–85

Technical control, 304–305, 305–306
Technical efficiency
  defined, 319
  profitability versus, 319–321
Technical progress. *See also* Technical change
  defined, 70
Technology, 236. *See also* Information technology
  defined, 70
  grain, used as example, 79–84
  lighting, 10–11
  permanent revolution in, 3–6
Technology gap, 363–364
Temporary Asset Recovery Program (TARP), 473
Temporary Assistance to Needy Families (TANF), 165
Temporary workers, 316
Texas Instruments, 211
Three-dimensional economics
  competition, 235
  horizontal dimension, 51–52, 184, 235
  time dimension, 53, 54
  vertical dimension, 52–53, 54, 68, 235
Time, accumulation and, 135–136
Time, clock and natural, 17, 53
Time-Warner, 266, 268
Tire industry, 310–311
Tocqueville, Alexis de, 41
Too-big-to-fail firms, 456
  AIG, 471, 473
  breaking up firms, 487
  moral hazard, 483
Total cost of investment, defined, 256
Total demand. *See* Aggregate demand
Total output, measuring, 391
Total product, 74–75, 79
  defined, 75
  and surplus product, 76
Total return from investment, 255
Totals, list of variables, 529–530
Townsend, Joseph, 186
Trade, 439–440
  evolution of, 25
  globalization, 25
Trade-offs, 32
Traffic congestion costs, 199–200
Tragedy of the commons, 203–205
Transfer payments, 326
Transfer pricing, 377
Transnational capitalism, 134, 145, 147, 425
  deregulation and financialization, 163–165
  globalism, 155–165
Transnational corporations (TNCs), 118–119, 147
  economies of scale, 361–362
  foreign government decisions, 520
  governments and, 374–375
  investment decisions by, 375–376
  investment in tax havens, 354
  investor-state dispute settlement (ISDS), 494, 520
  supranational rule changes, 519–524
  tax havens, 377
  tax holiday, 163, 164, 167, 516–517
  taxation and regulation, 516–517
  trade deficits and, 437
  transfer pricing, 437
Transportation, capitalist development and, 4–5
Triangle Shirtwaist Factory fire, 380
Trip recorders, 312
Trough, of business cycle, 408, 420
Trucking companies, 312

U

UK. *See* United Kingdom
Ultimatum game, 34–36, 37–38, 39, 41
Underconsumption, 476
Undernourishment, 100–101
Unemployment, 273, 431, 448–452
  and aggregate demand, 383–421
  and the business cycle, 408
  costs of, 400, 422–423
  and demand, 386
  dependency on jobs, 15
  during 1990s boom, 452
  effect on profit rate, 431
  as excess supply of labor, 389
  globalization, 25
  and government fiscal policy, 400–407
  government policy and, 400
  Great Recession, 473
  and high-employment profit squeeze, 425, 431–434
  insurance, 286–287, 296–297
  interest rate and Federal Reserve policy, 416
  and international competition, 435
  international trade and macroeconomic policy, 439–442
  in macroeconomic equilibrium, 398
  measuring, 386–389
  monetary policy, 415–417
  output and, 388–389
  quality of life, 400
  and social distress, 423
  two official rates, 474
  U.S. rate of, 1890–2014, 385
  wages and aggregate demand, 417–420
Unemployment compensation (*ui*), defined, 423
Unemployment duration (*ud*), 287
Unemployment-inflation trade-off. *See* Inflation-unemployment trade-off
Unemployment insurance (*ui*), 428
  as built-in stabilizer, 409
  effects of, 296–297
  and workers' fallback wage, 428

Unemployment rate, 404
rate too low for Wall Street, 422
Uneven development, 354, 359
in auto and truck production, 362
successes and failures, 373, 374
Unions, 147–152, 299–300, 314–317
collective bargaining rights, 454
defined, 314
globalization, 303, 315
history, 146
industrial, 148, 149
managing threat of replacement, 315–316
membership, 314–315, 452
in the United States, 149–151, 149–152
promoting social changes to benefit workers, 317
rise and fall of, 147–151, 147–152
types of activities, 315
in U.S., 452
Unit cost (*uc*), 432–433
Unit labor cost (*ulc*), 227–228, 277, 280–281, 291, 293, 307
high employment labor cost push, 426–430
reducing, 309–310
Unit materials cost (*umc*), 227, 426
defined, 224
United Arab Emirates, 369
United Kingdom (UK), 372, 373, 424
United Mine Workers, 280
United States
class structure, 139–141
immigration, 157–158
income distribution, 358–359
income distribution, 1966–2014, 334
leisure time compared, 330
living standards compared to other countries, 355–356
medical systems compared, 491–492, 494–496
net worth by race, gender, education, and income (2011), 337
unemployment rate, 424
Universal basic income (UBI), 523
Universal pension and health programs, 120, 498–500
Urbanization, 13–15
U.S. Army, 321
USSR. *See* Soviet Union (former USSR)

V

Value added, 222
defined, 76
use of concept, 392
Value of net output (*Y*), defined, 222
Value of the dollar, 437–439
Value of total output, 220
Values in political economy, 57–64
Vertical dimension, 52–53
Vertical economic interdependence, defined, 69
Vicious circles, 354
auto industry, 362
feedback loops as, 480
low productivity, 359, 361
technology gap, 363–364
Virtuous circles, 354
auto industry, 362
feedback loops as, 480
Volcanic events, 19
Voluntary, 56
Volvo, 303
Voter suppression, 526
Voters and voting, 126–128, 198
consumers as "voters," 191
turnout among nations, 127
Voting Rights Act, 526

W

Wage(s), 308
aggregate demand, and unemployment, 417–420
and bureaucratic control, 306
determination of, 279–280
during 1990s boom, 428–430
gender gap, 346–348, 346–351
in Great Britain, 1264–1999, 8
high-employment wage push, 426–430
manufacturing, 376
mining, 299–300
real wage, 428
too low for employer, 399–400
unemployment, and bargaining power of workers, 426–430
unionized labor, 149–151
women's and men's, 346–348, 347
Wage gap, 113, 327
Wage labor, 96, 108
defined, 108
wage trends, 124
Wage-led employment situation, 418, 419–420
defined, 419
Wage rate (*w*), 251, 286, 289
consumer demand, 386
defined, 221
Wage workers. *See* Wage labor; Working class
Wagner Act (1935), 148–149, 454
Wall Street, 422
average bonuses, 485
bonuses, 487–488
War on Drugs, 509
War on Poverty, 113
Wealth
composition of holdings in 2010, 335–336
concentration in U.S., 338
concentration of, and crises, 489–490
defined, 104
distribution, measuring, 337
distribution in quintiles, 337
highest quintile (U.S.), 337
income share of top group, 331, 333
inequality of, 327, 334–338

Wealth (*continued*)
kinds of, 335
ownership, by households, 335
and political influence, 525
as stock, 334
Wealthy
conflict over rules, 522
economic elite opinion versus average citizen opinion, 518
policy preferences, 493
Well-being, 84, 328–331
determinants (chart), 329
goods and services and well-being, 330
measuring, 328–331
non-commodity goods and services, 330
respect and, 329
West Bengal, 40
Western world, 7
Westinghouse Corporation, 253
White-collar workers, 313
Women
discrimination against, 153
earnings gaps, 346–351
elected political leaders, 349
income inequality, 327
work of, and men's work, 346–348, 347
Work, 51, 271–298
changing nature of, 15–16
contract specifying, defined, 284
effect of capitalist development on, 15–16
effort, 276–277, 279
extraction of, from workers, 272, 282–283, 290
intensity of, 301
length of work week, 330
and pay norms, 286
social organization of, 303
women's and men's, 346–348, 347
Work and pay norms, 286
Work effort (e), 236, 272, 276–277, 279, 297, 315
defined, 228
determination of, 286–296
Work week compared, 330
Worker cooperatives, 102–103
Workers, 107–108, 207–213, 309–310
accumulation process, 138
average growth of real GDP per worker, 370
bargaining process with, 281
conflict between employers and, 276–281, 309–310
defined, 107
discouraged, 387
divided interests over government policy, 125
extraction of work from, 272, 282–283, 290
fraction of wages consumed (c), 394–395
health and safety, 16, 59, 125, 227, 306
management, 309
part-time, 387–388
shifts in what they produce, 159–161
skills, 368
social organization of the workplace, 301–309
temporary, 316
and unions, 314
white-collar, 313
and work incentive, 448
Working class, 139, 141
defined, 107, 326
Workplace, 299–300, 309–314
conflict in, 309–310
rapid change, 16
social organization of, 301–309
socialization for, 42–43, 45
World Economic Forum, 26
World Health Organization, 361
World Trade Organization (WTO), 26, 130
intellectual property rights, 365
investor-state dispute settlement, 494
overview, 519–520
productivity curbed, 354
rules of the game, 162, 364
technology transfer restricted, 364

Offers an introduction to economics that pays extensive attention to inequality, the global economy, contemporary public issues, and more

"*Understanding Capitalism* is a great introduction to political economy and an excellent alternative to standard texts for introducing students to economics in general. The text's framework and analysis help students develop a critical perspective on the economic and social world around them, like no other textbook on the market does."

—**Jonathan Cogliano, *Dickinson College***

Understanding Capitalism: Competition, Command, and Change, Fourth Edition, is an introduction to economics that explains how capitalism works, why it sometimes fails, and how it undergoes and brings about change. It discusses both the conventional economic model and the role of power in economic interactions. *Understanding Capitalism* analyzes how profit-seeking is a source of competition among firms, conflict, and change.

NEW TO THIS EDITION

- Data and figures have been thoroughly updated throughout
- Three new and four substantially rewritten chapters clarify the content and strengthen the book's international perspective
- A new Chapter 6, "Government and the Economy," explains how the establishment of rules like limited liability and corporate personhood have increased investment growth and consolidation of enterprises in the U.S., laying the foundation for transnational corporations
- A new Chapter 18, "Financial and Economic Crisis," describes the unfolding of the Great Recession, analyzes changes in the financial sector, and discusses the causes of economic crises in general
- A new Chapter 19, "Government and the Economy in Transnational Capitalism," considers issues in private vs. public provision of services, especially in health care and prison management
- Simpler algebra has been seamlessly integrated into Chapters 10, 11, and 12

ABOUT THE AUTHORS

Samuel Bowles is Research Professor at the Santa Fe Institute and Professor of Economics at the University of Siena, Italy.

Richard Edwards is Professor of Economics at the University of Nebraska–Lincoln.

Frank Roosevelt is Professor Emeritus of Economics at Sarah Lawrence College.

Mehrene Larudee is Visiting Associate Professor at Hampshire College.

OXFORD
UNIVERSITY PRESS
www.oup.com/us/he

Cover Photo: iStock
Cover Design: OUP Design

ISBN 978-0-19-061093-7
90000
9 780190 610937